# Assessing the informational Content and Network Structure of Financial Assets with an Application to Pricing

Ehsan Fazel

A Thesis
In the Department
of
Economics

Presented in Partial Fulfillment of the Requirements
For The Degree of
Doctor of Philosophy (Economics) at
Concordia University
Montreal, Quebec, Canada

February 2020

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Ehsan Fazel

Entitled: Assessing the informational Content and Network Structure of Financial Assets with an Application to Pricing

and submitted in partial fulfillment of the requirements for the degree of

**DOCTOR OF PHILOSOPHY (Economics)**

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

______________________________ Chair
Dr. Kregg Hetherington

______________________________ External Examiner
Dr. Abderrahim Taamouti

______________________________ External to Program
Dr. Alexey Lyubimov

______________________________ Examiner
Dr. Ian Irvine

______________________________ Examiner
Dr. Xintong Han

______________________________ Thesis Supervisor
Dr. Prosper Dovonon

Approved by ______________________________
Dr. Christian Sigouin, Graduate Program Director

March 24, 2020 ______________________________
Dr. André Roy, Dean
Faculty of Arts and Science

# ABSTRACT

**Assessing the informational Content and Network Structure of Financial Assets with an Application to Pricing**

**Ehsan Fazel, Ph.D.**
**Concordia University, 2020**

This thesis consists of three essays. Essay one adds to Chinn and Coibion (2014) by suggesting an alternative measure to investigate the informational content of futures. The paper utilizes the causality measure introduced by Dufour and Renault (1998) and Dufour and Taamouti (2010) to investigate the informational content of the futures' basis in predicting the price changes. Fourteen commodities and their corresponding prices are examined in three different time horizons and for each, the causality measure and its percentile bootstrap confidence interval are obtained. Results show that in general for energy and agricultural products, the contemporaneous basis of the futures has information to predict the price change in the market, whereas base and precious metals fail in this respect.

Essay two designs a network of realized volatility for the equity market consisted of firms and their customers collected from Compustat customer segment data covering from 1980 to 2017. In the same spirit as Herskovic (2018), two factors are derived from the network. The first is concentration which portrays the node characteristics of the network; whereas the second, sparsity, describes the evolution of the edges. Clustering is used to group the weekly volatility series into nodes and then calculate the concentration factor as the negative entropy of market shares. To complete the network, the causality measure of Dufour and Taamouti (2010) is employed to acquire the spillovers and sparsity. Factors are shown to reduce average pricing errors significantly in the APT setup. To explore further the asset pricing implications, sorted portfolios are created. In tercile portfolios, assets with low-concentration-beta have a 2.15 percent higher return than their counterpart in high-concentration beta annually. For sparsity, the spread between high and low is 0.83 percent.

While both high-minus-low investment strategies for two factors show significant returns, results for concentration are more robust to various control variables.

Finally, essay three accommodates the concentration and sparsity factors previously used in analyzing the equity market to address the network of 25 commodity return volatilities for 2000-2017. Similar to previous studies, I document clustering among commodities with a noticeable difference compared to industry categorization. I investigate four commodity-based factors by creating portfolios of futures assets to investigate concentration and sparsity along with the hedging pressure and the basis. The spreads in all long-short strategies are insignificant. Following the law of one price and market integration, I estimate the price of risk of a wide set of equity-based and commodity-based factors. None of the risk premia are significant which appears to be in line with segmented features of the commodity market. Time-series regressions of betas for each commodity and each factor highlights the heterogeneous nature of commodities.

# ACKNOWLEDGMENT

First, I would like to express my appreciation to my supervisor, Dr. Prosper Dovonon, for all the support, encouragement, and motivation. Without his guidance and constant feedback and assistance, this Ph.D. would not have been achievable.

I would like to offer my sincere thanks to the members of my thesis committee, Dr. Abderrahim Taamouti, Dr. Alexey Lyubimov, Dr. Ian Irvine, and Dr. Xintong Han, for their invaluable and critical comments, which enabled me to address the weaknesses of the essays and aim to improve them.

My sincere gratitude also goes to the former and current Graduate Program Directors, Dr. Effrosyni Diamantoudi, Dr. Szilvia Pápai, Dr. Damba Lkhagvasuren, and Dr. Christian Sigouin (current), and to the former and current Chairs, Dr. Greg LeBlanc and Dr. Jorgen Hansen (current).

I will always carry the thought of my dear friends being present and supportive during this long journey. My special thanks go to Puya Ghasvareh, Claire Jacmin-Park, Rouzbeh Ghouchani, Pegah Behmaram, Ghodsi Madani, Roselia St. Louis, Jean Pierre Goulet, Mirwais Alamy, Michael Aguessy, Shahab Shafe, and Yasi Eftekhari. I am very thankful to my beloved brother and sister, Vahid and Toktam, for their unconditional love and support for every single step of this adventure.

I wish to thank the staff members of the department of economics, Ms. Elise Melançon, Ms. Lucy Gilson, Ms. Lise Gosselin, Ms. Bonnie Janicki, Ms. Mary-Ann Jirjis, Ms. Sandra Topisirovic, Ms. Melissa Faisal, and Ms. Émilie Martel. Their help, patience, and diligence were always appreciated.

And finally, above all, I would like to thank my beloved parents, Masy and Mohammad. I have been constantly carrying their voice of love, encouragement, and wisdom in my mind and my heart. Their love from thousands of kilometers away was felt deeply in my heart.

*To Maman & Baba*

*“The tears of the world are a constant quantity. For each one who begins to weep somewhere else another stops. The same is true of the laugh.”*

Samuel Beckett, Waiting for Godot

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# CHAPTER I

# ON THE INFORMATIONAL CONTENT OF THE FUTURES PRICE IN THE COMMODITY FUTURES MARKET

## *1.1 Introduction*

The commodity futures market has been the subject of extensive literature since the beginning of the 21st century. One of the reasons that researchers are keen to investigate the dynamics of this market is its growing importance in investment portfolios in particular and in the economy in general. Participants in a commodity futures market agree to buy or sell a predetermined amount of a commodity at a pre-specified price at a stated date in the future. These terms and conditions make a futures contract. The pre-specified price in the contract is called the futures price and since the contract matures at a time in the future, the futures price should be the predictor of the price of the commodity at the maturity date - which is called the spot price. The primary rationale that participants engage in such markets is the volatility of commodity prices. For example, in case of an expected fall in the price of wheat, by entering into a futures contract, a producer of wheat would be able to fix the selling price of her product long before the delivery of the commodity to the buyer and avoid a loss. To efficiently manage the aforementioned process, which is known as hedging, the commodity producer investigates the price movements in the market. Such investigation in our example, helps the wheat producer speculate on the future price of her commodity or, in other words, *predict the future spot price* of wheat. The aforesaid setting highlights the key relationship between the futures price and the spot price of a single commodity. In pursuance of studying this relationship, the theory needs to be established first.

The association between the futures price and the spot price is built upon a theory which is called the efficient market hypothesis. It was introduced by Fama (1969) and Samuelson (1965). Both define efficiency as the impossibility of beating the market consistently; put differently, a market is said to be efficient if prices are unpredictable. The only difference between the two is that Fama builds efficiency on the random walk definition, whereas Samuelson employs the martingale. Later, Fama (1991) introduces three forms of efficiency conditional on different information sets available to the players in the market. In the strong efficient market, prices reflect all information, including instant insider information. In the semi-strong efficient market, prices reflect public information and instantaneous changes to them. Finally, in the weak efficient market, prices reflect only publicly available information. The weak form of the theory has been tested extensively for financial markets in the literature, and the outcome of these tests appears to be engaging. Even though the theory is well-established, its validation in empirical research has yielded mixed results. (For a review on efficiency, see Lo, 2008). Here I try to classify the different works done on the subject and I shall mention that although different methodologies have been used to analyze the dynamics of the commodity futures market, this paper inquires into the market through time series analysis.

The first strand of literature revolves around the investigation of efficiency employing co-integration analysis. The underlying assumptions are that the logarithm of spot and futures prices exhibit unit root and as a result, there exists the possibility of co-integration among the pair. The co-integration vector of $(1, -1)$ is defined and tested. Baillie and Bollerslev (1989) use the Phillips and Perron (1988) tests of unit root and find non-stationarity in the spot and forward exchange rates as well as co-integration in the premium. The outcome that verifies efficiency in the long-run. As they point out in their paper, similar results are likely to be true for other financial assets with alike characteristics such as commodity futures. McKenzie and Holt (2002) follow the same methodology, but they only consider agricultural commodities. They test market efficiency and unbiasedness (see section 2.1 for details regarding unbiasedness) by utilizing co-integration and error correction models and conclude that live cattle, hogs, corn, and soybeans are efficient in the long-run. Switzer and El-Khoury (2007) test the efficiency of the NYMEX's light sweet crude oil from 1986 to 2005 and by rejecting the hypothesis of no co-integration using Johansen (1991) tests, conclude the presence of the long-run relationship between spot and futures prices.

The second strand of literature challenges the previous approach and proposes a more general form to capture the kinetics of the market. One of the early studies in this category is Baillie and Bollerslev (1994). According to the authors, fractionally integrated models, namely $I(d)$, that capture the long memory characteristics of a time series, are better representations of the futures price. They assert that the "traditional" way of testing efficiency proves the existence of a co-integration vector of $(1,-1)$ between spot and futures prices; but their results demonstrate that instead, a co-integration vector of $(1,d)$ should be utilized in analyzing forward exchange market efficiency. Baillie and Bollerslev (1994) set the way for using fractional integration in modeling spot and futures prices. Later on, Cavaliere et al. (2015) provide bootstrap tests for market efficiency and the underlying assumption in their work is that log of spot and futures prices, as well as the basis (the difference between spot and futures), exhibit long memory. They estimate the long-memory parameter by applying auto-regressive fractionally moving average (ARFIMA) models and find inefficiency in the oil market. (See section 2.3 for details on ARFIMA).

The third strand that investigates the time-series properties of prices considers more sophisticated methods to account for long memory in a series. Cajueiro and Tabak (2007) test the efficiency of crude oil markets of Brent and West Texas Intermediate (WTI) by means of estimating the long memory parameter. They estimate the Hurst exponent by using rescaled range analysis and conclude that these markets become more efficient in time. Another study that concludes the crude oil market is consistent with the efficient market hypothesis in the long-run is Alvarez-Ramirez et al. (2008). They utilize detrended fluctuation analysis to estimate the Hurst exponent dynamics of returns. See Kristoufek and Vosvrda (2013) for a comprehensive review on the subject.

The fourth strand of literature ventures to add another perspective in inspecting the futures market and that is the informational content of the futures. Here, the central notion is not merely testing efficiency or unbiasedness, but to examine the relative predictive power of futures prices among a set of commodities as well. Chinn and Coibion (2014) examine the predictive power of futures prices for a set of fourteen commodities, including energy, base, and precious metals, and agricultural commodities by carrying out estimation and statistical inference on the relationship between the futures basis and ex-post price change. They find out that futures prices of precious metals are the least accurate predictors of subsequent prices. A fact that highlights the heterogeneity among commodities in terms of market structure and the corresponding participants. Bernard et al. (2015) examine two methodologies in studying the dynamics of the oil market, namely, equilibrium models and time series analysis. They test alternative models to examine which could produce more accurate forecasts and it turns out to be models that allow for time-varying convenience yield. In other words, models that incorporate both the price level and the distance between the price and spot price perform better in terms of predictability. Alquist and Kilian (2010) is another study that finds weak accuracy for futures in the oil market. They study oil futures spread and unravel high variability of futures about the spot price using a two-country, multi-period general equilibrium model of futures and spot markets.

Having discussed the 4 main approaches in the literature, there are a few points that need to be explained regarding this article. This paper is in alignment to the last approach, namely, investigating the predictability of futures prices and their informational content. As a matter of fact, it could be considered as a complementary work to Chinn and Coibion (2014) in covering a longer period in the empirical application and in introducing an alternative way of measuring the predictive power. In Chinn and Coibion (2014), the basis equation is the center of the attention. The basis, which is simply the difference between the futures price and the spot price, has information to explain the pattern in the price change of a commodity. Therefore, investigating this relationship could be helpful to shed more light on different commodity markets.

The core question that I address is how accurate the futures price is in predicting the spot price? If we reconsider our example of wheat producers, we will be able to understand the importance of the question. The players in the commodity futures market should be able to measure the information contained in the futures since they observe the futures price as a proxy of the future value of the underlying commodity. The main contribution of the paper is suggesting the causality measure, which is popular for its simplicity, as a tool to capture the accuracy of futures to predict subsequent spot prices.

The paper is structured as follows. Section 2 explains the methodology which provides a comprehensive explanation of the concepts in commodity markets, causality measure, and parametric long memory models. Section 3 delivers the empirical analysis which contains a description of the data, the basis regression analysis, stationarity, and long memory tests, and finally investigation of the causal relationship. Lastly, section 4 concludes.

## *1.2 Methodology*

The methodology is divided into three sections. The first section introduces the commodity futures market setting and its corresponding variables and the relationships among them. The second section presents the definitions of the causality and causality measure that will be used later in the third section to examine the informational content of commodity futures in a prediction model.

### 1.2.1 Commodity Futures Market

As discussed in the introduction, the futures price is a prediction of the spot price. As Mckenzie and Holt (1998) emphasize, the statement that the futures price is an *unbiased* predictor of the spot price is a joint hypothesis of market efficiency and risk neutrality. According to Brenner and Kroner (1995) “the assumptions of risk neutrality and rationality are so central in many financial models that their importance cannot be understated”. These statements are about the difference in the current futures and future spot prices. But we have another notion in commodities that is called *the basis*. Fama and French (1987) express the difference between the futures price and the current spot price of a commodity as the following:

$$f_{t,t+h} - s_t = E_t[\pi_{t,t+h}] + E_t[s_t - s_{t+h}] \tag{1}$$

Where the left-hand side is the basis, $h$ is the horizon by which contract expires, and therefore $f_{t,t+h}$ is the (log) time $t$ of a futures contract price that matures at time $t+h$, $s_t$ is the time $t$ spot price, $E_t[\pi_{t,t+h}]$ is the expected premium, and $E_t[s_t - s_{t+h}]$ is an expected change in the spot price. The expected premium is defined as the forecast error of the futures price as a predictor of the spot price:

$$E_t[\pi_{t,t+h}] = f_{t,t+h} - E_t[s_{t+h}] \tag{2}$$

To assess the predictability power of the futures, Fama and French (1987) introduce the following regression that is known as the basis regression:

$$s_{t+h} - s_t = \alpha + \beta(f_{t,t+h} - s_t) + u_{t+h} \tag{3}$$

$\beta$ expresses the informational content of futures. If $\beta = 0$, then the basis has no predictability power. Consequently, as noted by Chinn and Coibion (2014), the joint hypothesis of $\alpha = 0$ and $\beta = 1$ tests whether the basis is the optimal predictor of the change in the spot price.

Chinn and Coibion (2014) express the above relationship in terms of futures prices at different horizons which yields the following:

$$f_{t+h-1,t+h} - f_{t,t+1} = \alpha + \beta(f_{t,t+h} - f_{t,t+1}) + u_{t+h} \tag{4}$$

Equations (3) and (4) are equivalent. The left-hand side of (4) represents the ex-post price change (hereafter the price change) and the right-hand side is the contemporaneous basis (hereafter the basis) for any horizon $h$. Below, another notation of (4) is presented:

$$p(t+h) = \alpha + \beta[b(t+h)] + u(t+h) \tag{5}$$

Where $p(t+h)$ is the time series of price change, and $b(t+h)$ is the time series of the basis of a single commodity for time horizon $h$. This simplified notation will be used later in examining the information of futures by means of causality measure.

### 1.2.2 Concepts and Definitions of Causality Measure

The statistical concept of causality among two vectors dates back to Granger (1969). He discusses the causality as the predictability of a stationary variable $X$, by its own past and the past of another variable $Y$, in a bivariate setting. Later Geweke (1982, 1984) introduced the causality measure in the first horizon; the idea that was extended by Dufour and Renault (1998) to any arbitrary horizon $h$, with $1 \leq h \leq \infty$. Here I borrow the exposition of causality measure, which is introduced by Zhang et al. (2015) and Dufour and Taamouti (2010), and I accommodate the setting of the commodity futures market in them, namely, the basis $(b_i)$ and price change $(p_i)$ for a single commodity $i$.

Denote $L^2 \equiv L^2(\Omega, A, Q)$ a Hilbert space with finite second moments. Information set is denoted by $I(t) = \{I(t) : t \in \mathbb{Z}, t > \omega\}$ with $t < k \Longrightarrow I(t) \subseteq I(k)$ for all $t > \omega$. $I(t)$ is defined on the Hilbert subspace of $L^2$ and $\omega \in \mathbb{Z} \cup \{-\infty\}$ represents a "starting point", $\mathbb{Z}$ is the set of all integers. Now consider two multivariate stochastic processes:

$$
\begin{aligned}
p(t) &= (p_{1,t}, ..., p_{i,t})', i = 1, ...k_1, \quad k_1 \geq 1 \\
b(t) &= (b_{1,t}, ..., b_{j,t})', j = 1, ...k_2, \quad k_2 \geq 1
\end{aligned}
$$

$p(\omega, t]$ and $b(\omega, t]$ are the Hilbert spaces spanned by the components of $p_i(\tau)$ and $b_i(\tau)$ respectively with $\omega < \tau < t$. In other words, $p(\omega, t]$ and $b(\omega, t]$ represent information on the history of the price change and the basis. Regarding the information sets, the following can be stated:

$$
\begin{aligned}
I_p(t) &= I(t) + p(\omega, t] \\
I_{pb}(t) &= I_p(t) + b(\omega, t]
\end{aligned}
$$

For any information set, $I(t)$ and any positive integer $h$ (horizon), $P[p(t+h)|I(t)]$ indicates the best linear forecast of $p(t+h)$ given the information set $I(t)$, and the followings are the corresponding prediction error and variance:

$$
\begin{aligned}
u[p_i(t+h)|I(t)] &= p_i(t+h) - P[p_i(t+h)|I(t)] && (6) \\
\sigma^2[p_i(t+h)|I(t)] &= E\{u^2[p_i(t+h)|I(t)]\} && (7)
\end{aligned}
$$

For a vector of observations, below corresponds to the best linear forecast of $p(t+h)$ and variance-covariance matrix:

$$
\begin{aligned}
P[p(t+h)|I(t)] &= (u[p_1(t+h)|I(t)]', ..., u[p_{k_1}(t+h)|I(t)]' && (8) \\
\Sigma[p(t+h)|I(t)] &= E\{U[p(t+h)|I(t)]U[p(t+h)|I(t)]'\} && (9)
\end{aligned}
$$

where $U[p(t+h)]$ is the forecast error at horizon $h$.

Having defined the prediction vector and variance-covariance matrix, it is now time to define causality (non-causality). The following is the first part of the definition of non-causality provided in Dufour and Renault (1998) that has been modified for the setting of the commodity futures market:

**Definition 1** *Non-causality at horizon h, for* $h \geq 1$*:*

*the basis* $(b)$ *does not cause price change* $(p)$ *at horizon h given I [denoted* $b \underset{h}{\nrightarrow} p|I$*] if and only if:*

$$P[p(t+h)I_p(t)] = P[p(t+h)|I_{pb}(t)], \quad \forall t > \omega \tag{10}$$

*Where* $I_p(t) = I(t) + p(\omega, t]$ *and* $I_{pb}(t) = I_p(t) + b(\omega, t]$.

Another way of characterizing non-causality is by the following proposition presented in Dufour and Tammouti (2010)*:*

**Proposition 2** *Covariance characterization of non-causality at horizon h, for* $h \geq 1$*:*

*the basis* $(b)$ *does not cause price change* $(p)$ *at horizon h given I [denoted* $b \underset{h}{\nrightarrow} p|I$*] if and only if:*

$$\det \Sigma[p(t+h)|I_p(t)] = \det \Sigma[p(t+h)|I_{pb}(t)], \quad \forall t > \omega \tag{11}$$

*Where* $\Sigma[p(t+h)|.]$ *is defined by* $E\{U[s(t+h)|.]U[s(t+h)|.]\}$

After presenting the definition of non-causality, I assimilate commodity futures market variables into the definition of the causality measure defined by Dufour and Taamouti (2010*):*

**Definition 3** *Mean-square causality measure at horizon h relative to an information set I and for* $h \geq 1$ *is shown as:*

$$C(b \underset{h}{\longrightarrow} p|I) = \ln\left[\frac{\det \Sigma[p(t+h)|I_p(t)]}{\det \Sigma[p(t+h)|I_{pb}(t)]}\right] \tag{12}$$

The interpretation of the above definition is important. If the basis does not have informational significance to predict the price change, then the knowledge of the history of the basis will not help us predict the future price change. Therefore the two terms in the fraction cancel out and the measure equals zero. However, as the predictive power of basis increases, the determinant in the denominators shrinks relative to the numerator and the measure indicates a value greater than zero.

### 1.2.3 The Prediction Model

A group of models that can capture long memories in time series is auto-regressive fractionally integrated moving average ($ARFIMA$) models. This class of models was first introduced by Granger (1980), Granger and Joyeux (1980) and Hosking (1981). $A$ uni-variate $ARFIMA$ with $p$ as the auto-regressive order, $q$ as the moving average order, and $d$ as the order of integration is represented by:

$$\Phi(L)(1-L)^d y_t = \Theta(L)z_t \tag{13}$$

Where $L$ is the backward-shift operator and $z_t$ is $i.i.d(0,\sigma^2)$. Auto-regressive, moving average, and the fractional differencing operators are defined by:

$$\Phi(L) = 1 - \phi_1 L - ... - \phi_p L^p \tag{14}$$

$$\Theta(L) = 1 + \varphi_1 L + ... + \varphi_q L^q \tag{15}$$

$$(1-L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)L^k}{\Gamma(-d)\Gamma(k+1)} \tag{16}$$

Where $\Gamma$ denotes the gamma function and $d$ is allowed to be any real value in a general form of the process. For the process to be stationary and invertible, $d$ needs to be in the interval of $(-\frac{1}{2}, \frac{1}{2})$[1].

Now consider the following vector on $L^2$:

$$\delta(t) = (p(t)', b(t)') \tag{17}$$

$\delta(t)$ is defined by a stationary and invertible bi-variate ARFIMA. A bivariate class of models can be extracted from a multi-variate ARFIMA presented in Tsay (2010), and it can be shown as:

[1] In the case of $d = 0$, the model turns to Autoregressive-Moving-Average ($ARMA$).

$$\Phi(L)diag(\nabla^d)\delta(t) = \Theta(L)Z(t) \tag{18}$$

Where $diag(\nabla^d) = \begin{bmatrix} \nabla^{d_1} & 0 \\ 0 & \nabla^{d_2} \end{bmatrix}$, $\nabla = 1 - L$, and $Z(t) = (z_{1,t}, z_{2,t})'$. $Z(t)$ is $i.i.d.$ random with $E[Z(t)] = 0$, $E[Z(t)Z(s)] = \Sigma_Z$ for $t = s$ and $E[Z(t)Z(s)'] = 0$ for $t \neq s$. $\Phi(L)$ and $\Theta(L)$ are finite order matrix polynomials such that:

$$\Phi(L) = \Phi_0 - \Phi_1 L - ... - \Phi_p L^p \tag{19}$$

$$\Theta(L) = \Theta_0 + \Theta_1 L + ... + \Theta_q L^q \tag{20}$$

Where $\Phi_0 = \Theta_0 = Id_2$, and $Id_2$ is a two by two identity matrix. To model the causality measure from $b(t)$ to $p(t)$, we need to specify the structure of another process that only captures the price change, that is:

$$\delta_0(t) = p(t)' \tag{21}$$

This process follows a uni-variate stationary $ARFIMA$ that is represented by the following:

$$\Phi_0(L)(1 - L)^d \delta_0(t) = \Theta_0(L)z(t) \tag{22}$$

Where $z(t)$ are $i.i.d$ random with $E[z(t)] = 0$, $E[z(t)z(s)] = \sigma_z$ for $t = s$ and $E[z(t)z(s)] = 0$ for $t \neq s$. $\Phi_0(L)$ and $\Theta_0(L)$ are lag polynomials.

Under stationarity and invertibility assumptions, $\delta(t)$ has a $VMA(\infty)$ representation:

$$\delta(t) = \sum_{j=0}^{\infty} \Psi_j Z(t - j) \tag{23}$$

Where $\Psi_j$ are impulse response functions with $\Psi_0 = Id_2$.The linear forecast error of

$\delta(t+h)$ and its variance-covariance matrix are:

$$
\begin{aligned}
U_L[\delta(t+h)|I_\delta(t)] &= \sum_{j=0}^{h-1}\Psi_j Z(t+h-j) && (24)\\
\Sigma[\delta(t+h)|I_\delta(t)] &= \sum_{j=0}^{h-1}\Psi_j E[Z(t+h-j)Z(t+h-j)']\Psi_j = \sum_{j=0}^{h-1}\Psi_j \Sigma_Z \Psi_j' && (25)
\end{aligned}
$$

Thus the MSE for the linear forecast of $p(t+h)$ is:

$$
\sigma^2[p(t+h)|I_\delta(t)] = \sum_{j=0}^{h-1} J_1 \Psi_j \Sigma_Z \Psi_j' J_1' \tag{26}
$$

Where $J_1 = (1 \;\; 0)$.

Same as $\delta(t)$, $\delta_0(t)$ can be written as an MA($\infty$) such as:

$$
\delta_0(t) = \sum_{j=0}^{\infty} \psi_j z(t-j) \tag{27}
$$

And the forecast error for the linear forecast of $p(t+h)$ and its variance are:

$$
\begin{aligned}
U_L[\delta_0(t+h)|I_{\delta_0}(t)] &= \sum_{j=0}^{h-1}\psi_j z(t+h-j) && (28)\\
\Sigma[\delta_0(t+h)|I_{\delta_0}(t)] &= \sum_{j=0}^{h-1}\psi_j^2 E[z^2(t+h-j)] = \sigma_z^2 \sum_{j=0}^{h-1}\psi_j^2 && (29)
\end{aligned}
$$

Consequently, the causality measure that captures the predictive power of the basis to predict future price change in the commodity futures market is expressed as:

$$
C(b \underset{h}{\rightarrow} p) = \ln\left[\frac{\sigma^2[p(t+h)|I_{\delta_0}(t)]}{\sigma^2[p(t+h)|I_\delta(t)]}\right] = \ln\left[\frac{\sigma_z^2 \sum_{j=0}^{h-1}\psi_j^2}{\sum_{j=0}^{h-1} J_1 \Psi_j \Sigma_Z \Psi_j' J_1'}\right] \tag{30}
$$

Now the issue is estimating the above measure. To estimate $C$, the linear estimation approach proposed by Dufour and Taamouti (2010) is used. Under the invertibility assumption, $\delta(t)$ can be written as an infinite auto-regressive process such as the following:

$$\delta(t) = \sum_{i=1}^{\infty} \Pi_i \delta(t-i) + U(t) \tag{31}$$

Given the realization $\delta(1), ..., \delta(t)$, one can approximate the above by a finite order $VAR(k)$ model:

$$\delta(t) = \sum_{i=1}^{k} \Pi_{ik} \delta(t-i) + U_k(t) \tag{32}$$

And by using least squares, the coefficients of the fitted $VAR(k)$, $\Pi$ and the variance-covariance of the error terms, $\Sigma_{U|k}$ can be estimated. The order $k$ is selected by using the AIC. The similar procedure can be done for the uni-variate time series, $\delta_0(t)$ to enable us to estimate the measure as the following:

$$\widehat{C} = \ln\left[\frac{\widehat{\sigma_\epsilon} \sum_{j=0}^{h-1} \widehat{\psi}_j^2}{\sum_{j=0}^{h-1} J_1 \widehat{\Psi}_j \widehat{\Sigma_Z} \widehat{\Psi}'_j J'_1}\right] \tag{33}$$

After capturing the informational content of futures in estimating the measure, the relative predictive power of the basis for a set of commodities is presented in the following definition.

**Definition 4** *The Relative predictive power of the basis:*

*Based on the causality measure in multiple horizons and for a set of $K$ commodities, $K \in \mathbb{N}$, the following relation ranks the commodities based on the predictive power of their basis in predicting price changes:*

$$\widehat{C_1}(b \underset{h}{\rightarrow} p|I) < ... < \widehat{C_K}(b \underset{h}{\rightarrow} p|I) \tag{34}$$

The aforementioned definition provides a ranking among commodities based on the information available in their corresponding basis. To transform the causality measure into a benchmark measure between zero and one, I present the following definition:

**Definition 5** *The Predictive power of the basis:*

*Based on the causality measure presented for the predictive power of the basis to predict the price change, and for a given commodity $i$, the ensuing relation yields the predictive power of the basis:*

$$\eta = 1 - \frac{1}{1 + \widehat{C}_i} \tag{35}$$

The intuition of $\eta$ is pretty simple. It is a bounded measure between zero and one. If the futures price does not have the power to predict the spot price, the causality measure and $\eta$ indicate a zero value. As the information in futures increases, they become useful in forecasting the spot price and consequently, causality measure rises and $\eta$ converges to one.

## *1.3 Empirical Analysis*

### 1.3.1 Data

Similar to Chinn and Coibion (2014), a set of fourteen commodities is examined. The commodities are categorized into four groups, namely energy, base metals, precious metals, and agricultural commodities. The source of the data is Bloomberg, but due to the heterogeneous nature of the commodities, the numbers of observations and the covered periods are different across the data-set. Energy products that include West Texas Intermediate crude oil (hereafter WTI), natural gas, heating oil, and gasoline have contracts that expire in each month of the year and they are available from 1990 at 1-, 3-, 6-, and 12-month horizons. Gasoline contracts underwent a change in 2006, and therefore two tickers of HU and XB are combined[2] (See Table (8) in the Appendix). Base metals, including aluminum, copper, lead, nickel, and tin have the same maturity patterns, but they are available from 1997. Concerning agricultural commodities, they are not delivered in each month of the year. March, May, July, September, and December are the months of delivery for wheat and corn; and soybeans are delivered in January, March, May, July, August, September, and November.

[2]Unleaded gasoline had a change in contract in 2006 to a new Reformulated Gasoline Blendstock for Oxygen Blending (RBOB) contract. The data of the two contracts are combined to make the gasoline time-series.

Apart from the availability of the time series, this is how the two vectors of the basis and price change are built. For instance, using equation (4) for the 3-month horizon, the basis and price change are identified by $(f_{t,t+3} - f_{t,t+1})$ and $(f_{t+3-1,t+3} - f_{t,t+1})$, respectively. The basis is simply the difference between the 3- and 1- month horizon futures and the price change is the difference between the 1-month horizon futures forwarded 2 months ahead and 1-month horizon futures. The same is done for the two variables in the 6- and 12-month horizons. Chinn and Coibion (2014) state that since the correlation between the ex-post spot price and the 1-month futures is close to zero, this setting allows for the 1-month horizon to be used as the spot price.

### 1.3.2 The Basis Regression

Let us consider the basis regression to examine the predictability of the basis in different markets. Within energy products, the basis of gasoline and natural gas tend to be the most effective ones in predicting price changes. Crude oil and heating oil behave similarly. Table (1) exhibits the regression outcomes by OLS, for equation (4). The table is first introduced in Chinn and Coibion (2014) and here the null of $(\beta = 0)$ is added and regression is run using the new data-set. First, the relatively higher $R^2$ in energy and agricultural futures compared to base and precious metals is a criterion that can demonstrate higher predictability.

Second, testing the null hypothesis of $\beta = 0$ is reported in column $p$. Results show its rejection for all energy and agricultural products at 5 % level[3]. This is an indication that the basis has predictive content for the price change. The opposite is true for the base and precious metals. The null cannot be rejected[4] at 5 % level; a symptom of an uninformative basis. The test results are robust to heteroskedasticity and possible auto-correlation up to some lag since the standard error of $\beta$ is reported by the Newey-West method.

Third, column $Wald$ expresses the p-value of the joint hypotheses of $\alpha = 0$ and $\beta = 1$, which is known as testing for market efficiency. In a bigger picture, the null is rejected at 5

[3] The null is rejected at 1 % level for wheat in the 12-month horizon.

[4] For copper in the 3-month horizon, the null is rejected at 5 % but not at 1 % level.

% level for the majority of commodities. Copper and gold in 3-month horizons violate this pattern and show signs of efficiency. The absence of efficiency may create an opportunity to beat the market and earn unexpected profit by speculation. The test results presented here once again demonstrate that market efficiency which is widely accepted in theory is hard to prove.

Fourth, the unbiasedness hypothesis, namely $\beta = 1$, is mostly rejected for the base and precious metals. There are instances in energy and agricultural markets that the unbiasedness is rejected, but overall agricultural commodities seem to have better unbiased properties. Gold and lead are interesting cases. The point estimate of $\beta$ is always negative, which shows a negative relationship between the basis and price change. One explanation given by Chinn and Coibion (2014) is that since gold is used as a hedge against inflation, this might make it behave like an exchange rate rather than a commodity and the negative relationship can be similar to what is knows as forward discount anomaly observed in exchange rates by Engel (1996).

In sum, the regression brightens up the heterogeneity across commodities. It appears that in energy and agricultural markets the basis has information to predict the price change, whereas in the base and precious metals, the basis is uninformative. In a more restrictive definition, except for a few, market efficiency is consistently violated for all futures.

