'Financialization in Commodity Markets'

V.V. Chari and Lawrence J. Christiano

December 5, 2019

Not for Publication

Abstract

This appendix describes the data and empirical methods used in the paper.

Contents

Part I

Introduction

We obtained data from the CFTC on the volume of trade in futures contacts for 29 commodities. We obtained monthly data on the spot prices for these commodities. We also obtained spot prices for commodities that are not heavily traded on formal US commodity exchanges. Annual data for 107 such commodities, 40 from the US Geological Survey (USGS) and 67 from the Food and Agriculture Organization of the United Nations (FAOSTAT). We also obtained monthly data for non-traded commodities from data services such as the one provided by the Federal Reserve Bank of St. Louis' service,FRED, the International Monetary Fund (IMF), and the private services,Trading Economics andIndexMundi.

These data were processed using MATLAB code, to produce the figures and tables in our manuscript. The code and data, together with a readme file, are available in the form of a zip file, replication.zip. This appendix describes the details of the data how they were used in the analysis.

Part II

Four Datasets Constructed for the Analysis

We constructed four data sets using MATLAB. Each is in the form of a MATLABstructure: CFTC_stats, USGS_stats, FAO_stats, and non_tr_monthly_with_q. Briefly, CFTC_stats contains data on the 29 commodities that are traded in organized US futures exchanges.[1] The second and third contain annual data on commodities that are not traded on US exchanges. In particular, USGS_stats (taken from theUS Geological Survey) has non-traded commodities that are minerals and metals and FAO_stats contains non-traded 'softs', primarily agricultural products from FAOSTAT. The last structure, non_tr_monthly_with_q, contains monthly data on non-traded commodities.

In this section, we describe the format of these four structures. In later sections we provide a detailed description of the underlying data sources. The four structures are stored in the MATLAB binary file, stats.mat which appears in the subdirectory, results, of the directory, replication_files. This mat file was created by the MATLAB program, do_scatters.m.

1. CFTC_stats

We first turn to a discussion of the MATLAB structure, CFTC_stats. The structure was originally constructed using the MATLAB program, create_CFTCstats.m (though it was actually saved to stats.mat by do_scatters.m). The structure, CFTC_stats, contains 31 entries. Of these, the first 30 (not including item 28, which is empty) are traded commodities that appear in the CFTC database. Thus, CFTC_stats(1) contains data on wheat, so that CFTC_stats(1).name contains the MATLAB string, 'WHEAT'. Consistent with the information in Table1, CFTC_stats(1).datain contains a vector of 21 numbers with the ratio of annual data on net financial flows for the period 1992-2012, scaled by world production of wheat. The ratio is unit free, because both numerator and denominator are in the same

[1] For more discussion on what it means to be in the CFTC data, and what it means to be out of that data, see section5..

units, tonnes. The 31^{st} entry of CFTC_stats provides information on aggregates of the 29 commodities, where the weights sum to unity and (unless otherwise indicated) correspond to shares of world production (in the case of P, CFTC_stats(31).P is the log of the weighted sum of the individual prices, after they have been exponentiated. Information on the specific identity of the 29 commodities and our sources is provided below. The contents of that structure is described in Table1:

Table 1: Data on Traded Commodities, MATLAB Structure, CFTC_stats

Variable	Meaning
name	name of commodity
month	field containing monthly information on commodity
datains	annual index of net financial flows, computed using method 3, scaled by world production
datains1	same as datains, computed using method 1
dataoi	annual open interest, scaled by world production
dataSl	annual long contracts, non-commercial traders, computed using method 3, scaled by world production
dataSl1	same as dataSl, computed using method 1
imbalance	$=(S^L-S^s)/(S^L+S^s)$, computed using method 3
imbalance1	same as imbalance, computed using method 1
dataNRl	non-reported long contracts
dataNRs	non-reported short contracts
dates	dates of annual observations: 1992-2012
P	commodity price, logged after scaling by its first observation in 1992 and by the PCE deflator
P_unscaled	raw commodity price, dollars
P_dates	date range for P_unscaled
Punit	units of raw commodity price
output	quantity of world production for the commodity
mP	mean, ΔP, for pre- and post- 2000 annual observations
value	dollar value of annual world production
insurance	mean, net financial flows scaled by world production, pre- and post- 2000
openinterest	mean, open interest scaled by world production, pre- and post- 2000
long	mean, non-commercial traders' long contracts, pre- and post- 2000
beta	least squares coefficients, regression of ΔP on pre- 2000 constant and time trend, and post-2000 constant and time trend
volatility	standard deviation, least squares residual of ΔP on constant and time trend, pre- and post- 2000
category	1$\sim$softs, 2$\sim$metal, 3$\sim$fuel
volP	standard deviation of ΔP, pre- and post- 2000
volP_time	centered, rolling standard deviation, two years into the future and past
index	1$\sim$if commodity appears in one of the two major commodity indices (S&P GSCI or Dow Jones-UBS Commodity Index), 0 otherwise

Below the entry, 'index', there appear four additional variables, datains_av, datains1_av, dataNRl_av, dataNRs_av. These apply only to variable 31, which represents our aggregate measure of commodities, and are empty in the case of variables 1-30. The notation, _av, means that the index was computed using a cross-section equally weighted average. The number 1 means that method 1 was used to compute shorts and longs of traders. More detailed discussion of the CFTC data appears below. For example, the method 1 and method 3 referred to in the table are explained in section4.1.1..

Note the entry, 'month', in Table1. That is a field that includes the monthly analog of the information in Table1. Thus, for example, CFTC_stats(i).month contains the following monthly information for commodity i, for i=1,2,....,27,29,30 (recall, 28 is empty):

Table 2: Contents of CFTC_stat(i).month,$i = 1, 2, \cdots, 27, 29, 30$

Variable	Dimension of variable
dataoi	252×1
datains	252×1
datains1	252×1
dataSI	252×1
dataSI1	252×1
imbalance	252×1
imbalance1	252×1
dataNRl	252×1
dataNRs	252×1
dates	252×1
P	252×1
P_unscaled	406×1
P_dates	2×1
P_units	
output	252×1
mP	2×1
insurance	2×1
openinterest	2×1
long	2×1
beta	2×1
volatility	2×1
volP	2×1
volP_time	252×1

NOTE : See Table1.

In the case of output in the above table, this is annual data, interpolated to monthly.

2. USGS_stats and FAO_stats

Next, we describe the dataset, USGS_stats. This was created by the MATLAB file, do_scatters.m, located in a subdirectory, results, in the directory, replication_files. The dataset, USGS_stats, contains information on 67 commodities for which we have annual observations. The CFTC does not track the trading of these data on organized exchanges. The structure, USGS_stats has the following contents:

Table 3: Data on Traded Commodities, MATLAB Structure, USGS_stats

Variable	Meaning
name	name of commodity
dates	1992-2012
price	dollars per metric tonne
quantity	world production, in metric tonnes
P	log, price scaled by PCE deflator and first price observation
mP	mean,ΔP for pre- and post- 2000 annual observations
beta	least squares coefficients, regression ofΔP on pre- 2000 constant and time trend, and post-2000 constant and time trend
volatility	standard deviation, least squares residual ofΔP on constant and time trend, pre- and post- 2000
volP	standard deviation ofΔP pre- and post- 2000
volP_time	5 year, centered moving average standard deviation ofΔP

The data set, USGS_stats, contains data on 40 variables. The structure, FAO_stats has the same format as USGS_stats. It contains annual data on 67 non-traded variables.

3. Monthly Prices and Production for Non-Traded Data

This section describes the contents of the structure, non_tr_monthly_with_q, which contains our monthly data on non-traded commodities. The structure is stored in the MATLAB mat file, production_data. That file is contained in the subdirectory, replication_files/CFTC_commodities_husnu. The contents of non_tr_monthly_with_q is as follows.

Table 4: Monthly, non-CFTC Traded Commodities

Variable	subfields	Meaning
name		name of commodity
annual		name of field containing annual data
	quantity	data, dates and source for world production
	price	data, dates and source for dollar price
	value	data, dates and source for dollar value of world production
note		information (where available) about foreign futures markets
monthly		
	name	
	units	
	description	
	Spot_Price	log, price scaled by PCE deflator and first price observation
	Spot_Time	dates for price data
	unscaled_Price	raw, dollar spot price of the commodity

Details about each of these variables are provided in section5. below.

Part III

Data Sources and Description

4. Spot Price, Quantity and Value Data For Variables in the CFTC Database

We now present a detailed discussion of the sources of data for CFTC_stats. We begin by describing the spot price, quantity and value data corresponding to the commodities that are in the CFTC database.

4.1. Commodity Futures Exchange Commission

We obtained data on the volume of trade in commodity futures markets from the CFTC. The weekly 'Commitments of Traders Futures Only' reports file was obtained from the Commodity Futures Trading Commission (CFTC).[2] The data were downloaded in the file, deafut_xls_1986_2015.zip. This file contains three excel files, FUT86_06.xls, FUT07_14.xls and FUT15_15.xls (the file names indicate the years to which these data pertain). The annual data were obtained by adding over all the weekly observations in a year. The variables in the CFTC data are displayed in Table5:

Table 5: Commodities Included in CFTC Dataset

WHEAT
CORN
OATS
SOYBEANS
SOYBEAN OIL
CRUDE OIL
HEATING OIL
NATURAL GAS
COTTON
RICE
ORANGE JUICE
BUTTER
HOGS
PORK BELLIES
CATTLE
LUMBER
PROPANE
COCOA
PALLADIUM
PLATINUM
SUGAR
COFFEE
SILVER
COPPER
GOLD
GASOLINE
ALUMINUM
COAL
SOYBEAN MEAL

4.1.1. Description of CFTC Volume of Trade Data

The CFTC data are obtained from 'commitments of traders' (COT) reports. According tothis:

[2]http://www.cftc.gov/MarketReports/CommitmentsofTraders/HistoricalCompressed/index.htm

"The COT reports provide a breakdown of each Tuesday's open interest for markets in which 20 or more traders hold positions equal to or above the reporting levels established by the CFTC. The weekly reports for Futures-Only Commitments of Traders and for Futures-and-Options-Combined Commitments of Traders are released every Friday at 3:30 p.m. Eastern time.

Reports are available in both a short and long format. The short report shows open interest separately by reportable and nonreportable positions. For reportable positions, additional data is provided for commercial and non-commercial holdings, spreading, changes from the previous report, percents of open interest by category, and numbers of traders."

Following is a more detailed description of the data, taken from theCFTC website. When it's a direct quote, the sentences include"".

Open Interest

"Open interest is defined as the total of all futures and/or option contracts entered into and not yet offset by a transaction, by delivery, by exercise, etc. The aggregate of all long open interest is equal to the aggregate of all short open interest." We use the following notation:

$$oi = S^L + H^L = S^s + H^s,$$

whereS^i denotes the number of long, when$i = L$, and short, when$i = s$, contracts held by non-commercial traders (sometimes referred to as 'outsiders' or 'speculators'). The object,H^i, denotes the analogous objects held by commercial traders (also sometimes referred to as 'hedgers').

Reportable Positions

Clearing members, futures commission merchants, and foreign brokers (collectively called reporting firms) file daily reports with the Commission. Those reports show the futures and option positions of traders that hold positions above specific reporting levels set by CFTC regulations.*If, at the daily market close, a reporting firm has a trader with a position at or above the Commission's reporting level in any single futures month or option expiration, it reports that trader's entire position in all futures and options expiration months in that commodity, regardless of size.* The aggregate of all traders' positions reported to the Commission usually represents 70 to 90 percent of the total open interest in any given market. From time to time, the Commission will raise or lower the reporting levels in specific markets to strike a balance between collecting sufficient information to oversee the markets and minimizing the reporting burden on the futures industry."

Commercial and Non-commercial Traders

"When an individual reportable trader is identified to the Commission, the trader is classified either as "commercial" or "non-commercial." All of a trader's reported futures positions in a commodity*are classified as commercial if the trader uses futures contracts in that particular commodity for hedging as defined in CFTC Regulation 1.3, 17 CFR 1.3(z).*[3] A trading entity generally gets classified as a "commercial" trader by filing a statement with the Commission, on CFTC Form 40: Statement of Reporting Trader, that it is commercially "...engaged in business activities hedged by the use of the futures or option markets." To ensure that traders are classified with accuracy and consistency, Commission staff may exercise judgment in re-classifying a trader if it has additional information about the trader's use of the markets.*A trader may be classified as a commercial trader in some commodities and as a non-commercial trader in other commodities. A single trading entity cannot be classified as both a commercial and non-commercial trader in the same commodity. Nonetheless, a multi-functional organization that has more*

[3] *For additional discussion, see http://www.cftc.gov/IndustryOversight/MarketSurveillance/SpeculativeLimits/index.htm*

than one trading entity may have each trading entity classified separately in a commodity. For example, a financial organization trading in financial futures may have a banking entity whose positions are classified as commercial and have a separate money-management entity whose positions are classified as non-commercial."

Nonreportable Positions

"The long and short open interest shown as "Nonreportable Positions" is derived by subtracting total long and short "Reportable Positions" from the total open interest. Accordingly, for "Nonreportable Positions," the number of traders involved and the commercial/non-commercial classification of each trader are unknown."

Spreading

"For the futures-only report, spreading measures the extent to which each non-commercial trader holds equal long and short futures positions. For the options-and-futures-combined report, spreading measures the extent to which each non-commercial trader holds equal combined-long and combined-short positions. For example, if a non-commercial trader in Eurodollar futures holds 2,000 long contracts and 1,500 short contracts, 500 contracts will appear in the "Long" category and 1,500 contracts will appear in the "Spreading" category. These figures do not include inter-market spreading, such as spreading Eurodollar futures against Treasury Note futures. Also see the "Old and Other Futures" section, below."

To understand the above technical discussion, we found it useful to construct an example. Consider Table6. There, 'NC' mean non-commercial and 'C' means commercial. The 'longs' and 'shorts' columns indicate the actual futures market positions of the traders. The first three types of traders are 'reported' in the sense that it is reported whether they are commercial or non-commercial. Trader 4 is 'non-reported' in the sense that they don't meet the CFTC's reporting requirement.

At the level of individual non-commercial traders, the CFTC only reports the net positions (see the discussion above). If a trader is long on net (as trader 1), then the net longs are reported as longs,NC_L, and shorts for that trader are reported as zero. The missing shorts appear in the 'spread' column. In the case of trader 2, he is on net short, so a zero is reported in the long column for that trader and the net shorts,NC_S, are reported in the short column. The longs are reported in the 'spread' column. In the case of commercial traders, it appears that all positions are reported in gross terms,C_L andC_s, so there is no spread for them (see the zeros in the 'spread' column). In the case of 'non-reported' trades (i.e., by traders not identified as commercial or non-commercial), their gross long positions,NR_L, and gross short,NR_S, positions are reported. The data reported by the CFTC are:

$$NC_L, NC_S, C_L, C_s, NR_L, NR_S, Spread, oi$$

The data reported by the CFTC is A trader that is net long (trader 1) has his trades reported in net

Table 6: Illustrative Example of CFTC Data

	longs	shorts	CFTC reported long	CFTC reported short	spread
NC trader 1	2200	1800	$NC_L = 400$	0	1800
NC trader 2	1500	1700	0	NC_S =200	1500
C trader 3	1500	1000	C_L =1500	C_s =1000	
trader 4 (non-reported)	300	1000	NR_L =300	NR_S =1000	0
CFTC reported spread					Spread=3300
open interest = 5500					
CFTC total longs(reported and unreported) = 400+1500+300+3300 = 5500,					
reported shorts = 200+1000+1000+3300=5500					

The CFTC's spread concept may at first seem mysterious, so we provide a simple example of how it works. Suppose that there are two types of traders,$j = 1, 2$. Type$j = 1$ trader buys a quantity,l_j, of long contracts and a quantity, s_j, of short contracts in a given commodity. All these quantities are positive, of course. Suppose that for traders of type 1, $l_1 > s_1$ and for traders of type 2, $l_2 < s_2$. Then, the CFTC's variable, "NonComm_Positions_Long_All", is $l_1 - s_1 > 0$ and the CFTC's variable, "NonComm_Positions_Short_All", corresponds to $s_2 - l_2 > 0$. The CFTC also computes the residual, the smaller of the long or short position of the trader, whichever is the smaller. The CFTC sums over the residuals of each trader and reports the result as "NonComm_Positions_Spread_All", or, simply spread. The residual for trader 1 in the example is s_1 and the residual for trader 2 is l_2. Thus, the spread in the example is $s_1 + l_2$. The objects we wish to measure are the gross longs and the gross shorts of the commercial traders. To obtain the gross longs of traders, $l_1 + l_2$, one simply adds the spread to NonComm_Positions_Long_All. To obtain the gross shorts, $s_1 + s_2$, across all traders, one adds the spread to NonComm_Positions_Short_All. For example,

$$\text{NonComm\_Positions\_Long\_All+spread} = l_1 - s_1 + \overbrace{(s_1 + l_2)}^{\text{spread}} = l_1 + l_2.$$

To obtain the gross long and short positions of non-commercial trades from the CFTC data obviously requires some care because of the spread term. The CFTC reports gross longs and shorts for the commercial trades, and so no spread adjustment is required there.

Here are some numbers for one contract on one data set (described below) that we downloaded:

Table 7: CFTC Volume Measures for one Wheat Contract

WHEAT - CHICAGO BOARD OF TRADE, 12/30/97		
(1) oi=Open_Interest_All	468,710	
(2) NC_L=NonComm_Positions_Long_All	71,155	
(3) NC_S=NonComm_Positions_Short_All	127,665	
(4) Spread=NonComm_Positions_Spread_All	23,675	
(5) C_L = Comm_Positions_Long_All	219,290	
(6) C_S = Comm_Positions_Short_All	193,560	
(7) R_L =Tot_Rept_Positions_Long_All	314,120	(7)=(2)+(4)+(5)
(8) R_S =Tot_Rept_Positions_Short_All	344,900	(8)=(3)+(4)+(6)
(9) NR_L =NonRept_Positions_Long_All	154,590	(1)=(9)+(7)
(10) NR_S =NonRept_Positions_Short_All	123,810	(1)=(8)+(10)

The problem is how to find S^L, S^s, H^L, H^s such that

$$oi = S^L + H^L = S^s + H^s.$$

This means we have to apportion the non-reported positions in the right way and also the spread. We have

$$oi = NC_L + C_L + NR_L + spread = NC_S + C_S + NR_S + spread$$

One extreme specification (Method #1) assigns all non-reported trades to commercial traders:

$$S^L + H^L = \overbrace{NC_L + spread}^{S^L} + \overbrace{C_L + NR_L}^{H^L}$$

$$S^s + H^s = \overbrace{NC_S + spread}^{S^s} + \overbrace{C_S + NR_S}^{H^s}$$

A second extreme (Method #2) assigns all non-reported trades to non-commercial traders:

$$S^L + H^L = \overbrace{NC_L + spread + NR_L}^{S^L} + \overbrace{C_L}^{H^L}$$
$$S^s + H^s = \overbrace{NC_S + spread + NR_S}^{S^s} + \overbrace{C_S}^{H^s}$$

An intermediate method is to assign NR_L proportionally between commercial and non-commercial traders. Thus, compute:

$$\lambda_L = \frac{NC_L + spread}{NC_L + spread + C_L} = \frac{71,155 + 23,675}{71,155 + 23,675 + 219,290} = \frac{94830}{314120} = 0.30, \quad (1)$$
$$\lambda_S = \frac{NC_S + spread}{NC_S + spread + C_S} = \frac{127,665 + 23,675}{127,665 + 23,675 + 193,560} = 0.44 \quad (2)$$

and

$$S^L = NC_L + spread + \lambda_L NR_L$$
$$H^L = C_L + (1 - \lambda_L)\, NR_L.$$

Similarly for S^s and H^s. We found a few observations in the case of silver, when NC_L, spread and C_L are all zero (8/27/96-11/19/96), so that λ_L is not defined. We simply set the undefined λ_L's to their average value in the other periods.

Obviously, method #1 corresponds to setting the weights to zero and method #2 corresponds to setting the weights to unity. A Bayesian might set the weights according to method #3. The question is how to apportion NR_L between commercial and non-commercial traders. Might as well do it in the same proportion as the other part of shorts and longs where we know what the apportionment is. Then,

$$nff = S^L - S^s = NC_L - NC_S + \lambda_L NR_L - \lambda_S NR_S.$$

In principle, we have a total of 5 methods for computing nff. In practice, we consider only three:

method #1: λ_L=λ_S=0

method #2: λ_L=λ_S=1,

and method #3 uses equation 1.

Consider the implications for nff.

$$\frac{nff^{method\,3} - nff^{method\,1}}{oi} = \frac{NC_L - NC_S + \lambda_L NR_L - \lambda_S NR_S - [NC_L - NC_S]}{oi} \quad (3)$$
$$= \frac{\lambda_L NR_L - \lambda_S NR_S}{NC_L + spread + NR_L + C_L}$$
$$= \lambda_L \frac{NR_L}{NC_L + spread + NR_L + C_L} - \lambda_S \frac{NR_S}{NC_S + spread + NR_S + C_S}$$
$$= 0.3 \frac{154,590}{468,710} - 0.44_S \frac{123,810}{468,710}$$
$$= 0.3 * 0.33 - 0.44 * 0.2642 = -0.0172$$

4.1.2. Figures for CFTC Volume Measures

Nonreportables

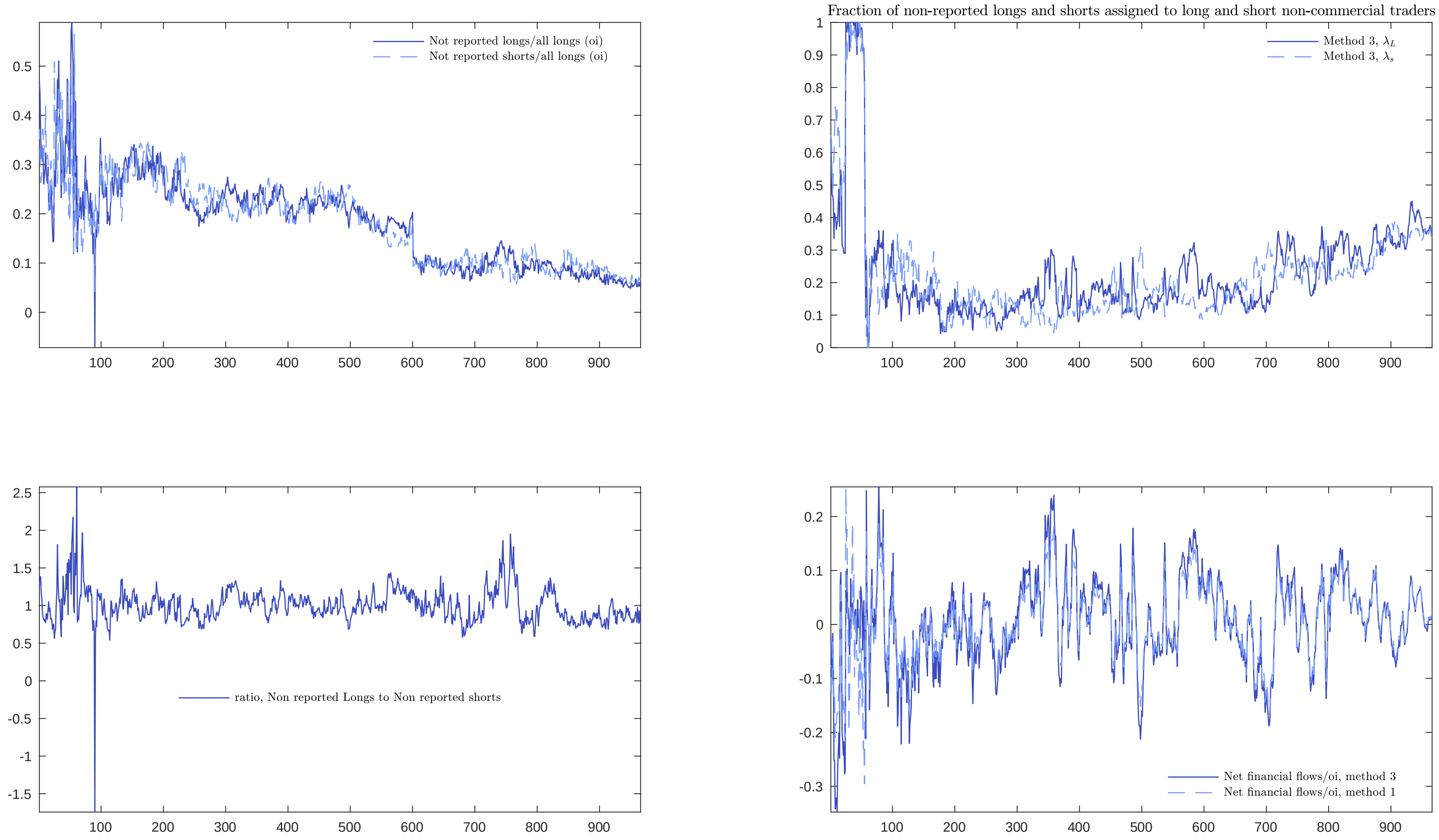
Figure1 displays properties of λ_L and λ_s, as well as non-reportedlongs, NR_L, and non-reported shorts, NR_s, for crude oil. Oil is a commodity that recorded as traded by the CFTC, and is the commodity in our dataset with the biggest share of world production. There are several things to note in the figure. First, non-reported longs and shorts are a non-trivial fraction of open interest, though they have been falling steadily over the years. The ratio falls from around 0.3 at the start of the data in 1986 (though it does jump as high as 0.6 then), and then falls steadily to about 0.05 towards the end. Second, the fractions, λ_L and λ_s are roughly equal in size. Third, on average, non-reported longs and shorts are similar orders of magnitude, and their ratio fluctuates in a range between 0.5 and 1.5. Because non-reportables and λ_L and λ_s are similar in magnitude, the difference between nff computed by methods 1 and 3 is expected to be small. This leads to the fourth observation about Figure1, which focuses on the 2,2 entry. Note that nff computed by the two methods produces results that are almost impossible to distinguish. One has to look closely to notice a difference, though the differences are someone more pronounced at the start of the sample.

Fifth, note that there is obvious measurement error in the data. In 2/27/87, non-reportable long positions, 'NonRept_Positions_Long_All, NR_L, is -8554 contracts of 1,000 barrels each of crude oil.[4] According to theCFTC, these positions are defined as open interest minus total reported long contracts, line (7) in Table7, which corresponds to the variable, R_L. Since long contracts by definition are positive, the non-reportable long positions being negative indicates that the measured number of long contracts is bigger than open interest, and therefore wrong, or open interest is too small, or both.

[4]Specifically, the contract is"CRUDE OIL, LIGHT 'SWEET' - NEW YORK MERCANTILE EXCHANGE", with contract market code 067651, CFTC commodity code, 067.

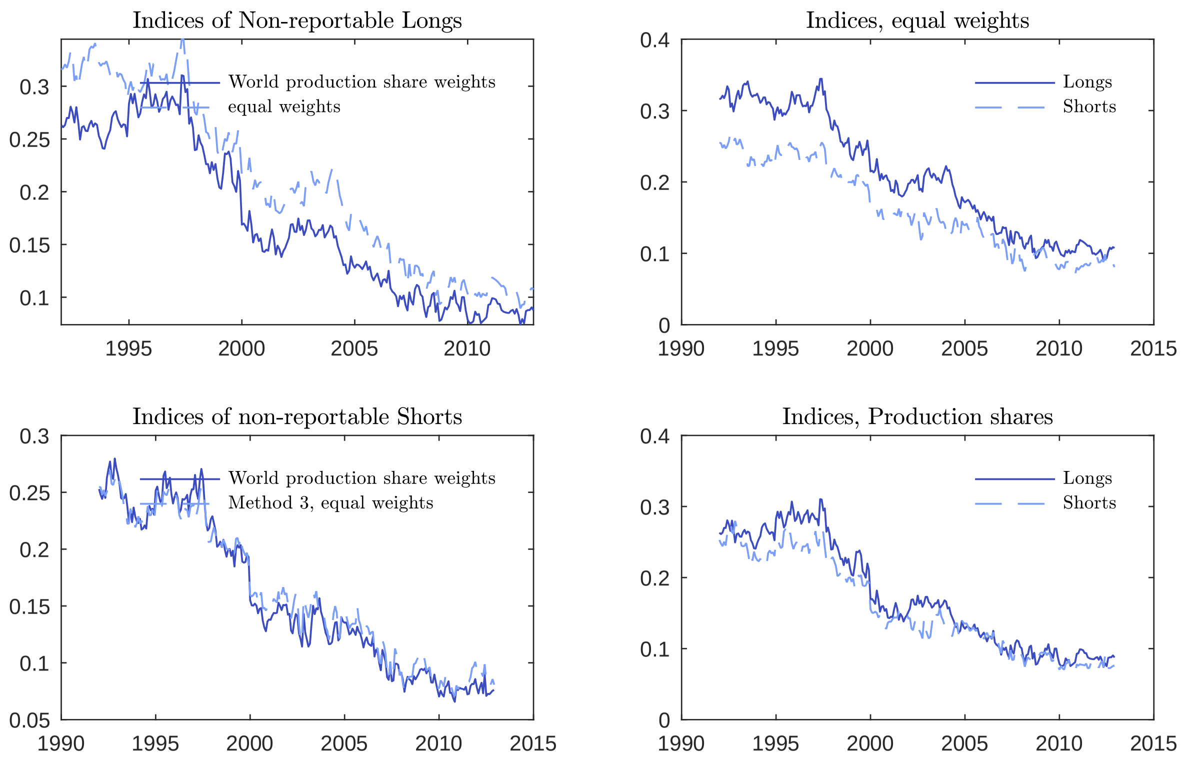
Figure 1: Data on Non-reported Longs and Shorts for Oil

CRUDE OIL



To get a sense of the size of non-reportable trades, overall, we computed two indices of these trades. One was computed using the shares of world production. But, these shares weight the fuels very heavily and we wanted something that weighted the non-reportables in other commodities as well. So, we also constructed an index which weights each of our 29 commodities equally.