Table 1: Regression Results of Ex-post Price Change on Contemporaneous Basis in Multiple Horizons

The table reports the regression results of ex-post price change on the contemporaneous basis in different time horizons of 3, 6, and 12 month. $\beta$ is the estimated coefficient in equation (4) by OLS. $p$ is the p-value of $H_0 : \beta = 0$. $SE(\beta)$ is the Newey-West standard error. $Wald$ indicates the p-value of $H_0 : \alpha = 0, \beta = 1$. $N$ is the number of observations. *, **, *** show significance levels in testing $H_0 : \beta = 1$.

| | 3-month horizon | | | | | | 6-month horizon | | | | | | 12-month horizon | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\beta$ | $p$ | $SE(\beta)$ | $Wald$ | $R^2$ | $N$ | $\beta$ | $p$ | $SE(\beta)$ | $Wald$ | $R^2$ | $N$ | $\beta$ | $p$ | $SE(\beta)$ | $Wald$ | $R^2$ | $N$ |
| **Energy** | | | | | | | | | | | | | | | | | | |
| WTI | 0.81 | 0.00 | (0.31) | 0.76 | 0.04 | 326 | 0.84 | 0.02 | (0.39) | 0.68 | 0.06 | 323 | 0.72 | 0.01 | (0.32) | 0.36 | 0.06 | 317 |
| Natural gas | 1.04 | 0.00 | (0.11) | 0.23 | 0.24 | 323 | 0.81* | 0.00 | (0.11) | 0.07 | 0.20 | 320 | 0.85 | 0.00 | (0.21) | 0.31 | 0.14 | 312 |
| Heating oil | 0.68* | 0.00 | (0.18) | 0.11 | 0.06 | 326 | 0.58** | 0.00 | (0.22) | 0.07 | 0.05 | 323 | 0.88 | 0.00 | (0.29) | 0.58 | 0.09 | 317 |
| Gasoline | 1.16 | 0.00 | (0.19) | 0.11 | 0.20 | 325 | 1.13 | 0.00 | (0.15) | 0.08 | 0.24 | 322 | 0.73 | 0.00 | (0.25) | 0.22 | 0.09 | 316 |
| **Base metals** | | | | | | | | | | | | | | | | | | |
| Aluminum | -0.10 | 0.90 | (0.77) | 0.37 | 0.00 | 236 | 0.02 | 0.98 | (1.05) | 0.64 | 0.00 | 233 | 0.05 | 0.94 | (0.87) | 0.42 | 0.00 | 227 |
| Copper | 0.55** | 0.03 | (0.22) | 0.04 | 0.02 | 236 | 0.17 | 0.77 | (0.73) | 0.31 | 0.00 | 233 | -0.84** | 0.24 | (0.91) | 0.02 | 0.03 | 227 |
| Lead | -0.12** | 0.80 | (0.47) | 0.05 | 0.00 | 236 | -0.54*** | 0.34 | (0.54) | 0.02 | 0.01 | 233 | -0.56*** | 0.30 | (0.61) | 0.01 | 0.01 | 224 |
| Nickel | 0.12 | 0.90 | (1.08) | 0.65 | 0.00 | 236 | 0.21 | 0.85 | (1.22) | 0.71 | 0.00 | 233 | 0.66 | 0.38 | (0.85) | 0.56 | 0.01 | 227 |
| Tin | 0.18 | 0.86 | (0.90) | 0.45 | 0.00 | 236 | -0.70* | 0.41 | (0.93) | 0.05 | 0.00 | 233 | -0.79* | 0.46 | (1.28) | 0.04 | 0.01 | 224 |
| **Precious metals** | | | | | | | | | | | | | | | | | | |
| Gold | -0.66** | 0.33 | (0.73) | 0.04 | 0.01 | 326 | -0.29** | 0.61 | (0.65) | 0.05 | 0.00 | 323 | -0.14** | 0.77 | (0.60) | 0.04 | 0.00 | 316 |
| Silver | -0.69* | 0.44 | (0.96) | 0.07 | 0.00 | 326 | -0.63** | 0.44 | (0.92) | 0.06 | 0.01 | 323 | 0.52* | 0.07 | (0.33) | 0.24 | 0.03 | 301 |
| **Agriculture** | | | | | | | | | | | | | | | | | | |
| Corn | 1.44 | 0.00 | (0.28) | 0.12 | 0.22 | 134 | 1.36 | 0.00 | (0.22) | 0.05 | 0.27 | 131 | 1.48** | 0.00 | (0.24) | 0.03 | 0.48 | 103 |
| Soybean | 1.21 | 0.00 | (0.23) | 0.55 | 0.14 | 189 | 1.42 | 0.00 | (0.26) | 0.24 | 0.26 | 186 | 1.05 | 0.00 | (0.33) | 0.28 | 0.23 | 164 |
| Wheat | 0.73 | 0.00 | (0.17) | 0.19 | 0.07 | 134 | 0.82 | 0.00 | (0.23) | 0.45 | 0.12 | 131 | 0.60 | 0.08 | (0.38) | 0.28 | 0.08 | 76 |

### 1.3.3 Testing for Stationarity and Long Memory

The augmented Dicky-Fuller test (hereafter ADF), with the null hypothesis of unit root or $I(1)$, is commonly used in the literature to examine the stationarity of a process under different options. The test is done for the basis and price change in level first[5]. For the basis, running the test without a constant term results in rejection[6] of the null for all commodities in all horizons at the 10% level, except gold in the 12-month horizon. Considering a 5 % level, gold is the only commodity that is non-stationary in the 3- and 12-month horizons. Taking into account other options, namely with intercept only and intercept and trend, the results show rejection of the null for all energy products. In other groups, the rejection happens less frequently, specifically for the base metals and gold. Copper is the one that behaves non-stationary in all horizons. The same is true for aluminum and lead in the 12- and corn in the 3-month horizon. The rejection frequency increases when testing the price change in level. In the 3-month horizon, all commodities show stationarity under all of the options. In the 6-month horizon, gold, corn, and wheat and in the 12-month horizon, copper, gold, silver, corn, and wheat seem to be non-stationary.

For robustness, unit root tests are accompanied by stationarity tests such as Kwiatkowski–Phillips–Schmidt–Shin (hereafter KPSS) test that considers the null hypothesis of $I(0)$. Same as ADF, KPSS is done for data in level first. Looking at the basis, among the energy products, WTI and natural gas reject the null with a constant term in all horizons at the 10% level. For base metals, all except tin show non-stationarity for all horizons considering a constant and a trend option[7]. Gold and silver display consistent rejection for all the horizons given a constant term; and for agricultural products, they seem to be generally stationary. As for the price change, the only regular rejection is for gold given a constant and a trend.

Table (2) summarizes the results of the two tests for variables in level at 5 % significance

[5] The detailed results are reported in the appendix.
[6] With a 10% significance level and less.
[7] The null of $I(0)$ is rejected for tin as well in the 12-month horizon with 10% significance level.

Table 2: Summary of the ADF and KPSS Tests for the Basis and Price Change in Level

The table demonstrates the results of Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests of stationarity for the basis and price change in level. R and CR stand for "reject" and "cannot reject" the tested hypothesis. The first element of (.,.) refers to ADF and the second to KPSS test results. For ADF we have $H_0 : I(1)$ or $H_0$ : the variable has unit root. For KPSS the null hypothesis is $H_0$ :the variable is stationary. The significance level is 5 percent.

| (a) Basis | 3-month horizon | | 6-month horizon | | 12-month horizon | |
|---|---|---|---|---|---|---|
| | Constant | Constant & trend | Constant | Constant & trend | Constant | Constant & trend |
| Energy products | | | | | | |
| WTI | (R, R) | (R, CR) | (R, R) | (R, CR) | (R, CR) | (R, CR) |
| Natural gas | (R, CR) | (R, CR) | (R, R) | (R, CR) | (R, R) | (R, CR) |
| Heating oil | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Gasoline | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Base metals | | | | | | |
| Aluminum | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (CR, CR) | (CR, R) |
| Copper | (CR, CR) | (CR, CR) | (CR, CR) | (CR, R) | (CR, CR) | (CR, R) |
| Lead | (R, CR) | (CR, R) | (R, CR) | (CR, R) | (CR, CR) | (CR, R) |
| Nickel | (R, CR) | (R, CR) | (CR, CR) | (R, R) | (R, CR) | (CR, R) |
| Tin | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Precious metals | | | | | | |
| Gold | (CR, R) | (CR, CR) | (CR, R) | (CR, CR) | (CR, R) | (CR, CR) |
| Silver | (R, R) | (R, CR) | (R, R) | (CR, CR) | (R, R) | (R, CR) |
| Agricultural products | | | | | | |
| Corn | (R, CR) | (CR, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Soybean | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Wheat | (R, CR) | (R, CR) | (R, CR) | (CR, CR) | (R, CR) | (R, CR) |

| (b) Price change | 3-month horizon | | 6-month horizon | | 12-month horizon | |
|---|---|---|---|---|---|---|
| | Constant | Constant & trend | Constant | Constant & trend | Constant | Constant & trend |
| Energy products | | | | | | |
| WTI | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Natural gas | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Heating oil | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Gasoline | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Base metals | | | | | | |
| Aluminum | (R, CR) | (R, CR) | (R, CR) | (CR, CR) | (R, CR) | (R, CR) |
| Copper | (R, CR) | (R, CR) | (R, CR) | (CR, CR) | (R, CR) | (CR, CR) |
| Lead | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (CR, CR) |
| Nickel | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Tin | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Precious metals | | | | | | |
| Gold | (R, CR) | (R, R) | (CR, CR) | (CR, R) | (CR, CR) | (CR, R) |
| Silver | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (CR, CR) | (CR, CR) |
| Agricultural products | | | | | | |
| Corn | (R, CR) | (R, CR) | (CR, CR) | (CR, CR) | (CR, CR) | (CR, CR) |
| Soybean | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (CR, CR) |
| Wheat | (R, CR) | (R, CR) | (CR, CR) | (CR, CR) | (CR, CR) | (CR, CR) |

Table 3: Summary of the ADF and KPSS Tests for the Basis and Price Change in First Difference

The table demonstrates the results of Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests of stationarity for the basis and price change in the first difference. R and CR stand for "reject" and "cannot reject" the tested hypothesis. The first element of (.,.) refers to ADF and the second to KPSS test results. For ADF we have $H_0$ : $I(1)$ or $H_0$ : the variable has unit root. For KPSS the null hypothesis is $H_0$ :the variable is stationary. The significance level is 5 percent. † shows 10 percent significance level.

| (a) Basis | 3-month horizon | | 6-month horizon | | 12-month horizon | |
|---|---|---|---|---|---|---|
| | Constant | Constant & trend | Constant | Constant & trend | Constant | Constant & trend |
| Energy products | | | | | | |
| WTI | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Natural gas | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Heating oil | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Gasoline | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Base metals | | | | | | |
| Aluminum | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Copper | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Lead | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Nickel | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Tin | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Precious metals | | | | | | |
| Gold | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Silver | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R†, CR) | (CR, CR) |
| Agricultural products | | | | | | |
| Corn | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Soybean | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Wheat | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |

| (b) Price change | 3-month horizon | | 6-month horizon | | 12-month horizon | |
|---|---|---|---|---|---|---|
| | Constant | Constant & trend | Constant | Constant & trend | Constant | Constant & trend |
| Energy products | | | | | | |
| WTI | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Natural gas | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Heating oil | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Gasoline | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Base metals | | | | | | |
| Aluminum | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Copper | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Lead | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Nickel | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Tin | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Precious metals | | | | | | |
| Gold | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Silver | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Agricultural products | | | | | | |
| Corn | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Soybean | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Wheat | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |

level. The outcomes of the two tests are presented with a pair. The first element of the pair refers to ADF, while the second corresponds to KPSS. Kwiatkowski et al. (1992) explain the combined results as the following. The pair (R, CR) which states one can reject the ADF and cannot reject the KPSS is an indication of stationarity, whereas (CR, R) refers to the unit root. (CR, CR) is the case that data are not sufficiently informative to differentiate between the two nulls; and for (R, R) there would be no clear conclusion. However, as Coakley et al. (2011) point out, the pair (R, R) could be a symptom of the presence of long memory in the variables. What can be emphasized from the two panels in table (2) is the relative presence of stationarity in the price change compared to the basis. Table (3) presents the outcome of two tests when the data is transformed by the first difference[8]. Both vectors for all commodities and in all horizons show stationarity except silver in 12-month when a constant and trend are considered.

The upshot of table (3) suggests that differencing the data can make them all stationary and suitable for fitting in a vector auto-regressive model. However, since an ARMA representation is one of the underlying assumptions in the ADF test statistic and there exist seemingly conflicting results between the unit root and the stationarity tests, it is imperative to account for more general cases. It should be noted that KPSS allows for the differencing parameter $d$ to be in the interval of $(-\frac{1}{2}, \frac{1}{2})$ and it is often used as a complement to the tests of long memory, but it would be desirable to consider tests that allow $d$ move beyond stationary interval[9]. One of the semi-parametric routines that estimates the long memory and tests unit root and stationarity is Phillips modified Geweke-Porter-Haduk (hereafter GPH) log periodogram regression estimator[10]. The regression is applied to the two vectors in level and a summary[11] of the results are presented in Table (4)[12]. Let us first consider the basis. The choice of

[8]The detailed results are reported in the appendix.

[9]Note that for $d \geq \frac{1}{2}$, the variance of the series is infinite.

[10]The test is a modified type of GPH to account for the unit root as well as testing a consistent estimate of $d$ against fractional alternatives of $d > 1$ and $d < 1$. The advantages of using semiparametric estimates of long memory are good asymptotic properties and high power. (Hauser (1997), Velasco (1999)).

[11]The detailed results are reported in the appendix

[12]The estimation is sensitive to the choice of the number of Fourier frequencies $\upsilon$ considered in the spectral regression. For robustness, two choices of powers of 0.5 and 0.7 are included in the regression where $\upsilon = \sqrt{T}$

the power considerably influences the estimation. It can be seen that generally power of 0.7 increases the point estimate of $d$. Although this is not supported for the base metals in the 3-month horizon. The hypothesis of $I(0)$ cannot be rejected for energy products in the 3- and 6-month horizons considering the power of 0.5. In general, looking at the power of 0.7, (R, R) is relatively more frequent. Looking at the price change, the majority of commodities fit into the stationary long memory, given 0.5 as the power. Note that at the power of 0.7 signs of $I(1)$ start to appear especially in the 12-month horizon.

Agiakloglou et al. (1993), Choi and Zivot (2007), and Coakley et al. (2011) argue that since GPH estimator has a finite sample bias in the presence of auto-regressive terms, they may be misleading. Therefore, the same as Coakley et al. (2011) the parametric ARFIMA representation with at least one auto-regressive term is fitted. The non-linear least square method is chosen since it does not put stationarity constraint on $d$. Results are reported in Table (6)[13]. Considering the ARFIMA estimation, all energy products are $I(0)$ in the 3- and 6-month horizons. Base metals demonstrate almost consistent long memory in all horizons. Gold can be classified as $I(1)$ in the 3- and 12-month horizons and the same can be said about silver in the 6-month horizon. Looking at the results for price change, nearly all estimations of the fractional differencing parameter are in the stationarity interval.

It is important to emphasize that long memory is present especially in the basis, but the fractional differencing parameter can still be found in the stationary interval. This is important since it lets us examine the causal channels in the futures market using the causality measure which is done in the following section.

and $\upsilon = T^{0.7}$.

[13]The estimation is done by OxMetrics 6. (See Doornik, 1999).

Table 4: Summary of the Phillips Modified GPH Test for the Basis and Price Change in Level

Panel (a) and (b) summarize the Modified Geweke-Porter-Hudak (GPH) test results of long memory for the contemporaneous basis and the ex-post price change in power levels of 0.5 and 0.7. R stands for "reject", and CR stands for "cannot reject" the hypothesis. The first element of (.,.) refers to $H_0 : d = 1$, and the second to $H_0 : d = 0$. The significance level is 5 percent.

| (a) Basis | 3-month horizon | | 6-month horizon | | 12-month horizon | |
|---|---|---|---|---|---|---|
| | $\sqrt{T}$ | $T^{0.7}$ | $\sqrt{T}$ | $T^{0.7}$ | $\sqrt{T}$ | $T^{0.7}$ |
| Energy products | | | | | | |
| WTI | (R, R) | (R, R) | (R, R) | (R, R) | (R, R) | (R, R) |
| Natural gas | (R, CR) | (R, CR) | (R, CR) | (R, R) | (R, R) | (R, R) |
| Heating oil | (R, CR) | (R, R) | (R, CR) | (R, R) | (R, CR) | (R, R) |
| Gasoline | (R, CR) | (R, R) | (R, CR) | (R, R) | (R, R) | (R, R) |
| Base metals | | | | | | |
| Aluminum | (R, R) | (R, R) | (R, R) | (R, R) | (CR, R) | (CR, R) |
| Copper | (R, R) | (R, R) | (CR, R) | (R, R) | (CR, R) | (CR, R) |
| Lead | (CR, R) | (R, R) | (CR, R) | (R, R) | (CR, R) | (R, R) |
| Nickel | (R, R) | (R, R) | (R, R) | (CR, R) | (R, R) | (CR, R) |
| Tin | (R, CR) | (R, R) | (R, CR) | (R, R) | (R, R) | (CR, R) |
| Precious metals | | | | | | |
| Gold | (CR, R) | (CR, R) | (CR, R) | (CR, R) | (R, R) | (R, R) |
| Silver | (CR, R) | (R, R) | (CR, R) | (CR, R) | (CR, R) | (R, R) |
| Agricultural products | | | | | | |
| Corn | (R, R) | (R, R) | (R, R) | (R, R) | (R, CR) | (R, R) |
| Soybean | (R, CR) | (R, R) | (R, R) | (R, R) | (R, CR) | (R, R) |
| Wheat | (R, CR) | (R, CR) | (R, CR) | (R, R) | (R, CR) | (R, R) |

| (b) Price change | 3-month horizon | | 6-month horizon | | 12-month horizon | |
|---|---|---|---|---|---|---|
| | $\sqrt{T}$ | $T^{0.7}$ | $\sqrt{T}$ | $T^{0.7}$ | $\sqrt{T}$ | $T^{0.7}$ |
| Energy products | | | | | | |
| WTI | (R, CR) | (R, CR) | (R, CR) | (R,R) | (R, CR) | (CR, R) |
| Natural gas | (R, CR) | (R, CR) | (R, CR) | (R, R) | (R, CR) | (CR, R) |
| Heating oil | (R, CR) | (R, CR) | (R, CR) | (R, R) | (R, CR) | (CR, R) |
| Gasoline | (R, CR) | (R, CR) | (R, CR) | (R, R) | (R, CR) | (R, R) |
| Base metals | | | | | | |
| Aluminum | (R, R) | (R, R) | (R, CR) | (R, R) | (R, CR) | (CR, R) |
| Copper | (R, CR) | (R, CR) | (R, CR) | (R, R) | (R, CR) | (CR, R) |
| Lead | (R, CR) | (R, CR) | (R, CR) | (R, R) | (R, CR) | (CR, R) |
| Nickel | (R, CR) | (R, CR) | (R, CR) | (R, R) | (R, CR) | (CR, R) |
| Tin | (R, CR) | (R, CR) | (R, CR) | (CR, R) | (R, CR) | (R, R) |
| Precious metals | | | | | | |
| Gold | (R, CR) | (R, CR) | (R, CR) | (R, R) | (R, CR) | (CR, R) |
| Silver | (R, CR) | (R, CR) | (R, CR) | (R, R) | (R, CR) | (CR, R) |
| Agricultural products | | | | | | |
| Corn | (R, CR) | (R, CR) | (R, CR) | (CR, R) | (CR, R) | (CR, R) |
| Soybean | (R, CR) | (R, CR) | (R, CR) | (R, R) | (R, R) | (CR, R) |
| Wheat | (R, CR) | (R, CR) | (R, CR) | (CR, R) | (CR, R) | (CR, R) |

Table 5: Summary of the Phillips Modified GPH Test for the Basis and Price Change in First Difference

Panel (a) and (b) summarize the Modified Geweke-Porter-Hudak (GPH) test results of long memory for the contemporaneous basis and the ex-post price change in power levels of 0.5 and 0.7. R stands for "reject", and CR stands for "cannot reject" the hypothesis. The first element of (.,.) refers to $H_0 : d = 1$, and the second to $H_0 : d = 0$. The significance level is 5 percent.

| (a) Basis | 3-month horizon | | 6-month horizon | | 12-month horizon | |
|---|---|---|---|---|---|---|
| | $\sqrt{T}$ | $T^{0.7}$ | $\sqrt{T}$ | $T^{0.7}$ | $\sqrt{T}$ | $T^{0.7}$ |
| Energy products | | | | | | |
| WTI | (R, CR) | (R, CR) | (R ,CR) | (R, CR) | (R, CR) | (R, CR) |
| Natural gas | (R, CR) | (R, R) | (R ,CR) | (R, R) | (R, CR) | (R, R) |
| Heating oil | (R, R) | (R, R) | (R, CR) | (R, R) | (R, CR) | (R, CR) |
| Gasoline | (CR, R) | (R, R) | (CR, R) | (R, R) | (R, R) | (R, R) |
| Base metals | | | | | | |
| Aluminum | (R, CR) | (R, R) | (R, CR) | (R, R) | (R, CR) | (R, CR) |
| Copper | (R, CR) | (R, R) | (R, CR) | (R, R) | (R, CR) | (R, CR) |
| Lead | (R, CR) | (R, R) | (R, CR) | (R, R) | (R, CR) | (R, R) |
| Nickel | (R, R) | (R, R) | (R, R) | (R, R) | (R, R) | (R, CR) |
| Tin | (R, CR) | (R, R) | (R, CR) | (R, R) | (R, CR) | (R, CR) |
| Precious metals | | | | | | |
| Gold | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Silver | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Agricultural products | | | | | | |
| Corn | (R, CR) | (R, R) | (R, CR) | (R, R) | (R, CR) | (R, CR) |
| Soybean | (R, CR) | (R, R) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Wheat | (R, R) | (R, R) | (R, CR) | (R, CR) | (R, CR) | (R, R) |

| (b) Price change | 3-month horizon | | 6-month horizon | | 12-month horizon | |
|---|---|---|---|---|---|---|
| | $\sqrt{T}$ | $T^{0.7}$ | $\sqrt{T}$ | $T^{0.7}$ | $\sqrt{T}$ | $T^{0.7}$ |
| Energy products | | | | | | |
| WTI | (R, CR) | (R, R) | (R, CR) | (R, CR) | (R, CR) | (R, R) |
| Natural gas | (R, R) | (R, CR) | (R, CR) | (R, CR) | (R, R) | (R, CR) |
| Heating oil | (R, CR) | (R, R) | (R, CR) | (R, CR) | (R, CR) | (R, R) |
| Gasoline | (CR, R) | (R, CR) | (R, R) | (R, CR) | (R, R) | (R, CR) |
| Base metals | | | | | | |
| Aluminum | (R, R) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, R) |
| Copper | (R, CR) | (R, R) | (R, R) | (R, R) | (R, CR) | (R, R) |
| Lead | (R, R) | (R, R) | (R, R) | (R, R) | (R, R) | (R, CR) |
| Nickel | (CR, R) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, R) |
| Tin | (R, CR) | (R, R) | (R, CR) | (R, CR) | (R, CR) | (R, R) |
| Precious metals | | | | | | |
| Gold | (R, CR) | (R, R) | (R, R) | (R, R) | (R, CR) | (R, CR) |
| Silver | (R, R) | (R, CR) | (R, CR) | (R, R) | (R, CR) | (R, CR) |
| Agricultural products | | | | | | |
| Corn | (R, CR) | (R, R) | (R, CR) | (R, CR) | (CR, CR) | (R, CR) |
| Soybean | (CR, R) | (R, CR) | (R, CR) | (R, CR) | (R, CR) | (R, CR) |
| Wheat | (R, CR) | (R, R) | (R, R) | (R, CR) | (R, CR) | (R, CR) |

Table 6: Non-linear Least Square Estimation of ARFIMA (p,d,q) for the Contemporaneous Basis and the Ex-post Price Change in Level

| (a) Basis | 3-month horizon | | | | | 6-month horizon | | | | | 12-month horizon | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $ARMA$ | $d$ | $SE$ | $t$ | $\tau$ | $ARMA$ | $d$ | $SE$ | $t$ | $\tau$ | $ARMA$ | $d$ | $SE$ | $t$ | $\tau$ |
| Energy products | | | | | | | | | | | | | | | |
| WTI | (1,0) | -0.06 | 0.11 | -0.57 | -8.91 | (1,0) | 0.07 | 0.13 | 0.54 | -7.05 | (1,0) | 0.11 | 0.12 | 0.88 | -7.05 |
| Natural gas | (1,0) | 0.14 | 0.09 | 1.52 | -9.29 | (1,1) | -0.11 | 0.14 | -0.75 | -7.92 | (1,0) | 0.28 | 0.11 | 2.44 | -6.14 |
| Heating oil | (1,1) | -0.21 | 0.19 | -1.07 | -6.36 | (1,0) | 0.06 | 0.10 | 0.65 | -9.24 | (1,0) | 0.76 | 0.08 | 8.95 | -2.69 |
| Gasoline | (1,0) | -0.08 | 0.11 | -0.76 | -9.25 | (1,0) | 0.11 | 0.09 | 1.26 | -9.26 | (1,0) | 0.55 | 0.08 | 6.28 | -5.03 |
| Base metals | | | | | | | | | | | | | | | |
| Aluminum | (1,0) | 0.23 | 0.11 | 2.11 | -6.97 | (1,0) | 0.45 | 0.13 | 3.29 | -3.88 | (1,0) | 0.69 | 0.10 | 6.61 | -2.95 |
| Copper | (1,0) | 0.37 | 0.07 | 5.15 | -8.68 | (1,0) | 0.61 | 0.06 | 9.33 | -5.72 | (1,0) | 0.83 | 0.06 | 12.3 | -2.74 |
| Lead | (1,0) | 0.36 | 0.08 | 4.44 | -7.71 | (1,0) | 0.57 | 0.08 | 6.58 | -4.83 | (1,0) | 0.69 | 0.08 | 8.06 | -3.60 |
| Nickel | (1,0) | 0.41 | 0.15 | 2.77 | -3.87 | (1,0) | 0.41 | 0.25 | 1.64 | -2.29 | (1,0) | 0.14 | 0.12 | 1.22 | -7.06 |
| Tin | (2,1) | -0.46 | 0.13 | -3.34 | -10.54 | (1,2) | -0.17 | 0.28 | -0.61 | -4.10 | (1,2) | -0.32 | 0.13 | -2.34 | -9.58 |
| Precious metals | | | | | | | | | | | | | | | |
| Gold | (1,0) | 1.18 | 0.08 | 14.3 | -2.19 | (1,0) | 0.05 | 0.09 | 0.54 | -17.60 | (1,0) | 0.85 | 0.07 | 12.0 | -1.96 |
| Silver | (1,0) | 0.74 | 0.08 | 8.59 | -2.96 | (1,0) | 0.89 | 0.13 | 6.58 | -0.84 | (1,0) | 0.66 | 0.06 | 10.6 | -5.41 |
| Agricultural products | | | | | | | | | | | | | | | |
| Corn | (1,0) | 0.18 | 0.22 | 0.85 | -3.68 | (1,0) | -0.08 | 0.26 | -0.32 | -4.15 | (1,0) | -0.05 | 0.38 | -0.15 | -2.76 |
| Soybean | (1,0) | 0.24 | 0.16 | 1.44 | -4.54 | (1,0) | 0.20 | 0.26 | 0.76 | -3.07 | (1,0) | 0.32 | 0.20 | 1.58 | -3.25 |
| Wheat | (1,0) | 0.21 | 0.15 | 1.45 | -5.16 | (1,0) | 0.00 | 0.18 | 0.02 | -5.31 | (1,0) | 0.20 | 0.46 | 0.44 | -1.73 |

| (b) Price change | 3-month horizon | | | | | 6-month horizon | | | | | 12-month horizon | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $ARMA$ | $d$ | $SE$ | $t$ | $\tau$ | $ARMA$ | $d$ | $SE$ | $t$ | $\tau$ | $ARMA$ | $d$ | $SE$ | $t$ | $\tau$ |
| Energy products | | | | | | | | | | | | | | | |
| WTI | (1,1) | -0.03 | 0.14 | -0.22 | -7.38 | (1,1) | 0.32 | 0.16 | 1.99 | -4.20 | (1,0) | 0.28 | 0.09 | 3.06 | -7.78 |
| Natural gas | (1,0) | -0.24 | 0.12 | -1.93 | -9.66 | (1,1) | -0.19 | 0.14 | -1.32 | -8.17 | (1,1) | -0.01 | 0.15 | -0.11 | -6.50 |
| Heating oil | (1,1) | -0.05 | 0.11 | -0.45 | -8.85 | (1,0) | 0.13 | 0.09 | 1.51 | -9.53 | (1,0) | 0.10 | 0.07 | 1.35 | -11.35 |
| Gasoline | (1,0) | -0.13 | 0.09 | -1.36 | -11.47 | (1,0) | 0.19 | 0.09 | 2.09 | -8.82 | (1,1) | -0.08 | 0.11 | -0.70 | -9.25 |
| Base metals | | | | | | | | | | | | | | | |
| Aluminum | (1,0) | -0.15 | 0.15 | -0.98 | -7.28 | (1,1) | 0.01 | 0.14 | 0.09 | -6.59 | (1,0) | 0.36 | 0.16 | 2.19 | -3.77 |
| Copper | (1,1) | 0.31 | 0.09 | 3.32 | -7.36 | (1,1) | 0.35 | 0.13 | 2.71 | -4.94 | (1,1) | 0.50 | 0.14 | 3.41 | -3.33 |
| Lead | (1,0) | -0.01 | 0.13 | -0.13 | -7.39 | (2,1) | -0.48 | 0.26 | -1.84 | -5.65 | (1,0) | 0.23 | 0.08 | 2.73 | -9.00 |
| Nickel | (1,1) | 0.06 | 0.10 | 0.57 | -8.75 | (1,1) | 0.13 | 0.17 | 0.76 | -4.88 | (1,1) | 0.41 | 0.16 | 2.43 | -3.48 |
| Tin | (1,0) | -0.03 | 0.15 | -0.21 | -6.84 | (1,0) | 0.22 | 0.10 | 2.18 | 0.03 | (1,0) | 0.18 | 0.07 | 2.57 | -11.33 |
| Precious metals | | | | | | | | | | | | | | | |
| Gold | (1,1) | 0.02 | 0.06 | 0.41 | -14.53 | (2,1) | 0.11 | 0.14 | 0.78 | -6.15 | (1,0) | 0.82 | 0.11 | 6.91 | -1.43 |
| Silver | (1,0) | -0.13 | 0.10 | -1.20 | -10.33 | (1,1) | -0.14 | 0.15 | -0.94 | -7.22 | (1,1) | 0.29 | 0.17 | 1.70 | -4.06 |
| Agricultural products | | | | | | | | | | | | | | | |
| Corn | (1,1) | -0.37 | 0.19 | -1.89 | -12.04 | (1,1) | -0.17 | 0.20 | -0.85 | -5.79 | (1,0) | 0.29 | 0.20 | 1.42 | -3.34 |
| Soybean | (1,0) | 0.00 | 0.16 | 0.01 | -6.00 | (1,0) | 0.19 | 0.17 | 1.09 | -4.62 | (1,0) | 0.14 | 0.15 | 0.94 | -5.56 |
| Wheat | (1,1) | -0.04 | 0.14 | -0.30 | -7.38 | (1,1) | -0.07 | 0.22 | -0.31 | -4.80 | (1,0) | 0.07 | 0.14 | 0.55 | -6.58 |

(i) Panel (a) and (b) show the $ARFIMA(p,d,q)$ estimation results of the contemporaneous basis and the ex-post price change respectively.
(ii) $ARMA$ indicates the order of the short memory.
(iii) $d$ corresponds to the estimation of the fractional differencing parameter by non-linear least square method.
(iv) $SE$ is the standard error for $d$.
(v) $t$ represents the test statistic $\frac{(d-0)}{SE}$.
(vi) $\tau$ shows the test statistic $\frac{(d-1)}{SE}$.
(vii) 1%, 5%, and 10% critical values are $\pm 2.57$, $\pm 1.96$, and $\pm 1.64$.

### 1.3.4 Causality Measure Estimation

Table (7) presents the Granger causality analysis done on the relationship between the basis and the price change. Since the predictability of the basis is the subject of our interest, the direction of causality is set from the basis to the price change (shown by $b \to p$). As it was discussed in the previous section, the basis of some commodities behaves non-stationary in some horizons. This can be taken care of by differencing the data in the first order. It is shown in the table when the value of $L$ is equal to 1. A $VAR(k)$ model is estimated by OLS with $k$ being identified by the BIC. After fitting the $VAR$ model, the Granger causality test is done by inspecting if the past values of the basis have useful information for predicting the price change. $Wald$ presents the p-value of the Granger causality test with the null hypothesis that the basis does not Granger-cause the price change. The null is rejected for all energy products in all the horizons and for all agricultural products in the 3- and 6-month horizons at 5 % level. For the base and precious metals, the null is not rejected except for copper in the 12-month and silver in the 6-month horizon.

$\mathbb{C}$ is the causality measure that is measured using equation (33) which captures the information in the basis. Across commodities, base and precious metals perform poorly in this respect, whereas energy commodities have better characteristics when it comes to unbiasedness theory. Across the time horizon, for energy products, the point estimate of the measure appears to be generally increasing as the horizon expands. For natural gas, for instance, $\mathbb{C}$ rises from 0.148 to 0.393 going from the 3- to 12-month horizons. The fact that can be seen in the confidence intervals as well. In the agricultural products, a similar pattern can be observed in the 3- and 6-month horizons. All the agricultural basis do not show signs of information in the 12-month horizon, but jumping from 3- to 6-month increases the predictability of the basis considerably.

The overall results indicate that the energy market possesses the most informative basis. And across this group of commodities, natural gas, heating oil, and gasoline appear to have more predictive basis than crude oil. Yet the oil basis as the least informative one among this group, performs better than the base and precious metals combined. Agricultural goods can be classified in between of the two.

A few points need to be emphasized though. The period of this study is not the same for all commodities. For the base metals, observations are available from 1997 only, whereas in the energy markets, the futures prices are consistently available from 1990. The other issue is the heterogeneity over the markets in terms of market structure and market participants. Gold is known as a popular hedge against inflation. When CPI is expected to soar, managers keep their assets in gold to avoid the loss. Even though traditionally gold is used as an inflation hedge, recent studies prove that copper could be a better option[14]. The red metal is sensitive to macroeconomic shocks and historically, in times of inflation, its value increased three times more than gold. Thus, the importance of macroeconomic shocks should be considered when one studies the behavior of futures price.

A possible extension to this paper is considering the indirect causality. In the setting of the commodity futures market, volume seems to be a proper candidate. Volume can be linked to liquidity in commodity futures. Traders prefer commodities with higher liquidity. Therefore, establishing an indirect link between the basis and the volume in one hand and the volume and the price, on the other, seems to be informative regarding the efficiency and unbiasedness of the markets. Another interesting area to discover is examining the causality measures jointly. Since some markets are closely related such as the gold and oil markets, conducting joint inference on the true causality measures seems to be enlightening.

[14] More Precious Than Gold? Copper's the Better Inflation Hedge. Susanne Barton. *Bloomberg*, June 2017.

Table 7: The Granger Causality Analysis

$L$ denotes the order of differencing. $Wald$ represents the p-value of $H_0$: No Granger causality from the contemporaneous basis to the ex-post price change. $\mathbb{C}$ indicates the causality measure which captures the informational content of the futures market. *Bootstrap confidence interval* is constructed at 95 percent level.

| | 3-month horizon | | | | 6-month horizon | | | | 12-month horizon | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $L$ | $Wald$ | $\widehat{C}$ | *Bootstrap confidence interval* | $L$ | $Wald$ | $\widehat{C}$ | *Bootstrap confidence interval* | $L$ | $Wald$ | $\widehat{C}$ | *Bootstrap confidence interval* |
| Energy products | | | | | | | | | | | | |
| WTI | 0 | 0.00 | 0.045 | [0.013, 0.134] | 0 | 0.00 | 0.247 | [0.186, 0.425] | 1 | 0.00 | 0.243 | [0.194, 0.442] |
| Natural gas | 0 | 0.00 | 0.148 | [0.122, 0.334] | 0 | 0.00 | 0.202 | [0.170, 0.398] | 1 | 0.00 | 0.393 | [0.327, 0.621] |
| Heating oil | 0 | 0.00 | 0.086 | [0.052, 0.212] | 0 | 0.00 | 0.104 | [0.096, 0.314] | 1 | 0.00 | 0.226 | [0.197, 0.466] |
| Gasoline | 0 | 0.00 | 0.074 | [0.035, 0.182] | 0 | 0.00 | 0.187 | [0.157, 0.410] | 1 | 0.00 | 0.202 | [0.174, 0.472] |
| Base metals | | | | | | | | | | | | |
| Aluminum | 1 | 0.23 | 0.001 | [0.000, 0.063] | 1 | 0.03 | 0.030 | [0.000, 0.140] | 1 | 0.84 | 0.000 | [0.000, 0.033] |
| Copper | 1 | 0.63 | 0.000 | [0.000, 0.069] | 1 | 0.12 | 0.019 | [0.000, 0.127] | 1 | 0.00 | 0.049 | [0.004, 0.158] |
| Lead | 1 | 0.13 | 0.009 | [0.000, 0.080] | 1 | 0.30 | 0.000 | [0.000, 0.045] | 1 | 0.68 | 0.000 | [0.000, 0.021] |
| Nickel | 1 | 0.06 | 0.011 | [0.000, 0.086] | 1 | 0.46 | 0.000 | [0.000, 0.028] | 1 | 0.73 | 0.000 | [0.000, 0.023] |
| Tin | 1 | 0.46 | 0.000 | [0.000, 0.079] | 1 | 0.40 | 0.002 | [0.000, 0.106] | 1 | 0.20 | 0.000 | [0.000, 0.049] |
| Precious metals | | | | | | | | | | | | |
| Gold | 1 | 0.77 | 0.000 | [0.000, 0.048] | 1 | 0.86 | 0.003 | [0.000, 0.061] | 1 | 0.93 | 0.003 | [0.001, 0.025] |
| Silver | 1 | 0.13 | 0.008 | [0.000, 0.073] | 1 | 0.04 | 0.021 | [0.000, 0.118] | 1 | 0.98 | 0.000 | [0.000, 0.016] |
| Agricultural products | | | | | | | | | | | | |
| Corn | 1 | 0.00 | 0.080 | [0.010, 0.309] | 1 | 0.00 | 0.204 | [0.114, 0.543] | 1 | 0.70 | 0.000 | [0.000, 0.079] |
| Soybean | 0 | 0.01 | 0.026 | [0.000, 0.142] | 0 | 0.00 | 0.073 | [0.029, 0.297] | 0 | 0.31 | 0.000 | [0.000, 0.075] |
| Wheat | 0 | 0.04 | 0.069 | [0.007, 0.308] | 0 | 0.00 | 0.185 | [0.119, 0.553] | 0 | 0.93 | 0.000 | [0.000, 0.099] |

## 1.4 *Concluding Remarks*

The commodity market has been the subject of ample research recently and they are gaining increasing importance in alternative investments due to their unique characteristics. In this paper, a new approach for investigating the informational content of commodity futures in the context of the causal relationship between the basis and the price change is suggested. Quantifying the causality from the basis towards the price change in the market can reveal the degree of unbiasedness in the futures. This has been done on a set of fourteen commodities and results indicate that energy and agricultural products seem to have information and therefore unbiased properties in their basis in terms of predicting the price change, whereas the base and precious metals seem to fail in this regard.

Yet another direction worthy of research scrutiny is the interconnection between the commodities. For instance, the gold market contains information that could help predict the price change in energy futures in general and crude oil in particular. This can be done by defining an indirect causality from the basis in gold futures to the price change in the oil market through the oil basis. Due to the linkage of these two markets, there might exist significant causality through which one basis in a market causes the price change in another and vice versa. This is currently under investigation by the author.

## 1.5 *Appendix*

In the appendix, Table (8) presents the mnemonics for the futures which includes the market, ticker, and time period. Table (9) shows the summary statistics of futures contracts which could be interesting for studying the price dynamics for each commodity. Table (10) demonstrates the same stats for the basis and the price change. Tables (11) and (12) exhibit Phillips modified GPH estimation results for the basis in level and first difference, respectively. Table (13) and (14) present the same outcomes for the price change. KPSS test results of stationarity are reported in Table (15) and (16) for the variables in level and first difference, respectively. Finally, Tables (17) and (18) display the results of the unit root tests. The appendix ends with the algorithm to obtain the bootstrap confidence intervals in Dufour and Taamouti (2010).

Table 8: Bloomberg Mnemonics for Futures Prices and Available Samples

| | | | Available sample | | | |
|---|---|---|---|---|---|---|
| | Market | Futures ticker | 1-month | 3-month | 6-month | 12-month |
| Energy products | | | | | | |
| WTI | NYMEX | CL | 1990:1-2017:3 | 1990:1-2017:3 | 1990:1-2017:3 | 1990:1-2017:3 |
| Natural gas | NYMEX | NG | 1990:5-2017:3 | 1990:5-2017:3 | 1990:5-2017:3 | 1990:7-2017:3 |
| Heating oil | NYMEX | HO | 1990:1-2017:3 | 1990:1-2017:3 | 1990:1-2017:3 | 1990:1-2017:3 |
| Gasoline | NYMEX | HU/XB | 1990:1-2017:3 | 1990:1-2017:3 | 1990:1-2017:3 | 1990:1-2017:3 |
| Base metals | | | | | | |
| Aluminum | CME | LA | 1997:7-2017:3 | 1997:7-2017:3 | 1997:7-2017:3 | 1997:7-2017:3 |
| Copper | CME | LP | 1997:7-2017:3 | 1997:7-2017:3 | 1997:7-2017:3 | 1997:7-2017:3 |
| Lead | CME | LL | 1997:7-2017:3 | 1997:7-2017:3 | 1997:7-2017:3 | 1997:7-2017:3 |
| Nickel | CME | LN | 1997:7-2017:3 | 1997:7-2017:3 | 1997:7-2017:3 | 1997:7-2017:3 |
| Tin | CME | LT | 1997:7-2017:3 | 1997:7-2017:3 | 1997:7-2017:3 | 1997:11-2017:3 |
| Precious metals | | | | | | |
| Gold | NYMEX | GC | 1990:1-2017:3 | 1990:1-2017:3 | 1990:1-2017:3 | 1990:3-2017:3 |
| Silver | NYMEX | SI | 1990:1-2017:3 | 1990:1-2017:3 | 1990:1-2017:3 | 1990:3-2015:3 |
| Agricultural products | | | | | | |
| Corn | LME | C | 1990:1-2017:3 | 1990:1-2017:3 | 1990:1-2017:3 | 1994:7-2016:5 |
| Soybean | LME | S | 1990:1-2017:3 | 1990:1-2017:3 | 1990:1-2017:3 | 1992:5-2017:3 |
| Wheat | LME | W | 1990:1-2017:3 | 1990:1-2017:3 | 1990:1-2017:1 | 2000:1-2016:11 |

Table 9: Summary Statistics of Futures Contracts for Multiple Horizons

| | 3-month horizon | | | | | 6-month horizon | | | | | 12-month horizon | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $N$ | $\mu$ | $\sigma$ | $\kappa_3$ | $\kappa_4$ | $N$ | $\mu$ | $\sigma$ | $\kappa_3$ | $\kappa_4$ | $N$ | $\mu$ | $\sigma$ | $\kappa_3$ | $\kappa_4$ |
| Energy products | | | | | | | | | | | | | | | |
| WTI | 328 | 47.21 | 30.82 | 0.73 | 2.24 | 328 | 47.22 | 31.03 | 0.70 | 2.16 | 328 | 46.94 | 31.07 | 0.66 | 2.06 |
| Natural gas | 325 | 4.02 | 2.41 | 1.51 | 5.48 | 325 | 4.11 | 2.45 | 1.34 | 4.50 | 323 | 4.16 | 2.40 | 1.15 | 3.60 |
| Heating oil | 328 | 136.79 | 92.87 | 0.76 | 2.24 | 328 | 137.33 | 93.76 | 0.74 | 2.21 | 328 | 137.31 | 94.16 | 0.68 | 2.06 |
| Gasoline | 327 | 134.94 | 85.14 | 0.72 | 2.19 | 327 | 133.54 | 83.92 | 0.66 | 2.03 | 327 | 131.64 | 83.84 | 0.65 | 2.00 |
| Base metals | | | | | | | | | | | | | | | |
| Aluminum | 238 | 1869.51 | 457.96 | 0.81 | 2.72 | 238 | 1881.35 | 459.53 | 0.81 | 2.72 | 238 | 1894.98 | 454.65 | 0.80 | 2.71 |
| Copper | 237 | 4761.72 | 2634.72 | 0.10 | 1.52 | 237 | 4741.33 | 2627.98 | 0.11 | 1.52 | 237 | 4690.57 | 2603.01 | 0.13 | 1.51 |
| Lead | 238 | 1425.31 | 809.49 | 0.23 | 1.86 | 238 | 1423.52 | 809.36 | 0.22 | 1.80 | 235 | 1426.83 | 803.92 | 0.18 | 1.68 |
| Nickel | 238 | 14580.69 | 8381.78 | 1.28 | 4.99 | 238 | 14449.08 | 8152.20 | 1.18 | 4.52 | 238 | 14132.76 | 7710.24 | 0.99 | 3.72 |
| Tin | 238 | 12949.25 | 7479.48 | 0.40 | 1.94 | 238 | 12916.11 | 7473.56 | 0.41 | 1.94 | 235 | 12949.01 | 7461.34 | 0.41 | 1.93 |
| Precious metals | | | | | | | | | | | | | | | |
| Gold | 328 | 690.41 | 452.69 | 0.88 | 2.34 | 328 | 697.40 | 452.10 | 0.88 | 2.34 | 327 | 713.38 | 452.98 | 0.86 | 2.32 |
| Silver | 328 | 11.24 | 8.80 | 1.46 | 4.72 | 328 | 11.35 | 8.80 | 1.45 | 4.67 | 301 | 10.77 | 8.53 | 1.66 | 5.39 |
| Agricultural products | | | | | | | | | | | | | | | |
| Corn | 328 | 341.97 | 139.41 | 1.44 | 4.36 | 328 | 346.82 | 125.47 | 1.21 | 3.52 | 264 | 355.84 | 123.15 | 0.79 | 2.23 |
| Soybean | 328 | 805.67 | 301.93 | 0.92 | 2.78 | 328 | 793.67 | 278.86 | 0.88 | 2.64 | 301 | 790.16 | 264.00 | 0.66 | 2.14 |
| Wheat | 328 | 456.31 | 174.66 | 1.07 | 3.14 | 325 | 466.93 | 175.00 | 1.08 | 3.12 | 205 | 541.21 | 202.68 | 0.32 | 1.67 |

$N$, $\mu$, $\sigma$, $\kappa_3$, and $\kappa_4$ represent the number of observations, mean, standard deviation, skewness, and kurtosis, respectively.

Table 10: Summary Statistics of the Basis and Price Change

| (a) Basis | 3-month horizon | | | | | 6-month horizon | | | | | 12-month horizon | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $N$ | $\mu$ | $\sigma$ | $\kappa_3$ | $\kappa_4$ | $N$ | $\mu$ | $\sigma$ | $\kappa_3$ | $\kappa_4$ | $N$ | $\mu$ | $\sigma$ | $\kappa_3$ | $\kappa_4$ |
| Energy products | | | | | | | | | | | | | | | |
| WTI | 328 | 0.00 | 0.01 | 0.00 | 4.83 | 328 | 0.00 | 0.02 | -0.14 | 3.60 | 328 | 0.00 | 0.04 | -0.17 | 3.25 |
| Natural gas | 325 | 0.01 | 0.04 | 0.02 | 8.91 | 325 | 0.02 | 0.07 | -0.38 | 5.07 | 323 | 0.02 | 0.07 | -0.04 | 3.54 |
| Heating oil | 328 | 0.00 | 0.02 | -2.82 | 14.07 | 328 | 0.00 | 0.03 | -1.65 | 7.40 | 328 | 0.00 | 0.04 | -0.55 | 4.09 |
| Gasoline | 327 | -0.00 | 0.02 | 0.03 | 4.78 | 327 | -0.00 | 0.04 | -0.00 | 2.95 | 327 | -0.01 | 0.05 | -0.08 | 3.76 |
| Base metals | | | | | | | | | | | | | | | |
| Aluminum | 238 | 0.00 | 0.00 | 0.95 | 11.85 | 238 | 0.00 | 0.00 | -0.66 | 3.39 | 238 | 0.00 | 0.01 | -0.65 | 3.25 |
| Copper | 237 | 0.00 | 0.01 | 0.14 | 13.61 | 237 | 0.00 | 0.01 | -0.37 | 6.47 | 237 | 0.00 | 0.02 | -1.05 | 4.36 |
| Lead | 238 | 0.00 | 0.00 | 0.03 | 7.54 | 238 | 0.00 | 0.01 | -1.17 | 5.48 | 235 | 0.00 | 0.02 | -1.40 | 5.77 |
| Nickel | 238 | 0.00 | 0.00 | -0.85 | 10.10 | 238 | 0.00 | 0.01 | -1.87 | 6.86 | 238 | -0.01 | 0.02 | -1.46 | 4.51 |
| Tin | 238 | 0.00 | 0.00 | -2.16 | 10.28 | 238 | 0.00 | 0.00 | -2.56 | 14.71 | 235 | 0.00 | 0.01 | -1.94 | 8.91 |
| Precious metals | | | | | | | | | | | | | | | |
| Gold | 328 | 0.00 | 0.00 | 0.34 | 2.07 | 328 | 0.01 | 0.00 | 0.31 | 1.99 | 327 | 0.02 | 0.01 | 0.25 | 2.20 |
| Silver | 328 | 0.00 | 0.00 | 0.77 | 3.25 | 328 | 0.01 | 0.00 | 0.67 | 2.76 | 301 | 0.01 | 0.03 | -2.14 | 10.41 |
| Agricultural products | | | | | | | | | | | | | | | |
| Corn | 136 | 0.01 | 0.02 | -2.51 | 11.69 | 136 | 0.02 | 0.04 | -1.10 | 4.08 | 110 | 0.02 | 0.07 | -0.86 | 3.65 |
| Soybean | 191 | 0.00 | 0.01 | -2.01 | 7.76 | 191 | 0.00 | 0.03 | -1.29 | 4.67 | 175 | -0.01 | 0.05 | -0.81 | 4.66 |
| Wheat | 136 | 0.01 | 0.02 | -1.44 | 5.14 | 135 | 0.02 | 0.04 | -0.53 | 2.86 | 86 | 0.03 | 0.07 | -0.71 | 4.51 |

| (b) Price change | 3-month horizon | | | | | 6-month horizon | | | | | 12-month horizon | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $N$ | $\mu$ | $\sigma$ | $\kappa_3$ | $\kappa_4$ | $N$ | $\mu$ | $\sigma$ | $\kappa_3$ | $\kappa_4$ | $N$ | $\mu$ | $\sigma$ | $\kappa_3$ | $\kappa_4$ |
| Energy products | | | | | | | | | | | | | | | |
| WTI | 326 | 0.00 | 0.06 | -0.28 | 5.66 | 323 | 0.00 | 0.10 | -0.99 | 6.47 | 317 | 0.01 | 0.13 | -0.45 | 3.52 |
| Natural gas | 323 | 0.00 | 0.09 | 0.07 | 3.39 | 320 | 0.00 | 0.12 | -0.19 | 3.10 | 314 | 0.00 | 0.16 | -0.44 | 3.99 |
| Heating oil | 326 | 0.00 | 0.06 | -0.14 | 4.83 | 323 | 0.00 | 0.09 | -0.71 | 4.85 | 317 | 0.01 | 0.13 | -0.26 | 3.45 |
| Gasoline | 325 | 0.00 | 0.07 | -0.56 | 5.36 | 322 | 0.00 | 0.10 | -0.84 | 5.63 | 316 | 0.01 | 0.12 | -0.25 | 3.14 |
| Base metals | | | | | | | | | | | | | | | |
| Aluminum | 236 | 0.00 | 0.03 | -0.45 | 5.24 | 233 | 0.02 | 0.11 | 0.18 | 2.70 | 227 | 0.00 | 0.09 | -0.97 | 5.85 |
| Copper | 237 | 0.00 | 0.05 | -0.88 | 9.06 | 234 | 0.00 | 0.09 | -0.97 | 7.89 | 228 | 0.02 | 0.12 | -0.04 | 4.38 |
| Lead | 236 | 0.00 | 0.05 | -0.68 | 5.28 | 233 | 0.01 | 0.08 | 0.05 | 4.66 | 227 | 0.02 | 0.13 | -0.12 | 4.92 |
| Nickel | 236 | 0.00 | 0.06 | -0.16 | 3.48 | 233 | 0.00 | 0.11 | 0.08 | 3.00 | 227 | 0.01 | 0.17 | 0.07 | 2.84 |
| Tin | 236 | 0.00 | 0.04 | -0.26 | 4.27 | 233 | 0.01 | 0.08 | -0.54 | 4.42 | 227 | 0.02 | 0.12 | 0.04 | 2.54 |
| Precious metals | | | | | | | | | | | | | | | |
| Gold | 326 | 0.00 | 0.02 | 0.34 | 3.67 | 323 | 0.00 | 0.03 | 0.36 | 3.76 | 317 | 0.01 | 0.05 | 0.23 | 2.94 |
| Silver | 326 | 0.00 | 0.04 | -0.06 | 3.88 | 323 | 0.00 | 0.07 | 0.09 | 4.45 | 317 | 0.01 | 0.10 | 0.57 | 3.70 |
| Agricultural products | | | | | | | | | | | | | | | |
| Corn | 134 | 0.00 | 0.08 | 0.00 | 3.09 | 131 | 0.00 | 0.11 | 0.32 | 3.04 | 125 | 0.01 | 0.15 | 0.27 | 3.06 |
| Soybean | 189 | 0.00 | 0.05 | -0.47 | 4.03 | 186 | 0.00 | 0.08 | 0.06 | 3.36 | 180 | 0.01 | 0.11 | 0.35 | 3.47 |
| Wheat | 134 | 0.00 | 0.07 | 0.60 | 3.80 | 131 | 0.00 | 0.10 | 0.53 | 3.17 | 125 | 0.01 | 0.15 | 0.30 | 3.32 |

(i) Panel (a) and (b) present summary statistics for the basis and the price change, respectively.
(ii) $N$, $\mu$, $\sigma$, $\kappa_3$, and $\kappa_4$ represent the number of observations, mean, standard deviation, skewness, and kurtosis, respectively.

Table 11: Phillips Modified Geweke-Porter-Haduk (GPH) Regression Estimation Results for the Basis in Level

| (a) Power = 0.5 | 3-month horizon | | | | | | 6-month horizon | | | | | | 12-month horizon | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $d$ | $SE$ | $t$ | $p_1$ | $z$ | $p_2$ | $d$ | $SE$ | $t$ | $p_1$ | $z$ | $p_2$ | $d$ | $SE$ | $t$ | $p_1$ | $z$ | $p_2$ |
| Energy products | | | | | | | | | | | | | | | | | | |
| WTI | 0.39 | 0.16 | 2.42 | 0.02 | -4.01 | 0.00 | 0.50 | 0.19 | 2.55 | 0.02 | -3.29 | 0.00 | 0.55 | 0.23 | 2.31 | 0.03 | -2.95 | 0.00 |
| Natural gas | 0.15 | 0.23 | 0.63 | 0.53 | -5.61 | 0.00 | 0.22 | 0.19 | 1.15 | 0.26 | -5.10 | 0.00 | 0.41 | 0.18 | 2.27 | 0.03 | -3.74 | 0.00 |
| Heating oil | 0.13 | 0.18 | 0.76 | 0.45 | -5.70 | 0.00 | 0.35 | 0.27 | 1.29 | 0.21 | -4.25 | 0.00 | 0.59 | 0.30 | 1.93 | 0.06 | -2.68 | 0.00 |
| Gasoline | 0.17 | 0.21 | 0.83 | 0.41 | -5.43 | 0.00 | 0.42 | 0.23 | 1.80 | 0.08 | -3.78 | 0.00 | 0.67 | 0.21 | 3.16 | 0.00 | -2.11 | 0.03 |
| Base metals | | | | | | | | | | | | | | | | | | |
| Aluminum | 0.61 | 0.24 | 2.47 | 0.02 | -2.35 | 0.01 | 0.63 | 0.18 | 3.46 | 0.00 | -2.21 | 0.02 | 0.78 | 0.16 | 4.64 | 0.00 | -1.30 | 0.19 |
| Copper | 0.51 | 0.21 | 2.45 | 0.02 | -2.90 | 0.00 | 0.96 | 0.26 | 3.64 | 0.00 | -0.22 | 0.82 | 1.12 | 0.19 | 5.76 | 0.00 | 0.75 | 0.44 |
| Lead | 0.71 | 0.18 | 3.84 | 0.00 | -1.70 | 0.08 | 0.73 | 0.15 | 4.79 | 0.00 | -1.61 | 0.10 | 0.71 | 0.16 | 4.19 | 0.00 | -1.74 | 0.08 |
| Nickel | 0.58 | 0.24 | 2.42 | 0.02 | -2.49 | 0.01 | 0.51 | 0.14 | 3.57 | 0.00 | -2.95 | 0.00 | 0.63 | 0.14 | 4.25 | 0.00 | -2.21 | 0.02 |
| Tin | 0.10 | 0.14 | 0.76 | 0.45 | -5.37 | 0.00 | 0.28 | 0.15 | 1.79 | 0.09 | -4.32 | 0.00 | 0.35 | 0.15 | 2.37 | 0.03 | -3.88 | 0.00 |
| Precious metals | | | | | | | | | | | | | | | | | | |
| Gold | 0.71 | 0.15 | 4.51 | 0.00 | -1.87 | 0.06 | 0.82 | 0.21 | 3.86 | 0.00 | -1.13 | 0.25 | 0.61 | 0.17 | 3.58 | 0.00 | -2.56 | 0.01 |
| Silver | 0.93 | 0.28 | 3.27 | 0.00 | -0.41 | 0.67 | 0.80 | 0.20 | 3.95 | 0.00 | -1.28 | 0.19 | 0.71 | 0.13 | 5.21 | 0.00 | -1.81 | 0.06 |
| Agricultural products | | | | | | | | | | | | | | | | | | |
| Corn | 0.52 | 0.16 | 3.14 | 0.00 | -2.43 | 0.01 | 0.58 | 0.16 | 3.59 | 0.00 | -2.15 | 0.03 | 0.58 | 0.44 | 1.31 | 0.21 | -2.04 | 0.04 |
| Soybean | 0.54 | 0.27 | 1.96 | 0.07 | -2.58 | 0.01 | 0.49 | 0.22 | 2.22 | 0.04 | -2.84 | 0.00 | 0.33 | 0.16 | 2.00 | 0.06 | -3.73 | 0.00 |
| Wheat | 0.38 | 0.20 | 1.85 | 0.09 | -3.16 | 0.00 | 0.45 | 0.27 | 1.61 | 0.13 | -2.83 | 0.00 | 0.37 | 0.24 | 1.51 | 0.16 | -2.93 | 0.00 |

| (b) Power = 0.7 | 3-month horizon | | | | | | 6-month horizon | | | | | | 12-month horizon | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $d$ | $SE$ | $t$ | $p_1$ | $z$ | $p_2$ | $d$ | $SE$ | $t$ | $p_1$ | $z$ | $p_2$ | $d$ | $SE$ | $t$ | $p_1$ | $z$ | $p_2$ |
| Energy products | | | | | | | | | | | | | | | | | | |
| WTI | 0.62 | 0.10 | 6.13 | 0.00 | -4.37 | 0.00 | 0.66 | 0.10 | 6.29 | 0.00 | -3.96 | 0.00 | 0.67 | 0.10 | 6.17 | 0.00 | -3.86 | 0.00 |
| Natural gas | 0.21 | 0.11 | 1.87 | 0.06 | -9.28 | 0.00 | 0.35 | 0.11 | 3.15 | 0.00 | -7.58 | 0.00 | 0.53 | 0.09 | 5.76 | 0.00 | -5.42 | 0.00 |
| Heating oil | 0.37 | 0.11 | 3.36 | 0.00 | -7.39 | 0.00 | 0.53 | 0.12 | 4.18 | 0.00 | -5.42 | 0.00 | 0.70 | 0.11 | 5.94 | 0.00 | -3.42 | 0.00 |
| Gasoline | 0.21 | 0.10 | 2.05 | 0.04 | -9.23 | 0.00 | 0.42 | 0.11 | 3.82 | 0.00 | -6.75 | 0.00 | 0.65 | 0.08 | 7.38 | 0.00 | -4.11 | 0.00 |
| Base metals | | | | | | | | | | | | | | | | | | |
| Aluminum | 0.37 | 0.10 | 3.61 | 0.00 | -6.56 | 0.00 | 0.79 | 0.13 | 6.06 | 0.00 | -2.15 | 0.03 | 0.91 | 0.09 | 9.23 | 0.00 | -0.90 | 0.36 |
| Copper | 0.27 | 0.12 | 2.18 | 0.03 | -7.55 | 0.00 | 0.52 | 0.11 | 4.43 | 0.00 | -4.92 | 0.00 | 0.82 | 0.11 | 7.12 | 0.00 | -1.79 | 0.07 |
| Lead | 0.48 | 0.08 | 5.98 | 0.00 | -5.43 | 0.00 | 0.66 | 0.08 | 8.13 | 0.00 | -3.53 | 0.00 | 0.75 | 0.07 | 10.64 | 0.00 | -2.55 | 0.01 |
| Nickel | 0.62 | 0.10 | 5.91 | 0.00 | -3.94 | 0.00 | 0.82 | 0.08 | 9.77 | 0.00 | -1.86 | 0.06 | 0.92 | 0.10 | 8.85 | 0.00 | -0.76 | 0.44 |
| Tin | 0.50 | 0.10 | 4.82 | 0.00 | -5.21 | 0.00 | 0.68 | 0.09 | 6.89 | 0.00 | -3.35 | 0.00 | 0.81 | 0.11 | 6.82 | 0.00 | -1.92 | 0.05 |
| Precious metals | | | | | | | | | | | | | | | | | | |
| Gold | 0.89 | 0.08 | 11.09 | 0.00 | -1.18 | 0.23 | 0.96 | 0.09 | 9.96 | 0.00 | -0.35 | 0.72 | 0.74 | 0.09 | 7.80 | 0.00 | -3.05 | 0.00 |
| Silver | 0.66 | 0.12 | 5.26 | 0.00 | -4.00 | 0.00 | 0.85 | 0.11 | 7.76 | 0.00 | -1.70 | 0.08 | 0.79 | 0.08 | 9.45 | 0.00 | -2.34 | 0.01 |
| Agricultural products | | | | | | | | | | | | | | | | | | |
| Corn | 0.36 | 0.12 | 2.98 | 0.00 | -5.54 | 0.00 | 0.72 | 0.12 | 5.66 | 0.00 | -2.35 | 0.01 | 0.52 | 0.24 | 2.14 | 0.04 | -3.74 | 0.00 |
| Soybean | 0.36 | 0.14 | 2.50 | 0.01 | -6.14 | 0.00 | 0.50 | 0.11 | 4.53 | 0.00 | -4.79 | 0.00 | 0.48 | 0.11 | 4.17 | 0.00 | -4.84 | 0.00 |
| Wheat | 0.35 | 0.21 | 1.70 | 0.09 | -5.56 | 0.00 | 0.66 | 0.15 | 4.23 | 0.00 | -2.86 | 0.00 | 0.52 | 0.10 | 5.10 | 0.00 | -3.44 | 0.00 |

(i) Panel (a) and (b) present the Phillips Modified GPH estimation results of the fractional difference parameter for powers 0.5 and 0.7, respectively.
(ii) $d$ is the estimated fractional difference parameter.
(iii) $SE$ is the standard error of estimated $d$.
(iv) $t$ is the t-statistic for $H_0 : d = 0$.
(v) $p_1$ is the p-value for $H_0 : d = 0$.
(vi) $z$ is the z-statistic for $H_0 : d = 1$.
(vii) $p_2$ is the p-value for $H_0 : d = 1$.

Table 12: Phillips Modified Geweke-Porter-Haduk (GPH) Estimation Results for the Basis in First Difference

| (a) Power = 0.5 | 3-month horizon | | | | | | 6-month horizon | | | | | | 12-month horizon | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $d$ | $SE$ | $t$ | $p_1$ | $z$ | $p_2$ | $d$ | $SE$ | $t$ | $p_1$ | $z$ | $p_2$ | $d$ | $SE$ | $t$ | $p_1$ | $z$ | $p_2$ |
| Energy products | | | | | | | | | | | | | | | | | | |
| WTI | 0.05 | 0.17 | 0.29 | 0.77 | -6.26 | 0.00 | -0.13 | 0.17 | -0.77 | 0.45 | -7.52 | 0.00 | -0.26 | 0.22 | -1.16 | 0.25 | -8.38 | 0.00 |
| Natural gas | -0.23 | 0.18 | -1.26 | 0.22 | -7.92 | 0.00 | -0.37 | 0.22 | -1.66 | 0.11 | -8.83 | 0.00 | -0.19 | 0.17 | -1.13 | 0.27 | -7.71 | 0.00 |
| Heating oil | -0.45 | 0.18 | -2.41 | 0.02 | -9.62 | 0.00 | -0.07 | 0.26 | -0.29 | 0.77 | -7.13 | 0.00 | -0.08 | 0.25 | -0.32 | 0.75 | -7.14 | 0.00 |
| Gasoline | 0.90 | 0.06 | 14.57 | 0.00 | -0.61 | 0.54 | 0.92 | 0.12 | 7.22 | 0.00 | -0.52 | 0.60 | 0.67 | 0.21 | 3.16 | 0.00 | -2.11 | 0.03 |
| Base metals | | | | | | | | | | | | | | | | | | |
| Aluminum | 0.11 | 0.28 | 0.41 | 0.68 | -5.31 | 0.00 | -0.34 | 0.25 | -1.36 | 0.19 | -8.14 | 0.00 | -0.17 | 0.20 | -0.87 | 0.39 | -7.11 | 0.00 |
| Copper | 0.17 | 0.18 | 0.91 | 0.37 | -5.00 | 0.00 | 0.36 | 0.17 | 2.05 | 0.05 | -3.85 | 0.00 | 0.36 | 0.27 | 1.36 | 0.19 | -3.81 | 0.00 |
| Lead | 0.04 | 0.18 | 0.23 | 0.81 | -5.77 | 0.00 | -0.10 | 0.13 | -0.79 | 0.43 | -6.69 | 0.00 | -0.31 | 0.16 | -1.83 | 0.08 | -7.91 | 0.00 |
| Nickel | -0.59 | 0.24 | -2.41 | 0.02 | -9.60 | 0.00 | -0.53 | 0.16 | -3.32 | 0.00 | -9.27 | 0.00 | -0.42 | 0.12 | -3.33 | 0.00 | -8.62 | 0.00 |
| Tin | -0.47 | 0.22 | -2.06 | 0.05 | -8.90 | 0.00 | -0.38 | 0.25 | -1.53 | 0.14 | -8.39 | 0.00 | -0.54 | 0.28 | -1.90 | 0.07 | -9.33 | 0.00 |
| Precious metals | | | | | | | | | | | | | | | | | | |
| Gold | -0.02 | 0.20 | -0.12 | 0.90 | -6.79 | 0.00 | -0.19 | 0.18 | -1.03 | 0.31 | -7.90 | 0.00 | 0.18 | 0.17 | 1.06 | 0.30 | -5.39 | 0.00 |
| Silver | 0.15 | 0.15 | 0.99 | 0.33 | -5.57 | 0.00 | 0.26 | 0.14 | 1.80 | 0.08 | -4.83 | 0.00 | 0.02 | 0.12 | 0.15 | 0.87 | -6.30 | 0.00 |
| Agricultural products | | | | | | | | | | | | | | | | | | |
| Corn | -0.27 | 0.15 | -1.72 | 0.11 | -6.58 | 0.00 | -0.37 | 0.18 | -1.97 | 0.07 | -7.08 | 0.00 | 0.29 | 0.26 | 1.13 | 0.28 | -3.45 | 0.00 |
| Soybean | -0.27 | 0.24 | -1.12 | 0.27 | -7.16 | 0.00 | 0.48 | 0.27 | 1.73 | 0.10 | -2.89 | 0.00 | 0.17 | 0.11 | 1.52 | 0.15 | -4.63 | 0.00 |
| Wheat | -0.77 | 0.25 | -3.00 | 0.01 | -9.17 | 0.00 | 0.04 | 0.23 | 0.18 | 0.85 | -4.93 | 0.00 | -0.24 | 0.10 | -2.23 | 0.05 | -5.81 | 0.00 |

| (b) Power = 0.7 | 3-month horizon | | | | | | 6-month horizon | | | | | | 12-month horizon | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $d$ | $SE$ | $t$ | $p_1$ | $z$ | $p_2$ | $d$ | $SE$ | $t$ | $p_1$ | $z$ | $p_2$ | $d$ | $SE$ | $t$ | $p_1$ | $z$ | $p_2$ |
| Energy products | | | | | | | | | | | | | | | | | | |
| WTI | -0.11 | 0.09 | -1.13 | 0.26 | -13.09 | 0.00 | -0.16 | 0.10 | -1.53 | 0.13 | -13.72 | 0.00 | -0.22 | 0.11 | -1.91 | 0.06 | -14.39 | 0.00 |
| Natural gas | -0.53 | 0.10 | -4.92 | 0.00 | -18.09 | 0.00 | -0.50 | 0.11 | -4.61 | 0.00 | -17.74 | 0.00 | -0.24 | 0.09 | -2.52 | 0.01 | -14.51 | 0.00 |
| Heating oil | -0.44 | 0.12 | -3.73 | 0.00 | -17.06 | 0.00 | -0.25 | 0.12 | -2.13 | 0.03 | -14.80 | 0.00 | -0.15 | 0.10 | -1.48 | 0.14 | -13.54 | 0.00 |
| Gasoline | 0.47 | 0.08 | 5.54 | 0.00 | -6.18 | 0.00 | 0.49 | 0.11 | 4.45 | 0.00 | -5.93 | 0.00 | 0.65 | 0.08 | 7.38 | 0.00 | -4.11 | 0.00 |
| Base metals | | | | | | | | | | | | | | | | | | |
| Aluminum | -0.46 | 0.12 | -3.74 | 0.00 | -15.28 | 0.00 | -0.26 | 0.11 | -2.30 | 0.02 | -13.20 | 0.00 | -0.09 | 0.09 | -0.91 | 0.36 | -11.41 | 0.00 |
| Copper | -0.45 | 0.12 | -3.76 | 0.00 | -15.27 | 0.00 | -0.31 | 0.11 | -2.80 | 0.00 | -13.75 | 0.00 | -.014 | 0.13 | -1.04 | 0.30 | -11.94 | 0.00 |
| Lead | -0.34 | 0.09 | -3.75 | 0.00 | -14.06 | 0.00 | -0.25 | 0.07 | -3.46 | 0.00 | -13.11 | 0.00 | -0.25 | 0.07 | -3.62 | 0.00 | -13.14 | 0.00 |
| Nickel | -0.44 | 0.10 | -4.16 | 0.00 | -15.10 | 0.00 | -0.21 | 0.08 | -2.61 | 0.01 | -12.72 | 0.00 | -0.11 | 0.09 | -1.21 | 0.23 | -11.62 | 0.00 |
| Tin | -0.37 | 0.11 | -3.25 | 0.00 | -14.33 | 0.00 | -0.22 | 0.10 | -2.08 | 0.04 | -12.83 | 0.00 | -0.18 | 0.14 | -1.32 | 0.19 | -12.40 | 0.00 |
| Precious metals | | | | | | | | | | | | | | | | | | |
| Gold | -0.05 | 0.09 | -0.60 | 0.54 | -12.45 | 0.00 | -0.01 | 0.08 | -0.21 | 0.82 | -11.99 | 0.00 | -0.01 | 0.09 | -0.11 | 0.91 | -11.89 | 0.00 |
| Silver | -0.17 | 0.11 | -1.53 | 0.12 | -13.82 | 0.00 | 0.02 | 0.08 | 0.32 | 0.74 | -11.43 | 0.00 | -0.01 | 0.06 | -0.25 | 0.79 | -11.66 | 0.00 |
| Agricultural products | | | | | | | | | | | | | | | | | | |
| Corn | -0.53 | 0.11 | -4.73 | 0.00 | -13.08 | 0.00 | -0.25 | 0.11 | -2.27 | 0.03 | -10.75 | 0.00 | 0.17 | 0.21 | 0.79 | 0.43 | -6.92 | 0.00 |
| Soybean | -0.53 | 0.14 | -3.79 | 0.00 | -14.99 | 0.00 | -0.10 | 0.13 | -0.78 | 0.44 | -10.78 | 0.00 | -0.14 | 0.12 | -1.15 | 0.25 | -10.85 | 0.00 |
| Wheat | -0.66 | 0.18 | -3.59 | 0.00 | -14.22 | 0.00 | -0.09 | 0.14 | -0.66 | 0.50 | -9.34 | 0.00 | -0.26 | 0.07 | -3.46 | 0.00 | -9.26 | 0.00 |

(i) Panel (a) and (b) present the Phillips Modified GPH estimation results of the fractional difference parameter for powers 0.5 and 0.7, respectively.
(ii) $d$ is the estimated fractional difference parameter.
(iii) $SE$ is the standard error of estimated $d$.
(iv) $t$ is the t-statistic for $H_0 : d = 0$.
(v) $p_1$ is the p-value for $H_0 : d = 0$.
(vi) $z$ is the z-statistic for $H_0 : d = 1$.
(vii) $p_2$ is the p-value for $H_0 : d = 1$.