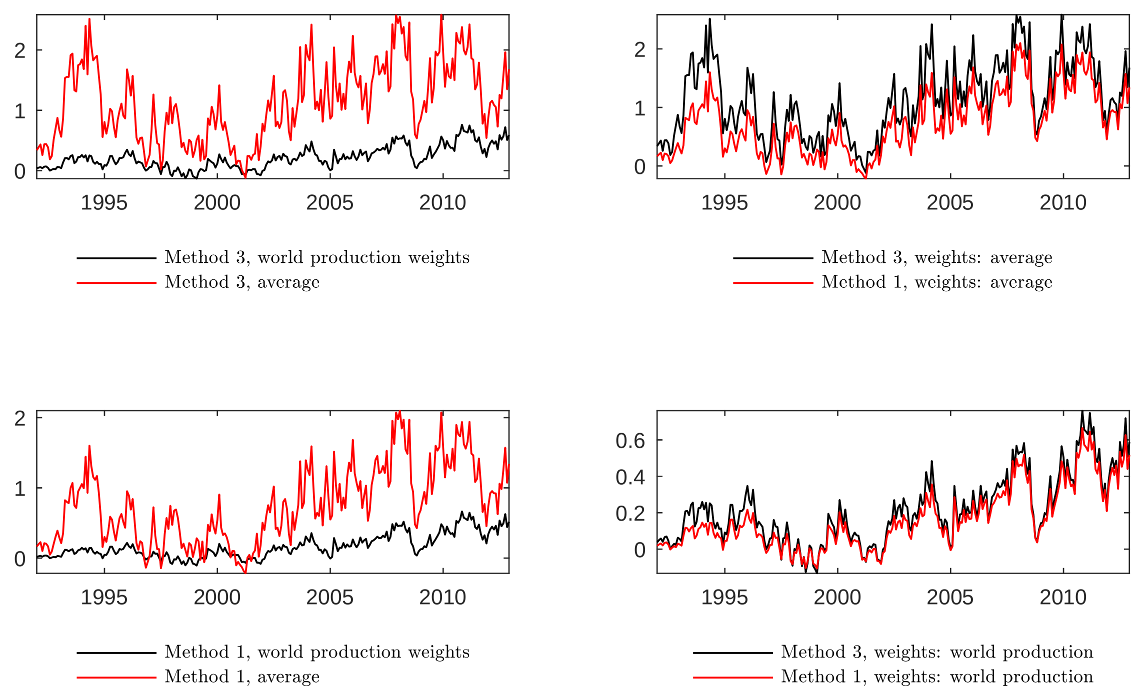
Figure 2: Non-reportables, Computed by Method 3



There are several features of Figure2 worth emphasizing. First, we see from all the figures that the non-reportables are have fallen substantially over the span of our data set. They start roughly at 30 percent of open interest (depending on which index) and end up at or below 10 percent. Apparently, the CFTC is acquiring more data from traders. Second, the second column indicates that, when weighted by share of world production, the nonreportable longs are a little smaller than when equally weighted. Thus, in the case of the longs, the CFTC has better data on the important (by world production) contracts. Third, the left column indicates that unreportable longs and shorts move together fairly closely, another reason why method 1 and method 3 does not produce big differences for net financial flows (see above).

Now consider aggregate measures of net financial flows, scaled by world production. Figure3 displays the results of two methods of constructing nff: method 1 and our baseline, method 3. In addition, we report the relationship between measure 1 and measure 3.

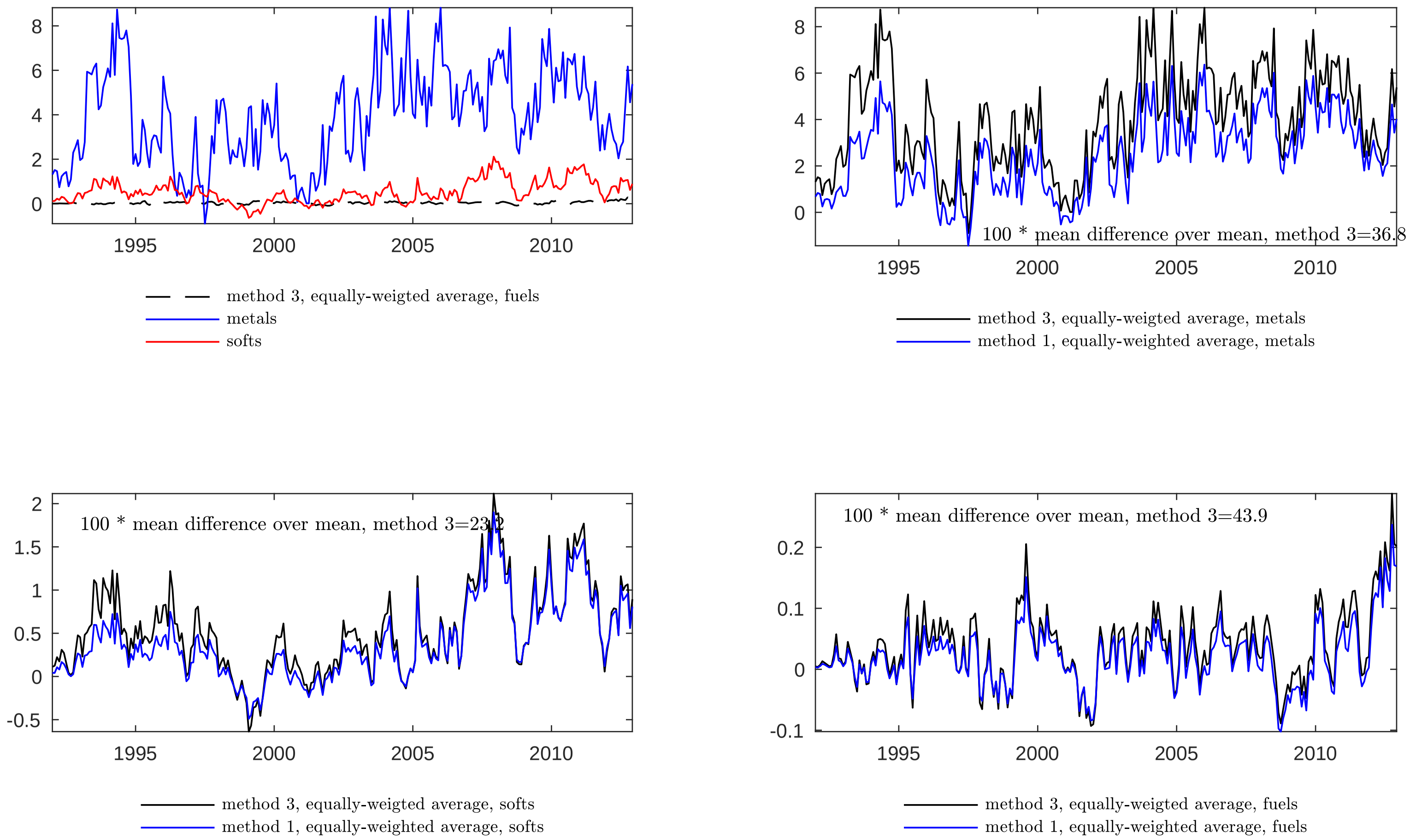
Figure 3: Two Measures of Net Financial Flows, and Two Aggregations



The most important message of Figure3 is that the equally weighted index (i.e., computed by averaging nff in the cross section of commodities) produces a substantially higher value of aggregate nff. The top 1,1 figure shows that when the nff of different commodities are equally weighted, then nff can be as large as 8 times world production. When the nff index is computed by using the share of world production for weights, then nff is nearly zero. Evidently, nff for fuels, which receive very high weight when production shares are used, behaves very differently than the nff of other commodities. A second message of Figure3 is that methods 1 and 3 provide very similar measures of nff, as suggested by the simple calculations above.

Next, we break down net financial flows into softs, metals and and fuels, in Figure4. In the top left panel of that figure, we see that nff for fuels is enormous, consistent with the observations in the previous paragraph. The other panels show that method 1 and method 3 produces roughly the same estimates for nff within each of the three categories. There is some difference in panel 1,2, however.

Figure 4: Net Financial Flows Different Commodity Categories



The figures display a measure of the difference in nff between methods 1 and 3. In each case the measure is the mean difference divided by the mean of method 3 nff. Division by the mean of nff was done (rather than the usual percent deviation) because nff can be extremely small and distort the results.

It is interesting to compare Hong and Yogo's measure of futures markets*imbalance*is the

> "... ratio of two variables. The numerator is the dollar value of short minus long positions held by commercial traders in the Commitments of Traders in Commodity Futures, summed across all commodities in that sector. The denominator is the dollar value of short plus long positions held by commercial traders, summed across all commodities in that sector. We then compute commodity market imbalance as an equally weighted average of futures market imbalance across the four sectors."

To exactly replicate the Hong-Yogo imbalance measure requires the prices of individual contracts. Here, we do something different that Hong and Yogo, which does not require the price of a contract, but we think it captures the spirit of their measure. For each commodity, compute it's imbalance:

$$\text{imbalance} = \frac{S^L - S^S}{s^L + S^S},$$

then, average over commodities. We can do this using the method 1 and 3 strategies for computing the quantity of non-commerical contracts. In some cases, the denominator of imbalance is zero, and the dates and commodities for which this happens are excluded from the averaging. The averaging was done over all CFTC-traded commodities, the softs, the fuels and the metals, separately. They were computed using method 1 and method 3. They appear in Figure5

Figure 5: Two Methods for Computing Imbalance

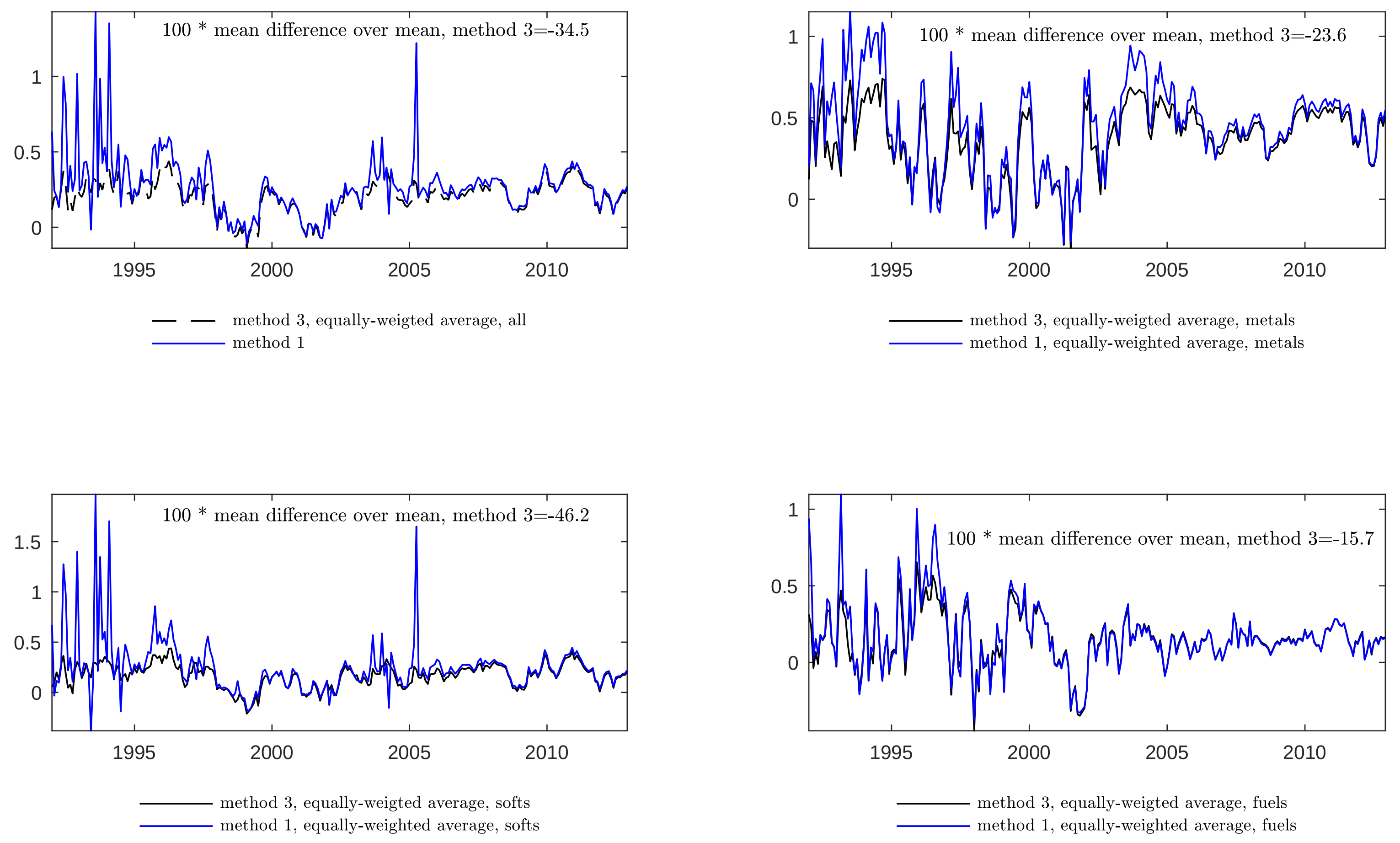
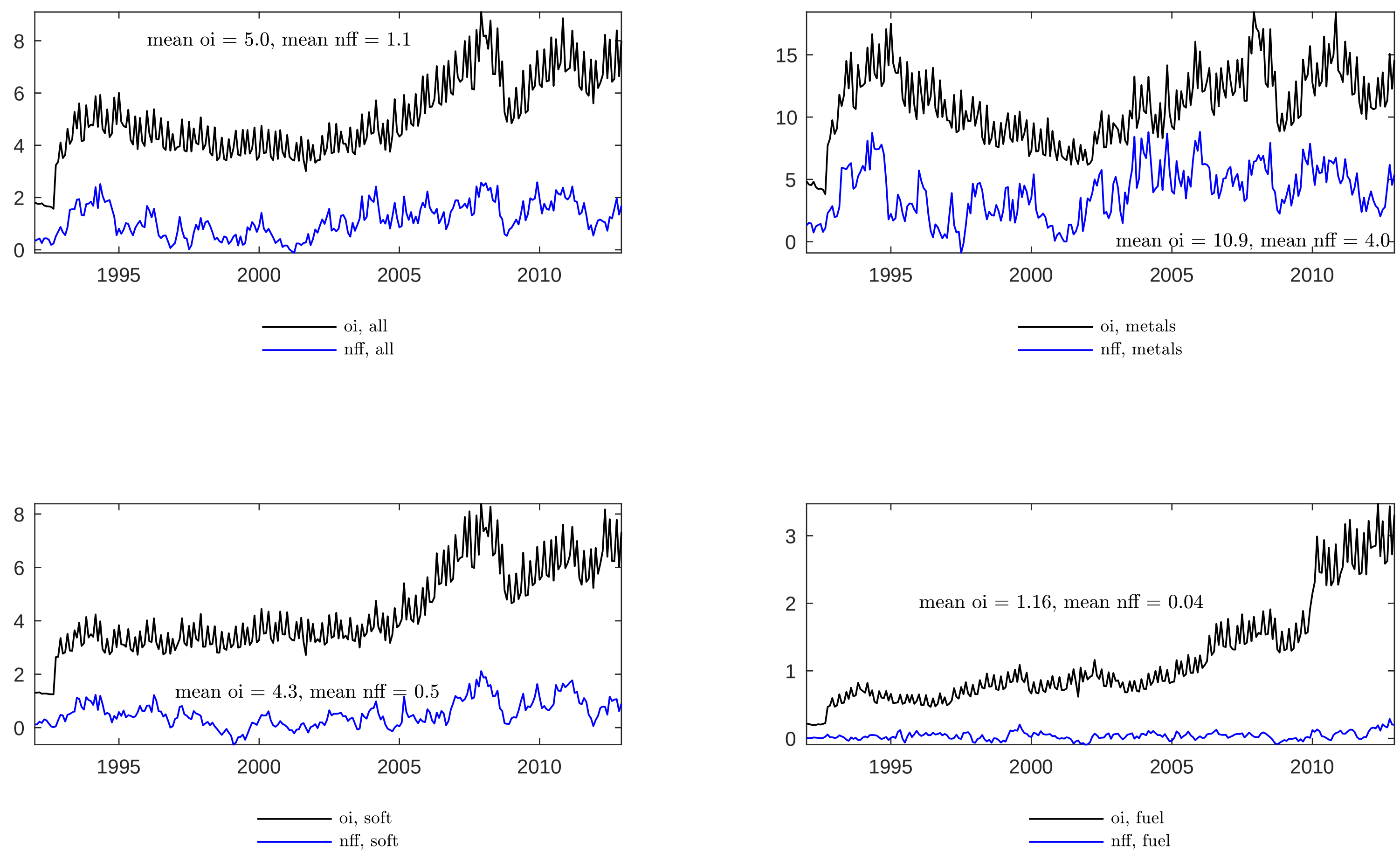


Figure6 compares nff and oi (both, scaled by world production), when averaged over subsets of commodities and over all traded commodities.

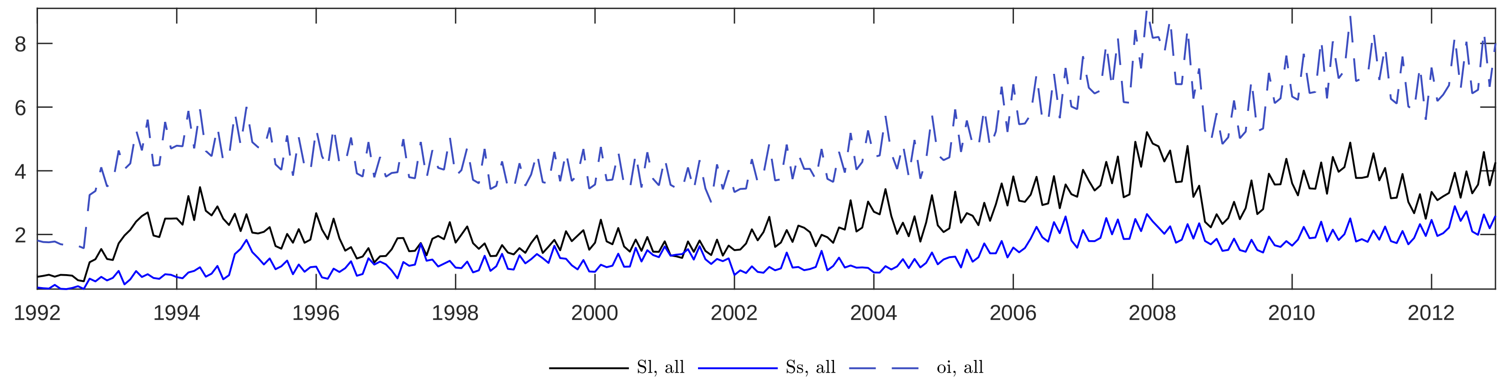
Figure 6: Open Interest and Net Financial Flows over Subsets of Traded Commodities



First, note how open interest on average rises from about four times world production to nearly 8. This rise hides considerable diversity among the commodities. Open interest for softs and fuel rise, but for metals open interest rises less. There is also considerable variation across commodity groups in the level of open interest. Net financial flows do not exhibit a trend, but here too there is considerable diversity in levels across categories. Net financial flows averages 0.04 times world production for fuels, but averages 10 times world production for metals. Finally, notice the very sharp volatility in open interest at the high frequencies.

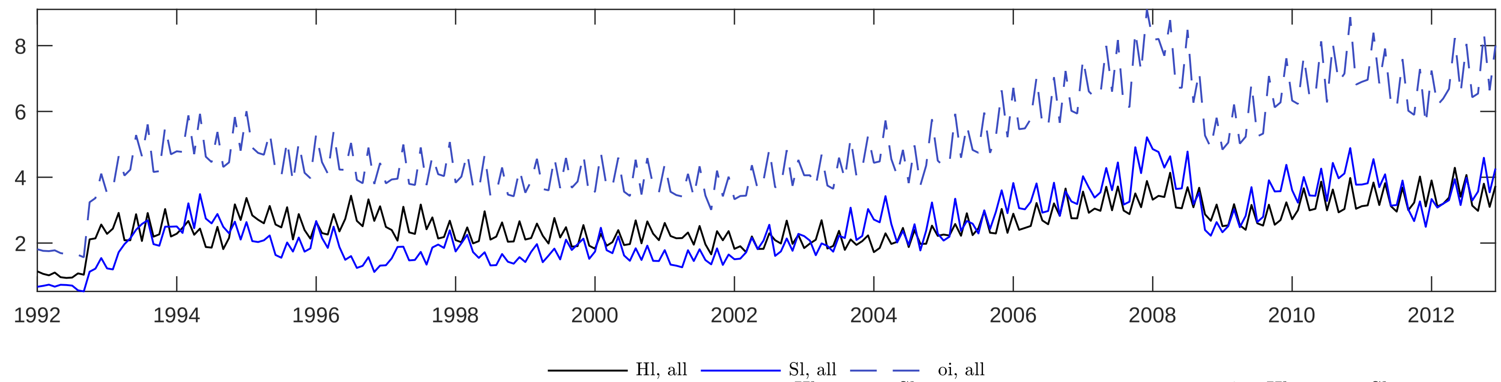
Figure7 explores the reasons underlying the patterns in Figure6. Figure7 displays the open interest and long and short positions for commercial and non-commercial traders, averaged over all traded commodities. Recall that Figure6 shows there was little increase in nff=Sl-Ss, which suggests the increase in oi was associated with an increase in trading within the group of outsiders and within the group of insiders, and that there wasn't an increase across groups.

Figure 7: Net Financial Flows and OI, Insiders Versus Outsiders: All Commodities



mean, 1992-1995: oi = 4.09, Sl = 1.88, Ss = 0.77, mean, 2009-2012: oi = 6.74, Sl = 3.52, Ss = 1.98

Δoi = 2.65, ΔSl = 1.64, ΔSs = 1.21, $\Delta nff/\Delta oi$ =0.16



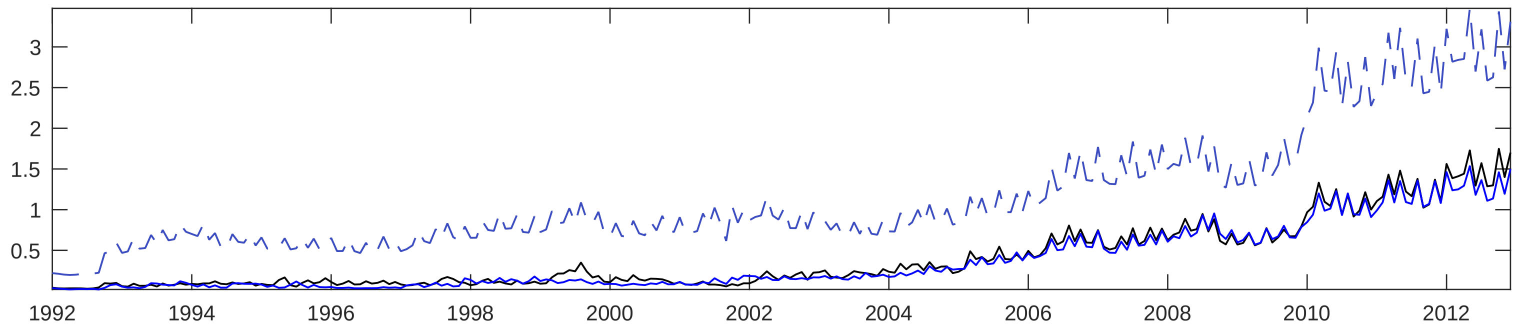
mean, 1992-1995: oi = 4.09, Hl = 2.21, Sl = 1.88, mean, 2009-2012: oi = 6.74, Hl = 3.22, Sl = 3.52

Δoi = 2.65, ΔHl = 1.01, ΔSl = 1.64, $\Delta Sl/\Delta oi$ =0.62

Figure7 seems to suggest that the increased activity occurred both among insiders and outsiders. Note from the bottom panel that both Hl and Ss have gone up by about the same amount, on average across all commodities.

Figure8 looks at the fuels.

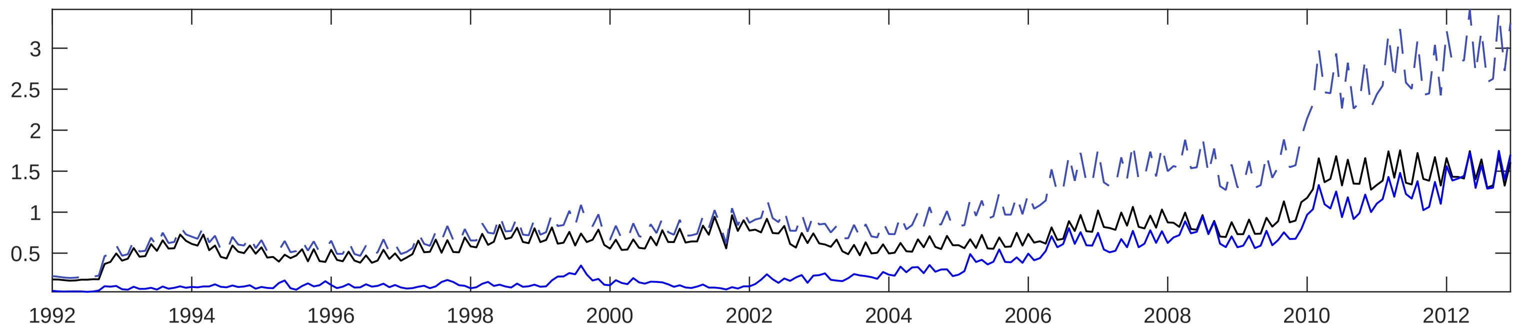
Figure 8: Net Financial Flows and OI, Insiders Versus Outsiders: Fuels



Sl, fuel Ss, fuel oi, fuel

mean, 1992-1995: oi = 0.53, Sl = 0.08, Ss = 0.06, mean, 2007-2011: oi = 2.44, Sl = 1.11, Ss = 1.04

Δoi = 1.90, ΔSl = 1.03, ΔSs = 0.98, $\Delta nff/\Delta oi$ =0.03



Hl, fuel Sl, fuel oi, fuel

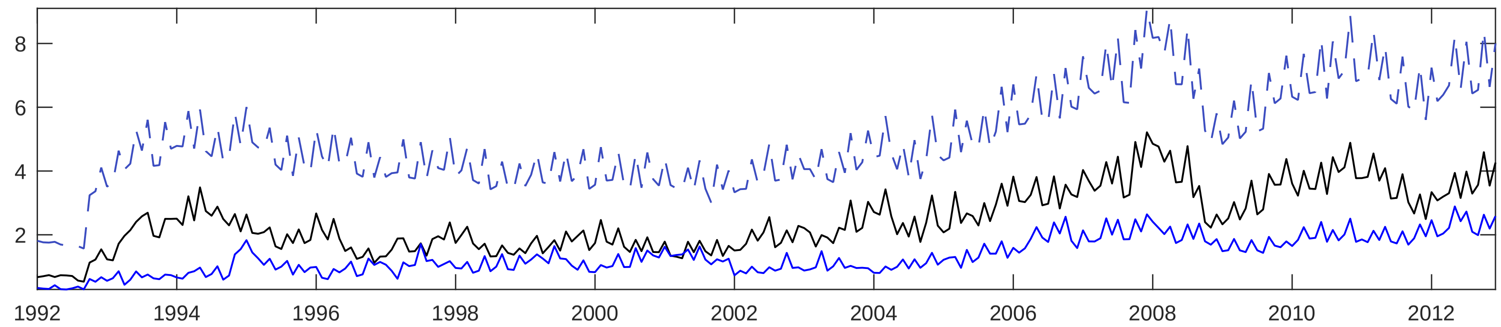
mean, 1992-1995: oi = 0.53, Hl = 0.45, Sl = 0.08, mean, 2009-2012: oi = 2.44, Hl = 1.32, Sl = 1.11

Δoi = 1.90, ΔHl = 0.87, ΔSl = 1.03, $\Delta Sl/\Delta oi$ =0.54

Note that, as we saw before, it is clear that there is very little nff (top and middle panel). The increase in open interest appears to have occurred among outsiders, with an increase in SI and Ss. Before the 2000's, the insiders represented almost all open interest, but the later increase in open interest occurred because of the increased participation among each other of insiders.