Table 13: Phillips Modified Geweke-Porter-Haduk (GPH) Estimation Results for the Price Change in Level

| (a) Power = 0.5 | 3-month horizon | | | | | | 6-month horizon | | | | | | 12-month horizon | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $d$ | $SE$ | $t$ | $p_1$ | $z$ | $p_2$ | $d$ | $SE$ | $t$ | $p_1$ | $z$ | $p_2$ | $d$ | $SE$ | $t$ | $p_1$ | $z$ | $p_2$ |
| Energy products | | | | | | | | | | | | | | | | | | |
| WTI | -0.08 | 0.16 | -0.55 | 0.58 | -7.20 | 0.00 | -0.24 | 0.17 | -1.40 | 0.17 | -8.01 | 0.00 | 0.01 | 0.18 | 0.07 | 0.94 | -6.33 | 0.00 |
| Natural gas | 0.18 | 0.15 | 1.18 | 0.25 | -5.23 | 0.00 | -0.26 | 0.21 | -1.24 | 0.23 | -8.15 | 0.00 | 0.17 | 0.20 | 0.86 | 0.39 | -5.28 | 0.00 |
| Heating oil | -0.01 | 0.19 | -0.07 | 0.94 | -6.70 | 0.00 | -0.22 | 0.22 | -0.99 | 0.33 | -7.86 | 0.00 | 0.03 | 0.19 | 0.19 | 0.85 | -6.19 | 0.00 |
| Gasoline | 0.20 | 0.14 | 1.41 | 0.17 | -5.28 | 0.00 | 0.08 | 0.11 | 0.74 | 0.46 | -5.87 | 0.00 | 0.18 | 0.19 | 0.95 | 0.35 | -5.25 | 0.00 |
| Base metals | | | | | | | | | | | | | | | | | | |
| Aluminum | -0.31 | 0.11 | -2.62 | 0.01 | -7.91 | 0.00 | -0.26 | 0.23 | -1.14 | 0.27 | -7.65 | 0.00 | 0.07 | 0.23 | 0.31 | 0.75 | -5.59 | 0.00 |
| Copper | 0.06 | 0.20 | 0.29 | 0.77 | -5.66 | 0.00 | 0.03 | 0.18 | 0.20 | 0.84 | -5.81 | 0.00 | 0.33 | 0.19 | 1.72 | 0.10 | -4.01 | 0.00 |
| Lead | -0.09 | 0.14 | -0.66 | 0.51 | -6.61 | 0.00 | -0.04 | 0.14 | -0.28 | 0.77 | -6.28 | 0.00 | 0.26 | 0.20 | 1.31 | 0.20 | -4.41 | 0.00 |
| Nickel | 0.21 | 0.26 | 0.81 | 0.43 | -4.76 | 0.00 | 0.02 | 0.24 | 0.09 | 0.92 | -5.90 | 0.00 | 0.36 | 0.24 | 1.47 | 0.16 | -3.82 | 0.00 |
| Tin | -0.03 | 0.24 | -0.12 | 0.90 | -6.22 | 0.00 | 0.00 | 0.16 | 0.03 | 0.97 | -6.00 | 0.00 | 0.23 | 0.22 | 1.06 | 0.30 | -4.59 | 0.00 |
| Precious metals | | | | | | | | | | | | | | | | | | |
| Gold | 0.36 | 0.20 | 1.77 | 0.09 | -4.23 | 0.00 | 0.34 | 0.22 | 1.52 | 0.14 | -4.20 | 0.00 | 0.51 | 0.14 | 3.43 | 0.00 | -3.12 | 0.00 |
| Silver | 0.09 | 0.30 | 0.31 | 0.75 | -5.99 | 0.00 | 0.15 | 0.18 | 0.81 | 0.42 | -5.46 | 0.00 | 0.30 | 0.16 | 1.87 | 0.07 | -4.45 | 0.00 |
| Agricultural products | | | | | | | | | | | | | | | | | | |
| Corn | 0.09 | 0.38 | 0.25 | 0.80 | -4.66 | 0.00 | 0.20 | 0.35 | 0.57 | 0.57 | -4.12 | 0.00 | 1.05 | 0.38 | 2.71 | 0.02 | 0.28 | 0.77 |
| Soybean | 0.03 | 0.24 | 0.12 | 0.90 | -5.44 | 0.00 | 0.24 | 0.25 | 0.97 | 0.34 | -4.22 | 0.00 | 0.50 | 0.20 | 2.45 | 0.02 | -2.80 | 0.00 |
| Wheat | 0.02 | 0.21 | 0.09 | 0.92 | -5.06 | 0.00 | 0.04 | 0.15 | 0.30 | 0.76 | -4.93 | 0.00 | 0.84 | 0.27 | 3.02 | 0.01 | -0.81 | 0.41 |

| (b) Power = 0.7 | 3-month horizon | | | | | | 6-month horizon | | | | | | 12-month horizon | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $d$ | $SE$ | $t$ | $p_1$ | $z$ | $p_2$ | $d$ | $SE$ | $t$ | $p_1$ | $z$ | $p_2$ | $d$ | $SE$ | $t$ | $p_1$ | $z$ | $p_2$ |
| Energy products | | | | | | | | | | | | | | | | | | |
| WTI | 0.04 | 0.10 | 0.43 | 0.66 | -11.25 | 0.00 | 0.36 | 0.11 | 3.13 | 0.00 | -7.52 | 0.00 | 0.83 | 0.11 | 7.16 | 0.00 | -1.95 | 0.05 |
| Natural gas | 0.13 | 0.09 | 1.38 | 0.17 | -10.15 | 0.00 | 0.35 | 0.13 | 2.57 | 0.01 | -7.56 | 0.00 | 0.85 | 0.12 | 6.60 | 0.00 | -1.68 | 0.09 |
| Heating oil | 0.07 | 0.10 | 0.77 | 0.44 | -10.85 | 0.00 | 0.42 | 0.12 | 3.36 | 0.00 | -6.74 | 0.00 | 0.87 | 0.11 | 7.37 | 0.00 | -1.43 | 0.15 |
| Gasoline | 0.11 | 0.11 | 1.01 | 0.31 | -10.45 | 0.00 | 0.45 | 0.12 | 3.51 | 0.00 | -6.33 | 0.00 | 0.80 | 0.11 | 6.84 | 0.00 | -2.23 | 0.02 |
| Base metals | | | | | | | | | | | | | | | | | | |
| Aluminum | 0.23 | 0.09 | 2.57 | 0.01 | -8.02 | 0.00 | 0.77 | 0.15 | 4.86 | 0.00 | -2.36 | 0.01 | 0.99 | 0.13 | 7.12 | 0.00 | -0.07 | 0.94 |
| Copper | 0.19 | 0.10 | 1.92 | 0.06 | -8.40 | 0.00 | 0.74 | 0.14 | 5.31 | 0.00 | -2.66 | 0.00 | 1.10 | 0.12 | 8.80 | 0.00 | 1.04 | 0.29 |
| Lead | 0.12 | 0.08 | 1.54 | 0.12 | -9.12 | 0.00 | 0.74 | 0.15 | 4.92 | 0.00 | -2.67 | 0.00 | 0.99 | 0.13 | 7.45 | 0.00 | -0.01 | 0.98 |
| Nickel | 0.20 | 0.11 | 1.85 | 0.07 | -8.27 | 0.00 | 0.71 | 0.14 | 4.97 | 0.00 | -2.99 | 0.00 | 0.99 | 0.13 | 7.53 | 0.00 | -0.00 | 0.99 |
| Tin | 0.21 | 0.11 | 1.80 | 0.07 | -8.22 | 0.00 | 0.82 | 0.15 | 5.18 | 0.00 | -1.84 | 0.06 | 1.22 | 0.17 | 7.06 | 0.00 | 2.31 | 0.02 |
| Precious metals | | | | | | | | | | | | | | | | | | |
| Gold | 0.09 | 0.09 | 0.92 | 0.35 | -10.70 | 0.00 | 0.43 | 0.09 | 4.46 | 0.00 | -6.63 | 0.00 | 0.87 | 0.13 | 6.59 | 0.00 | -1.46 | 0.14 |
| Silver | 0.05 | 0.13 | 0.42 | 0.67 | -11.11 | 0.00 | 0.44 | 0.09 | 4.76 | 0.00 | -6.54 | 0.00 | 0.83 | 0.12 | 6.70 | 0.00 | -1.86 | 0.06 |
| Agricultural products | | | | | | | | | | | | | | | | | | |
| Corn | 0.00 | 0.17 | 0.03 | 0.97 | -8.49 | 0.00 | 0.82 | 0.21 | 3.84 | 0.00 | -1.52 | 0.12 | 0.98 | 0.21 | 4.51 | 0.00 | -0.12 | 0.89 |
| Soybean | -0.01 | 0.11 | -0.17 | 0.86 | -9.93 | 0.00 | 0.78 | 0.20 | 3.79 | 0.00 | -2.07 | 0.03 | 0.92 | 0.15 | 6.07 | 0.00 | -0.73 | 0.46 |
| Wheat | 0.14 | 0.12 | 1.10 | 0.27 | -7.32 | 0.00 | 0.79 | 0.15 | 5.04 | 0.00 | -1.78 | 0.07 | 0.95 | 0.19 | 4.83 | 0.00 | -0.36 | 0.71 |

(i) Panel (a) and (b) present the Phillips Modified GPH estimation results of the fractional difference parameter for powers 0.5 and 0.7, respectively.
(ii) $d$ is the estimated fractional difference parameter.
(iii) $SE$ is the standard error of estimated $d$.
(iv) $t$ is the t-statistic for $H_0 : d = 0$.
(v) $p_1$ is the p-value for $H_0 : d = 0$.
(vi) $z$ is the z-statistic for $H_0 : d = 1$.
(vii) $p_2$ is the p-value for $H_0 : d = 1$.

Table 14: Phillips Modified Geweke-Porter-Haduk (GPH) Estimation Results for the Price Change in First Difference

| (a) Power = 0.5 | 3-month horizon | | | | | | 6-month horizon | | | | | | 12-month horizon | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $d$ | $SE$ | $t$ | $p_1$ | $z$ | $p_2$ | $d$ | $SE$ | $t$ | $p_1$ | $z$ | $p_2$ | $d$ | $SE$ | $t$ | $p_1$ | $z$ | $p_2$ |
| Energy products | | | | | | | | | | | | | | | | | | |
| WTI | 0.19 | 0.15 | 1.26 | 0.22 | -5.32 | 0.00 | 0.08 | 0.13 | 0.58 | 0.56 | -5.91 | 0.00 | -0.09 | 0.15 | -0.61 | 0.54 | -7.04 | 0.00 |
| Natural gas | 0.60 | 0.12 | 4.97 | 0.00 | -2.54 | 0.01 | -0.02 | 0.23 | -0.11 | 0.91 | -6.59 | 0.00 | -0.55 | 0.20 | -2.75 | 0.01 | -10.02 | 0.00 |
| Heating oil | 0.33 | 0.22 | 1.51 | 0.14 | -4.39 | 0.00 | 0.37 | 0.36 | 1.03 | 0.31 | -4.01 | 0.00 | 0.08 | 0.13 | 0.62 | 0.54 | -5.87 | 0.00 |
| Gasoline | 1.08 | 0.19 | 5.60 | 0.00 | 0.52 | 0.59 | 0.45 | 0.20 | 2.24 | 0.03 | -3.50 | 0.00 | -0.51 | 0.16 | -3.11 | 0.00 | -9.75 | 0.00 |
| Base metals | | | | | | | | | | | | | | | | | | |
| Aluminum | 0.41 | 0.15 | 2.67 | 0.01 | -3.55 | 0.00 | -0.51 | 0.24 | -2.11 | 0.05 | -9.12 | 0.00 | -0.21 | 0.16 | -1.26 | 0.22 | -7.32 | 0.00 |
| Copper | -0.40 | 0.33 | -1.21 | 0.24 | -8.47 | 0.00 | 0.32 | 0.14 | 2.18 | 0.04 | -4.09 | 0.00 | -0.18 | 0.17 | -1.06 | 0.30 | -7.13 | 0.00 |
| Lead | -0.68 | 0.25 | -2.69 | 0.01 | -10.19 | 0.00 | 0.37 | 0.16 | 2.33 | 0.03 | -3.77 | 0.00 | -0.53 | 0.23 | -2.25 | 0.03 | -9.28 | 0.00 |
| Nickel | 0.71 | 0.17 | 3.97 | 0.00 | -1.75 | 0.08 | 0.32 | 0.27 | 1.18 | 0.25 | -4.05 | 0.00 | -0.01 | 0.25 | -0.07 | 0.94 | -6.14 | 0.00 |
| Tin | -0.48 | 0.34 | -1.43 | 0.17 | -8.99 | 0.00 | -0.33 | 0.17 | -1.87 | 0.08 | -8.05 | 0.00 | -0.37 | 0.24 | -1.55 | 0.14 | -8.30 | 0.00 |
| Precious metals | | | | | | | | | | | | | | | | | | |
| Gold | 0.50 | 0.26 | 1.88 | 0.07 | -3.26 | 0.00 | 0.56 | 0.18 | 3.02 | 0.00 | -2.81 | 0.00 | 0.41 | 0.24 | 1.69 | 0.10 | -3.77 | 0.00 |
| Silver | 0.65 | 0.16 | 3.89 | 0.00 | -2.27 | 0.02 | -0.30 | 0.20 | -1.48 | 0.15 | -8.36 | 0.00 | -0.12 | 0.18 | -0.68 | 0.50 | -7.25 | 0.00 |
| Agricultural products | | | | | | | | | | | | | | | | | | |
| Corn | -0.14 | 0.32 | -0.43 | 0.66 | -5.91 | 0.00 | -0.25 | 0.25 | -1.01 | 0.33 | -6.51 | 0.00 | 0.64 | 0.35 | 1.81 | 0.09 | -1.84 | 0.06 |
| Soybean | 1.03 | 0.35 | 2.87 | 0.01 | 0.17 | 0.85 | -0.27 | 0.27 | -0.99 | 0.33 | -7.18 | 0.00 | -0.35 | 0.18 | -1.89 | 0.08 | -7.59 | 0.00 |
| Wheat | -0.07 | 0.22 | -0.34 | 0.73 | -5.57 | 0.00 | -0.61 | 0.13 | -4.49 | 0.00 | -8.34 | 0.00 | 0.42 | 0.25 | 1.70 | 0.11 | -2.96 | 0.00 |

| (b) Power = 0.7 | 3-month horizon | | | | | | 6-month horizon | | | | | | 12-month horizon | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $d$ | $SE$ | $t$ | $p_1$ | $z$ | $p_2$ | $d$ | $SE$ | $t$ | $p_1$ | $z$ | $p_2$ | $d$ | $SE$ | $t$ | $p_1$ | $z$ | $p_2$ |
| Energy products | | | | | | | | | | | | | | | | | | |
| WTI | -0.35 | 0.09 | -3.98 | 0.00 | -16.00 | 0.00 | -0.04 | 0.09 | -0.47 | 0.63 | -12.19 | 0.00 | 0.22 | 0.10 | 2.15 | 0.03 | -9.03 | 0.00 |
| Natural gas | -0.11 | 0.11 | -1.04 | 0.30 | -13.02 | 0.00 | -0.08 | 0.12 | -0.64 | 0.52 | -12.62 | 0.00 | 0.03 | 0.13 | 0.28 | 0.78 | -11.13 | 0.00 |
| Heating oil | -0.30 | 0.12 | -2.45 | 0.01 | -15.32 | 0.00 | 0.00 | 0.13 | 0.01 | 0.98 | -11.64 | 0.00 | 0.35 | 0.09 | 3.62 | 0.00 | -7.49 | 0.00 |
| Gasoline | 0.01 | 0.13 | 0.12 | 0.90 | -11.57 | 0.00 | -0.07 | 0.11 | -0.62 | 0.53 | -12.49 | 0.00 | -0.07 | 0.10 | -0.68 | 0.49 | -12.52 | 0.00 |
| Base metals | | | | | | | | | | | | | | | | | | |
| Aluminum | 0.14 | 0.11 | 1.30 | 0.19 | -8.90 | 0.00 | 0.18 | 0.15 | 1.15 | 0.25 | -8.57 | 0.00 | 0.53 | 0.15 | 3.45 | 0.00 | -4.82 | 0.00 |
| Copper | -0.60 | 0.12 | -4.94 | 0.00 | -16.80 | 0.00 | 0.52 | 0.11 | 4.39 | 0.00 | -4.96 | 0.00 | 0.34 | 0.11 | 2.89 | 0.00 | -6.79 | 0.00 |
| Lead | -0.70 | 0.10 | -6.73 | 0.00 | -17.82 | 0.00 | 0.42 | 0.13 | 3.29 | 0.00 | -5.96 | 0.00 | 0.16 | 0.16 | 0.99 | 0.32 | -8.66 | 0.00 |
| Nickel | 0.03 | 0.11 | 0.30 | 0.76 | -10.10 | 0.00 | 0.24 | 0.13 | 1.89 | 0.06 | -7.85 | 0.00 | 0.32 | 0.15 | 2.11 | 0.04 | -6.95 | 0.00 |
| Tin | -0.55 | 0.13 | -4.14 | 0.00 | -16.23 | 0.00 | 0.12 | 0.14 | 0.88 | 0.38 | -9.16 | 0.00 | 0.29 | 0.13 | 2.14 | 0.03 | -7.28 | 0.00 |
| Precious metals | | | | | | | | | | | | | | | | | | |
| Gold | -0.41 | 0.11 | -3.48 | 0.00 | -16.64 | 0.00 | 0.19 | 0.09 | 2.11 | 0.03 | -9.44 | 0.00 | 0.17 | 0.09 | 1.77 | 0.08 | -9.62 | 0.00 |
| Silver | 0.01 | 0.13 | 0.08 | 0.93 | -11.64 | 0.00 | -0.31 | 0.09 | -3.26 | 0.00 | -15.32 | 0.00 | 0.09 | 0.13 | 0.68 | 0.49 | -10.58 | 0.00 |
| Agricultural products | | | | | | | | | | | | | | | | | | |
| Corn | -0.70 | 0.16 | -4.24 | 0.00 | -14.53 | 0.00 | 0.26 | 0.22 | 1.16 | 0.25 | -6.28 | 0.00 | 0.17 | 0.21 | 0.79 | 0.43 | -6.92 | 0.00 |
| Soybean | -0.13 | 0.17 | -0.78 | 0.48 | -11.08 | 0.00 | 0.02 | 0.20 | 0.14 | 0.88 | -9.32 | 0.00 | 0.04 | 0.15 | 0.25 | 0.79 | -9.10 | 0.00 |
| Wheat | -0.35 | 0.14 | -2.39 | 0.02 | -11.59 | 0.00 | 0.17 | 0.21 | 0.83 | 0.40 | -7.03 | 0.00 | 0.25 | 0.19 | 1.30 | 0.20 | -6.22 | 0.00 |

(i) Panel (a) and (b) present the Phillips Modified GPH estimation results of the fractional difference parameter for powers 0.5 and 0.7, respectively.
(ii) $d$ is the estimated fractional difference parameter.
(iii) $SE$ is the standard error of estimated $d$.
(iv) $t$ is the t-statistic for $H_0 : d = 0$.
(v) $p_1$ is the p-value for $H_0 : d = 0$.
(vi) $z$ is the z-statistic for $H_0 : d = 1$.
(vii) $p_2$ is the p-value for $H_0 : d = 1$.

Table 15: Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test Statistics for the Basis and the Price Change in Level

| (a) Basis | 3-month horizon | | 6-month horizon | | 12-month horizon | |
|---|---|---|---|---|---|---|
| | Constant | Constant & trend | Constant | Constant & trend | Constant | Constant & trend |
| Energy products | | | | | | |
| WTI | 0.511** | 0.072 | 0.505** | 0.084 | 0.435* | 0.104 |
| Natural gas | 0.390* | 0.081 | 0.537** | 0.089 | 0.619** | 0.106 |
| Heating oil | 0.210 | 0.044 | 0.232 | 0.047 | 0.297 | 0.065 |
| Gasoline | 0.128 | 0.037 | 0.159 | 0.044 | 0.109 | 0.068 |
| Base metals | | | | | | |
| Aluminum | 0.148 | 0.137* | 0.285 | 0.140* | 0.303 | 0.153** |
| Copper | 0.282 | 0.123* | 0.197 | 0.159** | 0.197 | 0.190** |
| Lead | 0.188 | 0.172** | 0.180 | 0.181** | 0.213 | 0.214** |
| Nickel | 0.173 | 0.134* | 0.289 | 0.150** | 0.344 | 0.169** |
| Tin | 0.085 | 0.058 | 0.145 | 0.090 | 0.160 | 0.125* |
| Precious metals | | | | | | |
| Gold | 1.110*** | 0.078 | 1.140*** | 0.070 | 1.200*** | 0.070 |
| Silver | 1.260*** | 0.079 | 1.260*** | 0.136* | 1.100*** | 0.084 |
| Agricultural products | | | | | | |
| Corn | 0.111 | 0.109 | 0.103 | 0.103 | 0.107 | 0.102 |
| Soybean | 0.301 | 0.048 | 0.366* | 0.039 | 0.183 | 0.040 |
| Wheat | 0.296 | 0.072 | 0.229 | 0.066 | 0.126 | 0.095 |

| (b) Price change | 3-month horizon | | 6-month horizon | | 12-month horizon | |
|---|---|---|---|---|---|---|
| | Constant | Constant & trend | Constant | Constant & trend | Constant | Constant & trend |
| Energy products | | | | | | |
| WTI | 0.101 | 0.094 | 0.103 | 0.092 | 0.142 | 0.135* |
| Natural gas | 0.114 | 0.042 | 0.137 | 0.043 | 0.161 | 0.054 |
| Heating oil | 0.091 | 0.086 | 0.092 | 0.087 | 0.122 | 0.120* |
| Gasoline | 0.081 | 0.074 | 0.094 | 0.087 | 0.134 | 0.132* |
| Base metals | | | | | | |
| Aluminum | 0.059 | 0.058 | 0.066 | 0.059 | 0.094 | 0.060 |
| Copper | 0.128 | 0.123* | 0.134 | 0.121* | 0.183 | 0.132* |
| Lead | 0.080 | 0.082 | 0.084 | 0.085 | 0.105 | 0.097 |
| Nickel | 0.127 | 0.051 | 0.136 | 0.049 | 0.180 | 0.048 |
| Tin | 0.076 | 0.076 | 0.081 | 0.080 | 0.095 | 0.093 |
| Precious metals | | | | | | |
| Gold | 0.338 | 0.196** | 0.318 | 0.206** | 0.345 | 0.209** |
| Silver | 0.124 | 0.112 | 0.129 | 0.118 | 0.132 | 0.122* |
| Agricultural products | | | | | | |
| Corn | 0.077 | 0.077 | 0.081 | 0.083 | 0.099 | 0.102 |
| Soybean | 0.066 | 0.066 | 0.073 | 0.072 | 0.089 | 0.089 |
| Wheat | 0.090 | 0.088 | 0.102 | 0.092 | 0.107 | 0.107 |

(i) The table presents the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test for stationarity of a univariate time series.
(ii) The null hypothesis is that the variable is stationary.
(iii) Panel (a) shows the results for the basis and panel (b) corresponds to the price change.
(iv) Maximum lag length is selected by the Schwert criterion ($lag = [12.(\frac{T}{100})^{0.25}]$).
(v) For the KPSS test statistic with constant 1%, 5%, and 10% critical values are 0.739, 0.463, and 0.347 respectively.
(vi) For the KPSS test statistic with constant and trend 1%, 5%, and 10% critical values are 0.216, 0.146, and 0.119 respectively.
(vii) *, **, and *** show 10%, 5%, and 1% significance level respectively.

Table 16: Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test Statistics for the Basis and the Price Change in First Difference

| (a) Basis | 3-month horizon | | 6-month horizon | | 12-month horizon | |
|---|---|---|---|---|---|---|
| | Constant | Constant & trend | Constant | Constant & trend | Constant | Constant & trend |
| Energy products | | | | | | |
| WTI | 0.034 | 0.027 | 0.030 | 0.026 | 0.028 | 0.026 |
| Natural gas | 0.023 | 0.023 | 0.027 | 0.024 | 0.041 | 0.035 |
| Heating oil | 0.026 | 0.022 | 0.025 | 0.022 | 0.022 | 0.021 |
| Gasoline | 0.023 | 0.021 | 0.018 | 0.018 | 0.042 | 0.028 |
| Base metals | | | | | | |
| Aluminum | 0.059 | 0.041 | 0.106 | 0.058 | 0.079 | 0.058 |
| Copper | 0.036 | 0.033 | 0.050 | 0.047 | 0.067 | 0.065 |
| Lead | 0.041 | 0.042 | 0.046 | 0.043 | 0.055 | 0.044 |
| Nickel | 0.048 | 0.032 | 0.060 | 0.034 | 0.065 | 0.033 |
| Tin | 0.033 | 0.031 | 0.037 | 0.032 | 0.039 | 0.033 |
| Precious metals | | | | | | |
| Gold | 0.084 | 0.050 | 0.085 | 0.049 | 0.070 | 0.044 |
| Silver | 0.103 | 0.036 | 0.147 | 0.035 | 0.077 | 0.043 |
| Agricultural products | | | | | | |
| Corn | 0.049 | 0.043 | 0.047 | 0.046 | 0.052 | 0.053 |
| Soybean | 0.058 | 0.041 | 0.044 | 0.034 | 0.035 | 0.035 |
| Wheat | 0.040 | 0.037 | 0.052 | 0.047 | 0.158 | 0.070 |

| (b) Price change | 3-month horizon | | 6-month horizon | | 12-month horizon | |
|---|---|---|---|---|---|---|
| | Constant | Constant & trend | Constant | Constant & trend | Constant | Constant & trend |
| Energy products | | | | | | |
| WTI | 0.037 | 0.025 | 0.035 | 0.025 | 0.023 | 0.023 |
| Natural gas | 0.030 | 0.028 | 0.027 | 0.026 | 0.026 | 0.025 |
| Heating oil | 0.028 | 0.024 | 0.032 | 0.023 | 0.025 | 0.024 |
| Gasoline | 0.025 | 0.024 | 0.024 | 0.023 | 0.024 | 0.024 |
| Base metals | | | | | | |
| Aluminum | 0.035 | 0.032 | 0.041 | 0.041 | 0.045 | 0.044 |
| Copper | 0.060 | 0.035 | 0.067 | 0.039 | 0.059 | 0.046 |
| Lead | 0.027 | 0.027 | 0.033 | 0.025 | 0.029 | 0.030 |
| Nickel | 0.048 | 0.031 | 0.043 | 0.028 | 0.042 | 0.034 |
| Tin | 0.029 | 0.029 | 0.032 | 0.027 | 0.027 | 0.028 |
| Precious metals | | | | | | |
| Gold | 0.057 | 0.035 | 0.049 | 0.039 | 0.043 | 0.036 |
| Silver | 0.039 | 0.029 | 0.032 | 0.030 | 0.042 | 0.032 |
| Agricultural products | | | | | | |
| Corn | 0.046 | 0.045 | 0.044 | 0.043 | 0.045 | 0.041 |
| Soybean | 0.038 | 0.038 | 0.032 | 0.032 | 0.030 | 0.031 |
| Wheat | 0.062 | 0.053 | 0.062 | 0.043 | 0.059 | 0.043 |

(i) The table presents the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test for stationarity of a univariate time series.
(ii) The null hypothesis is that the variable is stationary.
(iii) Panel (a) shows the results for the basis and panel (b) corresponds to the price change.
(iv) Maximum lag length is selected by the Schwert criterion ($lag = [12.(\frac{T}{100})^{0.25}]$).
(v) For the KPSS test statistic with constant 1%, 5%, and 10% critical values are 0.739, 0.463, and 0.347 respectively.
(vi) For the KPSS test statistic with constant and trend 1%, 5%, and 10% critical values are 0.216, 0.146, and 0.119 respectively.
(vii) *, **, and *** show 10%, 5%, and 1% significance levels respectively.

Table 17: The Augmented-Dicky-Fuller (ADF) Test Statistics for the Basis and the Price Change in Level

| (a) Basis | 3-month horizon | | | 6-month horizon | | | 12-month horizon | | |
|---|---|---|---|---|---|---|---|---|---|
| | No intercept | Intercept | Intercept & trend | No intercept | Intercept | Intercept & trend | No intercept | Intercept | Intercept & trend |
| Energy products | | | | | | | | | |
| WTI | -5.84*** | -5.32*** | -5.68*** | -4.24*** | -4.22*** | -4.48*** | -3.33*** | -3.36** | -3.76** |
| Natural gas | -3.50*** | -4.38*** | -5.09*** | -3.32*** | -3.90*** | -4.78*** | -3.52*** | -3.99*** | -4.55*** |
| Heating oil | -6.83*** | -6.82*** | -6.92*** | -5.14*** | -5.14*** | -5.34*** | -5.15*** | -5.13*** | -5.28*** |
| Gasoline | -4.60*** | -5.11*** | -5.12*** | -4.76*** | -5.31*** | -5.33*** | -3.61** | -3.87*** | -3.87** |
| Base metals | | | | | | | | | |
| Aluminum | -3.24*** | -4.90*** | -4.89*** | -2.44** | -3.39** | -3.44** | -2.14** | -2.57 | -2.62 |
| Copper | -2.62*** | -2.69* | -2.48 | -2.42** | -2.53 | -2.43 | -2.24** | -2.46 | -2.38 |
| Lead | -3.17*** | -3.21** | -3.23* | -3.07*** | -3.06** | -3.05 | -2.61*** | -2.61* | -2.59 |
| Nickel | -3.97*** | -3.99*** | -4.03*** | -2.56** | -2.87* | -3.55** | -2.71*** | -3.00** | -3.35* |
| Tin | -4.30*** | -4.39*** | -4.39*** | -4.19*** | -4.32*** | -4.33*** | -3.98*** | -4.20*** | -4.20*** |
| Precious metals | | | | | | | | | |
| Gold | -1.73* | -2.33 | -3.02 | -2.25** | -2.77* | -3.00 | -1.60 | -2.14 | -2.76 |
| Silver | -1.69* | -2.90** | -4.12*** | -2.36** | -3.10** | -3.30* | -2.69*** | -3.20** | -4.38*** |
| Agricultural products | | | | | | | | | |
| Corn | -2.37** | -3.09** | -3.08 | -3.05*** | -3.73*** | -3.71** | -4.15*** | -4.30*** | -4.03*** |
| Soybean | -3.36*** | -3.39** | -3.49** | -3.85*** | -3.95*** | -4.51*** | -4.78*** | -4.97*** | -5.05*** |
| Wheat | -2.32** | -3.17** | -3.78** | -2.40** | -2.94** | -3.35* | -3.02*** | -3.38** | -3.46** |

| (b) Price change | 3-month horizon | | | 6-month horizon | | | 12-month horizon | | |
|---|---|---|---|---|---|---|---|---|---|
| | No intercept | Intercept | Intercept & trend | No intercept | Intercept | Intercept & trend | No intercept | Intercept | Intercept & trend |
| Energy products | | | | | | | | | |
| WTI | -5.72*** | -5.20*** | -5.22*** | -4.58*** | -4.63*** | -4.67*** | -3.89*** | -3.36** | -3.76** |
| Natural gas | -5.63*** | -5.63*** | -5.76*** | -5.17*** | -5.20*** | -5.55*** | -5.41*** | -3.99*** | -4.55*** |
| Heating oil | -5.06*** | -5.09*** | -5.10*** | -4.02*** | -4.04*** | -4.05*** | -3.90*** | -5.13*** | -5.28*** |
| Gasoline | -4.21*** | -4.27*** | -4.28*** | -4.30*** | -4.36*** | -4.37*** | -3.76*** | -3.81*** | -3.82*** |
| Base metals | | | | | | | | | |
| Aluminum | -5.06*** | -5.06*** | -5.09*** | -3.30*** | -3.31** | -3.37* | -3.67*** | -3.68*** | -3.74** |
| Copper | -3.66*** | -3.77*** | -3.87** | -2.88*** | -3.04** | -3.27* | -2.81*** | -2.96** | -3.12 |
| Lead | -3.81*** | -4.01*** | -4.02*** | -3.39*** | -3.64*** | -3.70** | -2.92*** | -3.16** | -3.21* |
| Nickel | -4.14*** | -4.15*** | -4.52*** | -4.00*** | -4.02*** | -4.48*** | -4.06*** | -4.06*** | -4.51*** |
| Tin | -3.82*** | -4.06*** | -4.06*** | -3.89*** | -4.12*** | -4.13*** | -3.31*** | -3.74*** | -3.75** |
| Precious metals | | | | | | | | | |
| Gold | -3.33*** | -3.56*** | -3.57** | -2.20** | -2.42 | -2.42 | -1.76* | -1.95 | -1.93 |
| Silver | -5.73*** | -5.86*** | -5.84*** | -3.98*** | -4.13*** | -4.13*** | -2.40** | -2.55 | -2.55 |
| Agricultural products | | | | | | | | | |
| Corn | -4.06*** | -4.05*** | -4.03*** | -1.68* | -1.66 | -1.51 | -1.78* | -1.77 | -1.63 |
| Soybean | -5.70*** | -5.71*** | -4.01*** | -4.41*** | -4.46*** | -4.44*** | -3.26*** | -3.30** | -3.30* |
| Wheat | -3.98*** | -3.98*** | -4.03*** | -2.81*** | -2.82* | -2.85 | -1.80* | -1.71 | -1.67 |

(i) The table presents the Augmented-Dicky-Fuller (ADF) test statistics under different options of no constant, with constant, and with constant and trend.
(ii) Panel (a) show the test results for the basis and panel (b) refers to the price change.
(iii) The null hypothesis is that the tested variable has a unit root.
(iv) The number of lags is selected by using the Akaike Information Criterion (AIC).
(v) The 1%, 5%, and 10% critical values for testing the variable with no constant are -2.58, -1.95, and -1.62 respectively.
(vi) The 1%, 5%, and 10% critical values for testing the variable with constant are -3.45, -2.87, and -2.57 respectively.
(vii) The 1%, 5%, and 10% critical values for testing the variable with constant and trend are -3.98, -3.42, and -3.13 respectively.
(viii) *, **, and *** show significance at the 10%, 5%, and 1% levels.

Table 18: The Augmented-Dicky-Fuller (ADF) Test Statistics for the Basis and the Price Change in First Difference

| (a) Basis | 3-month horizon | | | 6-month horizon | | | 12-month horizon | | |
|---|---|---|---|---|---|---|---|---|---|
| | No intercept | Intercept | Intercept & trend | No intercept | Intercept | Intercept & trend | No intercept | Intercept | Intercept & trend |
| Energy products | | | | | | | | | |
| WTI | -9.73*** | -9.72*** | -9.72*** | -5.38*** | -5.37*** | -5.37*** | -5.42*** | -5.41*** | -5.41*** |
| Natural gas | -7.23*** | -7.22*** | -7.21*** | -12.29*** | -12.27*** | -12.25*** | -6.18*** | -6.17*** | -6.15*** |
| Heating oil | -8.89*** | -8.88*** | -8.86*** | -10.43*** | -10.42*** | -10.40*** | -5.30*** | -5.29*** | -5.28*** |
| Gasoline | -4.74*** | -4.73*** | -4.73*** | -4.76*** | -4.75*** | -4.74*** | -12.19*** | -12.17*** | -12.15*** |
| Base metals | | | | | | | | | |
| Aluminum | -12.37*** | -12.35*** | -12.32*** | -6.86*** | -6.85*** | -6.83*** | -10.47*** | -10.44*** | -10.42*** |
| Copper | -3.64*** | -3.63*** | -3.73** | -3.60*** | -3.59*** | -3.64** | -3.36*** | -3.35** | -3.42** |
| Lead | -7.77*** | -7.75*** | -7.73*** | -11.61*** | -11.59*** | -11.56*** | -11.05*** | -11.03*** | -11.01*** |
| Nickel | -4.60*** | -4.60*** | -4.73*** | -4.52*** | -4.51*** | -4.58*** | -6.27*** | -6.25*** | -6.27*** |
| Tin | -6.87*** | -6.86*** | -6.85*** | -7.61*** | -7.60*** | -7.58*** | -6.09*** | -6.08*** | -6.06*** |
| Precious metals | | | | | | | | | |
| Gold | -4.46*** | -4.50*** | -4.51*** | -9.56*** | -9.60*** | -9.64*** | -8.18*** | -8.19*** | -8.17*** |
| Silver | -4.38*** | -4.39*** | -4.42*** | -12.34*** | -12.35*** | -12.42*** | -2.89*** | -2.82* | -2.75 |
| Agricultural products | | | | | | | | | |
| Corn | -5.67*** | -5.64*** | -5.62*** | -11.14*** | -11.10*** | -11.06*** | -7.11*** | -7.07*** | -7.03*** |
| Soybean | -5.72*** | -5.71*** | -5.72*** | -5.65*** | -5.64*** | -5.65*** | -7.30*** | -7.27*** | -7.26*** |
| Wheat | -5.03*** | -5.02*** | -4.96*** | -5.96*** | -5.94*** | -5.94*** | -10.13*** | -10.06*** | -10.05*** |

| (b) Price change | 3-month horizon | | | 6-month horizon | | | 12-month horizon | | |
|---|---|---|---|---|---|---|---|---|---|
| | No intercept | Intercept | Intercept & trend | No intercept | Intercept | Intercept & trend | No intercept | Intercept | Intercept & trend |
| Energy products | | | | | | | | | |
| WTI | -5.91*** | -5.90*** | -5.89*** | -7.37*** | -7.35*** | -7.34*** | -6.87*** | -6.86*** | -6.85*** |
| Natural gas | -7.30*** | -7.29*** | -7.28*** | -6.20*** | -6.19*** | -6.18*** | -7.81*** | -7.80*** | -7.79*** |
| Heating oil | -6.77*** | -6.76*** | -6.74*** | -5.96*** | -5.95*** | -5.95*** | -7.29*** | -7.28*** | -7.27*** |
| Gasoline | -6.01*** | -6.00*** | -5.99*** | -6.69*** | -6.68*** | -6.67*** | -7.56*** | -7.55*** | -7.54*** |
| Base metals | | | | | | | | | |
| Aluminum | -5.91*** | -5.91*** | -5.89*** | -5.08*** | -5.06*** | -5.04*** | -4.24*** | -4.23*** | -4.20*** |
| Copper | -6.12*** | -6.12*** | -6.12*** | -4.85*** | -4.84*** | -4.81*** | -4.75*** | -4.74*** | -4.71*** |
| Lead | -5.79*** | -5.78*** | -5.76*** | -4.98*** | -4.96*** | -4.95*** | -6.94*** | -6.93*** | -6.91*** |
| Nickel | -5.34*** | -5.32*** | -5.31*** | -4.42*** | -4.41*** | -4.38*** | -5.94*** | -5.92*** | -5.90*** |
| Tin | -5.43*** | -5.42*** | -5.41*** | -4.75*** | -4.74*** | -4.73*** | -4.42*** | -4.41*** | -4.40*** |
| Precious metals | | | | | | | | | |
| Gold | -8.27*** | -8.26*** | -8.25*** | -6.21*** | -6.21*** | -6.21*** | -7.37*** | -7.36*** | -7.35*** |
| Silver | -4.38*** | -4.39*** | -4.42*** | -6.49*** | -6.48*** | -6.47*** | -5.68*** | -5.67*** | -5.67*** |
| Agricultural products | | | | | | | | | |
| Corn | -4.55*** | -4.53*** | -4.64*** | -4.69*** | -4.68*** | -4.78*** | -5.34*** | -5.34*** | -5.45*** |
| Soybean | -6.51*** | -6.49*** | -6.47*** | -4.41*** | -4.41*** | -4.41*** | -5.06*** | -5.06*** | -5.04*** |
| Wheat | -6.96*** | -6.94*** | -6.89*** | -4.24*** | -4.24*** | -4.24*** | -5.38*** | -5.39*** | -5.40*** |

(i) The table presents the Augmented-Dicky-Fuller (ADF) test statistics under different options of no constant, with constant, and with constant and trend.
(ii) Panel (a) shows the test results for the basis and panel (b) refers to the price change.
(iii) The null hypothesis is that the tested variable has a unit root.
(iv) The number of lags is selected by using the Akaike Information Criterion (AIC).
(v) The 1%, 5%, and 10% critical values for testing the variable with no constant are -2.58, -1.95, and -1.62 respectively.
(vi) The 1%, 5%, and 10% critical values for testing the variable with constant are -3.45, -2.87, and -2.57 respectively.
(vii) The 1%, 5%, and 10% critical values for testing the variable with constant and trend are -3.98, -3.42, and -3.13 respectively.
(viii) *, **, and *** show significance at the 10%, 5%, and 1% levels.

For robustness, Dufour and Taamouti (2010) construct the bootstrap confidence interval for the causality measure and present its asymptotic validity. The algorithm to obtain the percentile bootstrap confidence intervals is the following:

- Step 1: estimate a $VAR(k)$ model, given the data realization and store the residuals: $\widetilde{u} = \delta(t) - \sum_{i=1}^{k} \widehat{\Pi}_{ik}\delta(t-i), \quad \text{for } t = k+1, ...T.$
  Where $\widehat{\Pi}_{ik}$ is estimated by OLS and $\Sigma_{u|k} = \sum_{t=k+1}^{T} \frac{\widetilde{u}(t)\widetilde{u}(t)'}{(T-k)}$ is the covariance matrix with $\widetilde{u}(t) = \widehat{u}(t) - \sum_{t=p+1}^{T} \frac{\widehat{u}(t)}{(T-k)}$ and $\widehat{u}(t) = \delta(t) - \sum_{i=1}^{k} \widehat{\Pi}_{ik}\delta(t-i)$.
- Step 2: Generate $(T-k)$ bootstrap residuals $u^*(t)$ by random sampling with replacement from the $VAR$ residuals $\widehat{u}(t)$, $t = k+1, ..., T$.
- Step 3: Set aside the vector of initial observation up to $k$, $\delta^*(0) = (\delta(1)', ..., \delta(k)')'$.
- Step 4: Given $\widehat{\Pi}_{ik}$, $\{u^*(t)\}_{t=k+1}^{T}$, and $\delta^*(0)$, generate $(T-k)$ the bootstrap sample form the equation
  $\delta^*(t) = \sum_{i=1}^{k} \widehat{\Pi}_{ik}\delta^*(t-i) + u^*(t), \quad \text{for } t = k+1, ..., T.$
- Step 5: Estimate a $VAR(k)$ model with OLS using the bootstrap sample generated in step 4.
- Step 6: By employing the bootstrap sample of the price change in $\{\delta^*(t)\}_{t=1}^{T}$, estimate an $ARIMA(k, 0, 0)$ model and save the residuals.
- Step 7: Using the bootstrap sample, calculate the causality measure by $\widehat{C}^{(j)*}(b \underset{h}{\rightarrow} p|I)$ in equation (30).
- Step 8: Choose $B$ such that $\frac{1}{2}\alpha(B+1)$ be an integer and repeat steps (2) to (7) $B$ times.
- Step 9: Construct the 95% percentile confidence interval by sorting $\widehat{C}^{(j)*}$.

# CHAPTER II

# ASSET PRICING IMPLICATIONS OF GROUPED PATTERNS OF HETEROGENEITY

## *2.1 Introduction*

Firms are connected through different types of links. Some links are well-established and observable, whereas some are less transparent and ambiguous. As an example of the former, one can consider input-output accounts data (also known as I-O tables) which are available on an annual basis by the Bureau of Economic Analysis (BEA). Depending upon examination by rows or columns, the table reveals links between economic sectors in terms of their production. Furthermore, input-output links can be compared to a different setting in which firms are grouped, but not necessarily due to the usual links of tradable goods and services. One example is the volatility spillover among firms and groups of firms. In this new setting, the focus will be on recovering the so-called opaque links and that introduces new challenges such as the criteria by which the firms are grouped and the implications of those criteria for economics and finance.

This paper attempts to find answers to the above questions by designing a network of realized volatility of common stocks as a candidate of the financial network and then investigating its asset pricing implications. Thanks to the recent popularity of network theory and application, there exists a rich literature in the field[1]. Networks are simple to understand and they can easily be defined by two sets of features, namely nodes and edges. Nodes in my proposed network are groups of firms that have similar volatility series. Edges are volatility spillovers among nodes. Two factors are derived in turn that describe the network. The first

[1]Related work includes Cohen and Frazzini (2008), Acemoglu, Ozdaglar, and Tahbaz-Salehi (2015), and Allen and Babus (2009).

is the concentration factor that narrates the evolution of nodes. The second is sparsity which unfolds the transformation of edges.

Herskovic (2018) is the first to propose concentration and sparsity as factors to examine networks of production. Ahern (2013) considers only the node property and concludes that industries with central positions earn higher returns than others. Later on, Herskovic (2018) completes this picture by introducing another property to take into account the edges that connecting firms; hence creating a complete network for real activities in the economy. In a general equilibrium model, Herskovic (2018) drives two factors that describe the input-output relationship among the economic sectors. Concentration narrates the evolution of nodes (sectors) and sparsity reports the evolution of edges (commodity flows among sectors). Both factors are tested and they show to have statistically significant risk premia for a diverse set of assets.

Concerning the volatility networks in particular and the financial networks in general, this paper is motivated by Billio et al. (2012), Bianchi et al. (2015), Deibold and Yilmaz (2014), Dufour and Jian (2016), and Barigozzi and Brownlees (2016). The research of Dufour and Jian (2016) is comparatively related to this paper. They define networks of volatility for firms in the S&P 100 by utilizing the multiple horizon causality measures of Dufour and Taamouti (2010). One feature of this paper, which makes it distinguishable from Dufour and Jian (2016) and the other works in the literature, is the fact that I try to recover financial factors from the real connections among firms (that can be named as the real economics or network) and compare their pricing powers with factors that were derived in Herskovic (2018). In other words, this research is a comparative study of financial and real networks.

Since this research is inspired by Herskovic (2018) and many notions are similar in definition to the ones presented in his work, I present an example to explain concentration and sparsity.

Table 19: Example of Network Economies

Network (A), (B), and (C) are all representatives of three-sector economies. The matrices in the second row represent different networks. Network A is symmetric; whereas, in B and C, links are not equally distributed among sectors. The last row presents the output share for each sector given the total output in each sample economy is 1. In A and B shares are equal, but in C sector 1 plays the central role.

| Network (A) three-sector economy | Network (B) three-sector economy | Network (C) three-sector economy |
|---|---|---|
| $w_1 = \begin{bmatrix} 0.33 & 0.33 & 0.33 \\ 0.33 & 0.33 & 0.33 \\ 0.33 & 0.33 & 0.33 \end{bmatrix}$ | $w_2 = \begin{bmatrix} 0.1 & 0.45 & 0.45 \\ 0.45 & 0.1 & 0.45 \\ 0.45 & 0.45 & 0.1 \end{bmatrix}$ | $w_3 = \begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0.6 & 0.2 & 0.2 \\ 0.6 & 0.2 & 0.2 \end{bmatrix}$ |
| $\delta_1 = \begin{bmatrix} 0.33 & 0.33 & 0.33 \end{bmatrix}$ | $\delta_2 = \begin{bmatrix} 0.33 & 0.33 & 0.33 \end{bmatrix}$ | $\delta_3 = \begin{bmatrix} 0.6 & 0.2 & 0.2 \end{bmatrix}$ |

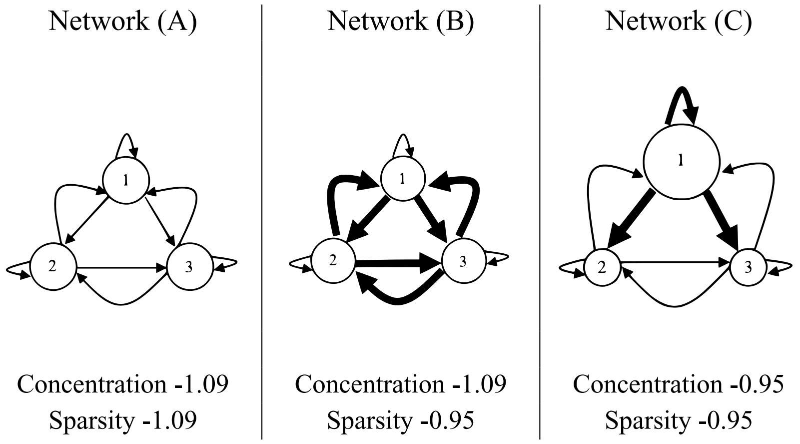
### 2.1.1 Network and its Descriptive Factors: Example

Let us explore the two factors in an example. In Table (19) there are three different networks that are representatives of three-sector economies. Network (A), (B), and (C) are shown by the square matrices $w_1$, $w_2$, and $w_3$, respectively. The total output of the economy is 1 and $\delta$ shows the output shares. (A) is an example of a symmetric network. If we read the matrix by row, sector 1 purchases 33% of its needed input from its own, 33% from sector 2, and 33% from sector 3. Shares are also equally distributed. In (B), the links are differently shaped, but the shares remain the same. Sector 1 purchases 10% of its needed input from its own, 45% from sector 2, and 45% from sector 3. If we look at the other two rows, we can see that the outflows of input from different sectors are stronger than the diagonal elements of the self-consuming inputs. In terms of real activities, (B) represents an economy with higher specialization than (A).

In network (C), sector 1 becomes highly influential. This is evident from $\delta_3$. Sector 1 produces 60% of the total output and because of this, the outflow of inputs from sector 1 is relatively stronger than other links, a fact that can be observed in $w_3$. To distinguish networks

Figure 1: Example of Network Economies: Graphs

The graphs visualize different networks. Network A is a symmetric, network B has a higher sparsity relative to network A. Finally, network C has a higher concentration compared to network B.




based on the characteristics of their nodes and edges, Herskovic (2018) drives the following two equations that represent the concentration and sparsity, respectively:

$$HNC(t) = \sum_{j=1}^{n} \delta_j(t) \log \delta_j(t) \tag{36}$$

$$HNS(t) = \sum_{i=1}^{n} \delta_i(t) \sum_{j=1}^{n} w_{ij}(t) \log w_{ij}(t) \tag{37}$$

where $n$ is the number of sectors, $\delta_j$ is the output share of sector $j$, and $w_{ij}$ is the entry of input-output table when sector $i$ purchases an input weight from sector $j$. $HNC$ is network concentration and $HNS$ refers to network sparsity (the notation that I will respect in this paper as well). Herskovic derives these two expressions through the conditions of a general equilibrium model. They both are negative. Concentration is the negative entropy of the output shares. If it shows rather a small number, then the output is dominated by only a few sectors' shares. On the other hand, sparsity indicates the thickness of the connections among sectors. If it rises, it is an indication of more specialization in the economy.

Now the numbers at the bottom of Figure (1) start to make sense. Network sparsity is the edge-characteristic of the network. By comparing (B) to (A), we can see that although the output shares are the same, the outflow of input becomes stronger and therefore (B) has a higher sparsity level. Furthermore, in (C) since sector 1 manipulates the most of economic activities and the network concentration is the node-property. Therefore, the network has a higher level of concentration relative to (A) and (B).

As was mentioned earlier, the transparent connections between firms based on real activities do not necessarily hold in the financial realm. The main objective of this paper is to shed more light on this subject by recovering those obscure links for a financial network.

The paper has the following structure. In the data description section, the data and their properties are presented. Then, in the network representation section, factors are defined. In evidence, the time series of concentration and sparsity and the appropriate asset pricing tests are explored. And finally, the last section concludes.

## 2.2 Data Description

The data-set contains the realized volatility series of firms listed as common stocks and their corresponding customers which are known as the U.S. firm-customer links. To the best of my knowledge, Cohen and Frazzini (2008)[2] are the first who gathered the information regarding the firm-customer supply links for the U.S. from 1980 to 2005. Looking at their data, each firm is spotted by a unique permanent number (PERMNO) that is identifiable in Compustat/CRSP. Going through Compustat customer historical segments, one can observe the name of all customers that purchase at least 10% of the final product of a firm in an annual frequency. The difficult part is matching the names of the customer with the set of PERMNOs in Compustat to identify the customer's ticker and other relative information. Herskovic (2018) went through the process to update the data-set in Cohen and Frazzini (2008) until 2013. I upgraded his set by tracking down the firm-customer links until 2017 and adding more common stocks that perhaps were missing. The result is a set of firm-customer links

[2]The original data-set which covers 1980 to 2005 is available in Andrea Frazzini's data library at http://people.stern.nyu.edu/afrazzin/data_library.htm

for the U.S. from 1980 to 2017. Table (20) presents a summary statistics of the data.

Table 20: Summary Statistics of the U.S.Firm-customer Links

To update the firm-customer dataset, the first step is acquiring all the common stocks with share codes 10, 11, and 12 listed in NASDAQ, AMEX, and NYSE from 2014 to 2017. The next step is getting the principal customers of these firms through the historical segment. After removing geographical and international customers, there remain business firms only as customers. The final step is matching the customers' names with PERMNO in our common stock list since the assumption is that the customer is a common stock as well. I could do that by running a fuzzy algorithm in Excel and match the customer names with the closest match and then, return its corresponding PERMNO. The final outcome is an annual list of the U.S. firm-customer links that covers 38 years.

Time Series (38 Annual Observations, 1980–2017)

| | Min | Max | Mean | SD | Median |
|---|---|---|---|---|---|
| Number of firms in the sample per year | 260 | 1218 | 806 | 289 | 825 |
| Number of customers in the sample per year | 156 | 1067 | 599 | 315 | 485 |
| Full sample % coverage of stock universe | 14 | 25 | 18 | 4 | 17 |

To extract realized volatility series, I follow Barunik et al. (2014). Suppose that the price process $p(t)$ is given by:

$$p(t) = p(0) + \int_0^t \sigma(s) dW(s) \tag{38}$$

where $W$ is the standard Brownian motion, and $\sigma$ is the strictly positive volatility process with the integrated variance of $\int_0^t \sigma^2(s) d(s)$. A measure for quadratic variance is presented by Barndorff-Nielsen (2002) and Andersen et al. (2001) which is the summation of squared daily returns. Therefore, the realized variance of t daily returns is given by:

$$RVar = \sum_{t=0}^{T} r^2(t) \tag{39}$$

The realized volatility is simply the square root of the realized variance:

$$RVol = \sqrt{RVar} \tag{40}$$

The square returns are added over a week ($T = 4$ for 5 working days starting from 0) and leave us with around 52 observations of realized volatility for each year. The return series are obtained from CRSP.

A possible extension to the data construction would be taking into consideration high-frequency data to estimate the volatility. In this new frequency, one can think of the possibility of jumps in the processes. Since the frequency of my data-set is weekly, I disregard jumps for now.

## *2.3 Network Representation*

This section introduces two factors that describe the realized volatility network. Concentration is obtained by employing the concept of grouped patterns of heterogeneity, (also known as grouping the data or clustering). Sparsity is derived by utilizing theories of causality measure as a method to capture the volatility spillover among groups of firms.

### 2.3.1 Concentration Factor

There are dimensions in the data that are unknown to the econometrician. This is called heterogeneity. Manresa and Bonhomme (2014) introduce a theoretical framework to take into account heterogeneities and efficiently estimate them. To recall their model, consider the following:

$$Y_i(t) = x_i(t)\theta + c_{g_i}(t) + u_i(t), \qquad i = 1, ..., N, \quad t = 1, ..., T \tag{41}$$

Where $Y_i$ is the panel of dependent variables, $x_i'$ is the set of co-variates, and $c_{g_i}$ is the group-specific time effect for $G$ groups $g \in \{1, ..., G\}$. There are N features and T data points in time. The choice of $G$ is something that I discuss later in this section and it is very important in designing the network.

Without the presence of co-variates (i.e. $\theta = 0$), the model coincides with the well-known *kmeans* algorithm (Forgy, 1965 and Steinley, 2006). The clustering algorithms allow for patterns to emerge from the unlabeled data and therefore they are classified as unsupervised learning algorithms. Going back to the realized volatility network, I can identify $Y_i(t)$ as the weekly realized volatility series of firm $i$, namely $RVol_i(t)$. For the sake of notational simplicity, I will show $RVol_i(t)$ as $V_i(t)$ for the rest of the paper.

To estimate $c_{g_i}(t)$, the following is the minimization problem that needs to be solved:

$$\widehat{c} = \arg\min_{c} \sum_{i=1}^{N} \Big( \min_{g \in \{1,\dots,G\}} \sum_{t=1}^{T} (V_i(t) - c_{g_i}(t))^2 \Big) \tag{42}$$

A critical question in grouping the data is how many groups ($G$) the researcher would choose to run the algorithm. $G$ in the network of realized volatility represents the number of nodes (nodes are groups of firms with similar volatility series). I use the notation in Manresa and Bonhomme (2015) that formalize a BIC-form information criterion. To define the criterion:

$$I(G) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (V_i(t) - \widehat{c}_g(t))^2 + G P_{NT} \tag{43}$$

$I$ denotes the information criterion and $P_{NT}$ is a penalty term. The optimal number of groups minimizes the below objective function:

$$G^* = \arg\min_{G \in \{1,\dots,G_{\max}\}} I(G) \tag{44}$$

where $G_{\max}$ is the maximum choice of $G$ and is set by the researcher. Following Manresa and Bonhomme (2015), I select $G_{\max}$ equal 15. The information criterion is in the form of the BIC and it has the following expression:

$$BIC(G) = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (V_i(t) - \widehat{c}_g(t))^2 + \widehat{\sigma}^2 \frac{GT + N}{NT} \ln(NT) \tag{45}$$

where $\widehat{\sigma}^2$ is the consistent estimator of the variance of $u_i$ and it is given by:

$$\widehat{\sigma}^2 = \frac{1}{NT - G_{\max} T - N} \sum_{i=1}^{N} \sum_{t=1}^{T} (V_i(t) - \widehat{c}_g(t))^2 \tag{46}$$

The group assignment that yields the minimum value of BIC corresponds to the optimal value of grouping, namely $G^*$. As was mentioned before, $G^*$ is the number of groups of firms with similar realized volatility patterns. Once $G^*$ is identified, the concentration time-series is described as the negative entropy of the nodes' shares:

$$NC(t) = \sum_{j=1}^{G^*} \delta_j(t) \log \delta_j(t) \tag{47}$$

where $\delta_j$ is defined as:

$$\delta_j(t) = \frac{Mcap_j}{\sum_{i=1}^{G^*} Mcap_i} \tag{48}$$

$Mcap_j$ is the market capitalization of group $j$ which is the sum of stock price multiply by the number of shares outstanding of firms within each group. Thus, the concentration factor measures the negative entropy of the capital shares' distribution.

### 2.3.2 Sparsity Factor

To complete the network, I need to define the edges that connect the nodes. A directed edge with causal interpretation is a suitable way to define the connections in an economic network. Therefore, I use the notions of multiple horizon causality measures introduced by Dufour and Taamouti (2010). I borrow the exposition of causality measure, introduced by Dufour and Taamouti (2010) and Dufour et al. (2015), and I accommodate the realized volatility market within them.

Suppose there exists a jointly stationary process $V(t)$ that includes three multivariate stochastic processes such that $V(t) = (V_1(t), V_2(t), V_3(t))$, where $V_1(t) = (v_{11}(t), ..., v_{1m_1}(t))'$, $V_2(t) = (v_{21}(t), ..., v_{2m_2}(t))'$, and $V_3(t) = (v_{31}(t), ..., v_{3m_3}(t))'$ with $m_1 + m_2 + m_3 = N$. Define the information set $I = \{I(t) : t \in \mathbb{Z}, t > \tau\}$ and $t < t' \Rightarrow I(t) \subseteq I(t')$ for all $t > \tau$, where $I(t)$ is defined on the Hilbert subspace of $L^2$, $\tau \in \mathbb{Z} \cup \{-\infty\}$ represents a starting point. If $V_1(\tau, t]$, $V_2(\tau, t]$, and $V_3(\tau, t]$ are the Hilbert spaces spanned by the components of $V_1$, $V_2$, and $V_3$ respectively then the information sets are the following:

$$
\begin{aligned}
I_{V_1}(t) &= I(t) + V_1(\tau, t] \\
I_{V_1V_2}(t) &= I(t) + V_1(\tau, t] + V_2(\tau, t]
\end{aligned}
$$

**Definition 6** *Granger non-causality at horizon $h$ : For any arbitrary forecast horizon $h$, $V_2$ does not Granger-cause $V_1$, given information set $I$, if and only if:*

$$
P[V_1(t+h)|I_{V_1}(t)] = P[V_1(t+h)|I_{V_1V_2}(t)] \tag{49}
$$

Here, $P[V_1(t+h)|I_{V_1}(t)]$ is the best linear forecast of $V_1(t+h)$, based on the information set $I_{V_1}(t)$.

To measure the Granger causality from the above definition of non-causality, Dufour and Taamouti (2010) present the below definition:

**Definition 7** *Granger causality measure at horizon $h$ : For any arbitrary forecast horizon $h$, such that $1 \leq h < \infty$,*

$$
GC(V_2 \underset{h}{\rightarrow} V_1|I) = \ln[\frac{\det\{\Sigma[V_1(t+h)|I_{V_1}(t)]\}}{\det\{\Sigma[V_1(t+h)|I_{V_1V_2}(t)]\}}] \tag{50}
$$

*GC is the mean-square Granger causality measure from $V_2$ to $V_1$ at horizon $h$ given, the information set $I$. The vector of prediction errors and the corresponding matrix of their second moments $\Sigma$ are:*

$$
\begin{aligned}
U[V_1(t+h)|I(t)] &= (u[v_{11}(t+h)|I(t)], ..., v_{1m_1}(t+h)|I(t)])' && (51) \\
where\ \ u[v_{1i}(t+h)|I(t)] &= v_{1i}(t+h) - P[v_{1i}(t+h)|I(t)] && (52) \\
\Sigma[V_1(t+h)|I_{V_1}(t)] &= E\{U[V_1(t+h)|I(t)]U[V_1(t+h)|I(t)]'\} && (53)
\end{aligned}
$$

#### 2.3.2.1 Causality Measure for VAR(p) Models

Dufour and Taamouti (2010) introduce the general framework to estimate the causality measure for the set of invertible processes which includes vector autoregressive (VAR) models. Consider the aforementioned process $V(t) = (V_1(t), V_2(t), V_3(t))$ as a vector autoregressive model of order $p$. To define the measure, the unconstrained and constrained models need to be defined. The unconstrained process $V(t) = (V_1(t), V_2(t), V_3(t))$ is stationary and has a $VAR(\infty)$ representation:

$$V(t) = \sum_{j=1}^{\infty} \Phi_j V(t-j) + u(t) \tag{54}$$

Having the realizations of $\{V(1), ..., V(T)\}$, the unconstrained process can be approximated by a finite order $VAR(k)$:

$$V(t) = \sum_{j=1}^{k} \Phi_{jk} V(t-j) + u_k(t) \tag{55}$$

To estimate the causality measure from $V_2(t)$ to $V_1(t)$, the constrained process $V_0(t) = (V_1(t), V_3(t))$ needs to be defined. Same as the unconstrained process, $V_0(t)$ has a $VAR(\infty)$ representation:

$$V_0(t) = \sum_{j=1}^{\infty} \overline{\Phi}_j V_0(t-j) + \varepsilon(t) \tag{56}$$

that can, in turn, be approximated by a finite order $VAR(k)$:

$$V_0(t) = \sum_{j=1}^{k} \overline{\Phi}_{jk} V_0(t-j) + \varepsilon_k(t) \tag{57}$$

To estimate the causality measure, 4 sets of variables should be estimated: the autoregressive coefficients of the unconstrained model, the autoregressive coefficients of the constrained model, the variance-covariance matrix of the error term $u_k(t)$, and the variance-covariance matrix of the error term $\varepsilon_k(t)$. Thus, we need the estimation of the following variables:

$$\Phi_{jk} = [\phi_{1k}, \phi_{2k}, ..., \phi_{kk}] \tag{58}$$

$$\Sigma_{u|k} \tag{59}$$

$$\overline{\Phi}_{jk} = [\overline{\phi}_{1k}, \overline{\phi}_{2k}, ..., \overline{\phi}_{kk}] \tag{60}$$

$$\Sigma_{\varepsilon|k} \tag{61}$$

In the unconstrained process, the estimated variance-covariance matrix of the forecast errors of $V(t+h)$ is:

$$\widehat{\Sigma}_k(h) = \sum_{j=0}^{h-1} \widehat{\Psi}_{jk} \widehat{\Sigma}_{u|k} \widehat{\Psi}'_{jk} \tag{62}$$

where $\widehat{\Psi}_{jk} = \widehat{\phi}^j_{1k}$, $\widehat{\phi}^0_{1k} = I_m$, $\widehat{\phi}^1_{1k} = \widehat{\phi}_1$, $\widehat{\phi}^{(j+1)}_{1k} = \widehat{\phi}^{(j)}_{2k} + \widehat{\phi}^{(j)}_{1k}\widehat{\phi}_{1k}$.

Similarly, for the constrained model:

$$\widetilde{\Sigma}_0(h) = \sum_{j=0}^{h-1} \widetilde{\Psi}_{jk} \widetilde{\Sigma}_{\varepsilon|k} \widetilde{\Psi}'_{jk} \tag{63}$$

Thus, the estimator of the causality measure from $V_2(t)$ to $V_1(t)$ is:

$$\widehat{GC^h}(V_2 \to V_1|I) = \ln[\frac{\det\{[J_0 \widetilde{\Sigma}_0(h) J'_0]\}}{\det\{[J_1 \widehat{\Sigma}_k(h) J'_1]\}}] \tag{64}$$

with $J_1 = [I_{m_1} 0 \;\; 0]$ and $J_0 = [I_{m_1} \;\; 0]$.

Due to the high dimension of the data (a large number of firms in each group), instead of estimating the measure for clustered volatility series, I use the estimated centroids $\widehat{c}_g(t)$ for each group. For example, in the case of $G^* = 10$, there are 10 estimated centroids

$\widehat{c}_g(t) = \{\widehat{c}_1(t), ..., \widehat{c}_{10}(t)\}$ that can be used in a $VAR(k)$ model. The centroids are the estimated realized volatility series that represent each group of firms. The vector autoregressive models are estimated by OLS, equation by equation. In the case of $G^* \geq 10$, LASSO (Tibshirani, 1996) is needed to shrink the dimension.

Now that I have the estimated causality measures, I can create a table that represents the linkage (also known as volatility spillover) between nodes. Therefore, the realized volatility network is:

$$\begin{bmatrix} \widehat{GC}^h_{11}(t) & \cdots & \widehat{GC}^h_{1G^*}(t) \\ \vdots & \ddots & \vdots \\ \widehat{GC}^h_{G^*1}(t) & \cdots & \widehat{GC}^h_{G^*G^*}(t) \end{bmatrix} \tag{65}$$

where $\widehat{GC}^h_{1G^*}(t)$ is the causality from group 1 to group $G^*$ at time horizon $h$ and it measures the volatility spillover (directed edge) from node 1 to node $G^*$ in the network. To obtain $GC^h_{ii}(t)$ for $i = 1, ..., G^*$ past information of node $i$ is removed from the information set. Thus, the causality from $i$ to $i$ seems to capture the relevance of the past observations in each node. Each row of the causality table shows the causality from one group to another. These measures are transformed into weighted measures by defining the following relationship:

$$w^h_{ij}(t) = \frac{\widehat{GC}^h_{ij}(t)}{1 + \widehat{GC}^h_{ij}(t)} \tag{66}$$

As a result, we have the following weighted volatility network:

$$\begin{bmatrix} w^h_{11}(t) & \cdots & w^h_{1G}(t) \\ \vdots & \ddots & \vdots \\ w^h_{G1}(t) & \cdots & w^h_{GG}(t) \end{bmatrix} \tag{67}$$

As mentioned earlier, the sparsity factor reveals the evolution of edges. Therefore, it needs to be obtained by the causality table. Similar to Herskovic (2018), the sparsity factor is specified as the following:

$$NS = \sum_{i=1}^{G^*} \delta_i(t) \sum_{j=1}^{G^*} w_{ij}(t) \log w_{ij}(t) \tag{68}$$

This factor measures the thickness of the edges ($w_{ij}$). In other words, sparsity is the negative entropy of weights, scaled by the share of each node.