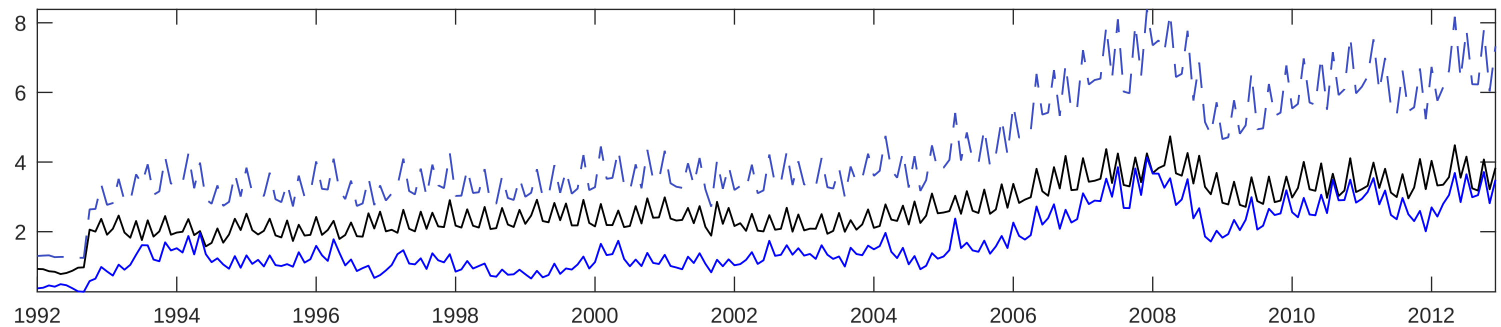
Figure9 displays results for softs. Open interest has increased substantially, from around 3.3 times world production in the early period to around 6.27 in the later period, peaking at times at around 8. Net financial flows have, in percent terms, increased by less.

Figure 9: Net Financial Flows and OI, Insiders Versus Outsiders: Softs



mean, 1992-1995: oi = 2.87, Sl = 1.05, Ss = 0.51, mean, 2009-2012: oi = 6.15, Sl = 2.77, Ss = 1.83

Δoi = 3.27, ΔSl = 1.72, ΔSs = 1.32, $\Delta nff/\Delta oi$ =0.12



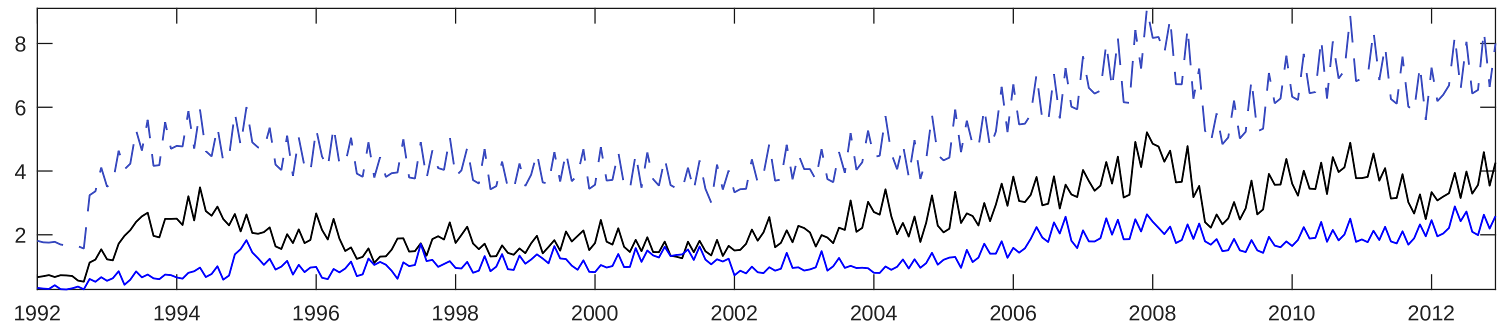
mean, 1992-1995: oi = 2.87, Hl = 1.82, Sl = 1.05, mean, 2009-2012: oi = 6.15, Hl = 3.38, Sl = 2.77

Δoi = 3.27, ΔHl = 1.56, ΔSl = 1.72, $\Delta Sl/\Delta oi$ =0.52

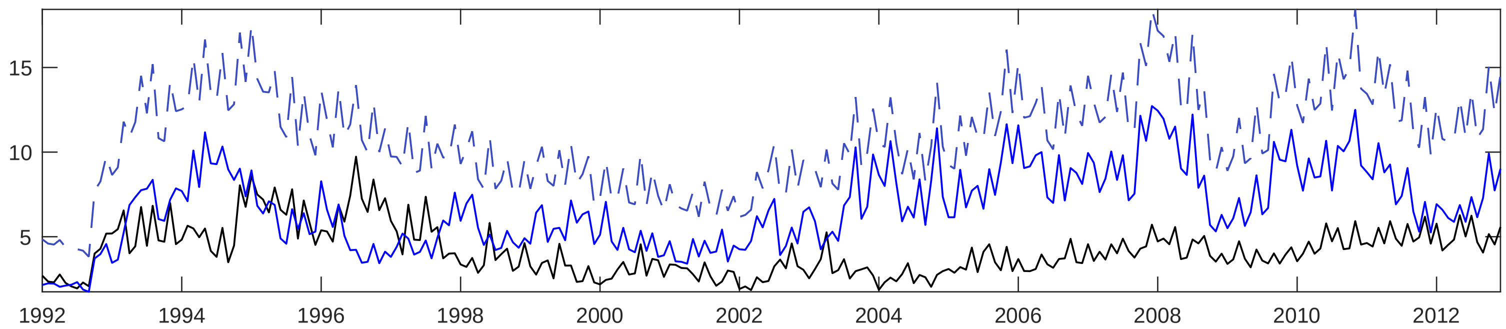
According to Figure9, open interest doubled from roughly 4 to nearly 8 times world production. Δoi\Delta oiConsistent with Figure6, the rise in nff seems modest (see first and second panels). The rise in open interest appears to be due primarily to increased activity by outsiders.

Figure10 displays results for Metals. In this case, net financial flows have increased even more than open interest. Outsiders have substantially increased their long positions, while insiders have cut back.

Figure 10: Net Financial Flows and OI, Insiders Versus Outsiders: Metals



mean, 1992-1995: oi = 11.09, Sl = 6.04, Ss = 2.21, mean, 2009-2012: oi = 12.71, Sl = 8.07, Ss = 3.35
Δoi = 1.62, ΔSl = 2.03, ΔSs = 1.14, $\Delta nff/\Delta oi$ =0.55



mean, 1992-1995: oi = 11.09, Hl = 5.05, Sl = 6.04, mean, 2009-2012: oi = 12.71, Hl = 4.64, Sl = 8.07
Δoi = 1.62, ΔHl = -0.41, ΔSl = 2.03, $\Delta Sl/\Delta oi$ =1.25

The paper describes a decomposition of oi into various forms of insurance. Table 8 provides that decomposition:

Table 8: Insurance

	Within insiders	Within outsiders	Between trader groups
All	0.63	0.27	0.1
Fuels	0.71	0.24	0.05
Metals	0.46	0.27	0.27
Softs	0.59	0.28	0.13

Note: Insider insurance is $\frac{\min\left\{H^L,H^s\right\}}{oi}$, Outsider insurance is $\frac{\min\left\{S^L,S^s\right\}}{oi}$,Insurance between trader groups: $\frac{|nff|}{oi}$

4.1.3. Spot Prices and Measures of World Production for Commodities in the CFTC Data

Table 9 reports on the commodity prices, quantities and values of production.

Table 9: Annual Spot Prices and World Production for Commodities in CFTC Data

Name	World Production (unless otherwise indicated)			Price			Value (millions $)	
	Range	Units	Source	Range	Units	Source	Range	Source
oil	1972-2014	thousands of barrels per day	BP[1]	1972-2014	dollars per barrel	BP[2]	1972-2014	derived
coal	1981-2014	millions of tonnes	BP	1990-2014	USD per tonne	BP[3]	1990-2014	derived
natural gas	1970-2014	billion cubic meters (bcm)	BP	1989-2014	USD per million BTU	BP[4]	1989-2014	derived
cotton	1965-2014	millions 480-lb bales	USDA[5]	1965-2014	cents per pound	USDA[6]	1965-2014	derived
roundwood	1961-2014	cubic meters	FAOSTAT[7]	1961-2014	dollars per cubic meter	FAOSTAT[8]	1961-2014	derived
sugar	1959-2015	1000 tonnes	USDA[9]	1916-2016	cents per pound	Trading Economics[10]	1959-2015	derived
pig crop	1960-2016	production, pig crop (in 1000 head)	USDA[11]	1991-2014	USD per tonne	FAOSTAT[12]	1991-2014	derived
calves	1960-2016	production, calf crop (in 1,000 head)	USDA[13]	1991-2014	USD per tonne	FAOSTAT[14]	1991-2014	derived
rice	1986-2015	millions of metric tons	FAOSTAT[15]	1991-2014	USD per tonne	FAOSTAT[16]	1991-2014	derived
cowmilk	1961-2013	tonnes	FAOSTAT[17]	1991-2013	USD per tonne	derived	1991-2013	FAOSTAT[18]
oats	1961-2013	tonnes	FAOSTAT[19]	1991-2014	USD per tonne	FAOSTAT[20]	1991-2013	FAOSTAT[21]
wheat	1961-2013	tonnes	FAOSTAT[22]	1991-2014	USD per tonne	FAOSTAT[23]	1991-2013	FAOSTAT[24]
soybeans	1961-2013	tonnes	FAOSTAT[25]	1991-2014	USD per tonne	FAOSTAT[26]	1991-2013	FAOSTAT[27]
coffee, green	1961-2012	tonnes	FAOSTAT[28]	1991-2014	USD per tonne	FAOSTAT[29]	1991-2013	FAOSTAT[30]
cocoabean	1961-2012	tonnes	FAOSTAT[31]	1991-2011	USD per tonne	FAOSTAT[32]	1991-2013	FAOSTAT[33]
oranges	1961-2012	tonnes	FAOSTAT[34]	1991-2014	USD per tonne	FAOSTAT[35]	1991-2013	FAOSTAT[36]
corn	1993-2015	millions of metric tons	FAOSTAT[37]	1991-2014	USD maize per tonne	FAOSTAT[38]	1991-2013	FAOSTAT[39]
gold	1900-2014	tonnes	USGS[40]	1900-2014	USD per tonne	USGS[41]	1900-2014	derived
silver	1900-2014	tonnes	USGS[42]	1900-2014	USD per tonne	USGS[43]	1900-2014	derived
copper	1900-2014	tonnes	USGS[44]	1900-2014	USD per tonne	USGS[45]	1900-2014	derived
platinum	1900-2014	tonnes	USGS[46]	1900-2014	USD per tonne	USGS[47]	1900-2014	derived
aluminum	1900-2014	tonnes	USGS[48]	1900-2014	USD per tonne	USGS[49]	1900-2014	derived
soybean oil	1961-2013	tonnes	FAOSTAT[50]	1980-2014	cents per pound in US	USDA[51]	1980-2014	derived
propane	1980-2012	thousands of barrels per day	EIA[52]	1992-2015	dollars per gallon	EIA[53]	1992-2012	derived
distillate fuel oil	1980-2012	thousands of barrels per day	EIA[54]	1986-2015	dollars per gallon	EIA[55]	1986-2012	derived
gasoline	1980-2012	thousands of barrels per day	EIA[56]	1986-2015	dollars per gallon	EIA[57]	1986-2012	derived
palladium	1900-2014	tonnes	USGS	1900-2014	USD per tonne	USGS[47]	1900-2014	derived
soybean meal	1964-2016	tonnes	IMF	1980-2015	USD per tonne	USDA	1980-2015	derived

NOTE : in the following, the excel file containing the original data source is reported.[1]all BP data in replication file, bp-statistical-review-of-world-energy-2015-workbook.xlsx;[2]1861-1944 US Average, 1945-1983 Arabian Light posted at Ras Tanura, 1984-2012 Brent dated.[3]US Central Appalachian coal spot price index;[4]US prices, Henry Hub, source: Source: Heren Energy Ltd.;[5]Appendix table 15–World cotton supply and use, 1965/66-2015/16, http://usda.mannlib.cornell.edu/MannUsda/viewDocumentInfo.do?documentID=1282, replication file USDA_Cotton.xls;[6]USDA Appendix table 1–U.S. cotton supply and use, 1965/66-2015/2016, replication file USDA_cotton_price.xlsx;[7]in Roundwood_m3_production.xlsx;[8]ratio of value of US exports to quantity of US exports, in Roundwood_FAO.xlsx;[9]Sugar, Centrifugal Total World Sugar Production, usda_Sugar.xlsx;[10]Data are monthly, 1914-1960 and daily thereafter, we worked with the sample mean of all available observations for each year, source: http://www.tradingeconomics.com/commodity/sugar, replication file: Sugar_TradEcon.xlsx;[11]World production, Swine,Swine.xlsx;[12]US producer price, Pig Live Weight, FAOPrices.xlsx;[13]World production, Cattle, Cattle.xlsx;[14]US producer price, Cattle Live Weight,FAOPrices.xlsx;[15]World Rice, Rough Production, Rice_milled_rough.xlsx;[16]US producer price, paddy rice, FAOPrices.xlsx;[17]Cow milk, whole, fresh (livestock primary), FAO_cow_milk.xlsx;[18]Gross Production Value (current million US$), World, FAOValues.xlsx;[19]World production, oats,FAOValues.xlsx;[20]US producer price, oats, FAOPrices.xlsx;[21]Gross Production Value (current million US$), World;[22]World Wheat Production, FAO_Quantity.xlsx;[23]US Producer Price, Wheat, FAOPrices.xlsx;[24]Gross Production Value (current million US$), wheat,FAOValues.xlsx;[25]World Soybean Production, FAO_Quantity.xlsx;[26]US Producer Price, Soybeans, FAOPrices.xlsx;[27]World Soybeans Gross Production (current million US$), FAOValues.xlsx;[28]World Coffee, green Production, FAO_Quantity2.xlsx;[29]Coffee, green US Producer Price (USD/tonne), FAOPrices.xlsx;[30]Coffee, green Gross Production Value (current million US$), FAOValues.xlsx;[31]World Cocoa beans Production, FAO_Quantity.xlsx;[32]Indonesia Cocoa beans Producer Price,FAOPrices.xlsx (Indonesia is the world's second largest producer of Cocoa beans, after Côte d'Ivoire; we use Indonesian data because the 2012 observation for the price of cocoa beans is available for Indonesia, but not Côte d'Ivoire);[33]Cocoa beans Gross Production Value (current million US$), FAOValues.xlsx;[34]World Oranges Production,FAO_Quantity2.xlsx;[35]US Oranges Producer Price (USD/tonne), FAOPrices.xlsx;[36]World Oranges Gross Production Value (current million US$),FAOValue.xlsx[37]World Corn production, FAO_corn_barley.xlsx;[38]US Maize Producer Price (USD/tonne), FAOPrices.xlsx;[39]World Maize Gross Production Value (current million US$), FAOValues.xlsx;[40]World production, ds140-gold.xls;[41]Unit value ($/t), ds140-gold.xls;[42]World production, ds140-silver.xls;[43]ds140-silver.xls;[44]ds140-coppe.xls;[45]ds140-coppe.xls;[46]ds140-plati.xls;[47]ds140-plati.xls;[48]ds140-alumi.xls;[49]ds140-alumi.xls;[50]World Soybean oil Production, FAO_QuantityProcessed.xlsx;[51]US price, crude, Decatur,soybeanoil_US_prodandprice.xlsx;[52]World Refinery Output of Liquefied Petroleum Gases (Thousand Barrels Per Day),Refinery_Output_of_Liquefied_Petroleum_Gases_(Thousand_Barrels_Per_Day).xlsx;[53]Mont Belvieu, TX Propane Spot Price FOB (Dollars per Gallon),Peroleum_Product_Prices.xlsx (page, Data 7);[54]World Refinery Output of Distillate Fuel Oil, Refinery_Output_of_Distillate_Fuel_Oil_(Thousand_Barrels_Per_Day).xlsx;[55]New York Harbor No. 2 Heating Oil Spot Price FOB Peroleum_Product_Prices.xlsx (page, 'Data 4');[56]World Refinery Output of Motor Gasoline, Refinery_Output_of_Motor_Gasoline_(Thousand_Barrels_Per_Day).xlsx;[57]New York Harbor Conventional Gasoline Regular Spot Price FOB, Gasoline_Prices.xlsx;

4.2. Prices of Monthly CFTC-Traded Commodities

4.2.1. List of Variables

Monthly observations on CFTC-traded commodities are summarized in Table10

Table 10: Monthly Spot Prices for Commodities in CFTC Data

Name	Range	Units	Source	Description
Crude_oil	1959/1-2015/12	US$/BBL	Trading Economics	NA
Coal	1980/1-2015/12	US$ per metric ton	IMF	Australian thermal coal, 12,000- btu/pound, Less than 1% sulfur, 14% ash, FOB Newcastle/Port Kembla
Natural_gas	1990/4-2015/12	US$/MMBtu	Trading Economics	NA
Cotton	1959/6-2015/12	Cents/lb	Trading Economics	NA
Soft Sawnwood	1980/1-2015/12	US$ per cubic meter	IMF	Average export price of Douglas Fir, U.S. Price, US$ per cubic meter
Sugar	1959/1-2015/12	Cents/LB	Trading Economics	NA
Swine (pork)	1980/1-2015/12	US cents per pound	IMF	51-52% lean Hogs, U.S. price, US cents per pound.
Beef	1980/1-2015/12	US cents per pound	IMF	Australian and New Zealand 85% lean fores, CIF U.S. import price
Rice	1981/4-2015/12	USD/CWT	Trading Economics	NA
raw_milk	1959/1-2015/12	index	FRED	Producer Price Index by Commodity for Farm Products: Raw Milk
Oat	1979/1-2015/12	US cents per Bushel	Trading Economics	NA
Wheat	1982/3-2015/12	US$/BU	Trading Economics	NA
Soybeans	1959/7-2015/12	US$/BU	Trading Economics	NA
Coffee	1972/7-2015/12	US cents per LB	Trading Economics	NA
Cocoa	1959/7-2015/12	US$ per metric ton	Trading Economics	NA
Oranges	1980/1-2015/12	US$ per metric ton	IMF	Miscellaneous oranges CIF French import price
Corn	1959/1-2015/12	US$ per BU	Trading Economics	NA
Gold	1968/1-2015/12	USD/t oz	Trading Economics	NA
Silver	1975/1-2015/12	USD/t oz	Trading Economics	NA
Copper	1980/1-2015/12	US$ per metric ton	IMF	Grade A cathode, LME spot price, CIF European ports
Platinum	1968/3-2015/12	USD per t oz	Trading Economics	NA
Aluminum	1986/3-2015/12	US$ and cents per metric ton	Index Mundi	NA
Soybean Oil	1980/1-2015/12	US$ per metric ton	IMF	Chicago Soybean Oil Futures (first contract forward) exchange approved grades
propane	1977/6-2015/12	index	FRED	Producer Price Index by Commodity for Fuels and Related Products and Power
heating_oil	1986/6-2015/12	index	FRED	No. 2 Heating Oil Prices: New York Harbor
gasoline	1967/1-2015/12	index	FRED	Consumer Price Index for All Urban Consumers: Gasoline (all types)
Palladium	1987/1-2015/12	US$ per t oz	Trading Economics	NA
Soybean Meal	1980/1-2015/12	US$ per metric ton	IMF	Chicago Soybean Meal Futures (first contract forward) Minimum 48 percent protein

In the case of data obtained from the FRED and the IMF, the detailed description of the spot price is provided. In the case of Trading Economics and Index Mundi, the detailed description is not so easy to see from their website. We obtained annual observations on world production for these variables by simply interpolating the annual world production data.

4.2.2. Figures of Prices and Volume

Figure11 displays the log indices for the prices of various subcatagories of commodities. In addition, we include the price of one commodity, crude oil. The basic patterns in the data are broadly similar. There is an increase in volatility in the 2000's and the trend in prices rises. But, there is considerable heterogeneity within this patter. The evolution of the price of softs is quite modest, and the overall rise in their price is around 1.2 percent per year over the 21 year sample. On the other end of the spectrum is fuel, and its primary component, oil. These rise nearly 5 percent per year over the sample. Moreover, since 1999 the rise was roughly twice that amount. Metals prices also highly volatile in the early 2000s, though their volatility was reduced somewhat in the later period, when oil fuel and oil became clearly the most volatile.

Figure 11: Log Aggregate Prices and Oil

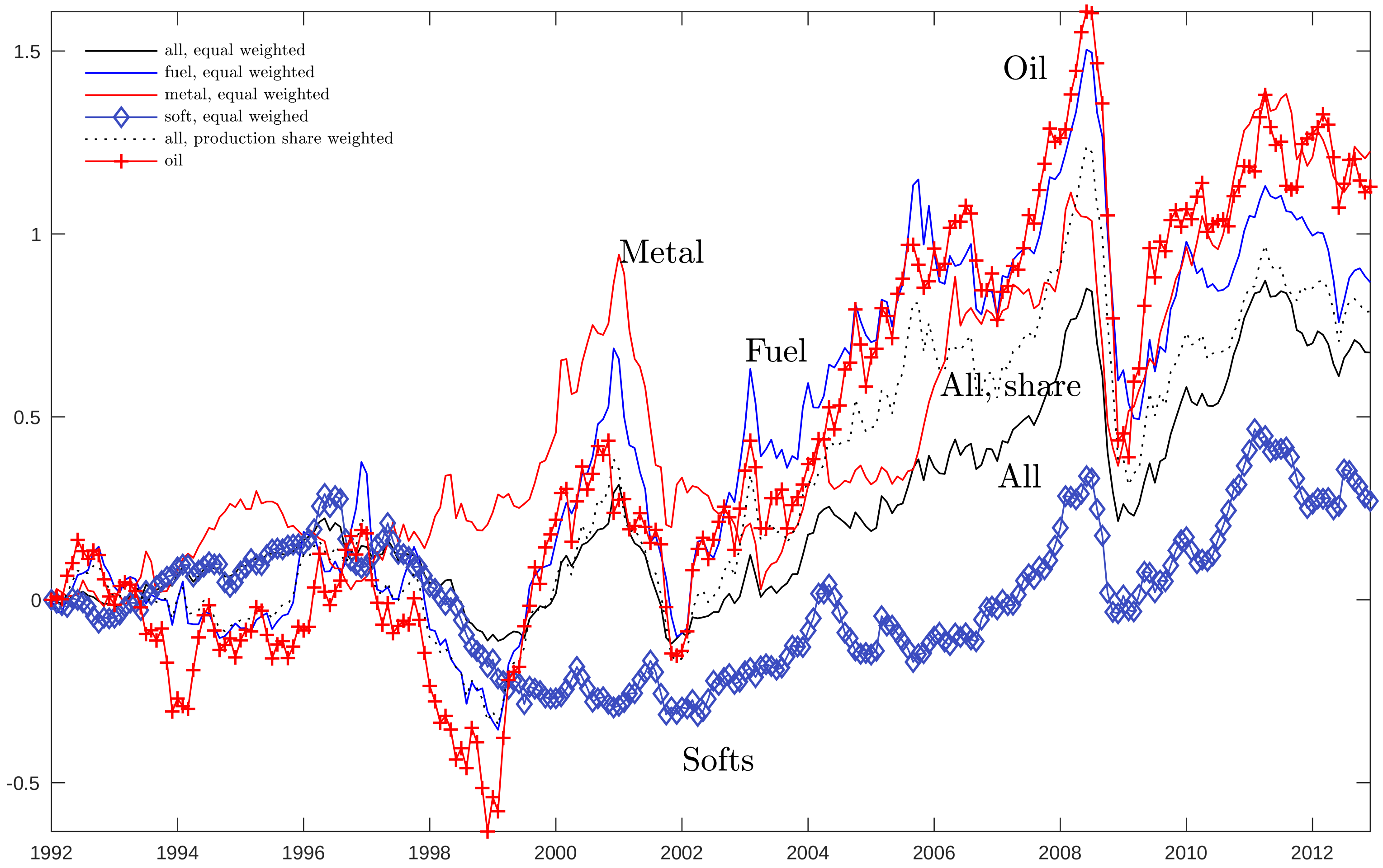


Figure13 shows prices and our two volumes of trade measures. All indices are represented based on equally-weighted shares and production shares. Here, there is a clear message. As noted before, the price of fuel rises by most, and yet the rise in open interest and nff is relatively low for those commodities. Another commodity, metals, shows a similar rise in spot prices, and it is heavily financialized according to our two measures. On average, that financialization - though at a high level - did not grow much, relative to the weighted measure. That does show some rise in nff and oi. Clearly, there is some heterogeneity inside the group of metals, which is worth examining more closely.

Metals prices are reported in Figure12. Not surprisingly, given the information in Figure13, the price of metals commodities is quite heterogeneous. One metal, Palladium, soared over 10% per year in the 1990s, and then fell back someone in the subsequent period. It's behavior is completely different from the other prices. The most important metal, in terms of share of world production of metals, is aluminum. By comparison to others, it's evolution is relatively moderate. It would be good to study the behavior of these commodity prices in relation to the associated measures of financialization.

Figure 12: Log, Metals Prices

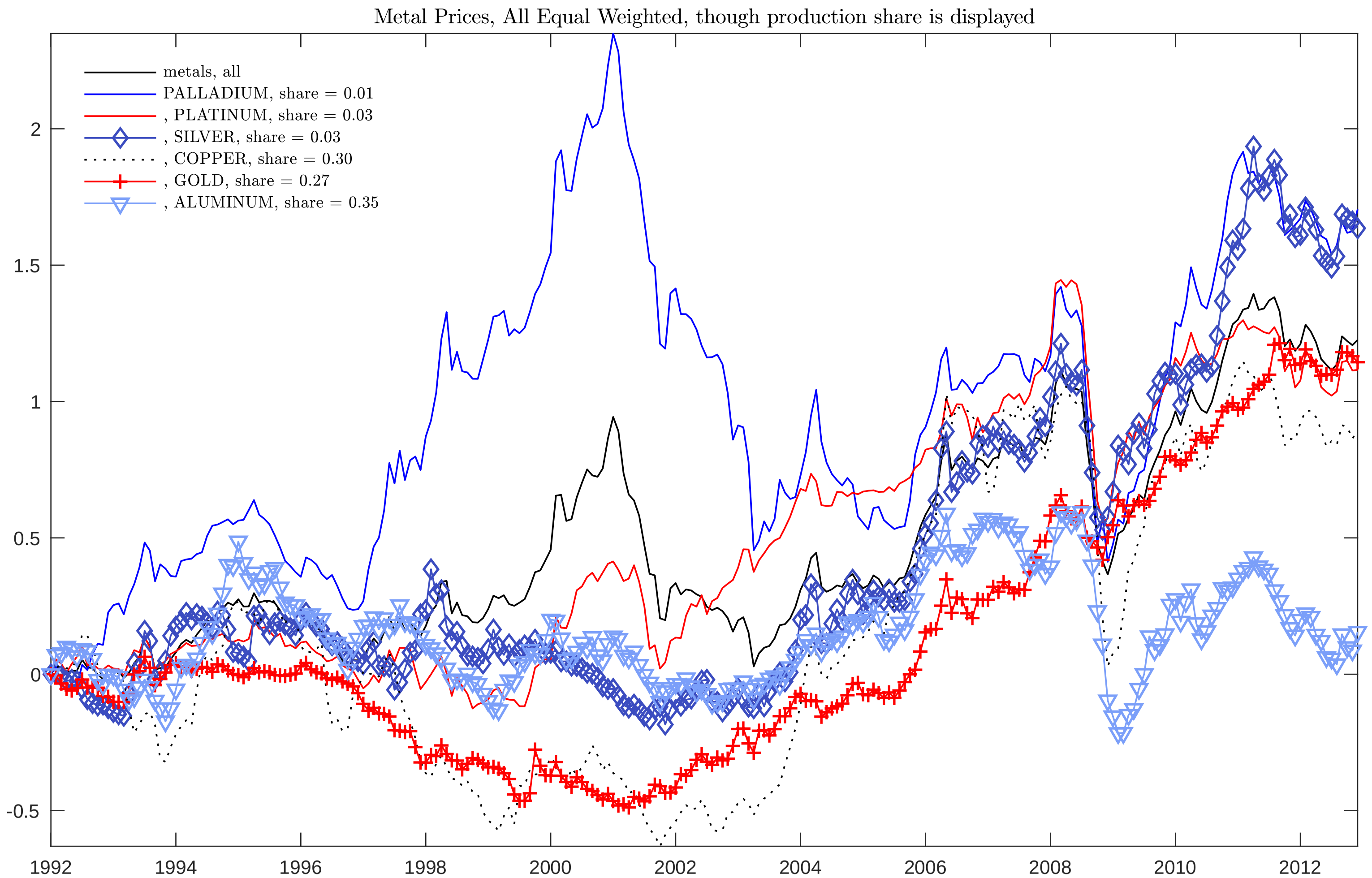


Figure 13: Prices and Trading Volume:
production weighted (left), equal weighted (right)