## *2.4 Evidence*

This section exhibits the main results. Section 4.1 studies the factors in the Arbitrage Pricing Theory and estimates pricing errors. Section 4.2 examines the sorted portfolios created by exposure of assets to the network factors. Section 4.3 estimates the price of risk by using two-stage regressions of Fama and Macbeth (1973). And finally, section 4.4 discusses the robustness checking.

Figure (2) shows the time series of concentration (panel 1) and sparsity (panel 2) computed by (47) and (68), respectively. Concentration appears to be responsive to the financial crisis as it is evident by a sharp drop in the graph during the 2007-2009 period. This can be explained by the loss of market capitalization and the fact that the network transforms into one with more evenly distributed node sizes. Sparsity is more difficult to interpret. The graph seems to be less volatile after 2000 with the exception of a hit in 2005. Network graphs that highlight the evolution of the factors during the financial crisis are presented in the appendix.

Figure 2: Concentration and Sparsity

The figure shows the time series of concentration and sparsity. Panel 1 is obtained by calculating (47), and panel 2 by calculating (68) in annual frequency. The time-domain starts in 1980 and ends in 2017. For concentration $G_{max}$ is 15, for sparsity $VAR$ is of order 1, and the first horizon is chosen to obtain the forecast errors.

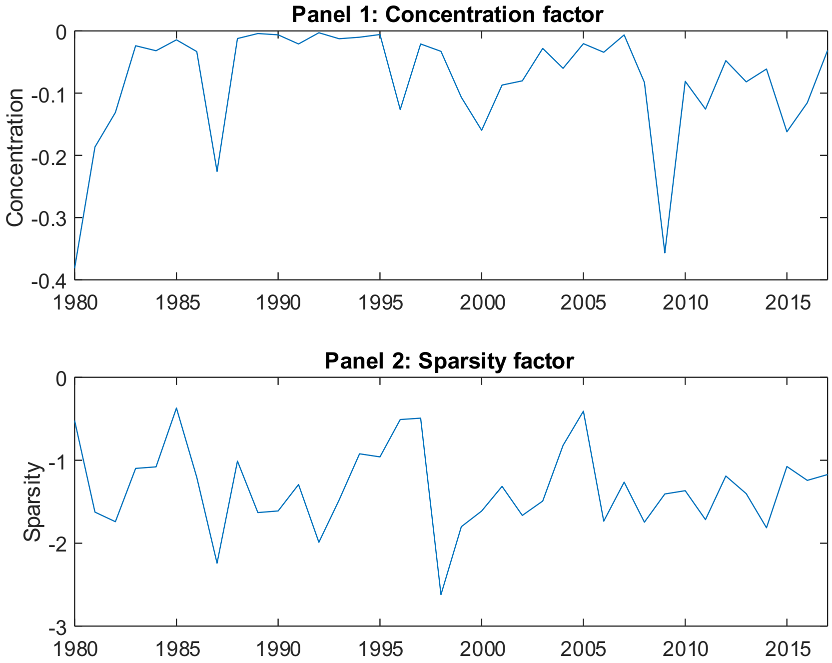


Table 21: Summary Statistics and Factor Correlation

Panel 1 is a summary statistics of concentration and sparsity. Autocorrelations in lag p and their p-values of the Ljung-Box Q-test are reported in front of AC(p) for p=1,2,3, and 4. The null hypothesis is no autocorrelation. Panel 2 shows the correlation coefficients and their p-values in parentheses for network factors. Panel 3 demonstrates the correlations between concentration and sparsity and the Fama-French 3 factors, namely, market factor (Mkt), size (smb), and book/market (hml). I add the two factors of concentration and sparsity presented in Herskovic (2018).

Panel 1

Summary Statistics of Network Factors

| | $NC$ | | $NS$ | |
|---|---|---|---|---|
| | $Level$ | $p-value$ | $Level$ | $p-value$ |
| $Mean$ | -0.07 | - | -1.32 | - |
| $S.D.$ | 0.08 | - | 0.50 | - |
| $AC(1)$ | 0.18 | (0.38) | 0.09 | (0.16) |
| $AC(2)$ | 0.08 | (0.50) | -0.26 | (0.20) |
| $AC(3)$ | -0.09 | (0.57) | -0.14 | (0.19) |
| $AC(4)$ | -0.10 | (0.68) | -0.17 | (0.29) |

Panel 2

Network Factors Correlation

| | $NC$ |
|---|---|
| $NS$ | 0.03 |
| | (0.83) |

Panel 3

All Factors Correlation

| | $Mkt$ | $smb$ | $hml$ | $HNC$ | $HNS$ |
|---|---|---|---|---|---|
| $NC$ | 0.04 | -0.15 | -0.21 | 0.14 | -0.35 |
| | (0.79) | (0.33) | (0.20) | (0.39) | (0.03) |
| $NS$ | 0.25 | -0.12 | 0.04 | -0.08 | 0.01 |
| | (0.11) | (0.44) | (0.78) | (0.59) | (0.95) |

Table (21) reports the summary statistics and correlations matrices. Upon examination of autocorrelations and p-values in panel 1, the null hypothesis of no autocorrelations in the Ljung-Box Q-test cannot be rejected. Network factors show no sign of correlation in panel 2, and the evidence of any co-movement between network factors and the Fama-French 3 factors, namely market, size, and book/market is low. I add Herskovic (2018) factors ($HNC$ and $HNS$) at the end. $NC$ demonstrates a correlation of -0.35 with $HNS$. The fact that $NS$ and $NC$ are not correlated might imply that they are two distinct sources of risk. In section 4.4, I will address the possible correlations by controlling for various factors.

### 2.4.1 The APT Pricing Errors

According to the Arbitrage Pricing Theory (APT), an asset return is a linear function of a set of factors and an error term. Here, I use the notation and description presented in Geweke and Zhou (1996). To show the APT model in vector form, consider the following:

$$\mathrm{r}(t) = \alpha + \beta \mathrm{f}(t) + \epsilon(t) \tag{69}$$

where $\mathrm{r}(t)$ is the vector of asset returns, $\alpha$ is the intercept or the expected return of any asset, $\mathrm{f}(t)$ is the set of explanatory factors, $\beta$ is the vector of factor loadings, and $\epsilon(t)$ is the idiosyncratic component. The APT has the following assumptions:

- $\mathrm{E}[\mathrm{f}(t)] = 0$
- $\mathrm{E}[\mathrm{f}(t)\mathrm{f}'(t)]=\mathrm{I}$
- $\mathrm{E}[\epsilon(\mathrm{t})|\mathrm{f}(t)] = 0$
- $\mathrm{E}[\epsilon(t)\epsilon'(\mathrm{t})|\mathrm{f}(t)] = \Sigma$

According to Ross (1977), in the absence of riskless arbitrage opportunities, the expected return of asset of $i$ is:

$$\alpha_i \approx \lambda_0 + \beta_{i1}\lambda_1 + \cdots + \beta_{ik}\lambda_k \tag{70}$$

The focus is on the pricing error which is an average of the squared pricing errors across assets. Therefore:

$$p^2 = \frac{1}{N}\sum_{i=1}^{N}(\alpha_i - \lambda_0 - \beta_{i1}\lambda_1 - \cdots - \beta_{ik}\lambda_k)^2 \tag{71}$$

where $p$ is the pricing error of asset $i$ and $\lambda_k$ is the risk premium on the k-th factor. As the number of assets approaches to zero, $p$ converges to zero as well. Geweke and Zhou (1996) further show that conditional of $\alpha$ and $\beta$, the minimized average pricing error is:

$$\mathrm{p}^2 = \frac{1}{N}\alpha'[\mathrm{I}_N - \beta^*(\beta^{*\prime}\beta^*)^{-1}\beta^{*\prime}]\alpha \tag{72}$$

where $\beta^* = (1_N, \beta)$ and $\beta$ is an $N \times K$ matrix of factor loadings.

Given the above description of the pricing errors, the accuracy of the factor model can be estimated. I consider 12 industry portfolio returns [3] (available in Kenneth French's data library) to examine the errors. For every portfolio, the excess returns are regressed on a constant and the network factors. Since the network has two factors, four regressions are considered. The first one has no factor. The second includes only the concentration factor. The third is similar to the second but with sparsity as the explanatory variable. And finally, the fourth regression considers both factors.

$$\begin{aligned}
r_i(t) &= \alpha_{0i} + \nu_i(t) && (73)\\
r_i(t) &= \alpha_{1i} + \beta_i^{NC} NC(t) + \xi_i(t) && (74)\\
r_i(t) &= \alpha_{2i} + \beta_i^{NS} NS(t) + \delta_i(t) && (75)\\
r_i(t) &= \alpha_{3i} + \beta_i^{NC} NC(t) + \beta_i^{NS} NS(t) + \epsilon_i(t) && (76)
\end{aligned}$$

To obtain a series of pricing errors, the estimation is done on a rolling window. For every $t$ and every industry portfolio, the regression is done over a 15-year window from t-14 to t. Then, the process is repeated until the last year of the sample.

[3] The summary statistics of the portfolios are available in the appendix.

Table (22) presents the results. With no factor (K=0), the pricing error is 0.1333. Once the concentration factor is entered into the model, the error drops to 0.0254. The same is true for sparsity. With the presence of sparsity only, the pricing error falls to 0.0259. However, when the two are put together in the regression, there does not seem to be a significant drop in the pricing error. This fact is in alignment with Geweke and Zhou's (1996) empirical findings which assert that introducing more factors does not necessarily reduce the pricing error by a significant amount.

Comparing the errors across a broader set of factors, I include concentration and sparsity in Herskovic (2018). $HNC$ seems to do a worse job compare to $NC$, as it drops the errors to 0.026. But $HNS$ is more successful than $NS$. $HNC$ and $HNS$ together appear to perform marginally better than $NC$ and $NS$.

Table 22: Average Pricing Errors for the Network Factors

The table reports the average pricing errors for different factors based on the Arbitrage Pricing Theory. The set of test assets is 12 industry annual portfolio returns available on Kenneth French's website. K indicates the number of factors in the model. p reports the pricing error, S.E. indicates the standard errors, and the last column demonstrates the 95 percent confidence interval.

| Industry returns (N = 12) | | | |
|---|---|---|---|
| K | p | S.E. | 95% C.I. |
| 0 | 0.1333 | 0.0329 | [0.1197, 0.1469] |
| $NC$ | 0.0254 | 0.0055 | [0.0232, 0.0277] |
| $NS$ | 0.0259 | 0.0055 | [0.0236, 0.0282] |
| $NC$ & $NS$ | 0.0244 | 0.0055 | [0.0221, 0.0266] |
| $HNC$ | 0.0262 | 0.0058 | [0.0238, 0.0286] |
| $HNS$ | 0.0236 | 0.0051 | [0.0213, 0.0255] |
| $HNC$ & $HNS$ | 0.0229 | 0.0049 | [0.0207, 0.0247] |

To investigate the proposed factors further, in the next section I examine portfolios created by exposure of stocks to the factors and then estimate the factor prices of risk and compare them to concentration and sparsity in Herskovic (2018).

### 2.4.2 Sorted Portfolios

Creating portfolios based on the sensitivities of asset returns to financial factors is a method to test if equities with different exposures to the common risk factors have different returns. To present this formally, consider the following:

$$r_i(t) = \alpha_{3i} + \beta_i^{NC} NC(t) + \beta_i^{NS} NS(t) + Controls + \epsilon_i(t) \tag{77}$$

This model is similar to (73) except that here, there exist control variables. The controls will be investigated in the robustness analysis later. $r_i(t)$ is the time series of excess returns of firm $i$. $NC(t)$ and $NS(t)$ are time series of concentration and sparsity, respectively. Finally, $\beta_i^{NC}$ and $\beta_i^{NS}$ measure the exposure of stock returns to the two factors.

For the equity returns, I consider all the common stocks with shares codes of 10, 11, and 12 in CRSP. Penny stocks (stocks with price less than \$5) are then removed from the series. To create portfolios, a rolling window regression is utilized. In a window length of 15 years and at every period $t$, I run the above regression from $t - 14$ to $t$. The equities are then sorted by each beta and portfolios are formed at $t + 1$. I repeat this process until the end of the period. The annual risk-free rate is downloaded from Kenneth French's data library.

Table (23) reports the statistics for tercile portfolios sorted by $\beta^{NC}$ and $\beta^{NS}$. Each tercile has an ID that is shown in front of "Rank". $H - L$ refers to the spread between the high beta portfolio (tercile 3) and the low beta portfolio (tercile 1). The next four rows report the mean, standard deviation, log of market capitalization, and book-to-market ratio of each tercile. In panel 1, the portfolios formed on high concentration beta stocks have a %2.15 lower return on average than portfolios formed on low concentration stocks (Column $(H - L)$). The t-statistics shows that the spread is significant and the negative sign affirms a negative price of risk for concentration. Upon examination of panel 2, I arrive at similar conclusions. There exists a significant spread of %0.83 between high sparsity and low sparsity portfolios annually.

$\alpha_{CAPM}$ and $\alpha_{FF3}$ measure the intercepts of the capital asset pricing and Fama-French three-factor models, respectively. In panel 1, both alphas are significant, meaning that the CAPM and Fama-French three-factor models are not able to explain the spread. The same is true for sparsity in panel 2. Notwithstanding a lower t-statistics compared to the spread in concentration, both models fail to explain the spread in high-minus-low investment strategies.

Table 23: Portfolio Sorting by Exposure to Concentration and Sparsity

The table reports the results of the univariate sorting by exposure to the concentration and sparsity factors. To estimate the betas the following regression is run: $ER_i(t) = \alpha_i + \beta_i^{nc} NC(t) + \beta_i^{ns} NS(t) + \epsilon_i(t)$. To construct the terciles, excess returns of common stocks are regressed on the sparsity and concentration factors from $t-14$ to $t$ . The betas are then sorted and portfolios are constructed at $t+1$ . Mean displays the average return for each portfolio. Std. Dev. reports the standard deviations of returns. Size is log of average market capitalization within portfolio. Book/Market is the average book to market value ratio. $\alpha_{CAPM}$ and $\alpha_{FF3}$ demostrate the $\alpha$ of the capital asset pricing and Fama-French 3 factor models, respectively. Pre-formation $\beta$ indicates the average $\beta$ of portfolio formation periods. H-L (high minus low) shows the long-short portfolio strategy. Student t-statistics are reported in square brackets.

Panel 1

Sorting by Concentration

| | Tercile | | | $H-L$ | |
|---|---|---|---|---|---|
| Rank | 1 | 2 | 3 | 3 - 1 | t-stat |
| Mean(%) | 4.55 | 3.65 | 2.39 | -2.15 | [-6.87] |
| Std. Dev. | 2.95 | 2.45 | 2.85 | - | - |
| Size | 4.56 | 5.16 | 5.22 | - | - |
| Book/Market | 0.70 | 0.71 | 0.70 | | |
| $\alpha_{CAPM}$ | 2.34 | 1.26 | 0.15 | -4.53 | [-10.12] |
| $\alpha_{FF3}$ | 2.14 | 1.15 | -0.07 | -2.19 | [-6.26] |
| Pre-Formation $\beta$ | -0.73 | -0.11 | 0.34 | - | - |

Panel 2

Sorting by Sparsity

| | Tercile | | | $H-L$ | |
|---|---|---|---|---|---|
| Rank | 1 | 2 | 3 | 3 - 1 | t-stat |
| Mean(%) | 3.98 | 3.47 | 3.14 | -0.83 | [-1.90] |
| Std. Dev. | 3.34 | 2.52 | 2.46 | - | - |
| Size | 4.87 | 5.21 | 4.96 | - | - |
| Book/Market | 0.72 | 0.71 | 0.70 | | |
| $\alpha_{CAPM}$ | 1.80 | 1.11 | 0.84 | -3.29 | [-5.83] |
| $\alpha_{FF3}$ | 1.49 | 1.02 | 0.70 | -0.95 | [-1.96] |
| Pre-Formation $\beta$ | -0.07 | -0.01 | 0.03 | - | - |

Following Ang et al. (2006), I report the pre-formation betas in the last row. These are the average values of coefficients ($\beta^{NC}$ and $\beta^{NS}$) in each rolling regressions in each tercile. For both panels, they are monotonically increasing. However, the evidence is more convincing for concentration as we observe an increase of 1.07 (from -0.73 in tercile 1 to 0.34 in tercile 3) compared to a rise of 0.1 (from -0.07 to 0.03) for sparsity.

### 2.4.3 Measuring the Price of Risk

Table (23) establishes that the spread between portfolios created by stocks with high and low concentration (sparsity) betas cannot be explained by the CAPM and Fama-French 3-factor models. To proceed with investigating these two factors further, the next step is estimating their corresponding prices of risk. To do so, I estimate the two-stage regression of Fama and Macbeth (1973). This is a well-known routine to estimate the risk premia for any risk factor that is assumed to have explanatory power in the cross-section of asset returns. The regression has two steps. In the first step, I regress each asset's returns on the risk factor to estimate the factor betas ($\beta$) in a time series regression. In the second step, I regress all assets' returns on the estimated betas obtained in the first step in a panel regression to estimate the price of risk for each factor ($\lambda$).

I create three sets of test assets to be used in the Fama-Macbeth regressions. The first set is the panel of 25 (5 by 5) portfolios double sorted on the market and concentration factor. Common stocks are regressed on the market and concentration factors. Then, quintile portfolios are created based on the betas of the market factor. Next, in each quintile, another set of quintile portfolios are created by the exposure of each group to the concentration factor. That creates a diverse set of portfolios that can be used in the estimation of the price of risk.

Table 24: The Price of Concentration Risk

The table shows the price of risk for the concentration factor using the two-stage regression of Fama-Macbeth. The test assets are 25 portfolios sorted on the market and concentration factors. Models (1) to (4) refer to different sets of explanatory variables in the Fama-Macbeth regressions. Model (1) has the market and concentration factor mimicking portfolios (Fm-NC) as explanatory variables. Model (2) adds Fama-French two factors of $smb$ and $hml$. Model (3) adds the network concentration of Herskovic, and finally (4) excludes the traditional Fama-French factors and examines the effects of concentration and sparsity in Herskovic (2018) on the test assets. The last row reports the adjusted R-square. Student t-statistics are in square brackets.

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Constant | 0.0032 | 0.0021 | 0.0048 | -0.0025 |
| | [0.48] | [0.28] | [0.62] | [-0.33] |
| $Mkt$ | 0.0136 | 0.0040 | -0.0346 | |
| | [0.14] | [0.04] | [-0.38] | |
| $smb$ | | 0.0272 | 0.0009 | |
| | | [0.95] | [0.02] | |
| $hml$ | | -0.0990 | -0.0850 | |
| | | [-1.78] | [-1.53] | |
| $HNC$ | | | 0.0319 | 0.0147 |
| | | | [0.44] | [0.28] |
| $HNS$ | | | | 0.0949 |
| | | | | [1.74] |
| $Fm\_NC$ | -0.0140 | -0.0148 | -0.0146 | -0.0157 |
| | [-4.09] | [-3.81] | [-3.75] | [-4.09] |
| $Adjusted\ R^2$ | 0.43 | 0.51 | 0.55 | 0.58 |

Table 25: The Price of Sparsity Risk

The table shows the price of risk for the sparsity factor using the two-stage regression of Fama-Macbeth. The test assets are 25 portfolios sorted on the market and sparsity factors. Models (1) to (4) refer to different sets of explanatory variables in the Fama-Macbeth regressions. Model (1) has the market and sparsity factor mimicking portfolios $Fm_NS$ as explanatory variables. Model (2) adds Fama-French two factors of $smb$ and $hml$. Model (3) adds the network concentration of Herskovic, and finally (4) adds network sparsity of Herskovic. The last row reports the adjusted R-square. Student t-statistics are in square brackets.

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Constant | 0.0007 | 0.0033 | 0.0001 | -0.0048 |
| | [0.08] | [0.38] | [0.01] | [-0.63] |
| $Mkt$ | -0.0090 | -0.0607 | -0.1126 | |
| | [-0.08] | [-0.57] | [-1.38] | |
| $smb$ | | -0.0124 | -0.0073 | |
| | | [-0.52] | [-0.33] | |
| $hml$ | | -0.0440 | -0.0073 | |
| | | [-0.69] | [-0.16] | |
| $HNC$ | | | | -0.0670 |
| | | | | [-1.33] |
| $HNS$ | | | 0.0685 | 0.0957 |
| | | | [1.20] | [2.10] |
| $Fm\_NS$ | -0.0143 | -0.0173 | -0.0195 | -0.0181 |
| | [-1.93] | [-1.93] | [-2.48] | [-2.78] |
| $Adjusted\ R^2$ | 0.20 | 0.28 | 0.29 | 0.29 |

Table (24) reports the Fama-Macbeth regression results for 25 portfolios sorted on the market and concentration. Each column corresponds to a different model specification. $Mkt$, $smb$, and $hml$ refer to the three factors in the Fama-French model ($Mkt$ alone in a model indicates the CAPM). $HNC$, and $HNS$ point out to the concentration and sparsity in Herskovic (2018), respectively. $Fm\_NC$ is the concentration factor mimicking portfolio I created by going long the portfolio that has high beta (tercile 3) and short the portfolio that has low beta (tercile 1). Model (1) presents the results of the regression of the test assets on $Mkt$ and $Fm\_NC$. Market is not significant, but the factor mimicking portfolio is and carries a negative sign, consistent with the negative spread found in table (23). Model (2) adds size and book/market. None is significant except for the concentration mimicking portfolio. The results do not change when I add the concentration factor of Herskovic (2018) in model (3). Finally, model (4) isolates only concentration and sparsity in Herskovic (2018) and examines them along with the concentration mimicking portfolio. Across all the models, the factor mimicking portfolio preserves its sign and appears to have the same magnitude and remains significant.

In table (25), I repeat the same procedure for the sparsity factor. Adjusted R-square is relatively lower than the previous case, inferring the lower explanatory power of sparsity mimicking portfolio. The value of the estimated coefficient of sparsity mimicking portfolio remains significant and negative in all the models. However, in model (4) $HNS$ appears significant with the correct sign. $HNC$ is not significant, but it has the correct negative sign in model (4) as well.

To explore further the price of risk, I construct more test assets. In table (26), the assets are 25 portfolios double sorted by $HNC$ and $HNS$. Model (1) regresses the assets on the market factor and concentration and sparsity mimicking portfolios. All variables appear to be significant with the correct sign. Model (2) adds Fama-French three factors into the model. Still, factor mimicking portfolios and market are significant, but size and book/market do not show explanatory powers. Model (3) keeps the market factor and puts together two concentration and two sparsity factors. My proposed factor mimicking portfolios are significant with negative signs. $HNC$ of Herskovic (2018) has a low t-value, but $HNS$ has the highest absolute t-value, confirming the results in Herskovic (2018).

Table 26: The Price of Risk: All Factors

The table reports the price of risk of factors for real and financial networks by using the two-stage regression model of Fama-Macbeth. The test assets are 25 portfolios sorted on the network concentration and sparsity introduced by Herskovic (2018). Models (1) to (4) refer to different sets of explanatory variables in the Fama-Macbeth regression model. Model (1) includes the market and concentration factor mimicking portfolio. Model (2) adds in the 3 factor model of Fama-French. Model (3) enters the concentration of Herskovic, and finally model (4) includes the sparsity factor of Herskovic as well. The final row reports the adjusted R-square. T-statistics are shown in square brackets.

| | (1) | (2) | (3) |
|---|---|---|---|
| Constant | -0.0060 | -0.0017 | -0.0081 |
| | [-0.89] | [-0.26] | [-1.07] |
| $Mkt$ | 0.2319 | 0.1811 | 0.3137 |
| | [3.00] | [2.64] | [3.53] |
| $smb$ | | 0.0048 | |
| | | [0.16] | |
| $hml$ | | -0.1065 | |
| | | [-2.18] | |
| $HNC$ | | | 0.0424 |
| | | | [1.39] |
| $HNS$ | | | 0.0466 |
| | | | [4.92] |
| $Fm\_NC$ | -0.0184 | -0.0148 | -0.0207 |
| | [-2.60] | [-2.01] | [-2.83] |
| $Fm\_NS$ | -0.0190 | -0.0235 | -0.0277 |
| | [-2.31] | [-2.74] | [-3.59] |
| $Adjusted\ R^2$ | 0.57 | 0.54 | 0.57 |

### 2.4.4 Robustness Checking

To check for the robustness of the main results, first, I examine the spread in long-short strategies for different choices of control variables in (77). Tables (28), (29), and (30) in the appendix present the portfolio sorting outcomes when I control for market, book/market, and size. Controlling for the market, the spread in the $H - L$ strategy declines slightly for both factors, but sparsity is no longer significant. However, the signs remain the same meaning that the high-beta-factor portfolios earn a lower return on average. The outcome is similar when controlling for book/market. Assets sorted by concentration betas keep their spread returns, pre-formation betas appear to be spread widely among terciles, and the alpha in long-short portfolios seems to be inexplicable by the CAPM and Fama-French three-factor models. Examining sparsity, the spread shrinks more and $H - L$ strategy is no longer significant. Although alpha in the spread cannot be explained by CAPM.

In table (30) the outcomes are similar to the original sorting portfolios in table (23). Both spreads are negative and significant, $\alpha_{CAPM}$ and $\alpha_{FF3}$ are not able to justify the presence of a negative return between the last and the first terciles. Overall, the evidence of the existence of annual average negative return in the spread is ample, although it is more convincing and robust for the network concentration.

Additionally, I check for different forecast horizons in the causality measure. Table (31) controls for $h$ in the causality table. The spread in portfolios sorted by sparsity seems to be more significant for the 5, 10, and 15 horizons, compared to 1 in my main results.

Finally, the last piece of robustness checking controls for the window length in the rolling window regressions. The length can take 16, 17, 18, 19, and 20 years. Panel 1 and 2 of the table (32) demonstrate the long-short portfolio returns ($H - L$) for different choices of the window. The strategies are relatively significant with the correct signs for both factors. As it is expected, portfolios sorted by concentration earn higher spread than sparsity across the different windows.

## *2.5 Concluding Remarks*

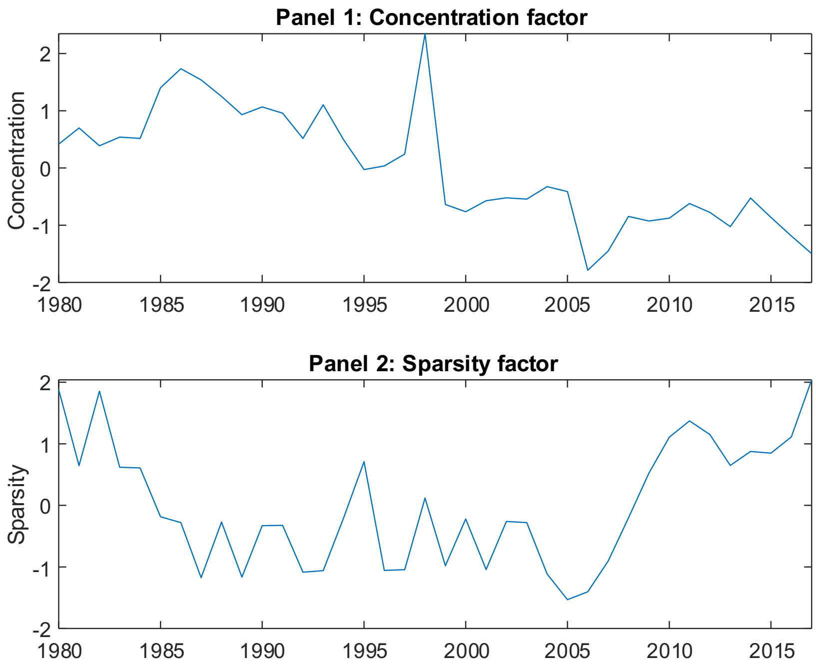
Firms are connected through links that are less transparent in financial markets. This study is an attempt to uncover such linkages in a network and study their properties. By using the U.S. firm-customer links data from Compustat and their corresponding return data from CRSP, I design a network of realized volatility from which two factors are recovered. The concentration factor depicts the node-property, whereas the sparsity factor describes the edge-characteristics. In the APT setting, both factors have considerable potential to reduce pricing errors. Upon further examinations of portfolios sorted by the exposures of common stocks to concentration and sparsity, results indicate significant negative average returns in high-minus-low investment strategies.

## 2.6 *Appendix*

### 2.6.1 Concentration and Sparsity in Herskovic (2018)

To replicate the factors in Herskovic, I used the updated data-set by the author which is the U.S. firm-customer links from 1980 to 2013, and then I upgraded the data until 2017. The data-set contains PERMNOs of the U.S. firms and their customers that purchase at least %10 of their final output. To calculate the factors, firms need to be aggregated by the North American Industry Classification System (NAICS). However, NAICS is quite a new system and most of the firms are still identified by the Standard Industrial Classification (SIC). Therefore, I converted SIC codes to NACIS and then aggregated the firms into two-digit sector codes[4]. Figure (3) shows the replicated time series obtained by (36) and (37).

Figure 3: Replication of Concentration and Sparsity in Herskovic (2018)




[4]For the direct relationship between SIC and NAICS visit https://www.census.gov/eos/www/naics/concordances/concordances.html

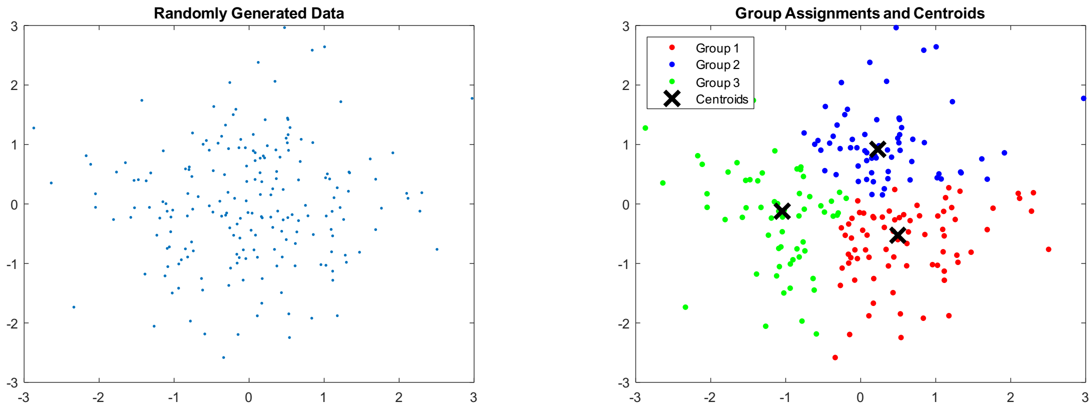
### 2.6.2 Kmeans estimation algorithm

The iterative algorithm that solves (42), alternates between two steps:

- Start with the initial values.
- Assignment step: given an initial set of groups, assign each observation to the closest (Euclidean distance) group assignment (centroid).
- update step: calculate the means of the observations in each group and call them the new centroids.
- Repeat until convergence (when group assignment stops changing).

Figure 4 sets an example for the iterative algorithm. The left panel demonstrates two-dimensional random data. After performing the kmeans algorithm for 1000 times, the final outcome is present in the right panel. Here, the pre-specified group number is 3 and the center of groups (centroids) are emphasized by a cross for each group (color). The estimation is done by the built-in MATLAB function, *Kmeans*.

Figure 4: Example of Using Kmeans to Group the Data



### 2.6.3 Complementary Tables and Figures

Table (27) reports the summary statistics of 12-industry portfolios that are used in calculating the pricing errors. Tables (28), (29), and (30) present the robustness checking for sorted portfolios, controlling for market, book/market, and size, respectively. Table (32) does robustness checking to the choice of the window length in the rolling window analysis that was carried out in the evidence. Figures (5) and (6) at the end of the appendix portray two network graphs that visualize the recent financial crisis.

Table 27: Summary Statistics of 12 Industry Portfolios

table reports statistics for 12 industry portfolios. The list of industries includes 1. consumer nondurables, 2. consumer durable, 3. manufacturing, 4. energy, 5. chemicals, 6. business equipment, 7. telecom, 8. utilities, 9. wholesale and retail, 10. healthcare, 11. finance, and 12. other.

| | | | Auto-correlation | | | |
|---|---|---|---|---|---|---|
| Variable | Mean | S.D. | AC(1) | AC(2) | AC(3) | AC(4) |
| Industry 1 | 14.71 | 22.53 | -0.155 | -0.260 | 0.289 | -0.065 |
| Industry 2 | 14.07 | 31.80 | -0.221 | -0.357 | 0.199 | 0.040 |
| Industry 3 | 15.10 | 23.52 | -0.255 | -0.327 | 0.221 | 0.035 |
| Industry 4 | 12.03 | 39.19 | -0.023 | -0.256 | 0.154 | 0.190 |
| Industry 5 | 15.34 | 23.50 | -0.279 | -0.280 | 0.318 | -0.121 |
| Industry 6 | 18.63 | 39.62 | -0.310 | -0.093 | -0.077 | 0.294 |
| Industry 7 | 19.02 | 41.12 | -0.307 | -0.099 | -0.172 | 0.273 |
| Industry 8 | 15.18 | 14.60 | -0.256 | -0.197 | 0.277 | -0.030 |
| Industry 9 | 14.57 | 28.22 | -0.193 | -0.195 | 0.096 | -0.119 |
| Industry 10 | 20.93 | 39.74 | -0.262 | -0.167 | -0.016 | 0.107 |
| Industry 11 | 17.04 | 22.98 | 0.108 | -0.057 | 0.102 | -0.061 |
| Industry 12 | 13.98 | 25.91 | -0.294 | -0.185 | 0.101 | -0.021 |

Table 28: Univariate Portfolio Sorting: Controlling for Market

The table reports the results of the univariate sorting by exposure to the sparsity and concentration factors after controlling for market factor. In order to estimate the betas, the following regression is run: $ER_i(t) = \alpha_i + \beta_i^{mkt} Mkt(t) + \beta_i^{nc} NC(t) + \beta_i^{ns} NS(t) + \epsilon_i(t)$. To construct the terciles, excess returns of common stocks are regressed on the factors from $t - 14$ to $t$ . The betas are then sorted and portfolios are constructed at $t + 1$ . Mean displays the average return for each portfolio. Std. Dev. reports the standard deviation of returns. Size is log of average market capitalization within portfolio. Book/Market is the average book to market value ratio. $\alpha_{CAPM}$ and $\alpha_{FF3}$ demostrate the $\alpha$ of the capital asset pricing and Fama-French 3 factor models, respectively. Pre-formation $\beta$ indicates the average $\beta$ of portfolio formation periods. H-L (high minus low) shows the long-short portfolio strategy. Student t-statistics are reported in square brackets.

Panel 1
Sorting by Concentration

| | Tercile | | | $H-L$ | |
|---|---|---|---|---|---|
| Rank | 1 | 2 | 3 | 3 - 1 | t-stat |
| Mean(%) | 4.49 | 3.53 | 2.57 | -1.91 | [-5.60] |
| Std. Dev. | 2.94 | 2.56 | 2.78 | - | - |
| Size | 4.61 | 5.16 | 5.20 | - | - |
| Book/Market | 0.73 | 0.72 | 0.70 | | |
| $\alpha_{CAPM}$ | 2.25 | 1.14 | 0.35 | -4.24 | [-7.61] |
| $\alpha_{FF3}$ | 2.03 | 1.00 | 0.16 | -1.89 | [-4.97] |
| Pre-Formation $\beta$ | -0.72 | -0.11 | 0.35 | - | - |

Panel 2
Sorting by Sparsity

| | Tercile | | | $H-L$ | |
|---|---|---|---|---|---|
| Rank | 1 | 2 | 3 | 3 - 1 | t-stat |
| Mean(%) | 3.83 | 3.48 | 3.27 | -0.56 | [-1.27] |
| Std. Dev. | 3.25 | 2.50 | 2.56 | - | - |
| Size | 4.91 | 5.20 | 4.93 | - | - |
| Book/Market | 0.70 | 0.71 | 0.70 | | |
| $\alpha_{CAPM}$ | 1.62 | 1.11 | 1.01 | -2.95 | [-4.89] |
| $\alpha_{FF3}$ | 1.33 | 1.01 | 0.86 | -0.61 | [-1.24] |
| Pre-Formation $\beta$ | -0.07 | -0.01 | 0.03 | - | - |

Table 29: Univariate Portfolio Sorting: Controlling for Book/Market

The table reports the results for the univariate sorting by exposure to the sparsity and concentration factors after controlling for the book/market factor. In order to estimate the betas, the following regression is run: $ER_i(t) = \alpha_i + \beta_i^{hml} hml(t) + \beta_i^{nc} NC(t) + \beta_i^{ns} NS(t) + \epsilon_i(t)$. To construct the terciles, excess returns of common stocks are regressed on the factors from $t - 14$ to $t$ . The betas are then sorted and portfolios are constructed at $t + 1$ . Mean displays the average return for each portfolio. Std. Dev. reports the standard deviation of returns. Size is log of average market capitalization within portfolio. Book/Market is the average book to market value ratio. $\alpha_{CAPM}$ and $\alpha_{FF3}$ demostrate the $\alpha$ of the capital asset pricing and Fama-French 3 factor models, respectively. Pre-formation $\beta$ indicates the average $\beta$ of portfolio formation periods. H-L (high minus low) shows the long-short portfolio strategy. Student t-statistics are reported in square brackets.

Panel 1
Sorting by Concentration

| | Tercile | | | $H-L$ | |
|---|---|---|---|---|---|
| Rank | 1 | 2 | 3 | 3 - 1 | t-stat |
| Mean(%) | 4.50 | 3.63 | 2.46 | -2.04 | [-5.41] |
| Std. Dev. | 2.82 | 2.41 | 3.12 | - | - |
| Size | 4.58 | 5.19 | 5.18 | - | - |
| Book/Market | 0.71 | 0.71 | 0.70 | | |
| $\alpha_{CAPM}$ | 2.27 | 1.22 | 0.26 | -4.36 | [-8.74] |
| $\alpha_{FF3}$ | 2.08 | 1.13 | -0.00 | -2.08 | -4.63 |
| Pre-Formation $\beta$ | -0.76 | -0.12 | 0.40 | - | - |

Panel 2
Sorting by Sparsity

| | Tercile | | | $H-L$ | |
|---|---|---|---|---|---|
| Rank | 1 | 2 | 3 | 3 - 1 | t-stat |
| Mean(%) | 3.79 | 3.50 | 3.31 | -0.48 | [-1.15] |
| Std. Dev. | 3.35 | 2.51 | 2.45 | - | - |
| Size | 4.92 | 5.20 | 4.92 | - | - |
| Book/Market | 0.74 | 0.73 | 0.73 | | |
| $\alpha_{CAPM}$ | 1.62 | 1.09 | 1.04 | -2.92 | [-4.57] |
| $\alpha_{FF3}$ | 1.33 | 1.02 | 0.85 | -0.47 | [-1.07] |
| Pre-Formation $\beta$ | -0.08 | -0.01 | 0.04 | - | - |

Table 30: Univariate Portfolio Sorting: Controlling for Size

The table reports the results for the univariate sorting by exposure to the sparsity and concentration factors after controlling for the size factor. In order to estimate the beta, the following regression is run: $ER_i(t) = \alpha_i + \beta_i^{smb} smb(t) + \beta_i^{nc} NC(t) + \beta_i^{ns} NS(t) + \epsilon_i(t)$. To construct the terciles, excess returns of common stocks are regressed on the factors from $t - 14$ to $t$ . The betas are then sorted and portfolios are constructed at $t + 1$ . Mean displays the average return for each portfolio. Std. Dev. reports the standard deviation of returns. Size is log of average market capitalization within portfolio. Book/Market is the average book to market value ratio. $\alpha_{CAPM}$ and $\alpha_{FF3}$ demostrate the $\alpha$ of the capital asset pricing and Fama-French 3 factor models respectively. Pre-formation $\beta$ indicates the average $\beta$ of portfolio formation periods. H-L (high minus low) shows the long-short portfolio strategy. Student t-statistics are reported in square brackets.

Panel 1
Sorting by Concentration

| | Tercile | | | $H-L$ | |
|---|---|---|---|---|---|
| Rank | 1 | 2 | 3 | 3 - 1 | t-stat |
| Mean(%) | 4.50 | 3.54 | 2.55 | -1.94 | [-5.00] |
| Std. Dev. | 2.97 | 2.55 | 2.84 | - | - |
| Size | 4.60 | 5.19 | 5.17 | - | - |
| Book/Market | 0.72 | 0.71 | 0.73 | | |
| $\alpha_{CAPM}$ | 2.34 | 1.17 | 0.23 | -4.46 | [-8.63] |
| $\alpha_{FF3}$ | 2.11 | 1.08 | 0.01 | -2.09 | [-4.60] |
| Pre-Formation $\beta$ | -0.68 | -0.07 | 0.40 | - | - |

Panel 2
Sorting by Sparsity

| | Tercile | | | $H-L$ | |
|---|---|---|---|---|---|
| Rank | 1 | 2 | 3 | 3 - 1 | t-stat |
| Mean(%) | 2.84 | 1.94 | 2.32 | -0.52 | [-2.04] |
| Std. Dev. | 3.22 | 2.16 | 2.69 | - | - |
| Size | 4.97 | 5.12 | 4.98 | - | - |
| Book/Market | 0.72 | 0.71 | 0.72 | | |
| $\alpha_{CAPM}$ | 0.70 | -0.40 | -0.00 | -3.05 | [-5.83] |
| $\alpha_{FF3}$ | 0.47 | -0.53 | -0.22 | -0.69 | [-2.44] |
| Pre-Formation $\beta$ | -0.07 | 0.00 | 0.05 | - | - |

Table 31: Univariate Portfolio Sorting: Controlling for Forecast Horizon

The table reports the results of long-short portfolio returns for different choices of the forecast horizon in the estimation of causality tables. The panel shows the average annual returns for high-minus-low investment strategies for sparsity. $\alpha_{CAPM}$ and $\alpha_{FF3}$ are the intercepts of the capital asset pricing and Fama-French 3-factor models. T-values are reported in square brackets.

Long-Short Portfolios Sorted by Sparsity

| $horizon$ | 5 | 10 | 15 |
|---|---|---|---|
| $H-L$ | -1.24 | -2.19 | -2.07 |
| | [-3.83] | [-5.10] | [-4.31] |
| $\alpha_{CAPM}$ | -2.76 | -3.80 | -3.61 |
| | [-6.66] | [-9.63] | [-8.41] |
| $\alpha_{FF3}$ | -1.12 | -2.24 | -2.18 |
| | [-3.06] | [-4.33] | [-3.94] |

Table 32: Univariate Portfolio Sorting: Controlling for Window Length

The table reports the results of long-short portfolio returns for different choices of window length in the rolling window regression used to sort portfolios. Panel 1 and 2 show the average annual returns for high-minus-low investment strategies for concentration and sparsity, respectively. $\alpha_{CAPM}$ and $\alpha_{FF3}$ are the intercepts of the capital asset pricing and Fama-French 3-factor models. T-values are reported in square brackets.

Panel 1: Concentration
Long-Short Portfolios

| $Window\ length$ | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|
| $H-L$ | -2.48 | -2.53 | -2.36 | -2.43 | -1.69 |
| | [-8.48] | [-8.82] | [-7.11] | [-6.97] | [-5.05] |
| $\alpha_{CAPM}$ | -4.70 | -4.56 | -4.20 | -4.11 | -3.12 |
| | [-9.84] | [-11.45] | [-11.15] | [-10.21] | [-7.59] |
| $\alpha_{FF3}$ | -2.49 | -2.35 | -2.25 | -2.44 | -1.60 |
| | [-7.12] | [-7.37] | [-6.19] | [-6.95] | [-4.28] |

Panel 2: Sparsity
Long-Short Portfolios

| $Window\ length$ | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|
| $H-L$ | -1.11 | -1.26 | -1.02 | -1.08 | -1.27 |
| | [-2.23] | [-2.49] | [-2.07] | [-2.17] | [-3.72] |
| $\alpha_{CAPM}$ | -3.37 | -3.33 | -2.98 | -2.86 | -2.87 |
| | [-5.78] | [-5.69] | [-5.30] | [-5.93] | [-6.64] |
| $\alpha_{FF3}$ | -1.00 | -1.10 | -0.78 | -0.81 | -1.17 |
| | [-1.83] | [-1.92] | [-1.37] | [-1.43] | [-3.02] |

Figure 5: Network Graph of 2007

The figure depicts the network graph for the year 2007. The number of groups ($G^*$) is 15. Each group is identified by an ID. Different sizes in nodes correspond to different market shares in (48). A thicker edge reveals a higher level of causality from a cluster to another in the realized volatility network. Groups 1 and 2 appear to have the highest and lowest shares in the network, respectively.

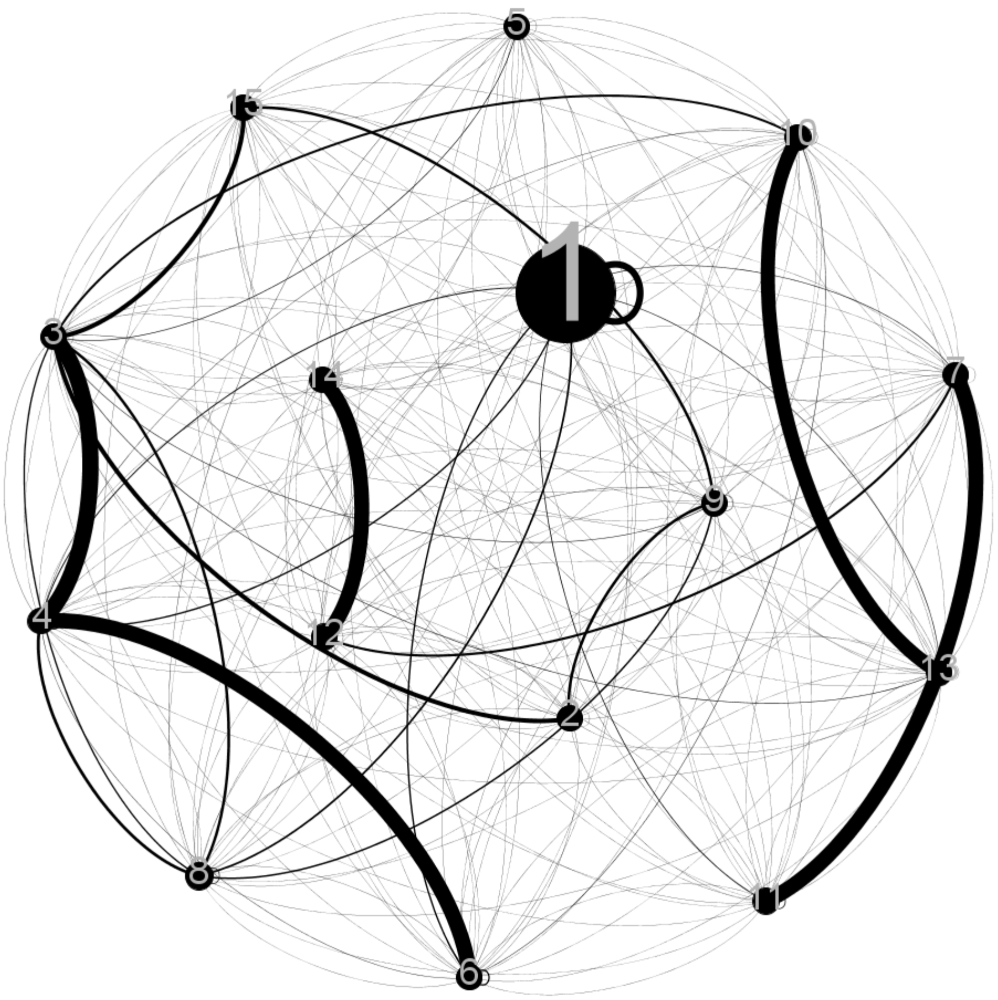
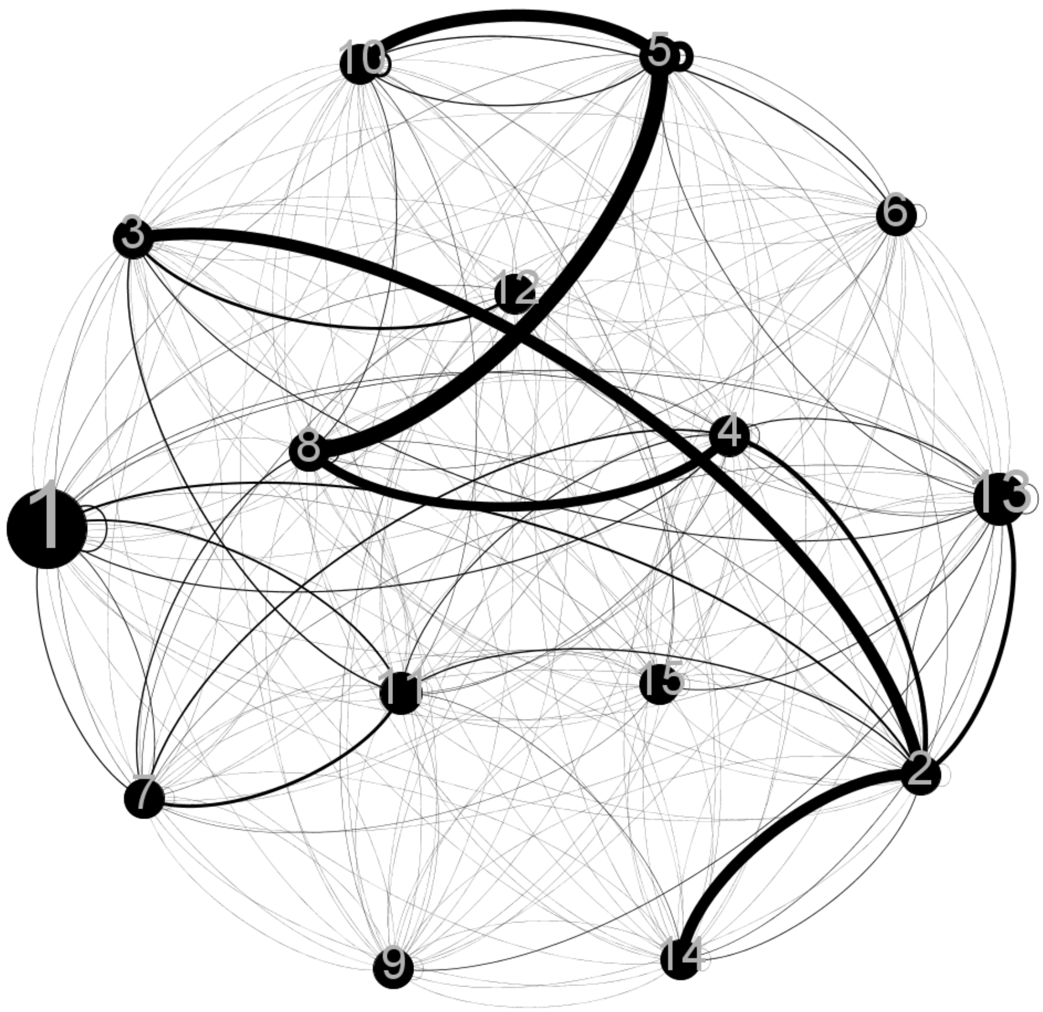


Figure 6: Network Graph of 2009

The figure depicts the network graph for the year 2009. The number of groups ($G^{*}$) is 15. Each group is identified by an ID. Different sizes in nodes correspond to different market shares in (9). A thicker edge reveals a higher level causality from a cluster to another in the realized volatility network.



# CHAPTER III

# A NETWORK OF COMMODITY REALIZED VOLATILITIES

## *3.1 Introduction*

Investigating the commodity futures market through the lens of the network theory is rather a new strand of literature. Commodity futures are a group of assets that have gained popularity in the 21st century due to the growth in their investment in recent years. Initially, they were considered as a segmented body of assets from the traditional markets such as the equities, but since more and more speculators are participating in commodity investing, some theories establish a close relationship between the commodities and the overall market movements. The theory of financialization explains the commodity price volatility due to the presence of speculators that spillover the effects from the equity market to the commodity market. Investigating the integration of these two markets will potentially contribute to the literature.

Another fairly new field of research is the network-based analysis of the financial markets. Assets are linked through unobservable linkages and spillover, and a network can help us uncover such links and explore their pricing implications. Networks are easy to understand and they can be visualized as graphs to help us better investigate the relationship among assets. In the commodity futures market, a node can be a single or a group of commodities. The connections among nodes are features that can, later on, be established based on the usual phenomena of a financial network such as volatility spillover.

This paper is an attempt to answer two main questions. First, do group patterns and spillover effects exist among the volatility series of various commodities in the 21st century? And second, are the factors describing the commodity network priced in the cross-section of commodity futures returns? Depending on the nature of each commodity, we might or might not encounter group patterns. In a wide set of commodities, it is logical to observe WTI and

Brent crude oil to be categorized in a group. But since each market has its fundamentals, we might surprisingly see a commodity such as sugar to join this group. Uncovering such group patterns and then investigating their pricing implications is an interesting subject.

Diebold et al. (2017) are closely related to this paper. By the variance decomposition from vector autoregressive models of commodity volatilities, they document clustering patterns in 19 subindices of the Bloomberg Commodity Price Index. The clustering patterns are evident in precious metals, grain, livestock, energy, and industrial metals. Softs such as cocoa and coffee show no sign of clustering, on the contrary. Although their connectedness measure captures the spillover effect, they do not recover factors that can fully explain the evolution of the network. Here, I contribute to their work by defining two factors that can formally explain the behavior of nodes and edges.

The paper also contributes to network-based factor pricing. Commodity network factors describe the commodity network as a segmented body from other financial markets such as the equity market. This will allow me to compare the performance of concentration and sparsity in the commodity network along with factors driven from theories that explain the price behavior of commodities such as the hedging pressure and the basis. Along the same lines as Herskovic (2018), I will investigate the pricing powers of the factors.

The rest of the paper is set out as follows: Section 2 reviews the data-set and obtains the desired realized volatility series. Section 3 presents the network factors and examines them in graphs. In section 4, asset pricing implications are investigated. And finally, section 5 concludes.

## *3.2 Commodity Data*

I consider 25 commodities whose data are available from January 2000 onwards in Bloomberg. Out of 25, 16 commodities are included in the sub-indices of the Bloomberg Commodity Price Index. Commodities are classified into five categories: energy, grains and oilseeds, livestock, metals, and softs. Table (33) presents an upshot of different futures contracts along with their exchange, ticker, delivery months, and unit.

Table 33: Commodity Futures Contracts

| Category | Commodity | Exchange | Ticker | Delivery Months | Unit |
|---|---|---|---|---|---|
| Energy | Brent crude oil | ICE | CO | 1:12 | 1,000 barrels |
| | Gasoil | ICE | QS | 1:12 | 100 tonnes |
| | Heating oil | NYMEX | HO | 1:12 | 1,000 barrels |
| | Natural gas | NYMEX | NG | 1:!2 | 10,000 mmbtu |
| | WTI crude oil | NYMEX | CL | 1:12 | 1,000 barrels |
| Grains & Oilseeds | Corn | CBOT | C | 3,5,7,9,12 | 5,000 bushels |
| | Soybeans | CBOT | S | 1,3,5,7,8,9,11 | 5000 bushels |
| | Wheat | CBOT | W | 3,5,7,9,12 | 5,000 bushels |
| Livestock | Feeder cattle | CME | FC | 1,3,4,5,8,9,10,11 | 40,000 lbs |
| | Lean hogs | CME | LH | 2,4,6,7,8,10,12 | 40,000 lbs |
| | Live cattle | CME | LC | 2,4,6,8,10,12 | 40,000 lbs |
| Metals | Aluminium | LME | LA | 1:12 | 25 metric tons |
| | Copper | LME | LP | 1:12 | 25,000 lbs |
| | Gold | COMEX | GC | 2,4,6,8,10,12 | 100 troy oz. |
| | Lead | LME | LL | 1:12 | 25 metric tons |
| | Nickel | LME | LN | 1:12 | 6 metric tons |
| | Platinum | COMEX | PL | 1,4,7,10 | 50 troy oz |
| | Silver | COMEX | SI | 1,3,5,7,9,12 | 5000 troy oz. |
| | Tin | LME | LT | 1:12 | 5 metric tons |
| | Zinc | LME | LX | 1:12 | 25 metric tons |
| Softs | Cocoa | ICE | CC | 3,5,7,9,12 | 10 metric tons |
| | Coffee | ICE | KC | 3,5,7,9,12 | 37,500 lbs |
| | Cotton | ICE | CT | 3,5,7,10,12 | 50,000 lbs |
| | Lumber | CME | LB | 1,3,5,7,9,11 | 1000 board feet |
| | Sugar | ICE | SB | 3,5,7,10 | 112,000 lbs |

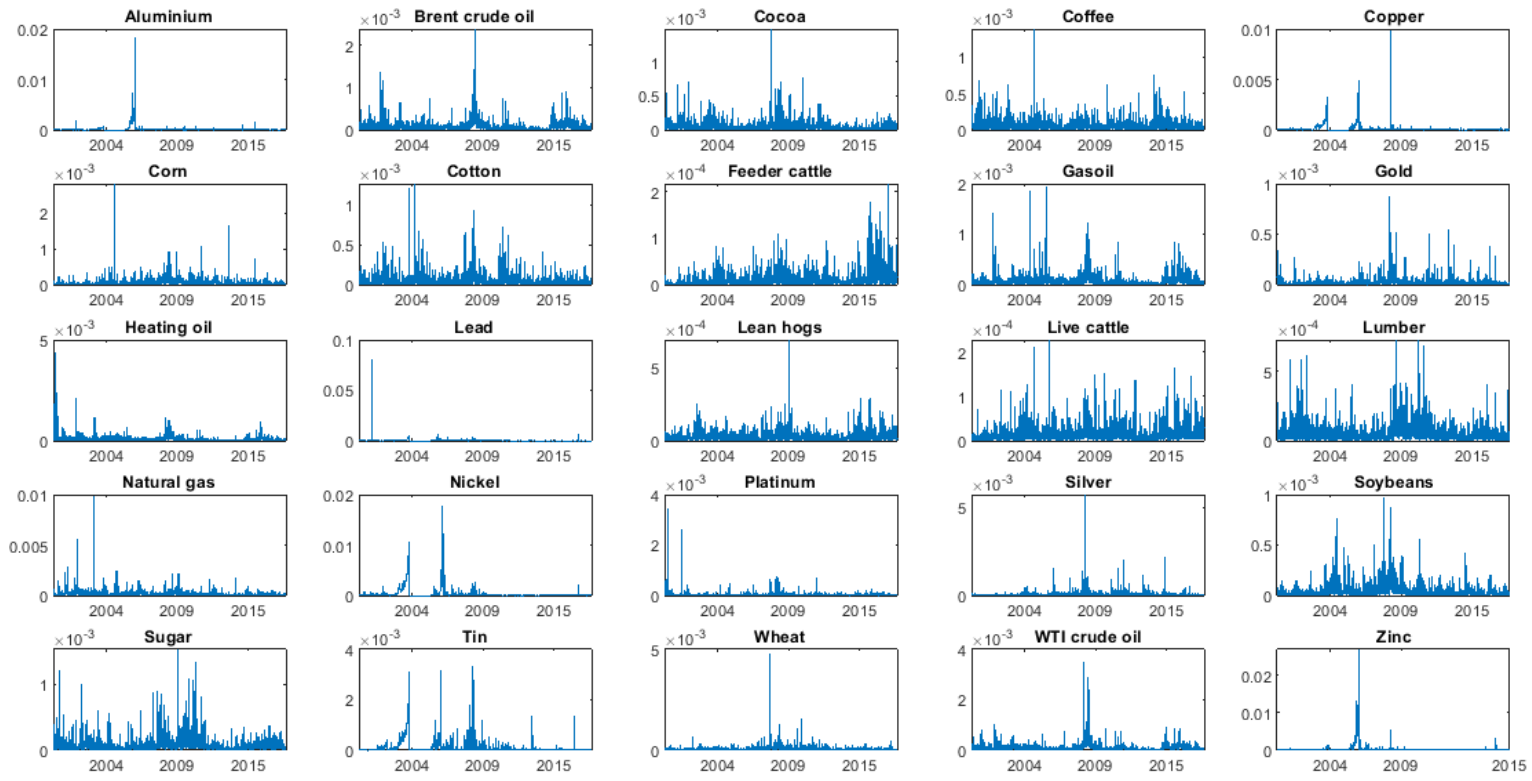
Following Diebold et al. (2017), I use the ranged-based volatility of Garman and Klass (1980). The ranged-based volatility is widely available and easy to calculate. The realized variance, $\widehat{\sigma}^2$, is defined as:

$$\begin{aligned}\widehat{\sigma}_i^2(t) &= 0.511(H_i(t) - L_i(t))^2 - 0.019[(C_i(t) - O_i(t))(H_i(t) + Li(t) - 2O_i(t)) \\ &\quad -2(H_i(t) - O_i(t))(L_i(t) - O_i(t))] - 0.383(C_i(t) - O_i(t))^2 \qquad (78)\end{aligned}$$

where $H_i(t)$, $L_i(t)$, $C_i(t)$, and $O_i(t)$ are log of high, low, closing, and opening prices for commodity $i$ at time $t$, respectively. Figure (7) plots the daily realized volatilities.The time-domain starts in 2000 and ends in 2017. One important feature which is evident in the plots is the autocorrelation in the time series. To take care of that, I take the first difference for all commodities before using them as inputs for the network design.

Figure 7: Realized Volatility Series

The figure shows the realized volatility series obtained by (78) for commodities listed in table (33). There are 5234 daily observations for each commodity, starting from January 2000 and ending in December 2017.



## 3.3 Commodity Network

As was discussed earlier, to design a network, the sets of nodes and edges should be present. Here, the desired network is perceived as the volatility spillover among clusters of commodities. So the first step is recovering clustering behavior in the market, if such patterns exist. Kmeans algorithm of Forgy (1965) and Steinly (2006) is a good candidate to group the realized volatilities. The key question in any clustering algorithm is the number of clusters. To avoid any pre-specification of the network, I use a Bayesian Information Criterion (BIC) to obtain the optimal number of groups ($G^*$) for each period. $G^*$ defines the number of nodes in the network and also captures the clustering patterns among commodities. By using the notations in Manresa and Bonhomme (2015), the following BIC will produce the optimal number of nodes:

$$I(G) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(V_i(t) - \widehat{c}_g(t))^2 + GP_{NT} \tag{79}$$

$I$ denotes the information criterion and $P_{NT}$ is a penalty term. $V_i(t)$ is the volatility series of commodity $i$ and $\widehat{c}_g(t)$ are the estimated centroids. The optimal number of groups minimizes the below objective function:

$$G^* = \underset{G \in \{1,\ldots,G_{\max}\}}{\arg\min}\ I(G) \tag{80}$$

where $G_{\max}$ is the maximum choice of $G$ and is set by the researcher. The information criterion is in the form of a BIC and it has the following expression:

$$BIC(G) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(V_i(t) - \widehat{c}_g(t))^2 + \widehat{\sigma}^2 \frac{GT+N}{NT}\ln(NT) \tag{81}$$

where $\widehat{\sigma}^2$ is the consistent estimator of the variance of $u_i$ and it is given by:

$$\widehat{\sigma}^2 = \frac{1}{NT - G_{\max}T - N}\sum_{i=1}^{N}\sum_{t=1}^{T}(V_i(t) - \widehat{c}_g(t))^2 \tag{82}$$

The group assignment that yields the minimum value of BIC corresponds to the optimal value of grouping, namely $G^*$. Once $G^*$ is established, the concentration time-series is described as the negative entropy of the nodes' shares. Thus, in the same spirit as Herskovic (2018), I define the commodity futures network concentration factor as the following:

$$NC_i^{Com}(t) = \sum_{j=1}^{G^*} \theta_j(t) \log \theta_j(t) \tag{83}$$

where

$$\theta_j(t) = \frac{\mathrm{Si}ze_j}{\sum\limits_{i=1}^{G^*} \mathrm{Si}ze_i} \tag{84}$$

And $\mathrm{Si}ze_j$ is the $\sum\limits_{k=1}^{K}(Open\ Interest)^k \times (Last\ \mathrm{Pr}ice)^k$ for $K$ commodities. Similar to the concentration factor for realized volatility network in the equity market, size captures the size of each node of the network in terms of market capitalization.

When the nodes are established, linkages among them are needed to complete the network. I utilize the multiple horizon causality measure of Dufour and Taamouti (2010) to capture the spillover effects. The estimator of the causality measure from $V_2(t)$ to $V_1(t)$ is:

$$\widehat{GC}^h(V_2 \to V_1|I) = \ln\left[\frac{\det\{[J_0 \widetilde{\Sigma}_0(h) J_0']\}}{\det\{[J_1 \widehat{\Sigma}_k(h) J_1']\}}\right] \tag{85}$$

with $J_1 = [I_{m_1} 0 \;\; 0]$ and $J_0 = [I_{m_1} \;\; 0]$. $\widetilde{\Sigma}_0(h)$ and $\widehat{\Sigma}_k(h)$ are variance matrices of forecast errors in restricted and unrestricted models. Once the measures are estimated, they are transformed into weighted measures as $\omega_{ij}^h(t) = \frac{\widehat{GC}_{ij}^h(t)}{1+\widehat{GC}_{ij}^h(t)}$ and therefore:

$$\begin{bmatrix} \omega_{11}^h(t) & \cdots & \omega_{1G}^h(t) \\ \vdots & \ddots & \vdots \\ \omega_{G1}^h(t) & \cdots & \omega_{G^*G^*}^h(t) \end{bmatrix} \tag{86}$$

To estimate the vector autoregressive models I use daily panels of data for three months. This will produce quarterly factors.

As mentioned earlier, the sparsity factor reveals the evolution of edges, so the factor needs to be obtained by the causality table. Similar to Herskovic (2018), the sparsity factor is specified as the following:

$$NS_i^{Com}(t) = \sum_{i=1}^{G^*} \theta_j(t) \sum_{j=1}^{G^*} \omega_{ij}(t) \log \omega_{ij}(t) \tag{87}$$

Figure (8) depicts the time series of concentration and sparsity. In panel (1) there seems to be an upward trend in concentration starting from 2005. For sparsity, the plot shows stationarity. However, the two series are seemingly correlated. Table (34) shows a significant correlation of 0.28 between the factors in level. Therefore to explore the pricing implications of concentration and sparsity, I consider the innovations in both series in the next section.

Table 34: Summary Statistics and Network Factors

Panel 1 is a summary statistics of concentration and sparsity. Autocorrelations in lag p and their p-values of Ljung-Box Q-test are reported in front of AC(p) for p=1,2,3, and 4. The null hypothesis is no autocorrelation. Panel 2 shows the correlation coefficients and their p-values in parentheses for network factors.

Panel 1
Summary Statistics of Network Factors

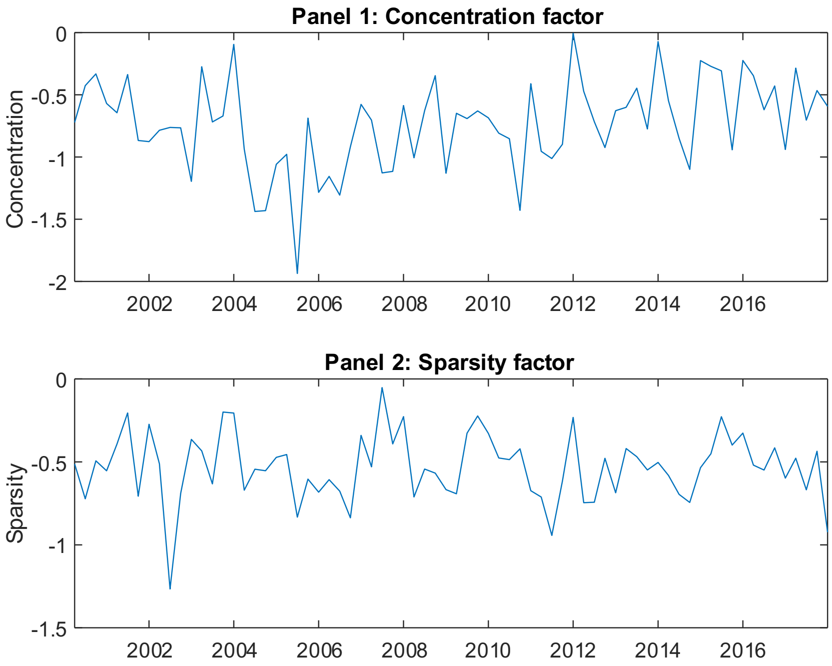
| | $NC$ | | $NS$ | |
|---|---|---|---|---|
| | $Level$ | $p-value$ | $Level$ | $p-value$ |
| $Mean$ | -0.73 | - | -0.53 | - |
| $S.D.$ | 0.36 | - | 0.20 | - |
| $AC(1)$ | 0.20 | 0.08 | 0.12 | 0.26 |
| $AC(2)$ | 0.18 | 0.06 | -0.01 | 0.53 |
| $AC(3)$ | 0.18 | 0.03 | -0.02 | 0.72 |

Panel 2
Network Factors Correlation

| | $NC$ |
|---|---|
| $NS$ | 0.28<br>(0.01) |

Figure 8: Concentration and Sparsity for the Commodity Network

The figure shows the time series of concentration and sparsity for the commodity networks. Panel 1 and 2 are obtained by calculating (83), and (87) respectively. The time-domain starts in January 2000 and ends in December 2017. The frequency is quarterly. For concentration $G_{max}$ is 10, for sparsity $VAR$ is of order 1. The first horizon is chosen to obtain the forecast errors.