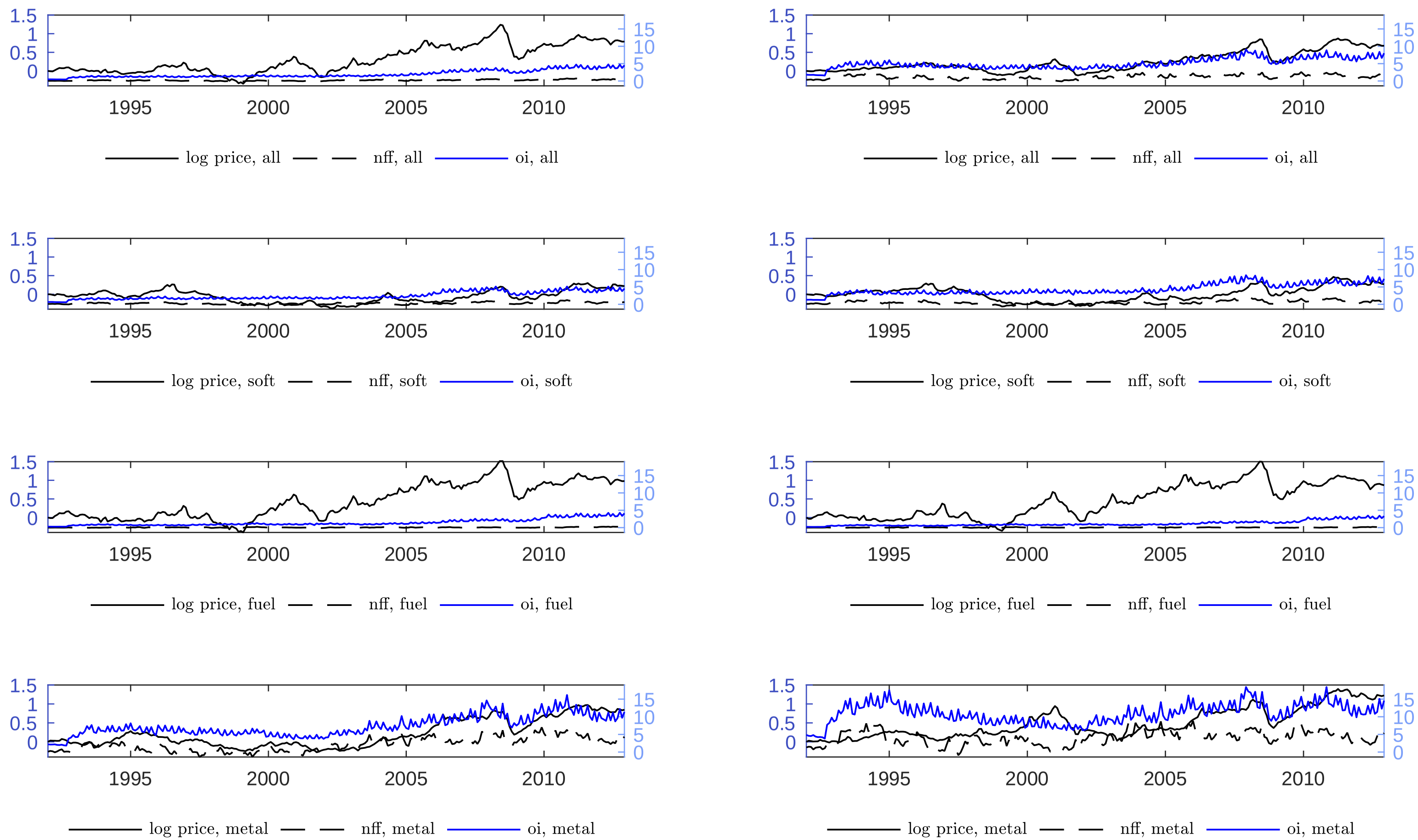
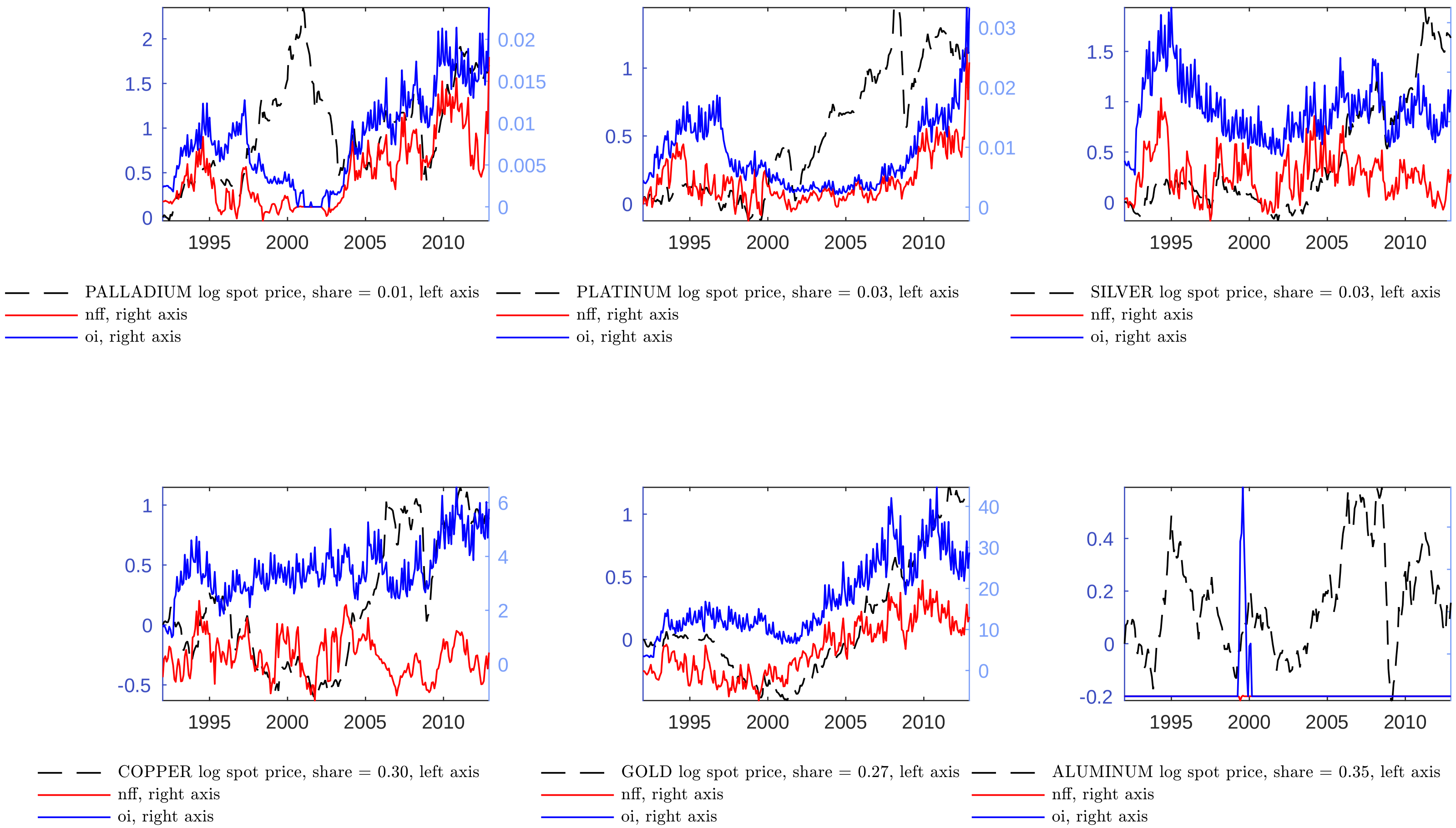


Figure14 displays prices and volume of trade for our six traded metals. Gold could be the poster-child of the proposition that speculators drive commodity prices. Open interest and net financial flows move closely with the pattern of gold prices. All three go down until 2000 (though slightly out of phase with each other) and then they rise afterward. On the other hand, silver prices display a similar temporal pattern to gold prices, and yet the volume of trade in futures markets is not at all the same. It is interesting how open interest and net financial flows in these two commodities is enormous. Open interest in silver is about 80 times world production at the end of the sample, and the corresponding number for gold is around 40. Palladium is another interesting example, because the price of the commodity moves very differently - at least around the turn of the millennium - from the flows in the futures market. Copper and Platinum are two more commodities where futures markets and volume of trade do not seem to co-move closely.

Figure 14: Metals



5. Monthly Data on Non-Traded Commodities

The following sections describe the data contained in the structure, non_tr_monthly_with_q, described in section3..

5.0.1. Fishmeal

The annual production data were obtained fromINDEXMundi. Data for the 16 biggest producer countries were obtained and world fishmeal production was approximated by the simple sum across the 16 countries.The 16 countries covered are, in order of largest to smallest producer: Peru, European Union, Chile, Thailand, China, USA, Japan, Russia, Vietnam, Norway, Iceland, Ecuador, South Africa, Canada, Malaysia and Mexico. The data are in 1000's of metric tons. All these data go to 2016, but they do not all start the same year. When observations were missing, they were replaced by the first observed data point. The program (get_fishmeal_m.m) also examined a method of 'interpolation by related variables' for filling in the missing observations, but this produced absurd results (some huge and some very negative numbers). In any case, for the period of the analysis, post 1992, it made very little difference which way of filling in the missing observations was used.

In principle, a third method for filling in data could have been used, namely to set the missing observations to zero. This would make sense if the reason the observations are missing for a particular country is that that country had not yet started to produce any fishmeal. But, a cursory examination of the data suggests this option is implausible. For example, the second biggest producer of fishmeal is the European Union, and those data do not start until 1999, when the EU was started. Of course, the countries in the EU must have been producing fishmeal before 1999. But, INDEXMundi does not report those data. Surely, setting those data to zero would be a mistake.

The total world production of fishmeal is taken as the simple sum of production in the above 16 countries. Implicitly, this assumes the price of fishmeal in each country is the same, which of course must not be literally true. But, we were not able to obtain country-level prices. Instead, we found monthly prices for Peruvian fishmeal in the commodity data set provided by the IMF. These data are "Fishmeal, Peru Fish meal/pellets 65% protein, CIF, US$ per metric ton". The IMF data set, External_Data.xls, can be found athere. These data can also be found in the IndexMundi database (seehere).

The World Bank also provides commodity price data. These can be found inthis zip file. The World Bank page where this file can be found ishere, and the excel file containing the monthly data is givenhere XXX this link does not work. Isthis correct one? XXX. According to that excel file, the price of fishmeal is "Fishmeal, $/mt, current$ - World. Fishmeal (any origin), 64-65%, c&f Bremen, estimates based on wholesale price, beginning 2004; previously c&f Hamburg". The word, 'World', suggests to us that the price is in some way representative of the price of the quantities produced in the various countries. We believe that c&f stands for 'cost and freight'. Before 2004, the price corresponded to cost and freight in Hamburg and after that it was in Bremen. The IMF and World Bank price series look similar at the low frequencies, though the World Bank series is a little more volatile. We compared the value of world production based on the two price series over the period, 1992-2012, and they look reasonably similar. We used the IMF price data.

Some historical observations on fishmeal: "In 1991 and 1992, fishmeal production collapsed partly because of the El Ninio phenomenon along the Pacific Coast, but also due to a collapse in Japanese pelagic catches and the dismantling of the former Soviet-Union's fishing fleet. ... Since 1993, there has been a sharp decline in fishmeal prices. This is partially due to an increase in Peruvian production" (seethe pdf file.)

This article suggests that there exist futures markets (maybe over-the-counter) in fishmeal: We could not find direct evidence of any organized futures markets in fishmeal, further suggesting that what markets exist are over-the-counter. The followingremark in a CME Group document reinforces the notion that there are no fishmeal futures contracts:"Hedging a cash commodity using a different but related futures contract when there is no futures contract for the cash commodity being hedged and the cash and futures markets follow similar price trends (e.g., using soybean meal futures to hedge fish meal).” Interestingly, this does raise another possibility, that you can in effect hedge a commodity that is not traded by ’cross hedging’, i.e., hedging another commodity that is closely related.

5.0.2. Palm Oil

The annual quantity of world production data were obtained fromINDEXMundi. Data for the 16 biggest producer countries were obtained and summed to obtain a measure of world production.The 16 countries covered are, in order of largest to smallest producer: Indonesia, Malaysia, Thailand, Colombia, Nigeria, Ecuador, Honduras, Papua New Guinea, Ghana, Guatemala, Côte D’ivoire, Brazil, Cameroon, Costa Rica, Congo and India. The data are in 1000’s of metric tons. The same strategy as for fishmeal was used to deal with the different start dates. Here, the choice of strategy made virtually zero difference.

The monthly price data are "Palm oil, Malaysia Palm Oil Futures (first contract forward) 4-5 percent FFA, US$ per metric ton". The data were taken from the IMF database cited in Fishmeal. Monthly prices for palm oil also appear in the World Bank database. They are called "Palm oil (Malaysia), 5% bulk, c.i.f. N. W. Europe". Although they have a different definition, when graphed they appear roughly the same. The IMF data are what is reported in IndexMundi, although IndexMundi reports the source as ’World Bank’. We used the IMF data.

Note the reference, in the name of the IMF variable to futures. This (together with information on the IndexMundi website, seehere) suggests that palm oil is in fact traded. It appears that Palm Oil has for a long time been traded in the Malay futures’ exchange, Bursa Malaysia Derivatives Berhad (BMD). It appears that the CME Group has formed a ’strategic’ partnership with BMD, giving CME customers access to Palm Oil futures using the CME Globex electronic trading platform. Seethe site.

We verified that Palm Oil does not appear in the CFTC data base reported inthe zip file. The url where the last zip file was foundhere. This does not mean that there are zero futures’ contracts in palm oil in the US. The Commitments of Traders (COT) reports provide a breakdown of each Tuesday’s open interest for markets in which 20 or more traders hold positions equal to or above the reporting levels established by the CFTC. Palm oil does not satisfy this requirement. The threshold for palm oil is 25 contracts. Of course, one or two traders could satisfy that threshold, say by trading one billion contracts each, and then Palm Oil would be excluded because you need at least 20 traders satisfying the threshold. (This is from a January 31, 2017 email from Jay Huhman, of the CFTC. It is not clear whether there even exists a complete data set, more complete that the COT reports.) There do exist Palm Oil futures exchanges, but they are not in the US. For example, Palm Oil futures have been traded for a long time in the Malaysian futures’ exchange, Bursa Malaysia Derivatives Berhad (BMD). Online, there is evidence that the CME Group has formed a ’strategic’ partnership with BMD, giving CME customers access to Palm Oil futures using the CME Globex electronic trading platform. Seethe website in CME. It is our understanding that these BMD contracts are not reported in the CFTC data because these contracts are not linked to identical contracts traded on US exchanges (by ‘identical’ we mean the contracts on BMD and the US exchange have the same futures price and that long and short contracts sum to zero across the two exchanges). The information onthis website suggests that not much palm oil is traded on the Globex system.

5.0.3. Tea and Rubber

World production of tea and rubber (’rubber, natural’) was taken fromFAOSTAT, and are measured in units of tonnes (i.e., metric tons). No explanation is offered about how different types of tea and rubber from different countries were aggregated (presumably, they did a simple sum). Monthly price data for tea in FAOSTAT appear to be limited to a few years. So, monthly price data were taken from the IMF commodity data set. The description of these data is: "Tea, Mombasa, Kenya, Auction Price, From July 1998, Kenya auctions, Best Pekoe Fannings. Prior, London auctions, c.i.f. U.K. warehouses, US cents per kilogram." To multiply price and quantity, we used the fact that 1 metric ton corresponds to 1,000 kg. An alternative works with the fact that FAOSTAT also provides data on the value of world production. The latter are used for purposes of constructing the index weights.

Monthly price data for rubber were obtained from the IMF dataset and is called: "Rubber, Singapore Commodity Exchange, No. 3 Rubber Smoked Sheets, 1st contract, US cents per pound." To multiply price and quantity, we used the fact that 1 metric ton corresponds to 2,204.62 pounds. But, as in the case of tea, FAOSTAT also provides annual data on the value of world production, and this is what was use to compute the weights in the index.

Tea appears not to be traded in US futures markets. It is not included in the CFTC data on Commitments of Traders. In addition, there is thisBloomberg article which reports that 'the first futures markets in tea may be opened in Kenya', which suggests, of course, that there are no such markets in the US.

Rubber also appears not to be traded in the US. According tothis WSJ article rubber appears to be traded in Asia. It is not clear that absence of tea and rubber futures contracts in US exchanges limits in any way US traders' opportunity to hedge or speculate in these contracts.

As in the case of Palm oil, it is not clear that US traders are constrained in any way by the fact that contracts are not traded on US soil. However, we found no evidence that tea and rubber contracts are available in CME's Globex trading platform.

5.0.4. Uranium

We obtained the annual quantity of worlduranium production from the United Nations' Energy Statistics Database. These data are in units of tonnes. We used the conversion factor, 1 tonne = 2,204.62 pounds. The data are provided at the level of countries, and we added over all countries for which data are available. We compared the resulting aggregate data with data from the World Nuclear Association and from the US Energy Information Administration (EIA). The UN data cover a longer period and over the subsamples in which they coincide, they are quite similar to the other data. That is why we used the UN data.

The monthly price series on uranium were obtained from the IMF commodity price data set and correspond to "Uranium, NUEXCO, Restricted Price, Nuexco exchange spot, US$ per pound".

The volume of trade in Uranium futures is not reported in the CFTC. However, Uranium futures areavailable on the CME's Globex trading platform.

Following is a discussion of the other sources of uranium production data. An analysis of these data is contained in the program, get_non_tr_monthly_with_annual_value_q.m.

TheWorld Nuclear Association has data on US and world production for the years 2004-2014. The U.S. Energy Information Administration (EIA) has data on US production of uranium, 1993-2015 (we backcasted the data to 1992 by extending a straight line fit to the 1993 and 1994 observations). These were taken fromhere. The production data are measured in millions of pounds ofU_3O_8.

We constructed an estimate of world production over the period 1992-2014 by starting with the EIA data on Uranium Concentrate Production. We computed the ratio of world to US uranium production for the period, 2004-2014 using the data from the World Nuclear Association. The ratio, over these years, is (after rounding),

$$46, 40, 23, 25, 31, 35, 32, 35, 37, 33, 29.$$

The mean of these observations is 32. To obtain an annual series on world production we simply multiply the EIA data by 32. As a check on the consistency of our two data sources, we computed the ratio of EIA data on US production to the World Nuclear Association data on US production for the period, 2004-2014, and found that the ratio averaged 1.15. We are not sure why these series differ by 15% on average. In computing the World Nuclear Association production data, we used the information on their website that indicates quantities are measured intU, which we interpret as tonnes (i.e., metric tons) of Uranium. We used the conversion factor, 1 tonne = 2,204.62

pounds.

5.0.5. Greasy Wool

We obtained annual data on the value, in millions of dollars, of world greasy wool production over the period 1992-2012 from FAOSTAT. We obtained the monthly price of wool (in dollars per 100Kg) fromTrading Economics.They were then integrated with other trading economics data into the MATLAB mat file, data_monthly. The latter data were in turn absorbed into the MATLAB mat file, Price_Monthly. This is the place from which the price is recovered when it is combined with the value data to obtain data on raw quantities. The quantities are in units of 100Kg.

The index construction obviously uses the dollar value data obtained from FAOSTAT. Greasy wool futures contracts are traded on the Australian Futures Exchange, seethe webpage. These contracts appear not to be available on the CME's Globex platform.

5.0.6. Bananas

The monthly price data were obtained from the IMF commodities database, and correspond to "Bananas, Central American and Ecuador, FOB U.S. Ports, US$ per metric ton". World production, in tonnes and dollars, was obtained from FAOSTAT.There do not appear to be any organized futures markets in bananas, anywhere in the world.

5.0.7. Barley

The monthly price data were taken from the IMF data base, and corresponds to "Barley, Canadian no.1 Western Barley, spot price, US$ per metric ton". World production, in tonnes and dollars, was obtained from FAOSTAT. Barley is traded in ICE Futures, Canada. Seethe website. But, it appears not to be traded on the CME's Globex trading platform.

5.0.8. Rapeseed

The monthly price data were taken from the IMF data base, and corresponds to "Rapeseed oil, crude, fob Rotterdam, US$ per metric ton". World production, in tonnes and dollars, was obtained from FAOSTAT and is on my mac pro laptop in FAO_nontraded_PQ_matched.xlsx. Rapeseed oil futures are traded in Euronext Paris,see. But, it appears not to be traded on the CME's Globex trading platform.

5.0.9. Groundnuts, with Shell

The monthly price data were taken from the IMF data base, and corresponds to "Groundnuts (peanuts), 40/50 (40 to 50 count per ounce), cif Argentina, US$ per metric ton". World production, in tonnes and dollars, was obtained from FAOSTAT and is on my mac pro laptop in FAO_nontraded_PQ_matched.xlsx. Barley is traded in Euronext Paris,see. But, it appears not to be traded on the CME's Globex trading platform. There appears not to be a futures market in this. According tothe article , "With no futures market, peanut producers face higher-than-normal farming risks." Although this article is hard to read, it suggests that even though there are no futures markets for peanuts there may be alternative ways to manage risk. For example, peanut producers can get 'crop insurance' on the weather. They find this useful when they go for a loan to finance their spring planting. Crop insurance is a partial substitute for futures market. An interesting question, from the point of view of the Chari-Christiano model, is whether the story in that model about whether a futures market exists or not for a particular commodity rings true for the peanut market.

5.0.10. Hides

Monthly price data were taken from the IMF database, and are called "Hides, Heavy native steers, over 53 pounds, wholesale dealer's price, US, Chicago, fob Shipping Point, US cents per pound". World production, in tonnes, was obtained from FAOSTAT and is on my mac pro laptop in hides.xlsx.

Evidently, there were futures trades in hides many years ago,see. See alsothis article. But, we have not been able to find any futures markets in hides in operation now. For example, hides are not listed in Globex.

5.0.11. Iron Ore

Monthly price data were taken from the IMF database, "Iron Ore, China import Iron Ore Fines 62% FE spot (CFR Tianjin port), US dollars per metric ton". The value of world production was obtained from USGS. There appear to be futures markets, but they seem to be in foreign countries. If there are futures markets in the US, then it must be that they don't satisfy the size requirements to get into the CFTC data. The contracts are traded onGlobex. The followingarticle suggests that Globex only started offering access to futures in iron ore since 2014.

5.0.12. Tin

Monthly price data wereobtained from Trading Economics. Tin Futures are available for trading in The London Metal Exchange (LME), but tin futures do not appear in the CFTC data. Of course, they might be traded in some US exchange, but not at a level that satisfies the CFTC size requirements. Tin futures are not available on Globex.

Annual value of world production is obtained from USGS. Quantity and price in metric tons are also provided by USGS.

5.0.13. Zinc

This is the same as Tin, except with 'Tin' replaced by 'Zinc'. In addition, Zinc is available on Globex.

5.0.14. Nickel

This is the same as Tin, except with 'Tin' replaced by 'Nickel'.

5.0.15. Lead

This is the same as Zinc, except with 'Zinc' replaced by 'Lead'.

5.0.16. Olive Oil

Monthly price data taken from IMF data base and is described as 'extra virgin less than 1% free fatty acid, ex-tanker price U.K.' Olive oil futures contracts do not appear in the CFTC. Either they are traded in a US exchange but don't meet the CFTC size requirements, or they cannot be traded in the US. Olive oil futures are traded in an exchange in Spain, the MFAO. Olive oil is not available on Globex. The annual quantity of world production, in thousands of tonnes, is obtained from the International Olive Council, seethis.

The value of world olive oil production is the product of the quantity obtained from the IOC, times the annual average of the monthly data.

5.0.17. Chicken

Monthly price data taken from IMF data base and is described as 'Poultry (chicken), Whole bird spot price, Ready-to-cook, whole, iced, Georgia docks, US cents per pound'. This commodity does not appear in the CFTC data. Chicken appears not to be available on the CME Globex platform. The annual quantity of world production, in tonnes, was obtained from the 'livestock primary' section ofFAOSTAT. The value of world production is just price times quantity. The 'value of agricultural production' section of FAOSTAT has the value of world production. Our value, which no doubt involves a slightly different price, is $41.26b in 1992 and the FAOSTAT number is $40.09b. The advantage of the IMF price data is that we have a precise definition of those data, and it's primarily the price data we're after. The value of world production is only for the purpose of computing the aggregate price index.

5.0.18. Lamb

Monthly price data were taken from the IMF data base are is described as 'Lamb, frozen carcass Smithfield London, US cents per pound'. This commodity does not appear in the CFTC data. Lamb appears not to be available on the CME Globex platform. The annual quantity of world production of sheep, in tonnes, was obtained from the 'Livestock Primary' section of FAOSTAT. The value of world production is reported by FAOSTAT in the 'Value of Agricultural Production' section of FAOSTAT. The value is $9.6b in 1992, and the correspond number, multiplying the IMF's lamb price times the quantity of sheep, is $17.9b. Evidently, the price of lamb is about twice the price of sheep that is implicit in the FAOSTAT data. The measurement in this case is clearly not the best.

We also considered using the quantity of goat production, but this seems worse than sheep. According tothis,"While sheep and goats have many similarities, their taxonomy (scientific clasification) eventually diverges. Each is a distinct species and genus. Sheep (Ovis aries) have 54 chromosomes, while goats (Capra aegagrus hircus) have 60. While sheep and goats will occasionally mate, fertile sheep-goat hybrids are rare. Hybrids made in the laboratory are called chimeras."

5.0.19. Sunflower Oil

Monthly price data were taken from the IMF data base and are described as"Sunflower oil, Sunflower Oil, US export price from Gulf of Mexico, US$ per metric ton". They appear not to be traded in CME Globex platform. The annual quantity of production, Sunflower Oil, is provided by FAOSTAT. FAOSTAT does not have value data for sunflower oil. They have it for sunflower seed.

5.0.20. Soybean Meal

This commodity is included among the traded commodities, and is not in the non-traded list. Monthly price data on soybean meal were taken from the IMF database, and it is called "Soybean Meal, Chicago Soybean Meal Futures (first contract forward) Minimum 48 percent protein, US$ per metric ton". Annual quantity produced, in 1000's of tonnes were obtained from the USDA. These data are in 1000's of metric tons of production for a large number of countries. We summed the quantity produced of all available countries to obtain a measure of world production (presumably, taking a simple sum across all these products mixes up different kinds of soybean meal). We almost decided to drop soybean meal from the non-traded list. According tothis, soybean meal is traded on the CME. But, its absence from the CFTC data set suggests that if it is traded, it is traded less than the amount on their cutoff. We did not see any volume measure on the CME website.

5.0.21. Hard Sawnwood

We obtained monthly data from the IMF, where it is called:"Hard Sawnwood, Dark Red Meranti, select and better quality, C&F U.K port, US$ per cubic meter". Section 20101 of the document,Chapter 201 Random Length Lumber Futures, provides the CME's statement of what sort of wood is traded in futures markets. They indicate that the wood must be 'grade stamped' Hem Fir (though they don't allow Hem Fir from any location, and the specify the locations carefully). A document from the Western Wood Products Association defines what Hem Fir is (see thisaddress, and select Hem Fir under the 'Species' option). It defines Hem Fir as a 'softwood species'. For this reason, we suppose that the Hard Sawnwood in the IMF data set is not included among the traded wood in the CFTC data set.

We obtained Sawnwood (NC), from the FAOSTAT website. The letters, NC, indicate"Non-Coniferous". We found information on the web that Non-Coniferous means"All woods derived from trees classified botanically as Angiospermae - e.g., maple (Acer), alder (Alnus), ebony (Diospyros), beech (Fagus), lignum vitae (Guiaicum), poplar (Populus), oak (Quercus), sal (Shorea), teak (Tectona), casuarina (Casuarina), etc. These are generally referred to as broadleaved or hardwoods.". We used the quantity of world production from FAOSTAT on Sawnwood (NC) as our measure of quantity of production. These data are from their 'domain', Forestry Production and Trade, and are measured in cubic meters.

The CME's Globex platform seems to trade insoftwood, not the sawnwood (NC).

5.0.22. Shrimp and Salmon

The monthly price of shrimp was obtained from the IMF, which calls it"Shrimp, No.1 shell-on headless, 26-30 count per pound, Mexican origin, New York port, US$ per kilogram." The monthly price of Salmon was also obtained from IMF, and has description,"Fish (salmon), Farm Bred Norwegian Salmon, export price, US$ per kilogram".

Shrimp and Salmon were each obtained from the Food and Agriculture Organization of the United Nations, Fisheries and Aquaculture Department. The data were accessed using theFishStatJ software. This allowed us to download the data on the value of production of these goods. The data are in 1000's of dollars. Neither Shrimp nor Salmon are available for futures market trading in CME's Globex.

In the case of shrimp, we simply added prawn and shrimp, which are apparently thesame thing.

Table 11: Monthly Non-Traded Spot Prices

	source of price data	Definition	source of production data
Fishmeal	IMF	Peru Fish meal/pellets 65% protein	IndexMundi, 16 biggest producer countries
Palm Oil	IMF	Malaysia Palm Oil	IndexMundi, 16 biggest producer countries
Rubber	IMF	Rubber, Singapore Commodity Exchange, No. 3 Rubber Smoked Sheets	FAOSTAT
Tea	IMF	Tea, Mombasa, Kenya, Auction Price, From July 1998, Kenya auctions, Best Pekoe Fannings.Prior, London auctions, c.i.f. U.K. warehouses	FAOSTAT
Uranium	IMF	Uranium, NUEXCO, Restricted Price	United Nations Energy Statistics Database
Wool	Trading Economics		Greasy wool, FAOSTAT
Bananas	IMF	Bananas, Central American and Ecuador, FOB U.S. Ports	FAOSTAT
Barley	IMF	Canadian no.1 Western Barley	FAOSTAT
Rapeseed	IMF	Rapeseed oil, crude, fob Rotterdam	FAOSTAT
Groundnuts	IMF	Groundnuts (peanuts), 40/50 (40 to 50 count per ounce), cif Argentina	FAOSTAT
Hides	IMF	Hides, Heavy native steers, over 53 pounds, wholesale dealer's price, Chicago	FAOSTAT
Iron Ore	IMF	Iron Ore, China import Iron Ore Fines 62% FE spot (CFR Tianjin port)	USGS
Tin	Trading Economics		USGS
Zinc	Trading Economics		USGS
Nickel	Trading Economics		USGS
Lead	Trading Economics		USGS
Olive Oil	IMF	extra virgin less than 1% free fatty acid, ex-tanker price U.K.	International Olive Council
Chicken	IMF	Poultry (chicken), Whole bird spot price, Ready-to-cook, whole, iced, Georgia docks	FAOSTAT
Lamb	IMF	Lamb, frozen carcass Smithfield London	FAOSTAT
Sunflower Oil	IMF	Sunflower Oil, US export price from Gulf of Mexico	FAOSTAT
Hard Sawnwood	IMF	Hard Sawnwood, Dark Red Meranti, select and better quality	Sawnwood (NC), FAOSTAT
Shrimp	IMF	Shrimp, No.1 shell-on headless, 26-30 count per pound, Mexican origin, New York port	Fisheries and Aquaculture Department, UN
Salmon	IMF	Farm Bred Norwegian Salmon, export price	Fisheries and Aquaculture Department, UN

6. Oil, Coal and Natural Gas

Data on oil, coal, and natural gas are obtained from British Petroleum's Statistical Review of World Energy 2015 workbook. In terms of oil prices, over the period 1861-1944, they represent the US average. Over the period, 1945-1983, they correspond to Arabian Light posted at Ras Tanura. Over the period, 1984-2014, they are Brent dated.

Figure 15

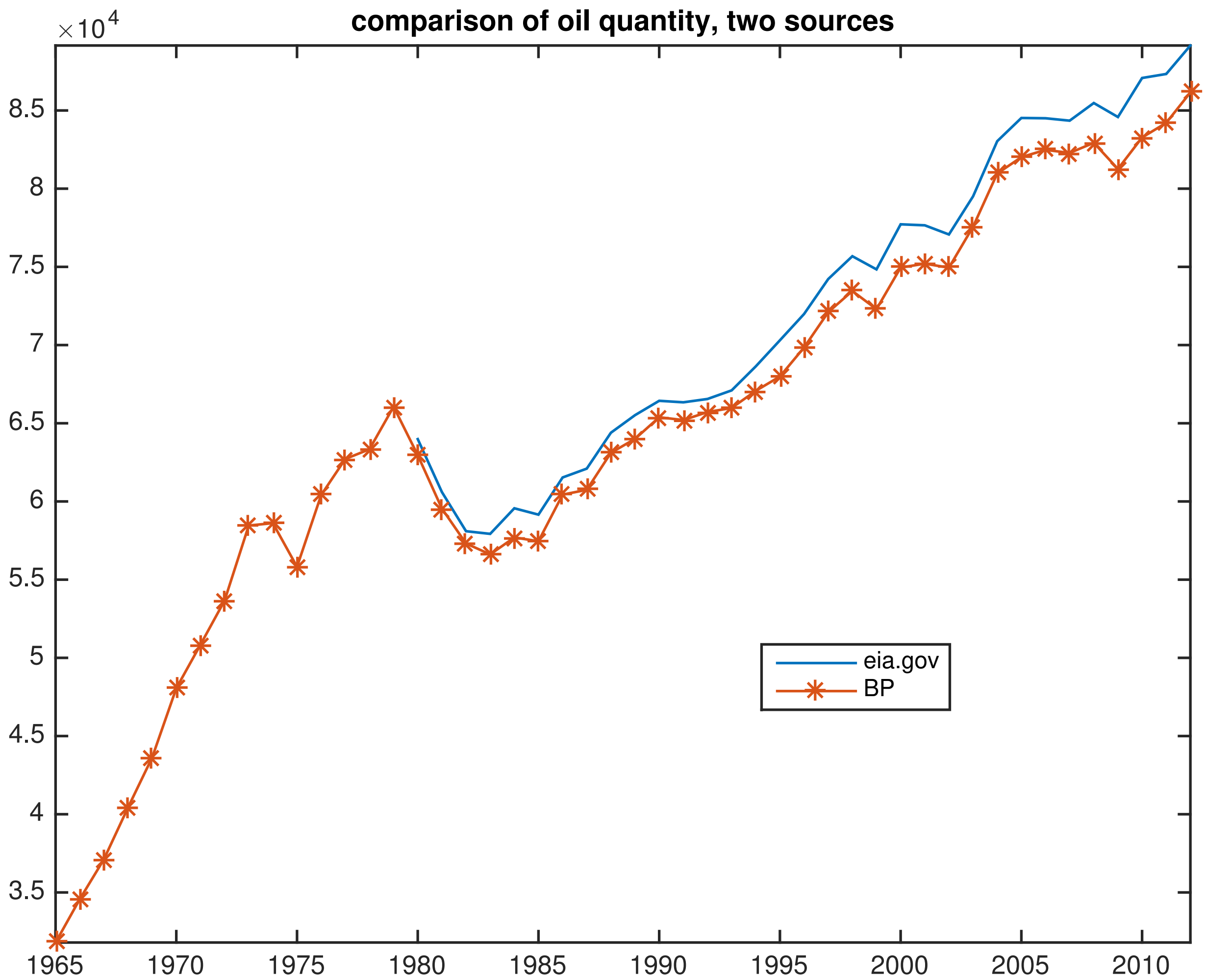
Figure xx graphs the real oil price data over the entire sample. The table indicates that the quantity of oil is measured in thousands of barrels per day and the price in dollars per barrel. That the 'source' for the total value of oil production is 'derived' indicates that it is obtained simply by multiplying the price and quantity columns.

Figure 16: US Share of World Oil Production

We also obtained data on the quantity of world oil production from the US Energy Information Administration (USEIA[5]. The data are displayed in

[5]See http://www.eia.gov/cfapps/ipdbproject/iedindex3.cfm?tid=5&pid=53&aid=1&cid=regions&syid=1980&eyid=2012&unit=TBPD

Figure 17: Two Measures of World Oil Production



Note that the two series are quite similar.

CFTC Data on Volume

Part IV

Figure Results: Annual Data

6.1. Definition of Index

We now describe several indices associated with our data. We construct an aggregate price index in the following way:

$$P_t = \sum_{i=1}^{N} w_i \frac{P_{it}/P_t^{PCE}}{P_{i,0}},$$

$$P_{it} \sim i^{th}\text{commodity price},$$

where P_t^{PCE} denotes the personal consumption expenditure deflator and N denotes the number of commodities in the dataset. In the case of annual data, $N = 136$ and in the case of monthly data, $N = 52$. Here, w_i denotes the share of world production of the i^{th} commodity:

$$w_i = \frac{1}{T} \sum_{t=1}^{T} \frac{P_{it} Y_{i,t}}{\sum_j P_{jt} Y_{j,t}}.$$

Also,

$$oi_t = \sum_{i=1}^{N^{CFTC}} \bar{w}_i \frac{S_{i,t}^L + H_{i,t}^L}{Y_{i,t}}, \quad nff_t = \sum_{i=1}^{N^{CFTC}} \bar{w}_i \frac{S_{i,t}^L - S_{i,t}^s}{Y_{i,t}}, \quad \bar{w}_i = \frac{w_i}{\sum_{j=1}^{N^{CFTC}} w_j}$$

where N^{CFTC} is 29, the number of commodities in our CFTC dataset.

6.2. Properties of Indices

Figure 18

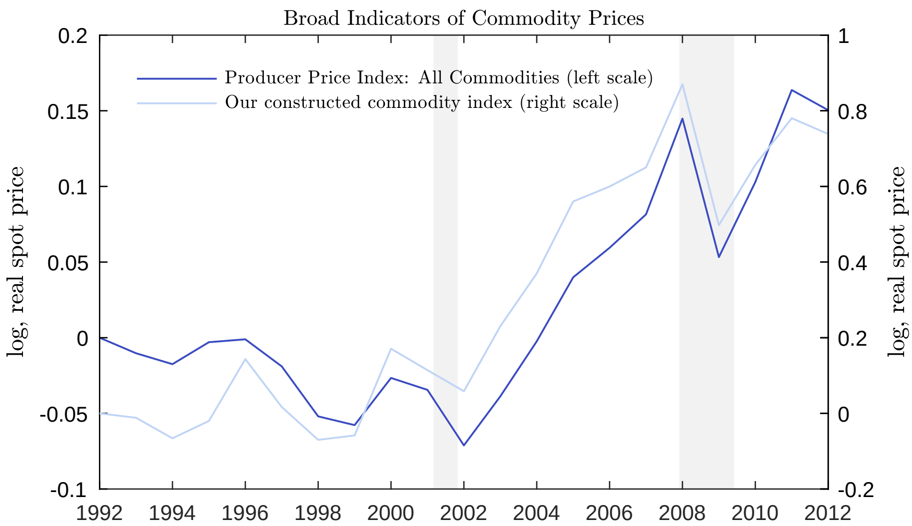


Figure 18 displays $\log\left(P_t/P_0\right)$ over our sample. In addition, the figure displays the producer price index (logged and scaled by the initial observation). The left scale corresponds to our price index and the right corresponds to the PPI. There are two things worth emphasizing about these prices. First, they comove over time. Second, comparing the left and right scales their values are very different. In particular, our price index displays substantially more growth over the sample. Much of the growth of our price index reflects the importance of oil prices. In particular, crude oil receives a weight of a little less than 30% in the index.

Figure 19: open interest

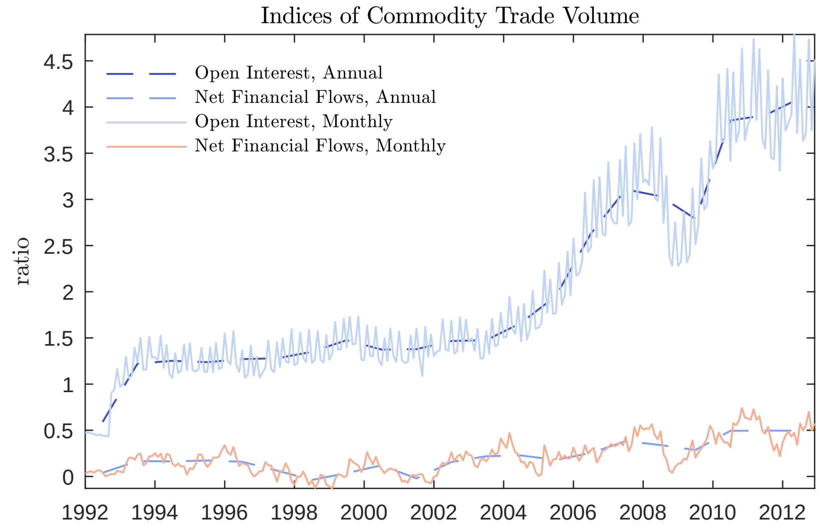


Figure19 displays the time series on$oi_t$and$n_{ff,t}$. Note that net financial flows are generally positive and grow very little. At the same time, open interest grows sharply, beginning around 2004, going from around 1.5 times world production to around 4 times world production. Evidently, while volume has increased sharply in commodity futures markets, the net flows between outsiders and insiders have changed very little.

Figure 20: price breakdown

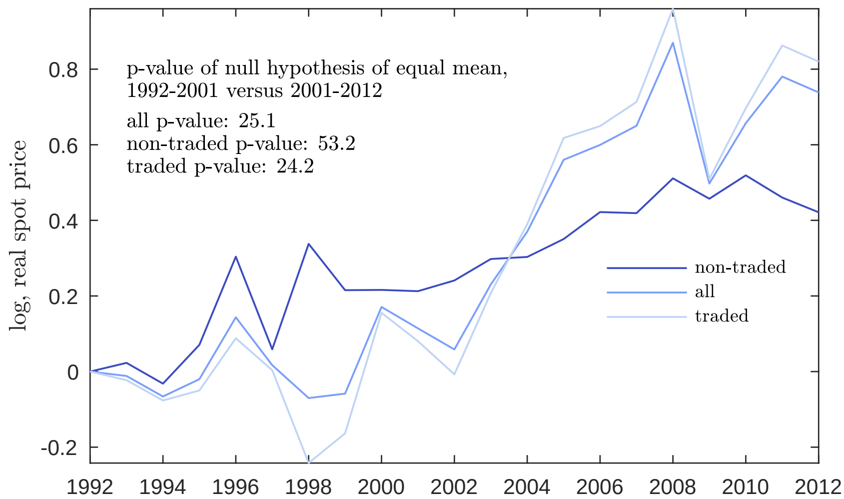


Figure20 shows the price index for different sub-categories of commodity. Note that whether we include all commodities or just the CFTC-traded ones, makes little difference. This is because the weights on the CFTC-traded commodities are the largest. Also, note that when the price index includes only the non-traded commodities, then its growth over the sample falls roughly in half. This reflects the importance of oil and the growth of its price over the sample, from around \$25 per barrel to a little over \$100 per barrel.

Figure 21: price breakdown

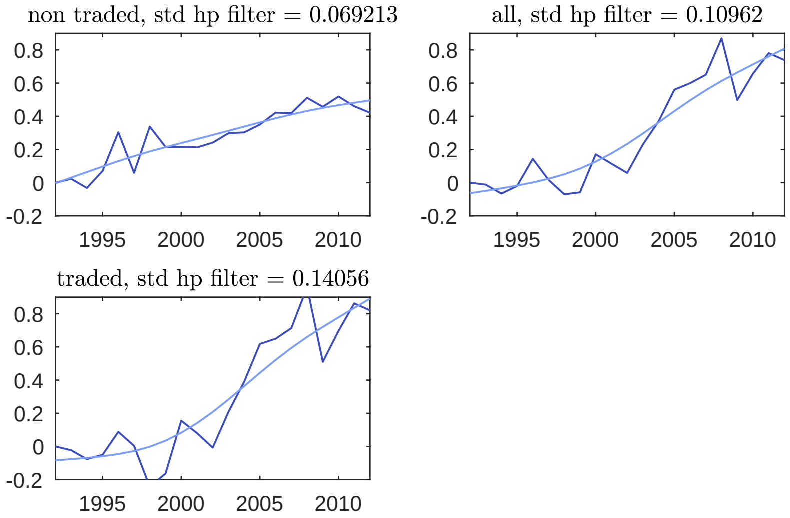


Figure21 exhibits the traded, non-traded and overall price indices, and shows that the traded are more volatile relative to their HP filter trend ($\lambda = 100$).

Figure 22: price breakdown

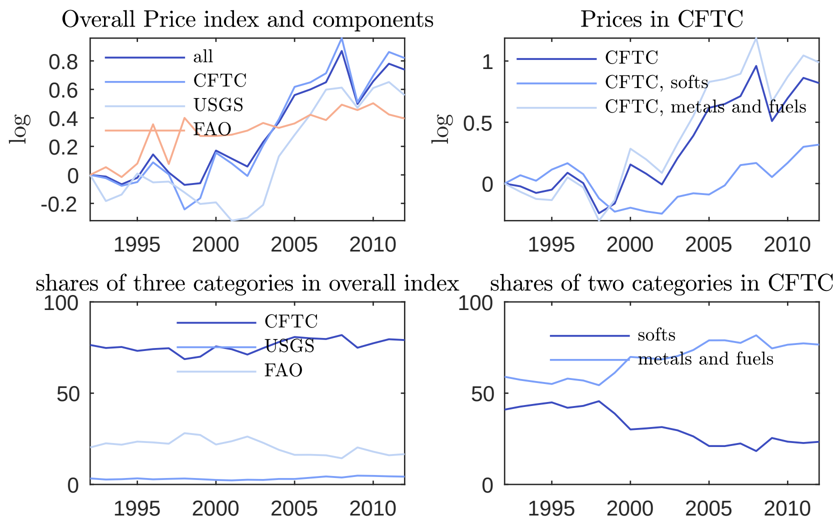


Figure20 has four panels. The 3,2 panel panel shows that, in terms of the overall price index, the CFT-traded commodities receive the greatest weight. 80% of the value of world production of all the 136 commodities in the annual data set is accounted for by CFTC traded commodities. Minerals come in second, a little less than 20% and softs are the smallest category. The 2,2 panel shows that minerals and fuels increased in importance, relative to softs, starting in the late 1990s. The 1,2 panel shows that the CFTC price index is primarily driven by the minerals and fuels component. The 1,1 panel shows that The CFTC and all, are basically the same. The CFTC-traded softs and FAO data (non-traded softs) display a similar pattern over time. They both display relative little growth of approximately 0.4 log points over the sample. CFTC-traded and nontraded minerals and fuels both fall before eventually rising, though the former trough CFTC hits a minimum in 1997, while USGS hits a minimum sooner.

6.3. Scatter Plots of Changes in Volatility Against Changes in Volume of Trade

Figure 23: open_interest

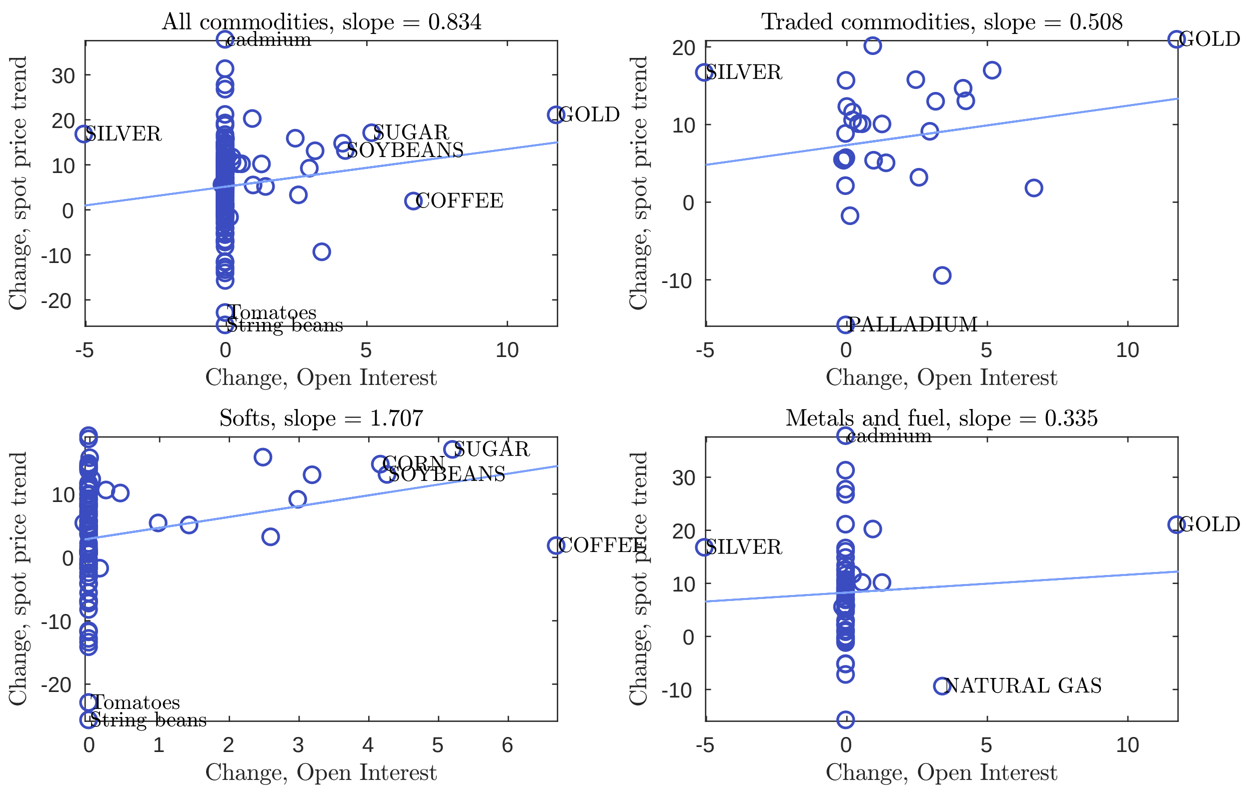


Figure 19 shows the scatter of the change in spot price trend against the change in open interest, when the change is between the value in 1992-2000 and 2001-2012. Outlier observations are labelled. Outliers were identified as follows. Let the i^{th} observation reported in a figure be denoted by (x_i, y_i) , where x_i denotes the value on the horizontal axis and y_i denotes the value on the vertical axis. Let σ_x denote the standard deviation of the x_i's and define σ_y similarly. Also, let $\overline{x}$ and $\overline{y}$ denote the sample mean of the two variables. Let q_i denote the 'distance' of the observation, (x_i, y_i), from the mean of the variables:

$$q_i = \left(\frac{x_i - \overline{x}}{\sigma_x}\right)^2 + \left(\frac{y_i - \overline{y}}{\sigma_y}\right)^2.$$

We order the q_i's from smallest to largest. The largest 5% are defined as the outlier observations and the name of the associated commodities are displayed. This method was used throughout this appendix.

In Figure 19, the change in the spot price trend is measured as the change in the slope of a regression of the spot price on a constant and a trend. The price is the log of the commodity price, after scaling by its initial price and the PCE deflator. Open interest is the average open interest (scaled by world production), over the early or late sub period.

Figure 24: net financial flows

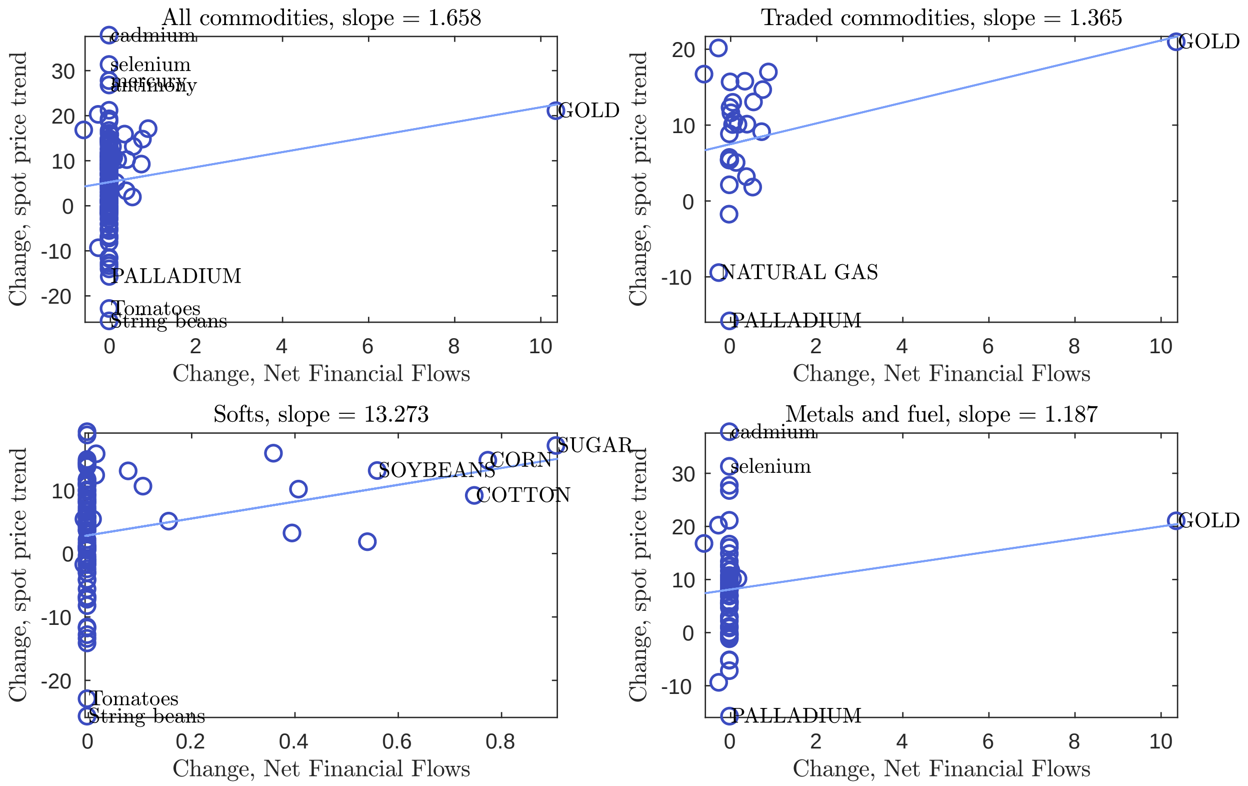


Figure 24 is the same as the previous figure, except that financialization is measured in the form of net financial flows, relative to output.

Figure 25: change in std-oi

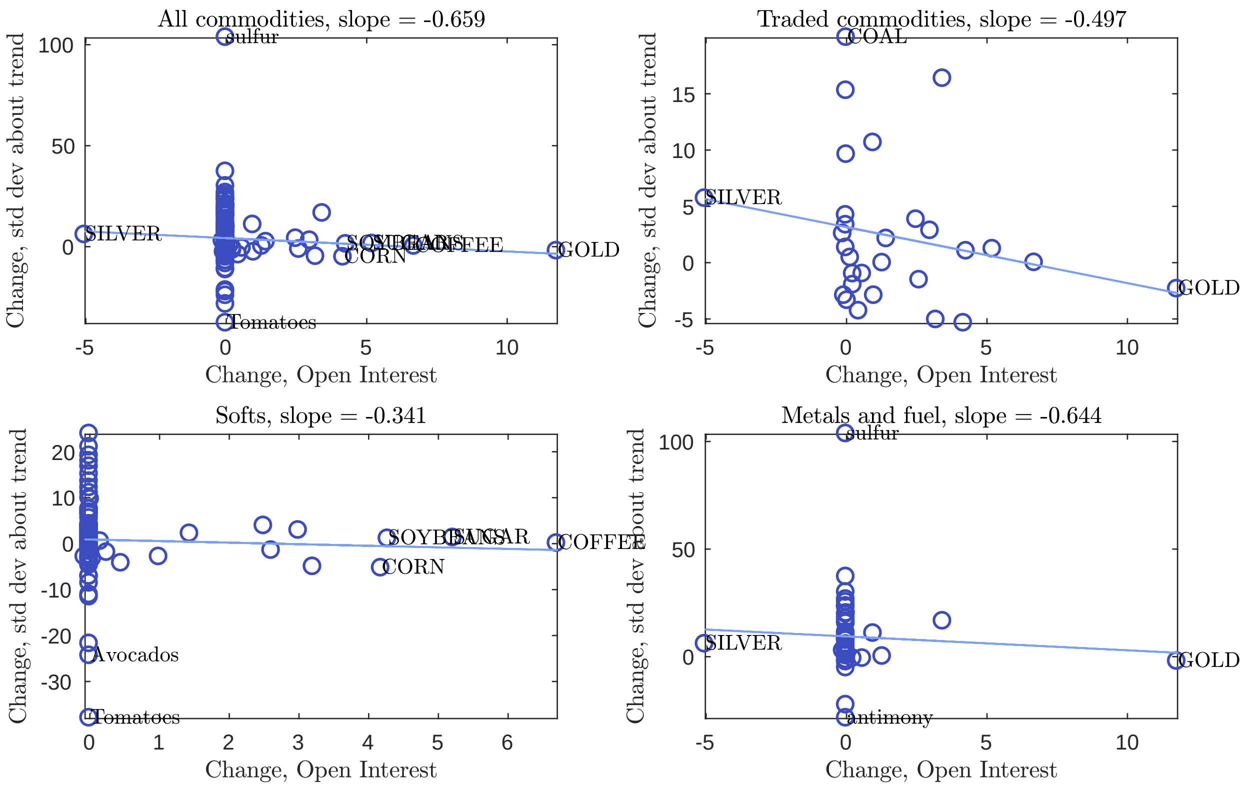


Figure25 displays the relationship between the change in the standard deviation of the residual in the regression of log price on a constant and time trend between early and late periods and financialization. Here, financialization is measured by open interest. In the next figure, it is measured by net financial flows.