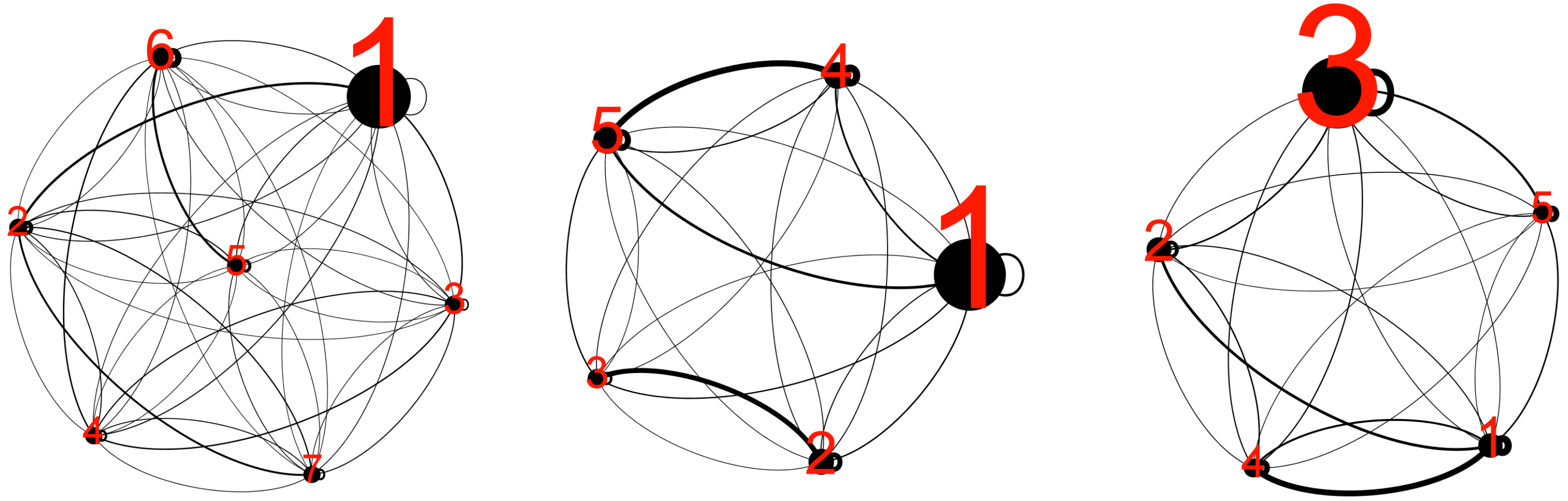



### 3.3.1 Network Graph

Figure (9) presents three network graphs representing the beginning, the middle, and the end of the series. (A) depicts the network for the first quarter of 2000, (B) portrays the first quarter of 2008, and (C) graphs the last quarter of 2017. As is evident in the figure, the number of groups drops from 7 in (A) to 5 in (B) and (C), emphasizing the fact that concentration is slightly rising. This is an indication that commodities are moving in larger and more dominant groups. The edges connecting the nodes are showing more thickness in recent years compared to the beginning of the period. Figure (9) also reveals considerable clustering patterns among commodities in all three sub-sample network graphs.

Figure 9: Network Graphs

The figure depicts three networks representing the beginning, middle, and end of the time-series of concentration and sparsity. (A) is the first quarter of 2000, (B) is the first quarter of 2008, and (C) is the last quarter of 2017. The following reveals the node specifications. For (A): 1: feeder cattle, lumber, tin, lead, gasoil, coffee, cotton, silver, gold, nickel, zinc, copper, aluminum, soybeans, corn, wheat, lean hogs, live cattle, 2: heating oil, 3: WTI crude oil, Brent crude oil, 4:sugar 5: platinum 6: cocoa, 7: natural gas. For (B): 1: feeder cattle, lumber, tin, cotton, gold, nickel, copper, aluminum, soybeans, corn, lean hogs, live cattle, 2: wheat, 3: lead, 4: cocoa, 5: platinum, gasoil, Brent crude oil, coffee, sugar, silver, zinc, heating oil, WTI crude oil, natural gas For (C): 1: gasoil, Brent crude oil, coffee, heating oil, WTI crude oil, 2: cocoa, 3: feeder cattle, platinum, tin, lead, cotton, sugar, silver, gold, nickel, zinc, copper, aluminum, soybeans, corn, wheat, lean hogs, live cattle 4: lumber, 5: natural gas. The graphs are produced by Gephi 0.9.2.




(A) (B) (C)

## 3.4 *Pricing Factors in Commodity Futures Network*

### 3.4.1 Returns in Commodity Futures

The first step in generating the commodity futures return data is creating the time series of quarterly futures prices. The second is step is calculating the percentage change of those prices.

Similar to de Roon and Szymanowska (2010), Gorton et al. (2012), and Sakkas and Tessaromatis (2018), I assume that the holder of the futures contract holds the first nearest-to-maturity contract until the beginning of the delivery month[1]. Since the holder of the contract has no interest in receiving the commodity, he rolls over to the second nearest-to-maturity contract. This pattern continues until I reach the last quarter of the time frame. The quarterly futures prices are considered from the first quarter of 2000 to the last quarter of 2017, a total of 72 observations per commodity. Table (39) in the appendix presents the summary statistics of the return data.

### 3.4.2 Equity-Based Factors

Based on the law of one price and market integration, a factor that is successful in explaining the cross-section of assets in the equity market should also be useful in pricing commodities in the futures market. The market factor as the average return on all firms in AMEX, NASDAQ, and NYSE is the first candidate. To move beyond the capital asset pricing models, I include the three-factor model of Fama and French (1993). In this mode, portfolio returns sorted on size and book-to-market are added to the market factor.

Regarding the network-based factors, the list includes the concentration and sparsity factors of Herskovic (2018). Concentration is the weighted average of output shares of economic sectors, whereas sparsity is the weighted average of input-output linkages among them. To expand upon Herskovic factors, I add concentration and sparsity presented in the second essay. There, the network reveals the volatility spillover among grouped firms that have similar

[1] Excluding the last trading price.

volatility patterns.

### 3.4.3 Commodity-Based Factors

To investigate the pricing powers of my proposed factors, I follow the approach in Daskalaki et al. (2014), by considering a wider set of factors and categorizing them into equity-based and commodity-based factors. Table (35) introduces the factors and their corresponding definitions. Based on the law of one price, any factor that is successful in explaining the cross-section of equity returns should have explanatory powers in commodity returns, given market integration between commodities and equities. To look into this, I consider the market, size, and book/market of Fama and French (1993). To include network-based factors, I add concentration and sparsity in Herskovic (2018), and the concentration and sparsity of the realized volatility network in the second essay to the existing factors.

Concerning the commodity-based factors, two key theories explain the price of commodity futures. The first is the theory of hedging pressure that was developed by Keynes (1930) and Cootner (1960). It states that if the demand for short hedging against risk exceeds the long position, then the long position should be compensated with a risk premium. To measure the hedging pressure, Roon et al. (2000), Szymanowska et al. (2013), Daskalaki et al. (2014), present the following measure:

$$HP_i(t) = \frac{number\ of\ short\ hedge\ positions_i(t) - number\ of\ long\ hedge\ positions_i(t)}{total\ number\ of\ hedge\ positions_i(t)} \tag{88}$$

To find the number of hedging positions, I consider non-commercial positions for each commodity. These traders are the ones that hedge against the risk and are not considered as speculators in the market. The data is available for free in Commodity Futures Trading Commission.

The second is the theory of storage which highlights the role of inventories. The theory of storage, developed by Working (1933) and Kaldor (1939), introduces the convenience yield as the benefit of keeping inventories. Thus, the storage costs and the convenience yield can help explain the futures price. Since the data for inventory is hard to obtain, the following measure is presented in Daskalaki et al. (2014) as an approximation to measure the basis for commodity $i$ :

$$Basis_i(t) = \frac{F_{i1}(t) - F_{i2}(t)}{F_{i1}(t)} \tag{89}$$

where $F_{i1}(t)$ and $F_{i2}(t)$ indicate the futures prices of the nearest and the next nearest-to-maturity contracts, respectively.

Table 35: Risk Factors and Definitions

The table lists factors that are considered in commodity futures pricing. They are classified into two sections, namely, equity-based and commodity based-factors. The market refers to the CAPM model. Size and book/market are two additional factors in Fama and French (1993). HNC and HNS are concentration and sparsity of Herskovic (2018). NC and NS are concentration and sparsity presented in the second essay. HP, the hedging pressure, and basis are motivated by the theory of normal backwardation and theory of storage, respectively. Finally, $NC^Com$ and $NS^Com$ are concentration and sparsity in commodity realized volatility network.

| Factor | Definitions |
|---|---|
| *Equity-Based Factor* | |
| $Market$ | Return on all NASDAQ, NYSE, and AMEX stocks. |
| *Size* | The average return on the three small portfolios minus the three big portfolios. |
| $Book/market$ | The average return on the two value portfolios minus the two growth portfolios. |
| $HNC$ | Negative entropy of the output shares' distribution for different sectors. |
| $HNS$ | Measurement of the distribution of sectoral linkages. |
| $NC$ | Negative entropy of size shares' distribution for different clusters of realized volatility in the network of equity realized volatilities. |
| $NS$ | Measurement of the distribution of volatility spillover among clusters of realized volatility in the network of equity realized volatilities. |
| *Commodity-Based Factor* | |
| $HP$ | The difference between the portfolio of commodity futures with positive and the portfolios with negative heading pressure. |
| $Basis$ | The difference between the portfolio of commodity futures with positive and the portfolios with negative basis. |
| $NC^{Com}$ | Negative entropy of size shares' distribution for different clusters of realized volatility in the network of commodity realized volatilities. |
| $NS^{Com}$ | Measurement of the distribution of volatility spillover among clusters of realized volatility in the network of commodity realized volatilities. |

Table (36) presents the average return for portfolios sorted on commodity-based factors. To construct the hedging pressure portfolios, at every period $t$, I separate the commodities with negative HP from the ones with positive HP. Then, at $t+1$, I consider the strategy that goes (long) the commodities with positive (negative) HP. In a similar approach, to create basis portfolios, at every period $t$, commodities with a positive basis are separated from the ones with a negative basis. Then at $t+1$, long-minus-short portfolios are created in the same manner. Concerning the commodity network factors, the approach is different. First, I regress the time series of commodity futures returns on both concentration and sparsity factor in the following regression:

$$r_i(t) = \alpha_{3i} + \beta_i^{NC} \Delta NC^{Com}(t) + \beta_i^{NS} \Delta NS^{Com}(t) + \epsilon_i(t) \tag{90}$$

In a rolling window regression of length 20, I regress the commodity futures return series on the network concentration and sparsity from $t-20$ to $t$, and then I sort them from low to high factor for each factor. Consequently, I create tercile portfolios in $t+1$. $HmL_{NC^{Com}}$ and $HmL_{NS^{Com}}$ reveal the average return of the strategy of going long the high network factor-beta and short the low factor-beta.

As the results indicate, none of the long-short strategies are significant for my data-set. The spreads for HP, basis, and concentration factors are positive, whereas sparsity demonstrates a negative price of risk. As robustness checks, I control for the window length and forecast horizon ($h$) in the causality matrix. The results are presented in the appendix.

To explore the factors further, Fama and Macbeth (1973) two-stage regression is used to estimate the risk premia. Table (37) demonstrates the estimated lambdas and their corresponding t-statistics. The table is divided into two parts of equity-based models and commodity-based models. Concerning the equity-based models, some factors including the market, size, concentration and sparsity in Herskovic (2018), and sparsity of realized volatility appear with the correct signs. However, risk premia are not statistically significant. The commodity-based models are incapable of explaining the cross-section of commodity futures returns as

Table 36: Sorting Portfolios

The table shows the average returns on long-minus-short investment strategies for commodity-based factors. The returns are annualized. T-statistics are reported in square brackets.

Portfolios Sorted by Commodity-Based Factors

| | Average Return | Standard Deviation |
|---|---|---|
| **Hedging Pressure** | | |
| High HP Portfolio | 9.24% | 41.79% |
| Low HP Portfolio | 7.01% | 34.60% |
| $HmL_{HP}$ | 2.12% | 28.46% |
| t-stat | [0.64] | |
| **Basis** | | |
| High basis portfolio | 9.43% | 33.32% |
| Low basis portfolio | 7.72% | 43.26% |
| $HmL_{basis}$ | 1.61% | 30.39% |
| t-stat | [0.52] | |
| **Concentration** ($NC^{Com}$) | | |
| High concentration portfolio | 9.10% | 50.77% |
| Medium concentration portfolio | 7.65% | 40.17% |
| Low concentration portfolio | 4.83% | 41.48% |
| $HmL_{NC^{Com}}$ | 4.27% | 33.01% |
| t-stat | [0.97] | |
| **Sparsity** ($NS^{Com}$) | | |
| High sparsity portfolio | 9.33% | 42.03% |
| Medium sparsity portfolio | 6.85% | 45.29% |
| Low sparsity portfolio | 5.29% | 44.95% |
| $HmL_{NS^{Com}}$ | -4.04% | 34.20% |
| t-stat | [-0.88] | |

well. The signs are in line with my findings in sorted portfolios for the basis and commodity volatility network factors of concentration and sparsity.

The insignificant prices of risk for various factor models confirm the results in Daskalaki et al. (2014) for a different period. They also indicate that the commodity market is a segmented market from the equities. Commodity prices are driven by market fundamentals that are exclusive to each commodity. Therefore, the traditional market and network factors are not able to capture them. To explore the heterogeneous nature of commodities further, I perform time series regression and obtain individual betas for each commodity. Table (38) reports the results. As it is evident from the table, equity factors perform poorly in general and commodities respond differently to various factors. In commodity-based factors, the outcomes are relatively more significant. Commodities are most responsive to hedging pressure but some betas are not significant even within each commodity group. Overall, table (38) confirms the results in Daskalaki et al. (2014). However, commodities seem to be more heterogeneous in the 21st century compared to the period considered in their research.

Table 37: Estimating Price of Risk

The table reports the estimation results of risk premia, using the two-stage regression of Fama and Macbeth (1973). T-statistics are reported in square brackets. Model (1) to (4) refer to the equity-based factors. Models (5) to (7) incorporate the commodity-based factors.

| | Equity_Based Models | | | | Commodity-Based Models | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Constant | 0.0194 | 0.0186 | -0.0066 | -0.0010 | 0.0192 | 0.0201 | 0.0030 |
| | [2.08] | [2.10] | [-0.30] | [-0.04] | [2.22] | [2.41] | [0.12] |
| $Mkt$ | 0.1585 | 0.2156 | | | | | |
| | [0.19] | [0.21] | | | | | |
| Size | | 0.2541 | | | | | |
| | | [0.26] | | | | | |
| Book/Market | | -0.2503 | | | | | |
| | | [-0.36] | | | | | |
| $HNC$ | | | -0.0019 | | | | |
| | | | [-0.25] | | | | |
| $HNS$ | | | 0.0048 | | | | |
| | | | [0.65] | | | | |
| $NC$ | | | | 0.0001 | | | |
| | | | | [0.01] | | | |
| $NS$ | | | | -0.0073 | | | |
| | | | | [-0.76] | | | |
| $HP$ | | | | | -0.0058 | | |
| | | | | | [-0.55] | | |
| $Basis$ | | | | | | 0.0014 | |
| | | | | | | [0.06] | |
| $NC^{Com}$ | | | | | | | 0.0218 |
| | | | | | | | [0.43] |
| $NS^{Com}$ | | | | | | | -0.0362 |
| | | | | | | | [-1.06] |
| $R^2$ | 0.01 | 0.08 | 0.10 | 0.06 | 0.24 | 0.00 | 0.07 |

Table 38: Time Series Estimation of Beta

This table reports individual regression of commodity futures return on various factors. The numbers are estimated betas for each commodity and each factor. *, **, and *** show significance at 10, 5, and 1 percent level.

| | | Equity-Based Factors | | | | | | | Commodity-Based Factors | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Category | Commodity | Market | Size | Book/Market | HNC | HNS | NC | NS | HP | Basis | NC(Com) | NS(Com) |
| Energy | Brent crude oil | 0.0008 | 0.0085 | -0.0103 | 0.2586 | 0.3431 | -6.6987 | 3.5833 | -1.4835*** | 0.3414 | 0.5771* | 0.1723 |
| | Gasoil | 0.0091 | 0.0092 | -0.0188** | 0.9475 | -0.0739 | -0.3800 | -0.7633 | -0.4575 | 0.1142 | 0.1295 | 0.2022 |
| | Heating oil | 0.0047 | -0.0001 | -0.0021 | 0.7821 | -0.2813 | -6.0343 | 4.0034 | -1.2904*** | 0.2489 | 0.5366* | 0.1709 |
| | Natural gas | 0.0123 | -0.0062 | 0.0016 | 5.5684** | -4.7240* | -1.8053 | 3.0387 | -0.6993 | 0.4010 | 0.8469* | 0.0762 |
| | WTI crude oil | 0.0048 | 0.0050 | -0.0092 | 1.1164 | -0.3808 | -6.1120 | 2.5885 | -1.2703*** | 0.2345 | 0.6395* | 0.2636* |
| Grains | Corn | 0.0033 | -0.0097* | 0.0145** | -0.3231 | 0.1774 | 3.8334 | 0.3845 | 0.5833** | 0.2747 | 0.1259 | 0.1127 |
| & Oilseeds | Soybeans | 0.0039 | -0.0059 | 0.0049 | 0.8125 | -0.5089 | 5.0109* | 0.3497 | 0.3760 | 0.5504** | -0.2023 | 0.1771 |
| | Wheat | 0.0016 | -0.0020 | 0.0004 | -0.2586 | -0.3842 | 3.3267 | -2.4829 | 0.4837* | 0.3376 | -0.2104 | 0.2379* |
| Livestock | Feeder cattle | 0.0020 | 0.0027 | -0.0019 | 0.7355 | -0.4631 | -1.7501 | 0.6898 | -0.4501*** | 0.0457 | 0.2797* | -0.1055* |
| | Lean hogs | 0.0069 | 0.0012 | -0.0033 | 3.9496** | -2.7490 | 0.7034 | 0.2845 | -0.0743 | 0.2572 | -0.1628 | -0.1304 |
| | Live cattle | 0.0043 | -0.0034 | 0.0039 | 1.7519 | -1.7831 | 1.8675 | 1.2413 | 0.0083 | 0.0093 | 0.4054** | -0.0894 |
| Metals | Aluminium | 0.0068** | 0.0029 | -0.0035 | 0.8799 | -0.9170 | -0.3118 | 0.3715 | 0.0619 | 0.1583 | 0.6241*** | 0.0350 |
| | Copper | 0.0101** | 0.0047 | -0.0022 | 1.5664 | -1.5508 | -0.3118 | 0.1698 | -0.0854 | 0.2004 | 0.6635** | -0.0213 |
| | Gold | -0.0007 | 0.0015 | -0.0070** | 0.2236 | 0.0671 | 0.7370 | 0.1813 | 0.1404 | 0.1701 | 0.0391 | 0.0069 |
| | Nickel | 0.0192*** | -0.0079 | 0.0071 | 1.3643 | -0.0272 | -3.5411 | -1.5327 | -0.0867 | 0.1826 | 0.5957 | -0.0570 |
| | Platinum | 0.0100*** | -0.0015 | -0.0043 | 1.4195** | -1.0152 | 0.3151 | -1.2079 | 0.0290 | 0.4131** | 0.3036 | 0.0332 |
| | Silver | 0.0110** | -0.0033 | -0.0002 | -0.0427 | 1.4439 | -1.1428 | 0.0638 | 0.2302 | 0.2768 | 0.3099 | 0.2699** |
| | Tin | 0.0093* | 0.0025 | -0.0076 | -0.7369 | 1.7560 | -4.2503 | 0.3126 | -0.0664 | 0.3217 | 0.4120 | 0.1489 |
| | Zinc | 0.0136*** | 0.0067 | 0.0007 | 0.0056 | -0.4458 | -2.5119 | -0.1396 | 0.3216 | 0.0647 | 0.2515 | 0.0187 |
| Softs | Cocoa | -0.0040 | -0.0002 | 0.0018 | 1.8294 | -1.7675 | 0.6118 | -3.1570* | 0.0855 | -0.0561 | 0.1558 | 0.0645 |
| | Coffee | 0.0036 | 0.0074 | -0.0111 | 0.5328 | -0.0196 | -7.4744** | 2.7287 | -1.3434*** | 0.2756 | 0.5910* | 0.1488 |
| | Cotton | 0.0068 | 0.0001 | -0.0080 | -0.1639 | -0.0514 | -0.4338 | -0.5662 | -0.0796 | 0.6521** | 0.0471 | 0.0796 |
| | Lumber | 0.0144*** | 0.0062 | -0.0004 | -1.0174 | 1.3051 | -4.1191 | -1.4896 | 0.4857* | 0.0749 | -0.0121 | 0.1810 |

## 3.5 Concluding Remarks

This paper uncovers clustering behavior in the commodity futures market. In the first part, a network of realized volatility is defined with two factors that represent the evolution of edges and nodes, namely concentration and sparsity. In the second part, the asset pricing implications of a wide set of factors are explored. Generally, factors fit into two groups known as equity-based and commodity-based factors. In sorting portfolios, none of the commodity-based factors generate significant high-minus-low investment strategies for the period of the study. Upon further examination, risk premia are not significant. However, they are in accord with positive or negative high-minus-low spreads. By performing individual time-series regressions, the heterogeneous nature of each commodity is established. The fact that may explain the unsuccessful pricing factors in this particular market.

## 3.6 *Appendix*

Table (39) presents the annualized average returns for each commodity. Table (40) and (41) demonstrate the robustness analysis of the high-minus-low investment strategies, controlling for the window length and forecast horizon, respectively.

Table 39: Summary Statistics of Futures' Returns

The table reports annualized average returns for the commodity futures market. Due to the unavailability of the data, lead and silver are removed from the list.

| Category | Commodity | Ticker | Return |
|---|---|---|---|
| Energy | Brent crude oil | CO | 11.42% |
| | Gasoil | QS | 11.27% |
| | Heating oil | HO | 11.40% |
| | Natural gas | NG | 14.23% |
| | WTI crude oil | CL | 10.27% |
| Grains & Oilseeds | Corn | C | 7.32% |
| | Soybeans | S | 8.75% |
| | Wheat | W | 6.77% |
| Livestock | Feeder cattle | FC | 4.95% |
| | Lean hogs | LH | 6.42% |
| | Live cattle | LC | 4.65% |
| Metals | Aluminium | LA | 3.49% |
| | Copper | LP | 12.69% |
| | Gold | GC | 9.57% |
| | Nickel | LN | 8.71% |
| | Platinum | PL | 5.75% |
| | Tin | LT | 11.46% |
| | Zinc | LX | 9.46% |
| Softs | Cocoa | CC | -7.76% |
| | Coffee | KC | -3.39% |
| | Cotton | CT | 6.74% |
| | Lumber | LB | 6.94% |
| | Sugar | SB | 11.57% |

Table 40: Sorting Portfolios: Controlling for Window Length

The table reports the high-minus-low annualized returns for different choices of the window length.

Portfolios Sorted by Network-Based Factors

Panel 1: Concentration ($NC^{Com}$)

| Window Length | $HmL_{NC^{Com}}$ | $t-stat$ |
|---|---|---|
| 25 | 2.07% | [0.44] |
| 35 | 0.69% | [0.14] |
| 45 | -1.49% | [-0.27] |

Panel 2: Sparsity ($NS^{Com}$)

| Window Length | $HmL_{NS^{Com}}$ | $t-stat$ |
|---|---|---|
| 25 | -3.59% | [-0.81] |
| 35 | -3.10% | [-0.63] |
| 45 | -1.17% | [-0.25] |

Table 41: Sorting Portfolios: Controlling for Forecast Horizon

The table reports the high-minus-low annualized returns for portfolios sorted on sparsity controlling for the forecast horizon h in the causality table in (86).

Portfolios Sorted by Sparsity

| Forecast Horizon ($h$) | $HmL_{NS^{Com}}$ | $t-stat$ |
|---|---|---|
| 5 | -6.20% | [-1.2] |
| 10 | -9.57% | [-1.89] |
| 15 | -3.39% | [-0.58] |

# Bibliography


[1] Acemoglu, D., Ozdaglar, A. & Tahbaz-Salehi, A. (2015). Systemic risk and stability financial networks. American Economic Review, 105(2), 564–608.

[2] Agiakloglou, C., Newbold, P. & Wohar, M. (1993). Bias in an estimator of the fractional difference parameter. Journal of Time Series Analysis, 14, 235-246.

[3] Ahern, K. (2013). Network centrality and the cross section of stock returns. Working paper.

[4] Alquist, R. & Kilian, L. (2010). What do we learn from the price of crude oil futures? Journal of Applied Econometrics, 25, 539-573.

[5] Allen, F. & Babus, A. (2008), Networks in finance. Wharton Financial Institutions Center Working Paper.

[6] Alvarez-Ramirez, J., Alvarez, J. & Rodriguez, E. (2008). Short-term predictability of crude oil markers: a detrended fluctuation analysis approach. Energy Economics, 30, 2645-2656.

[7] Andersen, T., T. Bollerslev, F. Diebold, & P. Labys (2001). The distribution of realized exchange rate volatility. Journal of the American Statistical Association 96 (453).

[8] Ang, A., Hodrick, R., Xing, Y. & Zhang, X. (2006). The cross-section of volatility and expected returns. The Journal of Finance, 61(1), 259-299.

[9] Baillie, R. T. & Bollerslev, T. (1989). Common stochastic trends in a system of exchange rates. The Journal of Finance, 44(1), 167-181.

[10] Baillie, R. T. & Bollerslev, T. (1994). The long memory of the forward premium. Journal of International Money and Finance, 13(5), 565-571.

[11] Barigozzi, M. & Brownlees, C. T. (2016). Nets: network estimation for time series. Working paper.

[12] Barndorf-Nielsen, O. E.(2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64(2), 253–280.

[13] Barunik, J., Kocenda, E. & Vacha, L. (2014). Asymmetric connectedness of stocks: How does bad and good volatility spill over the U.S. stock market?. Working paper.

[14] Bernard J.-T., Khalaf L., Kichian M. & S. McMahon (2015). The Convenience yield and the informational content of the oil futures price. Energy Journal, 36, 29-46.

[15] Bianchi, D., Billio, M. & Casarin, R. (2015). Modeling contagion and systemic risk, Working paper.

[16] Billio, M., Getmansky, M., Lo, A. W., & Pelizzon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. Journal of Financial Economics, 104(3), 535 – 559.

[17] Bonhomme, S. & Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. Econometrica, 83(3), 1147-1184.

[18] Brenner, R. J., & Kroner, K. F. (1995). Arbitrage, co-integration, and testing the unbiasedness hypothesis in financial markets. Journal of Financial and Quantitative Analysis, 30, 23-42.

[19] Cajueiro, D. O. & Tabak, B. M. (2007). Are the crude oil markets becoming weakly efficient over time? a test of time-varying long-range dependence in prices and volatility. Energy Economics, 29, 28-36.

[20] Cavaliere, G., Nielsen, M. O. & Taylor, A. M. R. (2015). Bootstrap score tests for fractional integration in heteroskedastic ARFIMA models, with an application to price

dynamics in commodity spot and futures markets. Journal of Econometrics, 187, 557-579.

[21] Chinn, M. D. & Coibion, O. (2014). The predictive content of commodity futures. The Journal of Futures Market, 34(7), 607-636.

[22] Choi, K., & Zivot, E. (2007), Long memory and structural breaks in the forward discount: An empirical investigation. Journal of International Money and Finance, 26, 342–363.

[23] Cohen, L. & Frazzini, A. (2008). Economic links and predictable returns. The Journal of Finance, 63(4), 1977-2011.

[24] Cootner, P. H. (1960). Returns to speculators: Telser versus Keynes. Journal of Political Economy, 68, 396-404.

[25] Daskalaki, C., Kostakis, A., & Skiadopoulos, G. (2014). Are there common factors in individual commodity futures returns?.Journal of Banking and Finance, 40, 346-363.

[26] de Roon, F., Nijman, T. E. & Veld, C. (2000). Hedging pressure effects in futures markets. Journal of Finance, 55, 1437-1456.

[27] de Roon, F. & Szymanowska, M. (2010). The cross-section of commodity futures returns. Working paper.

[28] Diebold, F. X. & Yilmaz, K. (2014). On the network topology of variance decompositions: measuring the connectedness of financial firms. Journal of Econometrics, 182(1), 119 – 134.

[29] Diebold, F. X., Liu, L. & Yilmaz, K. (2017). Commodity connectedness. Working paper.

[30] Doornik, J. (1999). Object-orientated matrix programming using Ox (3rd ed.). London: Timberlake Consultants Press and Oxford.

[31] Dufour, J.-M., Galbraith, J. W & Zhang, H. J. (2015). Exchange rates and commodity prices: measuring causality at multiple horizons. Journal of Empirical Finance, 36, 100-120.

[32] Dufour, J.-M. & Jian, B. (2016). Multiple horizon causality in network analysis: measuring volatility interconnections in financial markets. Working paper.

[33] Dufour, J-M. & Renault, E. (1998). Short-run and long-run causality in time series : theory. Econometrica, 66, 1099-1125.

[34] Dufour, J.-M. & Taamouti, A. (2010). Short and long run causality measures: theory and inference. Journal of Econometrics, 154, 42-58.

[35] Engel, C. (1996). The forward discount anomaly and the risk premium: A survey of recent evidence. Journal of Empirical Finance, 3, 123–192.

[36] Fama, E. F. (1969). Efficient capital markets: a review of theory and empirical work. The Journal of Finance, 25(2), 28-30

[37] Fama, E. F. (1969). Efficient capital markets: II. The Journal of Finance, 46(5), 1575-1617

[38] Fama, E. F. & French, K. R. (1987). Commodity futures prices: some evidence on forecast power, premiums, and the theory of storage. The Journal of Business, 60(1), 55-73.

[39] Fama, E. F. & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. Journal of Financial Economic,. 33: 3–56.

[40] Fama, E. & Macbeth, J. D. (1973). Risk, return, and equilibrium: empirical tests. Journal of Political Economy, 81(3), 607–636.

[41] Forgy, E.W. (1965). Cluster analysis of multivariate data: efficiency vs. interpretability of classifications. Biometrics, 21, 768–769.

[42] Garman, M. B. & Klass, M. J. (1980). On the estimation of security price volatilities from historical data. Journal of Business, 53, 67-78.

[43] Geweke, J. (1982). Measurement of linear dependence and feedback between multiple time series. Journal of the American Statistical Association, 77, 304-313.

[44] Geweke, J. (1984). Measures of conditional linear dependence and feedback between time series. Journal of the American Statistical Association, 79, 907-915.

[45] Geweke, J. (1984). Inference and causality in economic time series models. Chapter 19 in Handbook of Econometrics, vol. 2, 1101-1144.

[46] Geweke, J., & Porter-Hudak, S. (1983). The estimation and application of long memory time series models. Journal of Time Series Analysis, 4, 221–238.

[47] Geweke, J. & Zhou, G. (1996). Measuring the price of the Arbitrage Pricing Theory. The Review of Financial Studies, 9(2), 557-587.

[48] Gorton, G. B., Hayashi, F. & Rouwenhorst, K. G. (2012). The fundamentals of commodity futures returns. Review of Finance, 17, 25-105.

[49] Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. Econometrica, 37(3), 424-438.

[50] Granger, C. W. J. (1980). long memory relationships and the aggregation of dynamic models. Journal of Econometrics, 14, 227-238.

[51] Granger, C. W. J. & Joyeux, R. (1980). An introduction to long memory time series models and fractional differencing. Journal of Time Series Analysis, 1(1), 15-29.

[52] Herskovic, B. (2018). Network in production: asset pricing implications. Journal of Finance. Vol. 73 (4), pp. 1785–1818

[53] Hosking, J. R. M. (1981). Fractional differencing. Biometrika, 68(1), 165-176.

[54] Hauser, M. A. (1997). Semi-parametric and nonparametric testing for long memory: A Monte Carlo study. Empirical Economics, 22, 247–271.

[55] Johansen, S. (1991). Estimation and hypothesis testing of co-integration vectors in Gaussian vector auto-regressive models. Econometrica, 59(6), 1551-1580

[56] Kaldor, N. (1939). Speculation and economic stability. Review of Economic Studies, 7(1), 1-27.

[57] Keynes, J. M. (1930). A treatise on money. Macmillan, London.

[58] Kristoufek, L. & Vosvrda, M. (2013). Commodity futures and market efficiency. Working paper

[59] Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. Journal of Econometrics, 54, 159-178.

[60] Lo, A. W. (2008). Efficient market hypothesis. The New Palgrave Dictionary of Economics, second edition. Eds. Steven N. Durlauf and Lawrence E. Blume. Palgrave Macmillan.

[61] McKenzie. A. M. & Holt, M. T. (2002). Market efficiency in agricultural futures markets. Applied Economics, 34(12), 1519-1532

[62] Phillips, P. C. B. & Perron, P. (1988). Testing for a unit root in time series regression. Biomekrika, 75(2), 335-346

[63] Phillips, P. C. B. (1999). Unit root log periodogram regression. Unpublished working paper No. 1244, Cowles Foundation for Research in Economics, Yale University.

[64] Ross, S. A., (1976). The arbitrage theory of capital asset pricing. Journal of Economic Theory, 13, 341-360.

[65] Samuelson, P. A. (1965). Proof that properly anticipated prices fluctuate randomly. Industrial Management Review, 6(2), 41-49.

[66] Sakkas, A. & Tessaromatis, N. (2018). Factor-based commodity investing. Working paper.

[67] Steinley, D. (2006). K-means clustering: a half-century synthesis. British Journal of Mathematical and Statistical Psychology, 59, 1–34.

[68] Switzer, L. N. & El-Khoury, M. (2007). Extreme volatility, speculative efficiency, and the hedging effectiveness of the oil futures markets. The Journal of Futures Markets, 27(1), 61-84.

[69] Szymanowska, M., de Roon, F.A., Nijman T. E., & Van den Goorbegh, R. (2013). An anatomy of commodity futures risk premia. Journal of Finance, 69(1), 453-482.

[70] Tsay, W., (2010). Maximum likelihood estimation of stationary multivariate ARFIMA processes. Journal of Statistical Computation and Simulation, 80(7), 729-745.

[71] Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. Journal of the Royal Statistical Society, 58(1), 267-288.

[72] Velasco, C. (1999). Non-stationary log-periodogram regression. Journal of Econometrics, 91, 325–327.

[73] Working H. (1949). The theory of price of storage. American Economic Review, 39, 1254-1262.