Figure 26

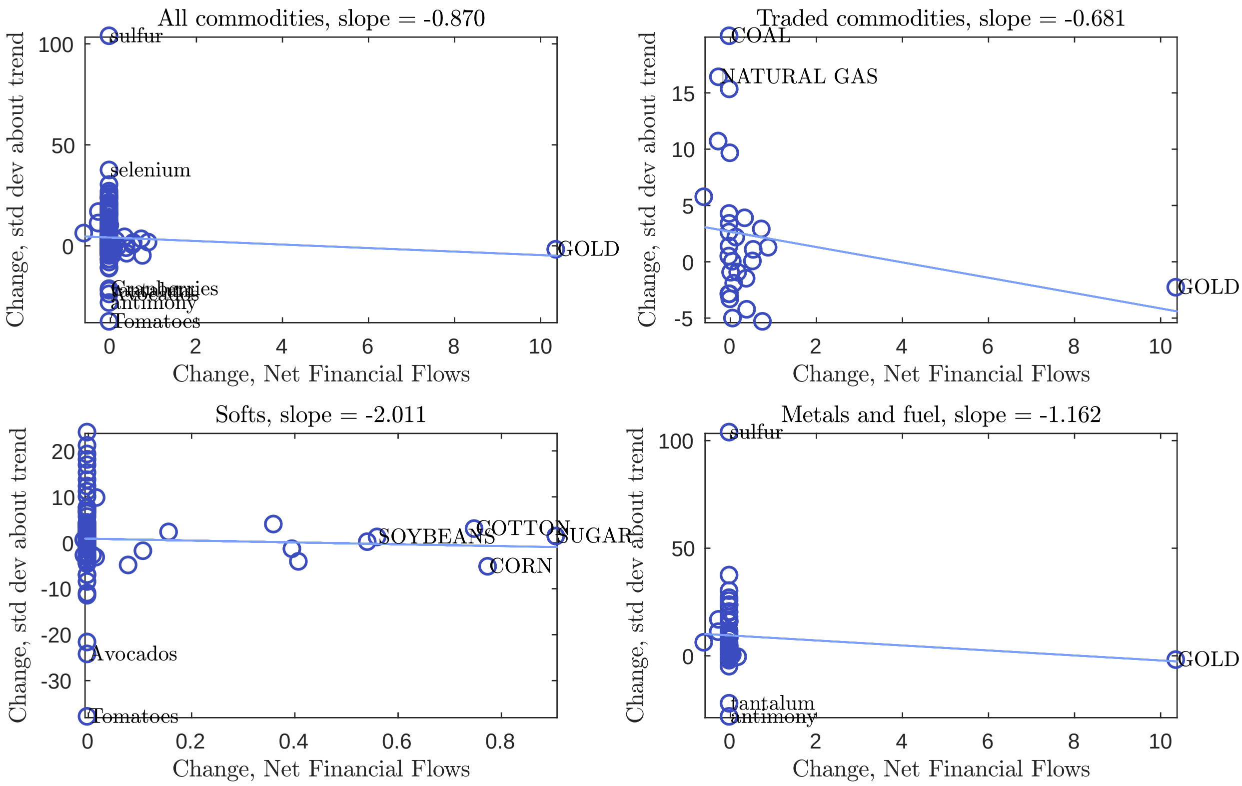


Figure 27: indexed-non-indexed

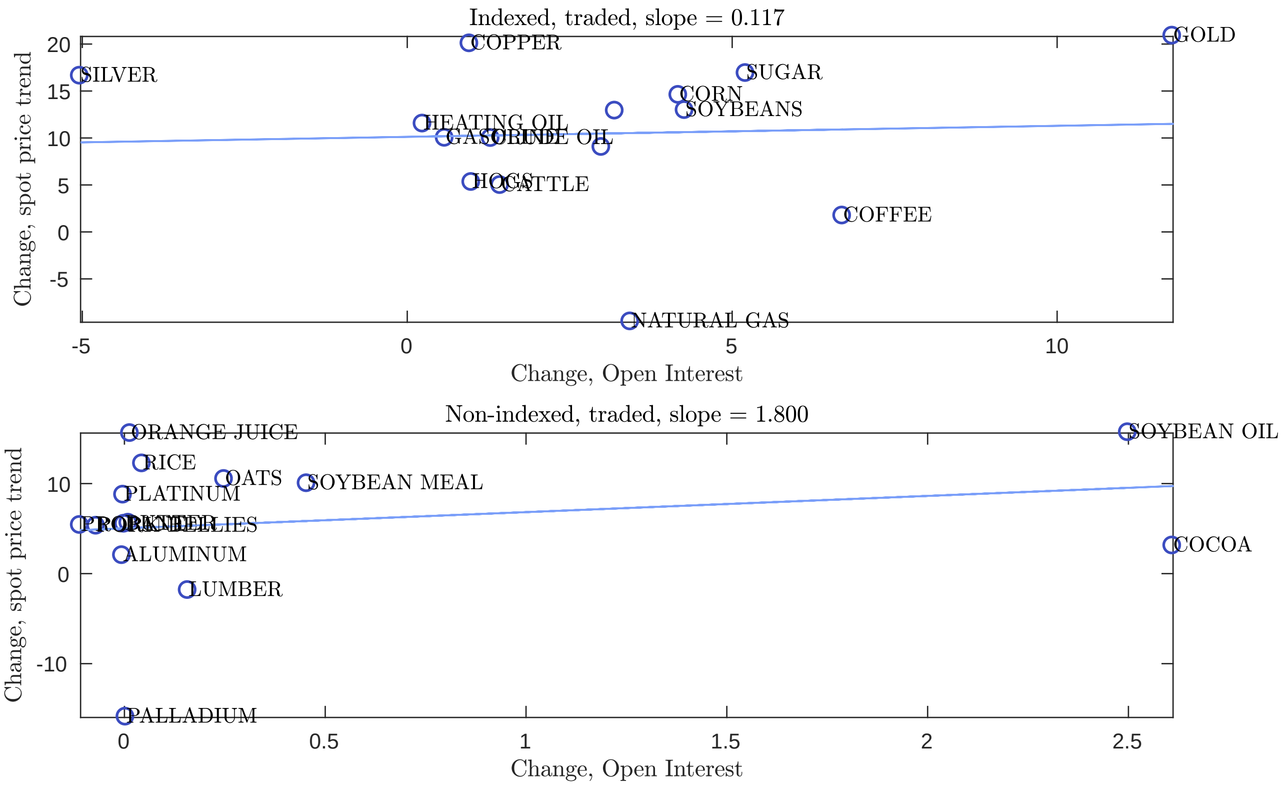


Figure27 compares the association between financialization and the change in the spot price trend, between traded commodities that appear in both commodity index funds and commodities that appear in neither fund. This is a small number of observations, since the total number of commodities involved is 29. In this figure, financialization is measured by open interest. The next figure shows the same calculations, measured by net financial flows.

Figure 28

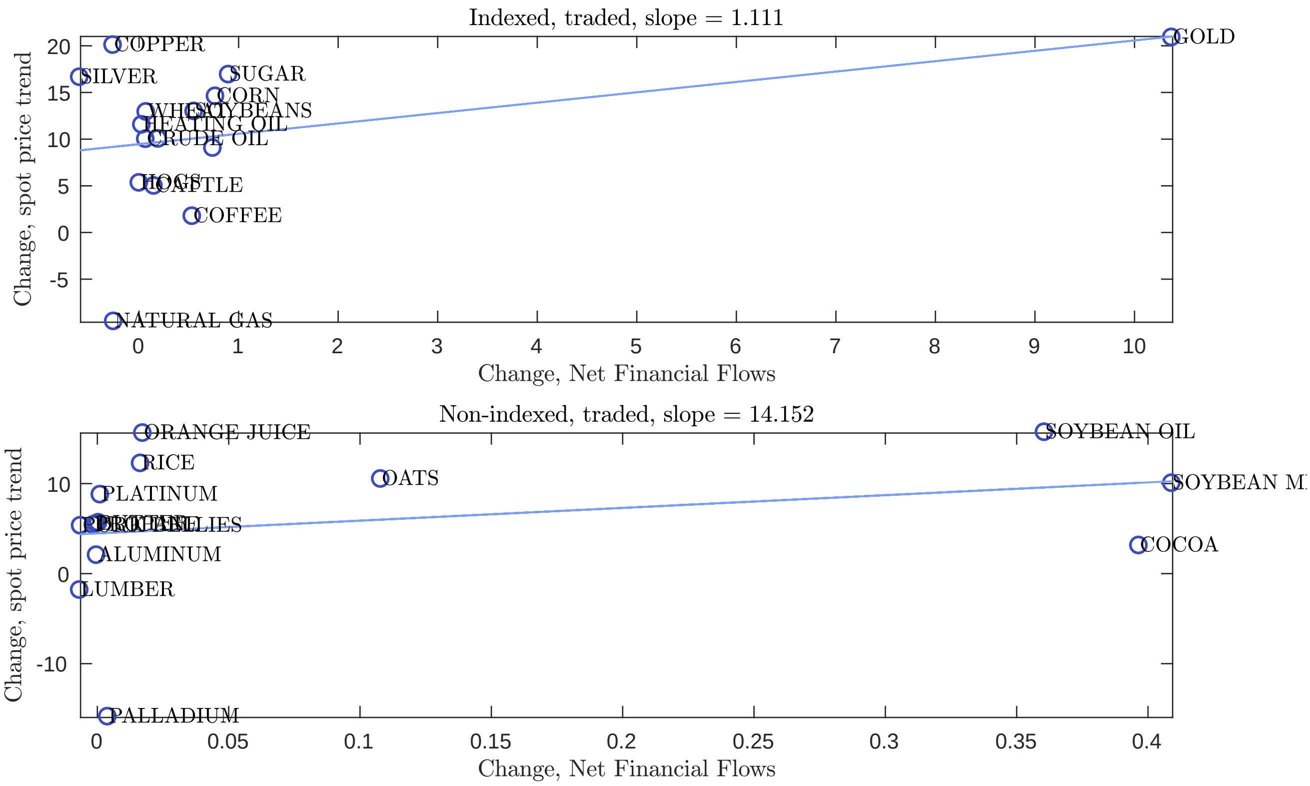


Figure 29: bar_chart_comparison

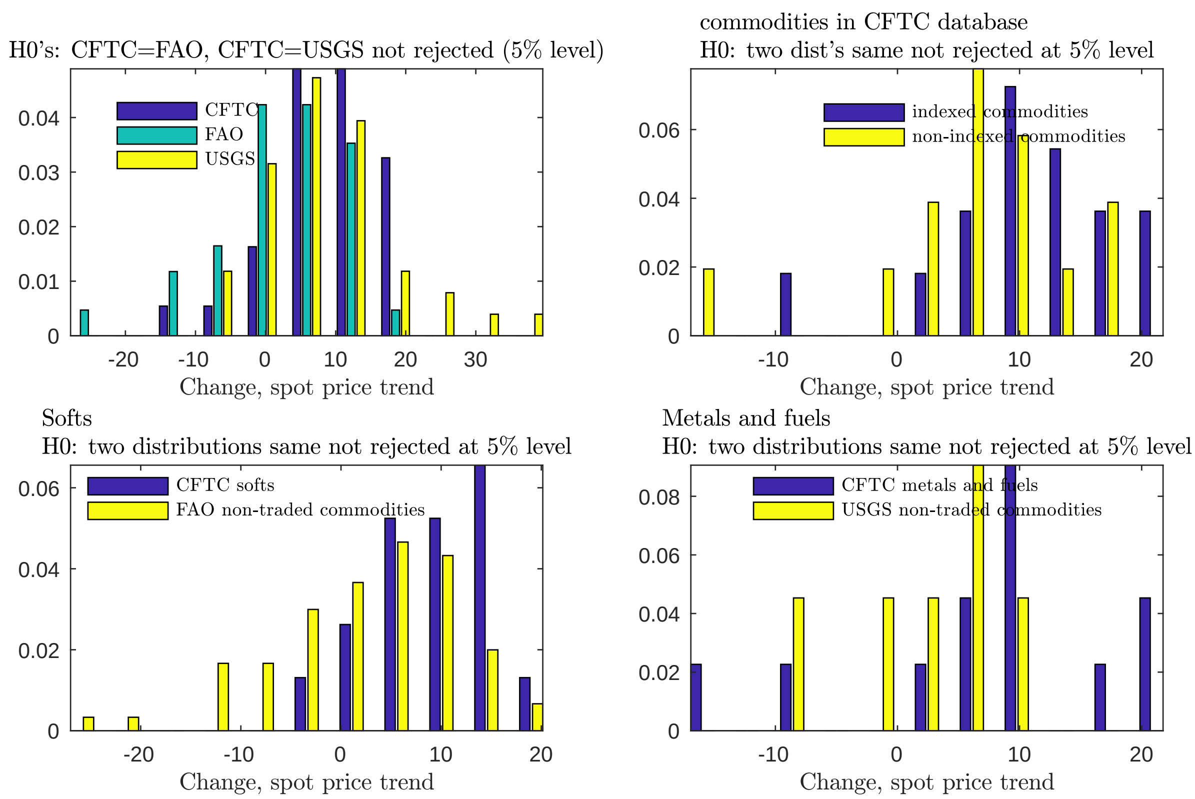


Figure29 displays the empirical density (the sum of the areas for each color of bars equals unity) of the regression - commodity by commodity - of the change in the spot price trend. The 1,1 panel does so for CFTC traded commodities, non-traded commodities from the FAO and non-traded commodities from USGS. The panel also reports the results for a test of the null hypothesis that the two densities are the same. The test is the Kolmogorov-Smirnov goodness-of-fit hypothesis test, performed by Matlab code kstest2.m. The test takes two empirical density functions and tests the null hypothesis that the two distributions are the same. In panel 1,1, the test fails to be rejected at the 5% significance level. In the 1,2 panel results are reported across indexed and non-indexed commodities among the CFTC-traded commodities, and the two densities are not significantly different at the 5% level. The 2,1 panel compares the results for CFTC softs and FAO softs and the 2,2 panel compares results for CFTC minerals and fuels and USGS minerals and fuels. In all cases, the null hypothesis that the underlying densities are the same is not rejected.

Figure 30: Change in volatility and change in financialization

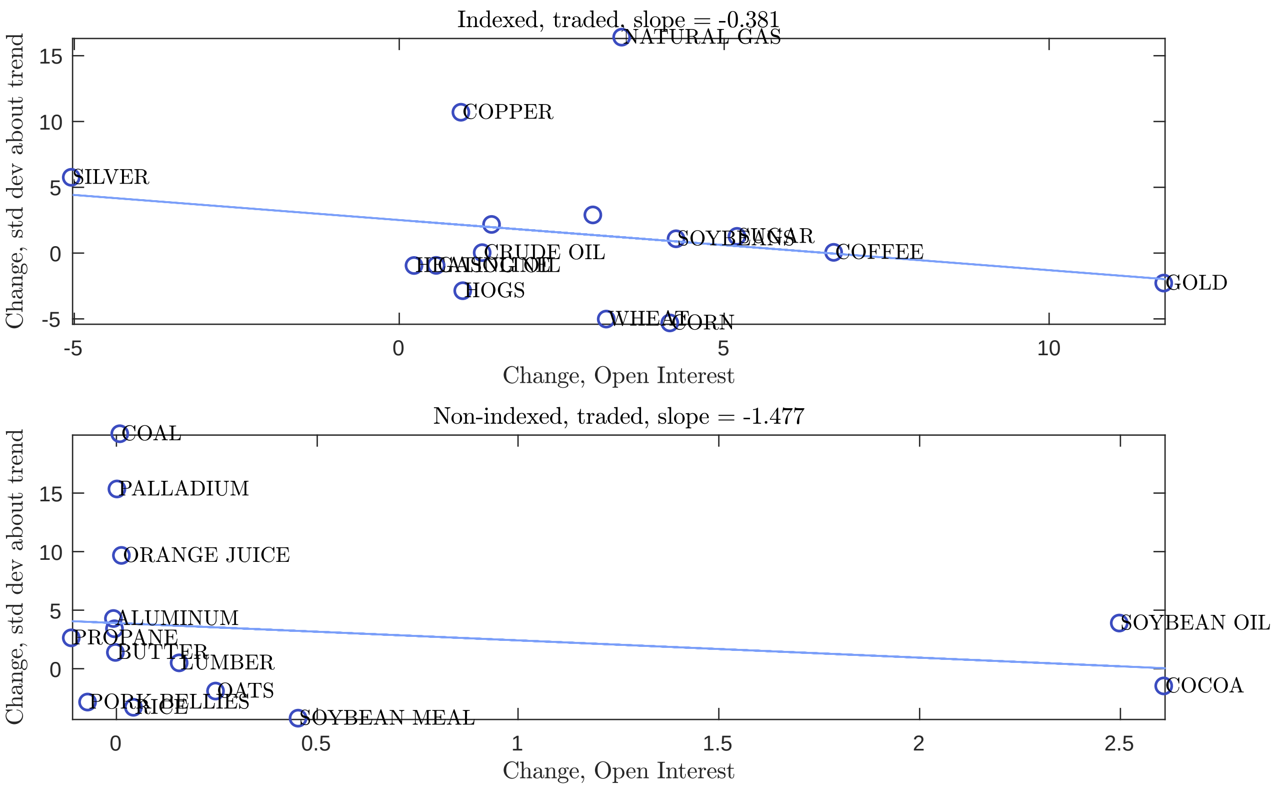


Figure30 displays the scatter of the change in standard deviation of the regression residual against the change in financialization. It does so for indexed versus non-indexed commodities and for the case when financialization is measured by open interest. The regression is the one where the commodity price (logged, and scaled by its initial price and the PCE deflator) is regressed on a constant and time trend in the early and late period. The following figure reports the same results for the case where open interest is measured by net financial flows.

Figure 31

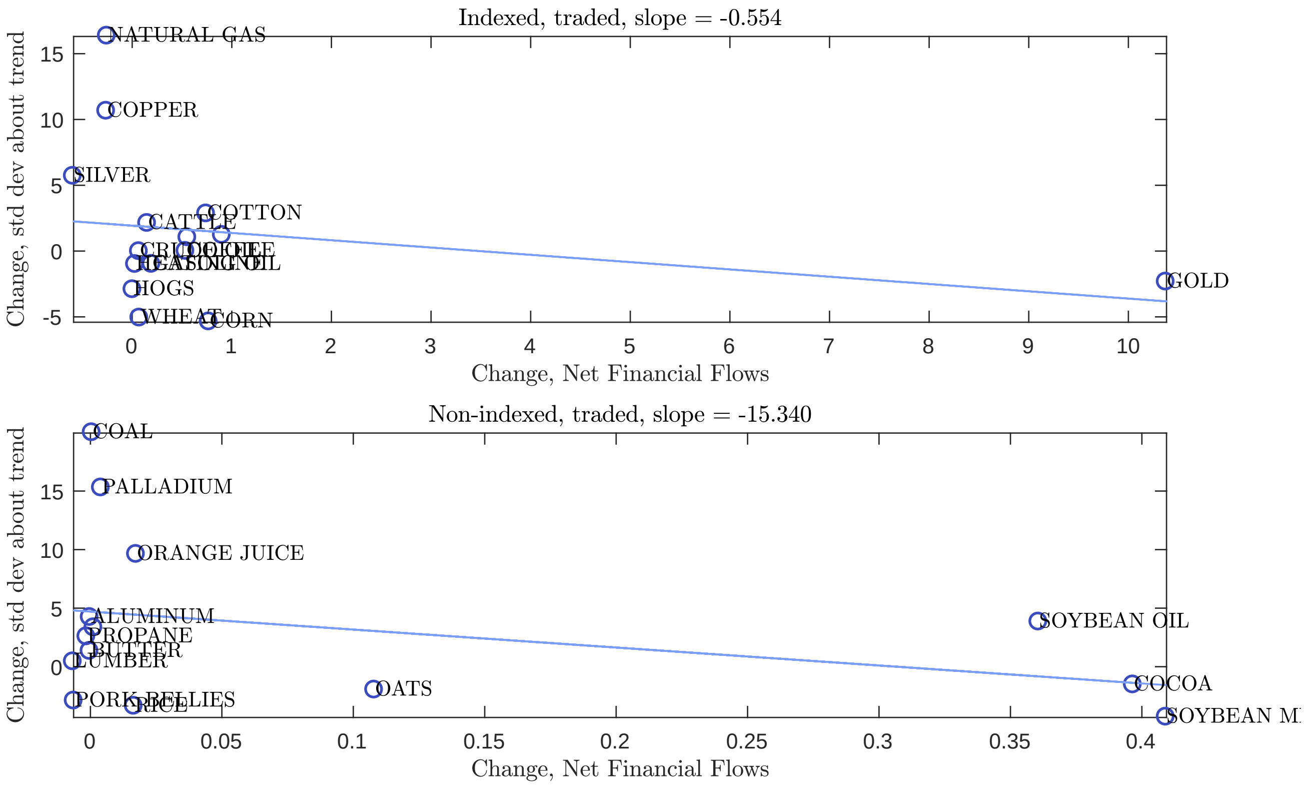
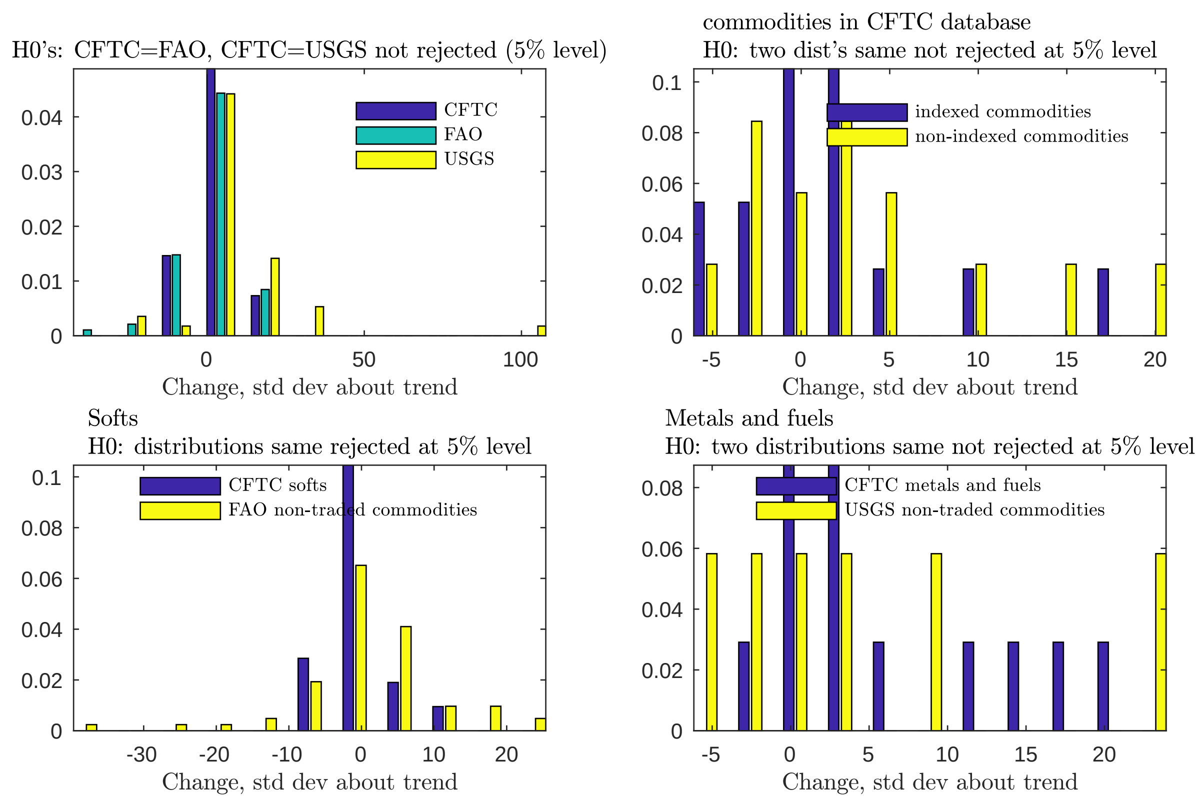


Figure 32: Change in volatility



This figure displays the empirical density of the change in the standard deviation of the regression residual analyzed in Figure30. It compares the densities for different categories of commodities. With one exception, the null hypothesis of no difference is not rejected. The exception is the 2,1 panel, where the density of FAO softs appears to be significantly shifted to the right, relative to the density of the CFTC softs.

Figure 33: Time Series of Volatility on Volume

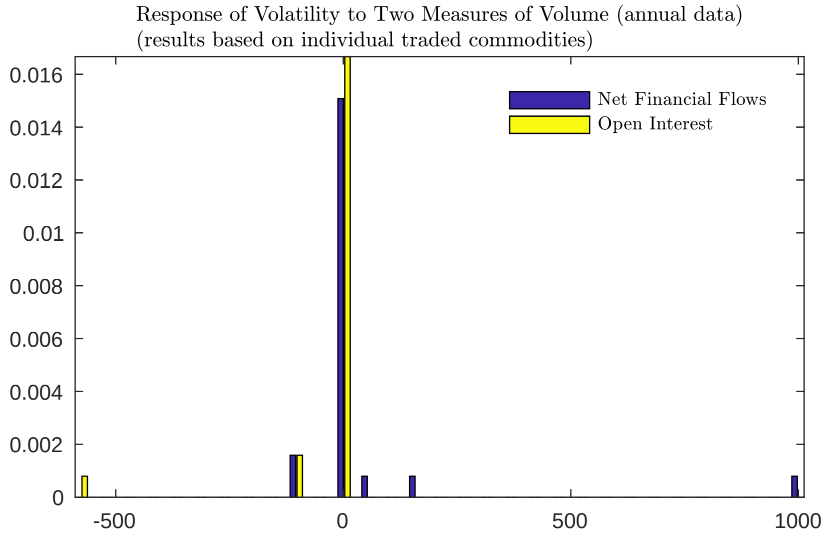


Figure33displays the distribution of slope coefficients, commodity by commodity, in regressions of commodity price (logged and scaled by initial price and PCE deflator) on the indicated measure of volatility. The measure of volatility is the standard deviation of the first difference of the commodity price, computed using a centered moving average of data. That is, the observation$t$ measure of volatility is the standard deviation, based on data two years before and after$t$ of the observations. We have 21 observations from 1992 to 2012 on commodity prices for each commodity. Thus, we lose two initial observations, one to first differencing and the other two to the two-year window and we lose two last observations. So, each regression coefficient is based on 21-5=16 time series observations. In doing these regressions, we dropped four commodities, Butter, Propane, Aluminum and Coal, because there was at least one date when they had zero open interest. We did this to avoid variables that produced extreme observations on the regression coefficient.

6.4. Scatter Plots of Volatility Time Series Against Volume of Trade Time Series

Figure 34: Volatility and Financialization

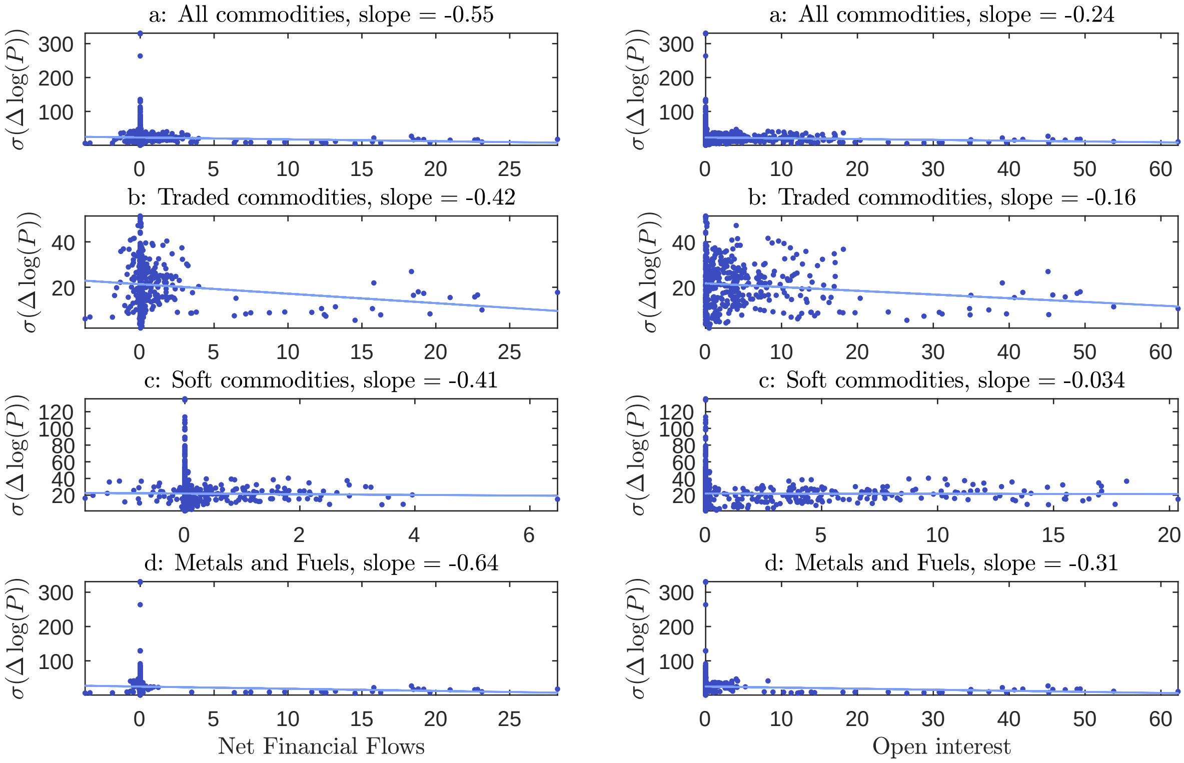


Figure34displays scatter plots of the standard deviation of volatility (measured by the two-year centered time series of standard deviations of commodity price growth) against volume. Each standard deviation is multiplied by 100, to convert the numbers into percent terms. The left column of panels measures volume by the level of net financial flows and the right column of results measures volume as the level of open interest. Row a, 'all commodities' reports results when all time series observations for different commodities are reported. There are 136 commodities, with 21-5=16 observations each (16 are lost to first differencing and the centered moving average). The other rows display results for subsets of commodities. The numbers reported in the headers are the slope term in the regression of volatility on volume. There is no evidence that high volume is associated with high volatility. If anything, there is some modest evidence that financialization decreases volatility. In this respect, the message of this figure is slightly different from that of Figure21.

Note that although the relationship between volatility and volume of trade appears negative, the point estimates are small. Increasing open interest by one times world production reduces volatility by 0.24 in annual percent terms. All these slope terms are very small.

We also consider the data when they are replaced by the residual after doing a commodity-specific regression of volatility and volume on a time trend. The scatter plot of these residuals are displayed in Figure 35. Note that the coefficients of the regressions across residuals are very similar in magnitude to what we found in the raw data.

Figure 35: Volatility and Financialization, Allowing for Different Constants and Trends

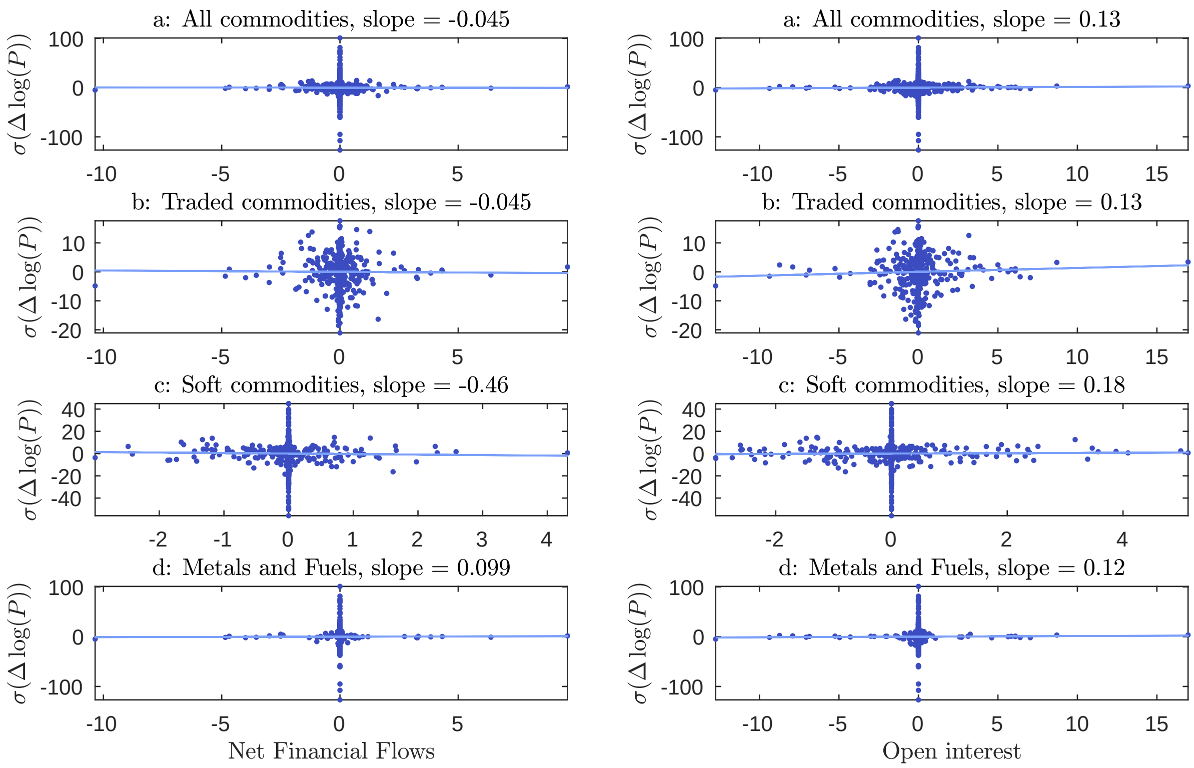


Figure35 is similar to Figure34. The difference is has to do with constants and trends. In Figure34 the constant term is the same for each commodity. Figure35 is different in two respects. First, the constant term is fixed over time for each commodity, but is allowed to be different for each commodity. Second, a trend is permitted for each commodity, and the coefficient on the trend is allowed to be different for each commodity. Of course, the slope term on financialization is the same for each commodity. That slope term is what is reported in the headers. The data displayed in the scatter plot are the errors in the regression of the raw data on a constant term and time trend. The slope coefficient that is reported is the coefficient on financialization

Figure 36: Net Financial Flows, Open Interest and Volatility

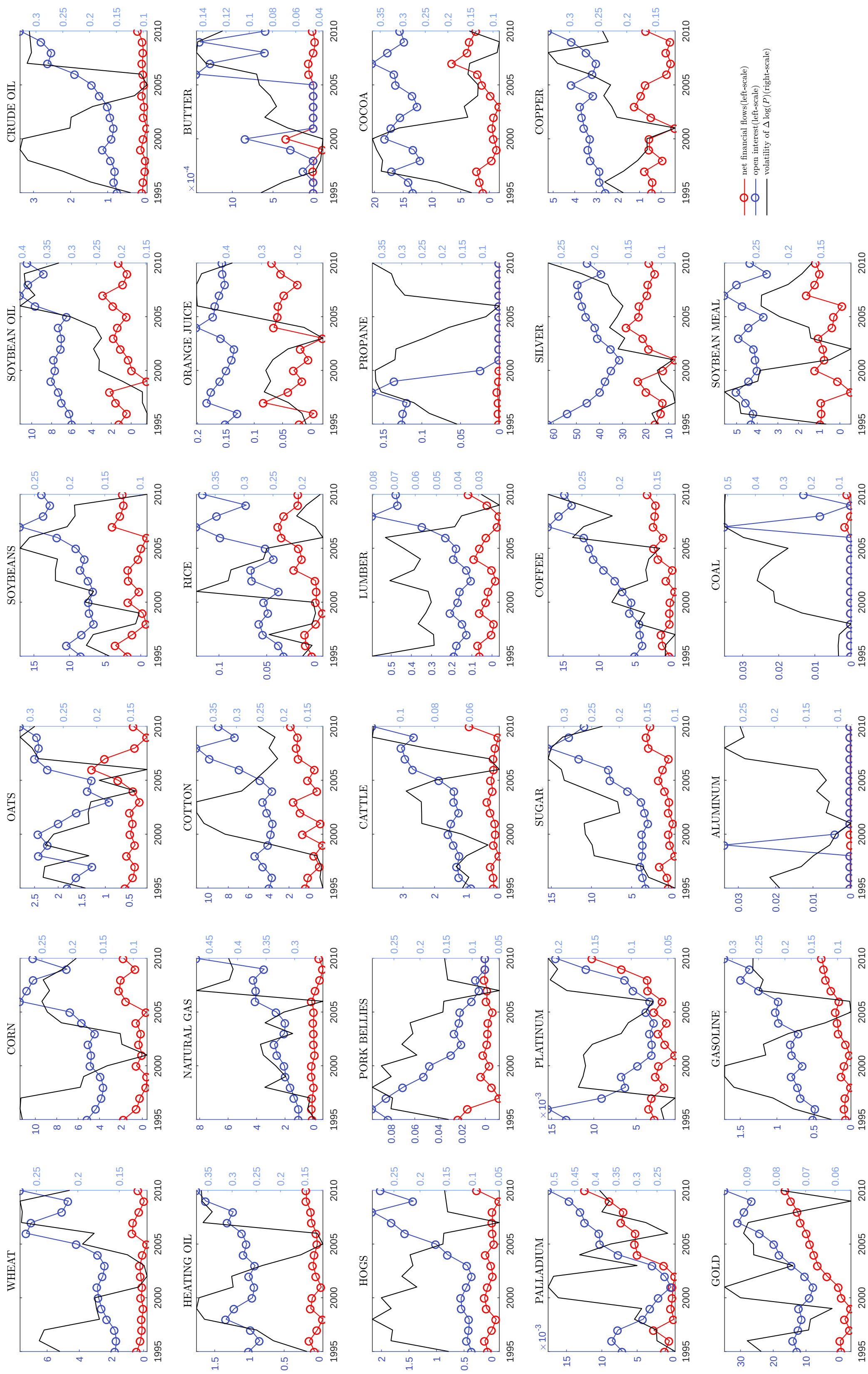


Figure36 displays annual data for the CFTC-traded commodities. The volatility data are based on a centered, five year moving window of data. Note how very volatile the data are.

Part V

Figure Results: Monthly Data

We now turn to the monthly data. We display essentially the same figures displayed in the previous section. We now have fewer commodity price series. As before, we have 29 series that are traded according to the CFTC. In addition, we have 23 more monthly series that are not CFTC-traded.

6.6. Indices

Figure 37: Aggregate Price Indices

(a) Annual (b) Monthly

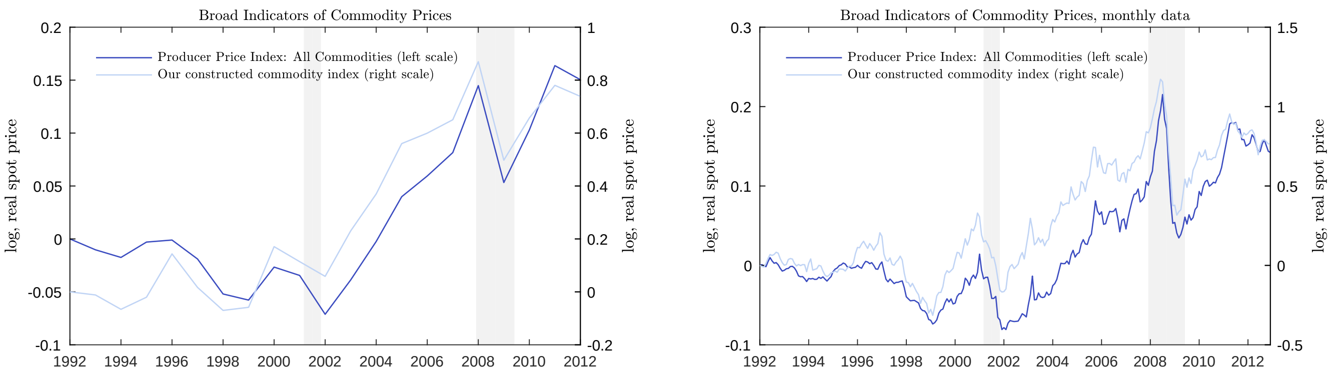


Figure37 compares our constructed$\log{(P_t/P_0)}$ with a corresponding empirical price index, for monthly and annual data. The annual data repeats figure18. The results are similar.

Figure 38

(a) Annual (b) Monthly

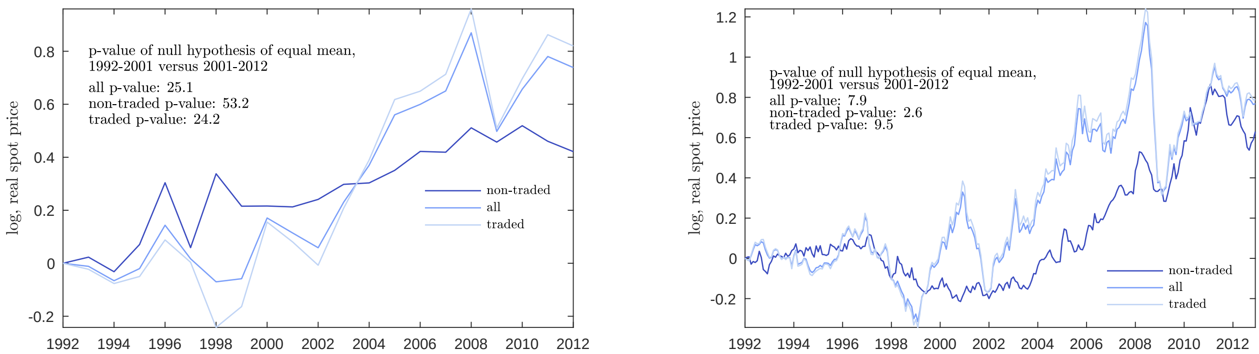
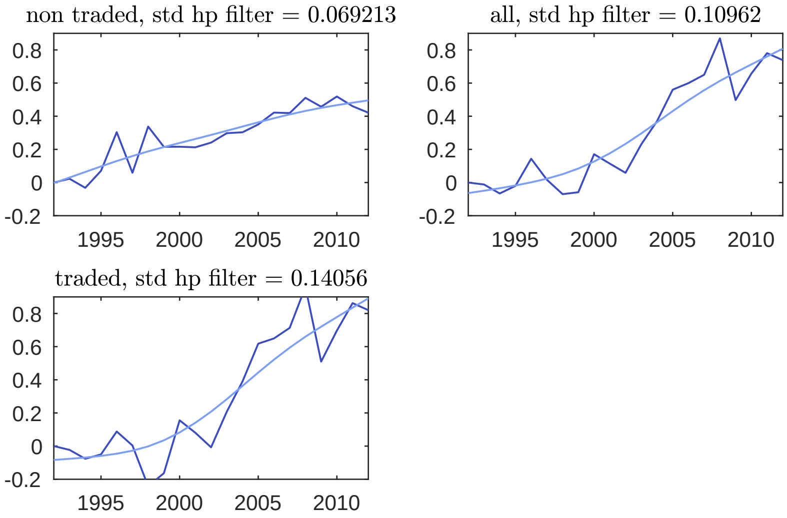
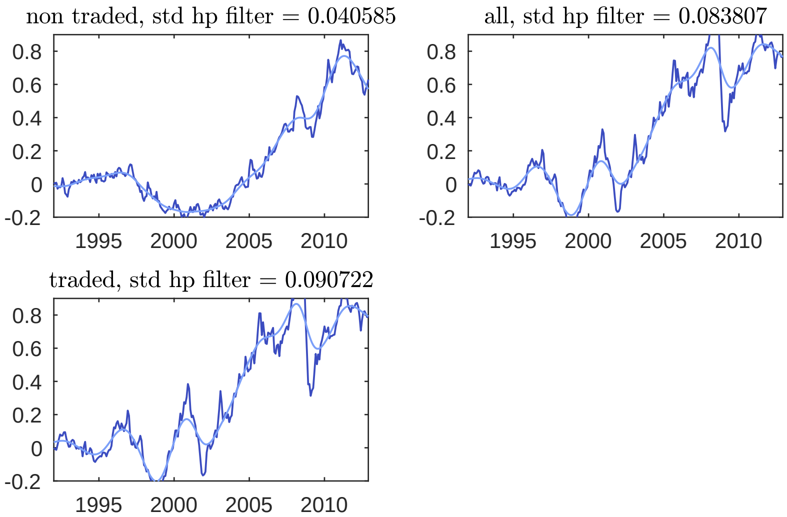


Figure 39

(a) Annual



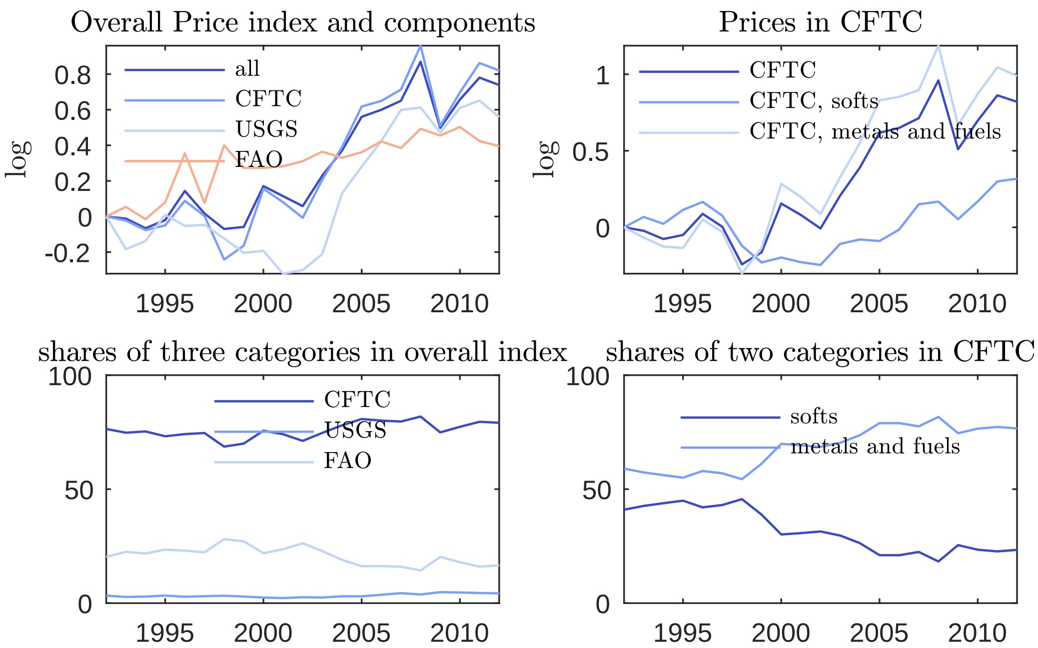


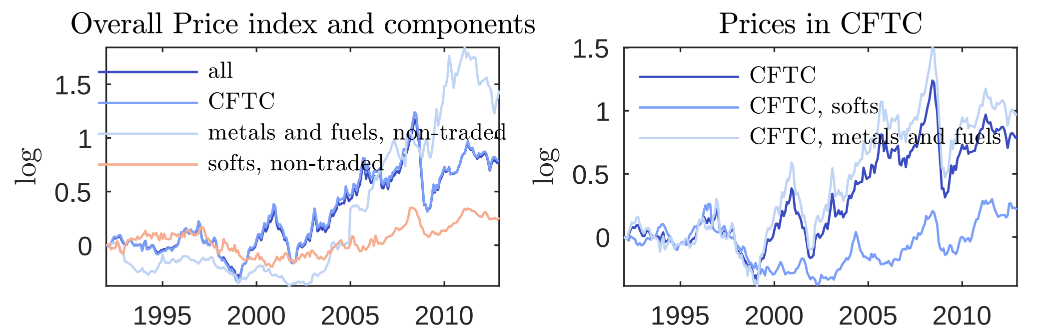
(b) Monthly

Panel a repeats Figure21 for convenience. The non-traded goods are somewhat different, but it must be remembered that these are different commodities. The choice of monthly and annual commodities was dictated by data availability. We have 107 non-traded annual commodities and only 23 monthly.

Figure 40

(a) Annual





(b) Monthly

The first panel in the above figure reproduces, for convenience, Figure20. The bottom panel does not contain the price index weights because we use the same weights for the monthly data. The interesting result in this figure is that the CFTC data look very similar in monthly annual frequencies, but they non-traded look different. This is because they are actually different commodities.

6.7.

Figure 41: Log, Scaled Commodity Prices, Monthly and Annual

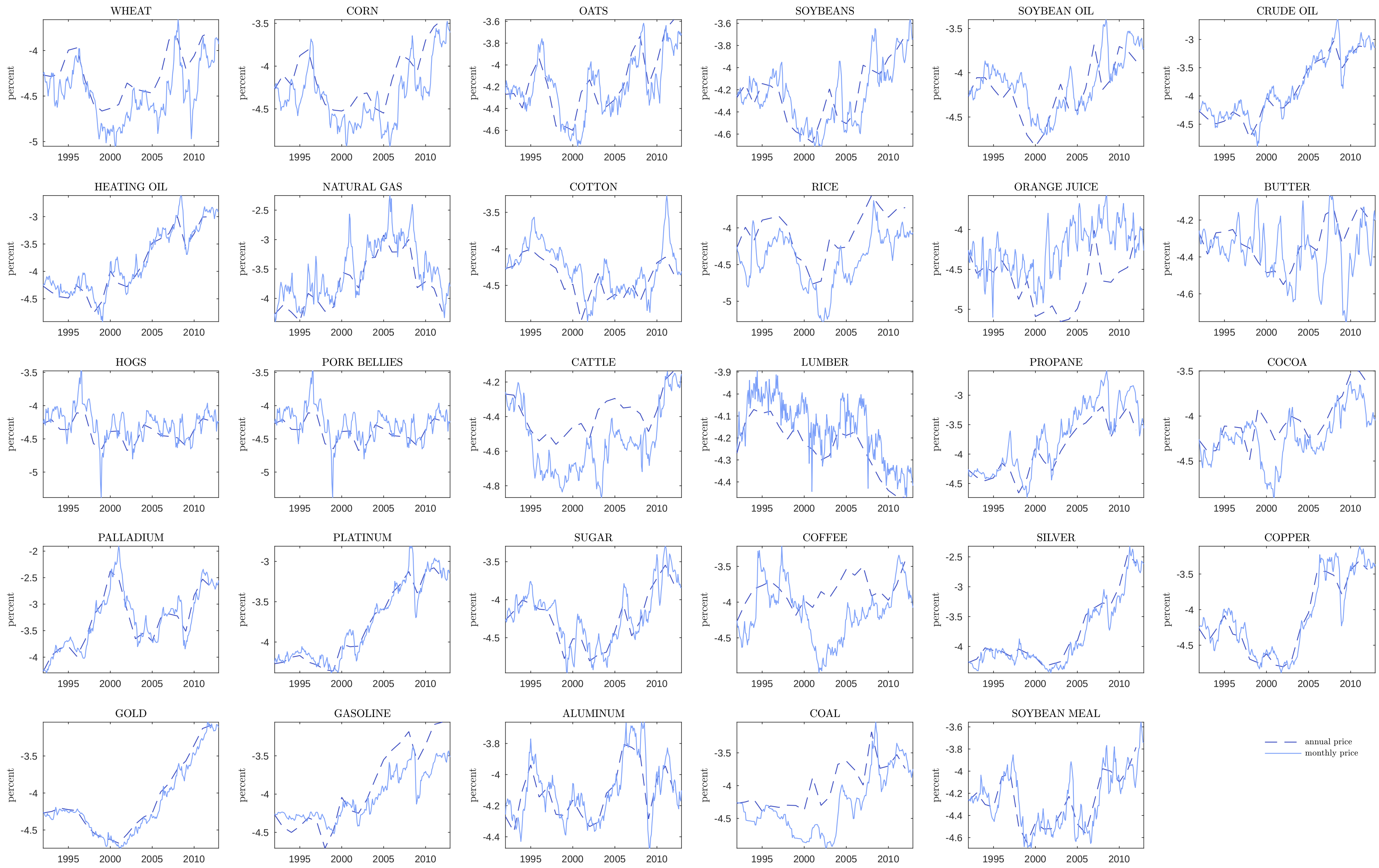


Figure41 displays CFTC-traded commodity prices after scaling by the PCE deflator and the initial observation and taking the log of the result. This is done for the annual and monthly data. In most cases, the data look similar. Exceptions are orange juice, coffee, coal, orange juice, cattle and gasoline.

Figure 42: Volatility: Annual versus Monthly

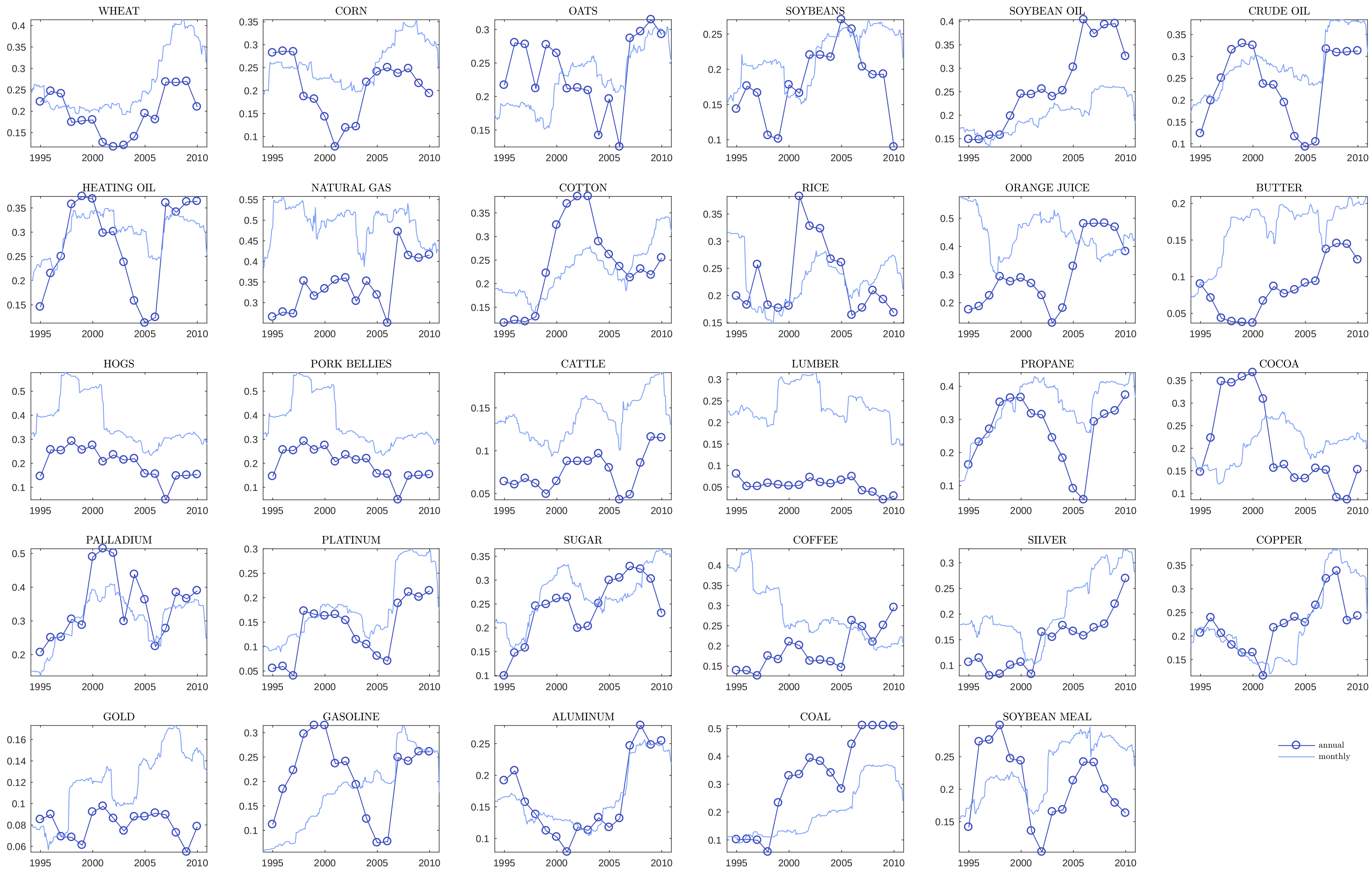


Figure42 compares the annual versus monthly data on the centered volatility of the first difference of log commodity prices. The monthly data have been multiplier by$\sqrt{12}$. In some cases, e.g., orange juice and butter, the results are very different.

Figure 43: Share of World Production, Among CFTC-traded Commodities

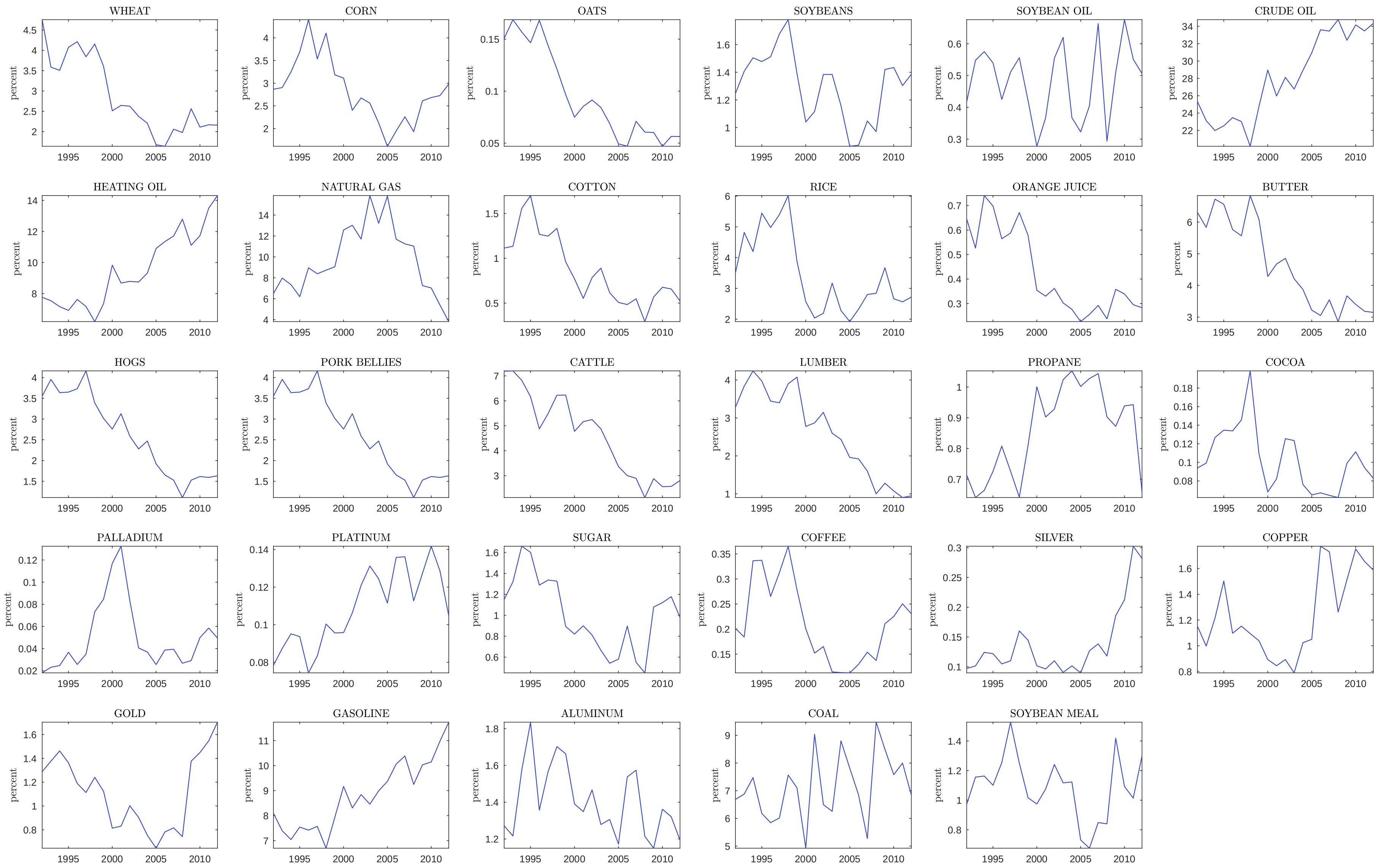


Figure 43 displays the share of world production by commodity. For each year, t, and commodity, i, the numerator is the value of world production in year t for commodity i, and the denominator is the sum, across all CFTC-traded commodities, of the value of world production.

Figure 44: Open interest: Annual and Monthly

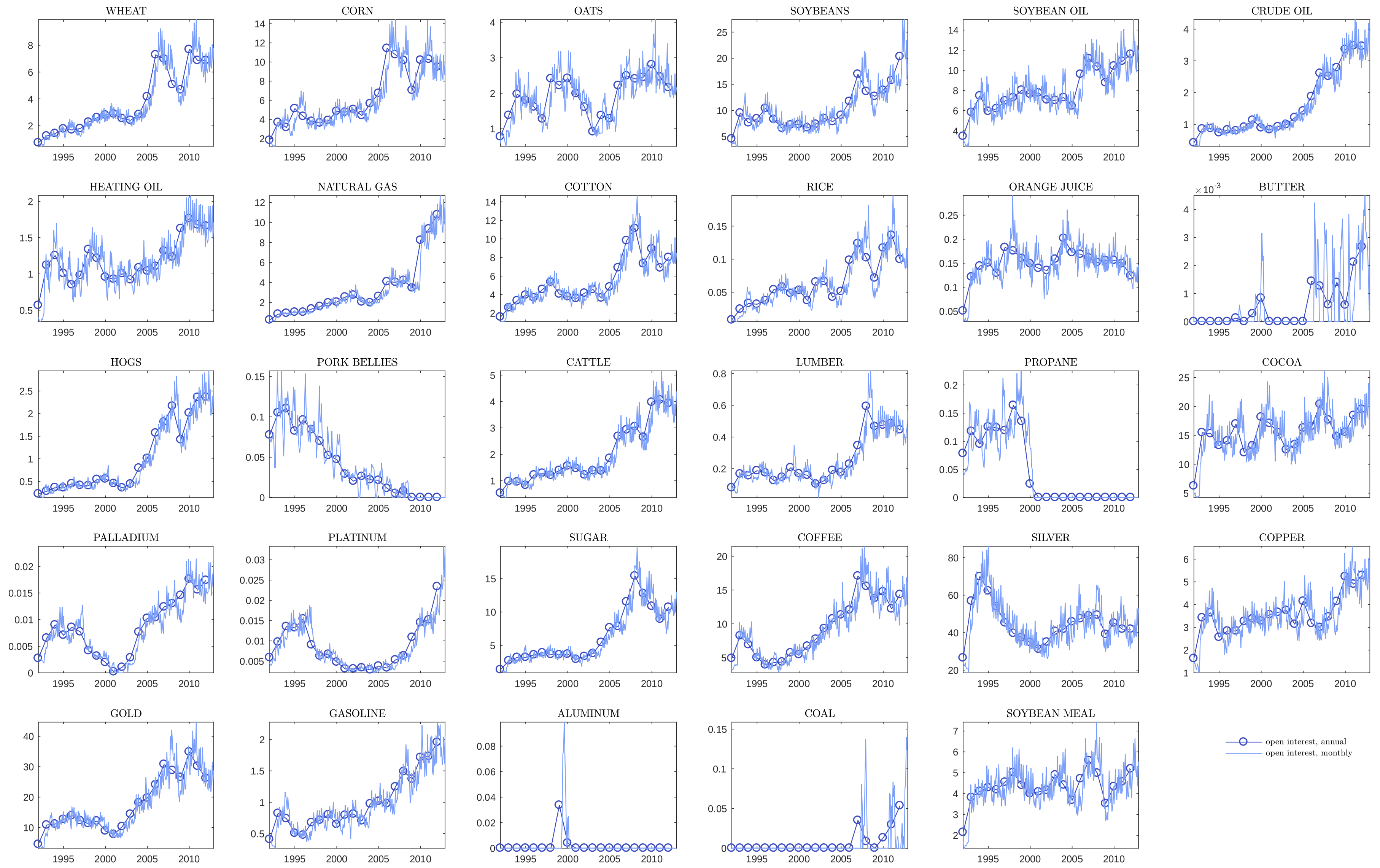


Figure44 displays our monthly and annual open interest data, after scaling by world production. Daily open interest data from the CFTC is the basis for the monthly and annual open interest numbers. The monthly production data are simply the annual data, interpolated. The monthly and annual data appear to be similar in all cases. However, we were concerned by the sharp, high frequency movement in the monthly data. So, we looked at the most important (in terms of world production) commodity, oil.

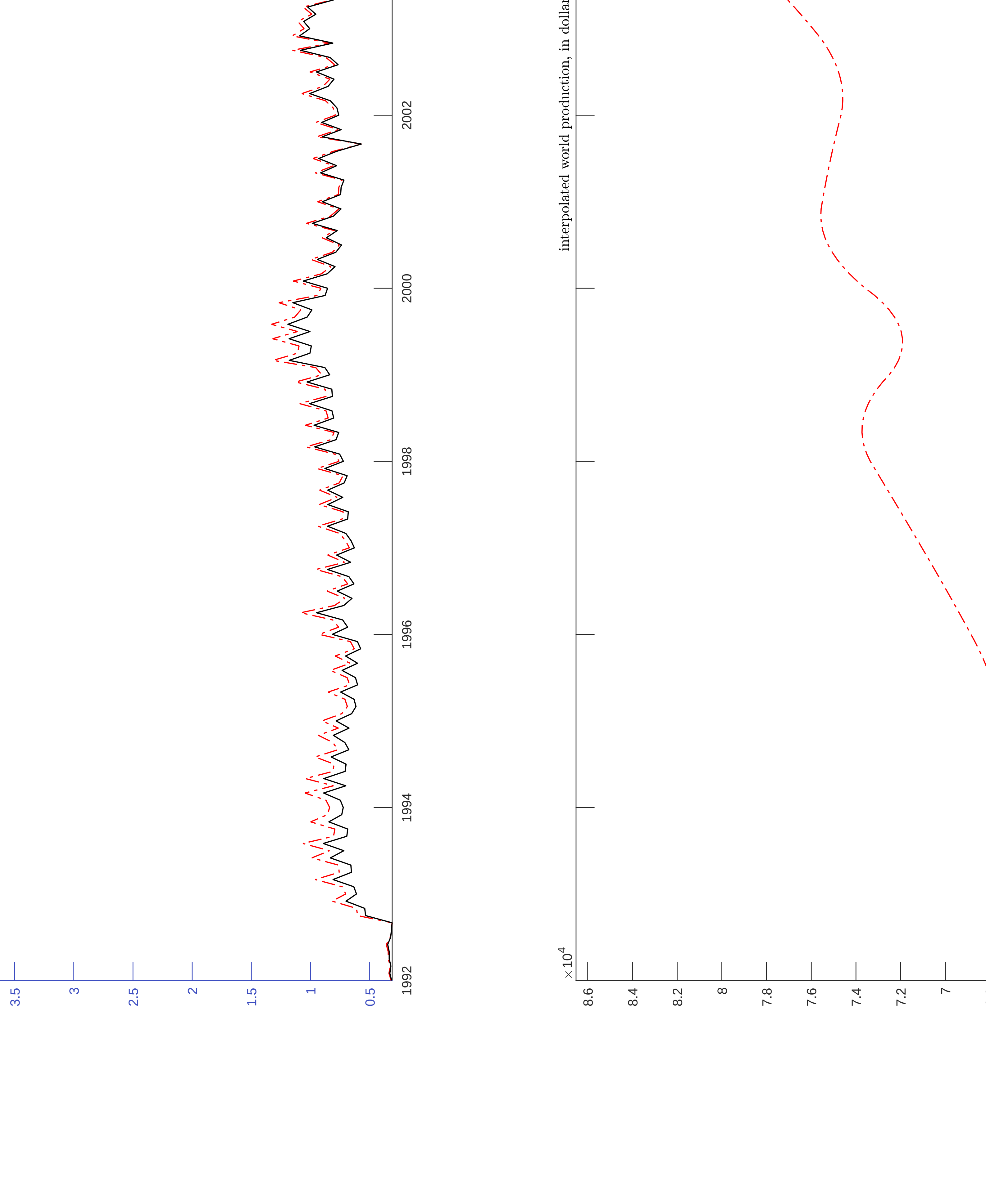

interpolated world production, in
1992
1994
1996
1998
2000
2002
0.5
1
1.5
2
2.5
3
3.5
$\times 10^4$
7
7.2
7.4
7.6
7.8
8
8.2
8.4
8.6

Figure23shows that open interest in oil exhibits the high frequency behavior noticed in Figure44. The upper panel in Figure23displays our open interest measure scaled by world output (left scale) and open interest, not scaled by output (right scale). Note that the two series are roughly the same, suggesting that the reason for the high frequency movements is the open interest data itself. Consistent with this view, the interpolated world production data in the bottom panel of Figure23 show that those series are not the source of the high frequency feature of the scaled oi data.

Figure 46: Open interest and 'seasonal'

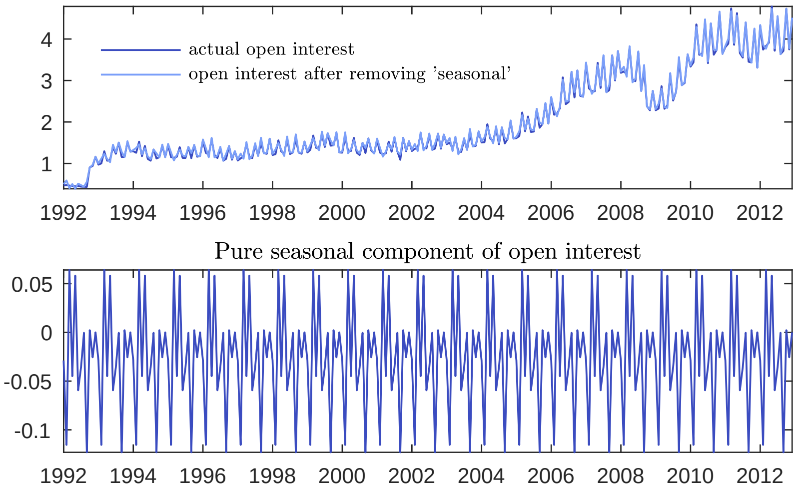


Figure46 explores the high frequency component in open interest some more. The top panel displays the monthly index of open interest (scaled, of course), as well as a version of the monthly oi with its 'seasonal component' removed. This was accomplished by regressing oi on 11 monthly dummies, a constant terms and a trend term. The second line in the top panel subtracts the component of the regression with the 11 monthly dummies from the data ('seasonally adjusts the data') from the raw data. Thus, the second line is the 'seasonally adjusted' data. Note that the two series are virtually indistinguishable, suggesting that the the seasonal is very small. A slightly more formal approach was taken in the bottom part of Figure46. That graphs the seasonal component itself. We did an F test for the null hypothesis that the 11 monthly coefficients are all zero:

$$F = \frac{\left(SSR_r - SSR_{ur}\right)/q}{SSR_{ur}/\left(\text{T-}(q+2)\right)} = 0.26,$$

where$T = 252$ denotes the number of observations,q denotes the 11 monthly coefficients and$q + 2$ denotes the regressors in the unrestricted regression, equal to 13. Also,SSR_r andSSR_{ur} denote the restricted and unrestricted sum of square residuals, respectively. Under the null hypothesis that the coefficients on the monthly dummies are zero, the$F$ statistic is drawn from an$F$ distribution with$q$ numerator and$T -$ q+2 denominator degrees of freedom. Under the null hypothesis, the probability of drawing an$F$ higher than 0.26 is 99%. So, the seasonal does not appear to be statistically significant. Another way to see how small this$F$ statistic is, is to compare the meanSSR_r andSSR_{ur}, which are 0.2521 and 0.2491, respectively. Thus, dropping the monthly dummies increases the variance of the regression error by only 1.2 percent.

Figure 47: Net financial flows: Annual and Monthly

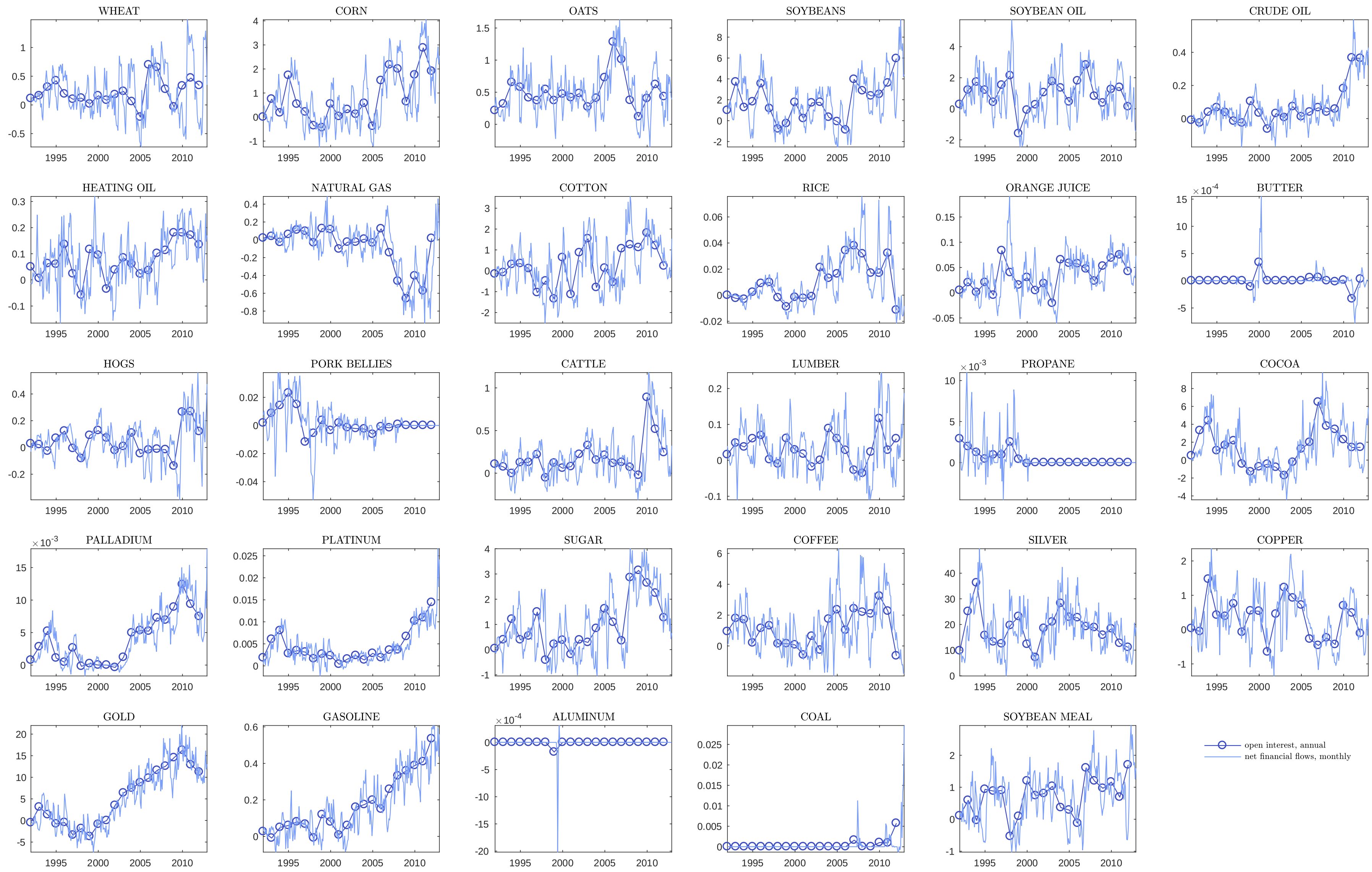
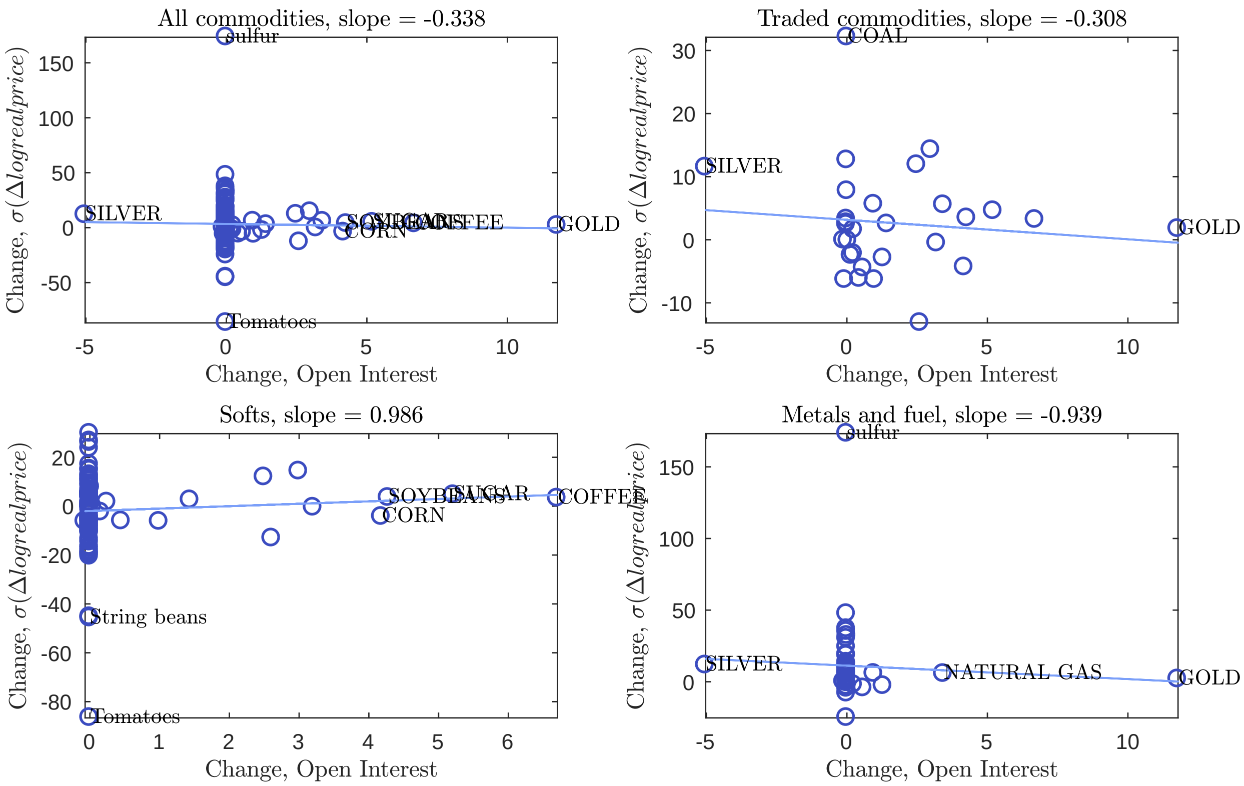
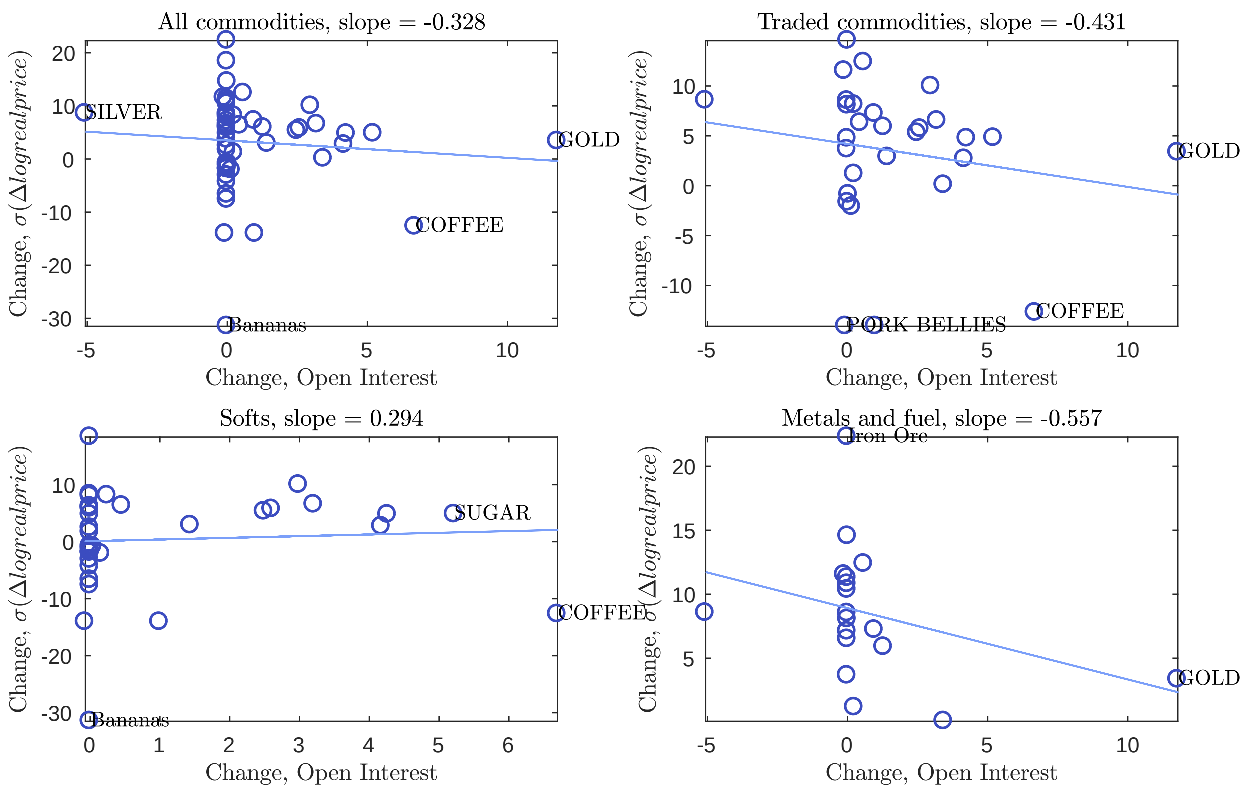


Figure47 displays net financial flows. As in the previous figure, these data have been deflated by world production. In the monthly case, the measure of world production is the annual data, interpolated.

Figure 48: Change in Standard Deviation of Commodity Price Growth Against Change in Open Interest

(a) Annual (change multiplied by 100)



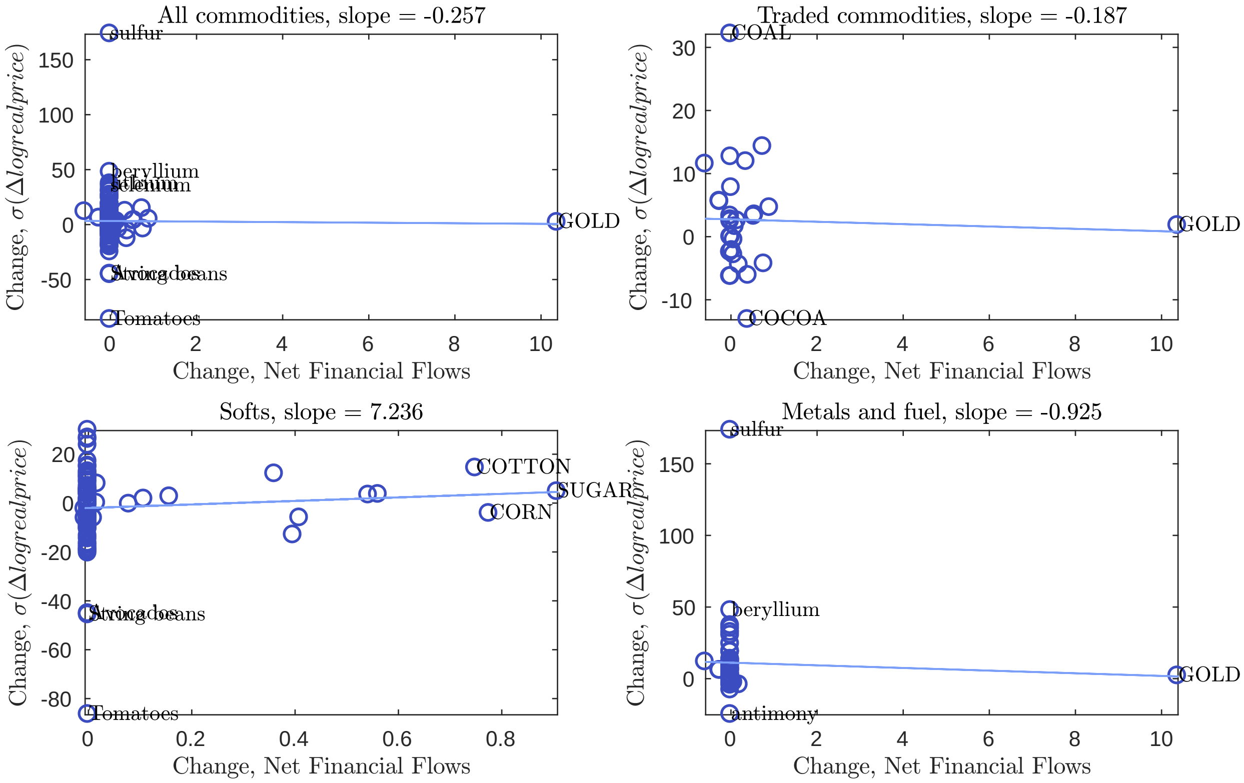


(b) Monthly (change multiplied by 1200)

Figure 48 displays the scatter of the change in the standard deviation commodity price growth against the change in open interest.

Figure 49: Change in Standard Deviation of Commodity Price Growth Against Change in Net Financial Flows

(a) Annual (change multiplied by 100)



(b) Monthly (change multiplied by 1200)

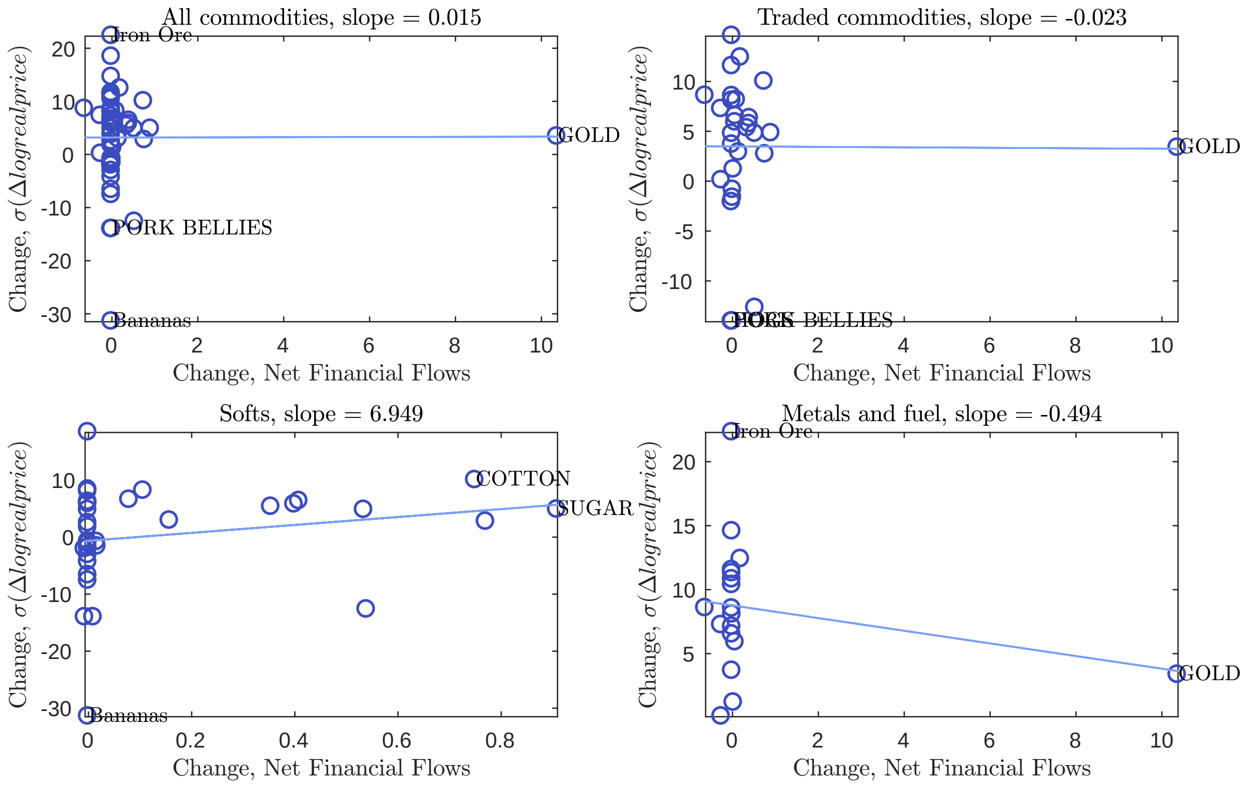
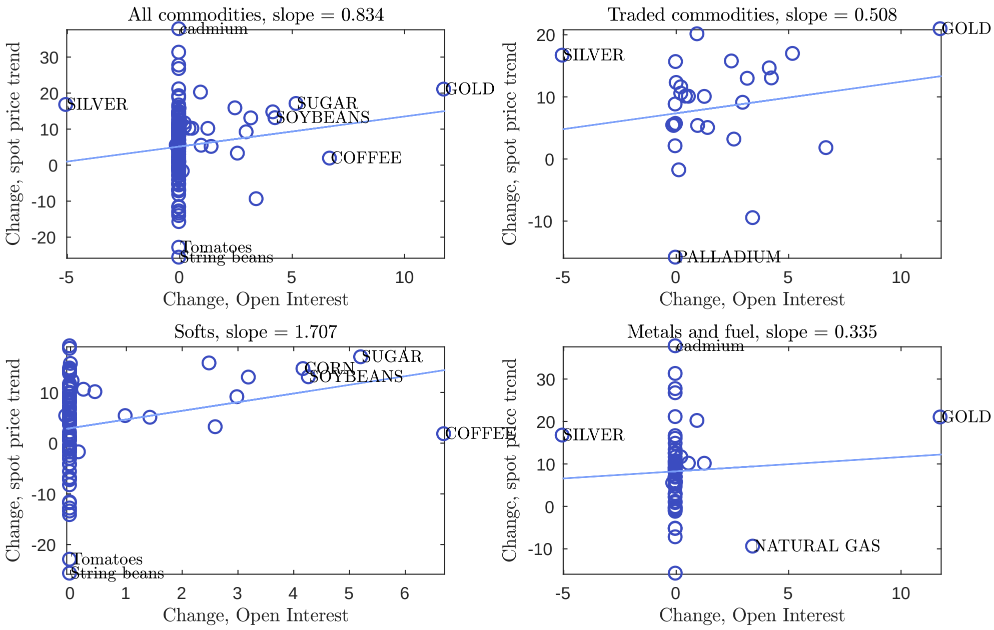
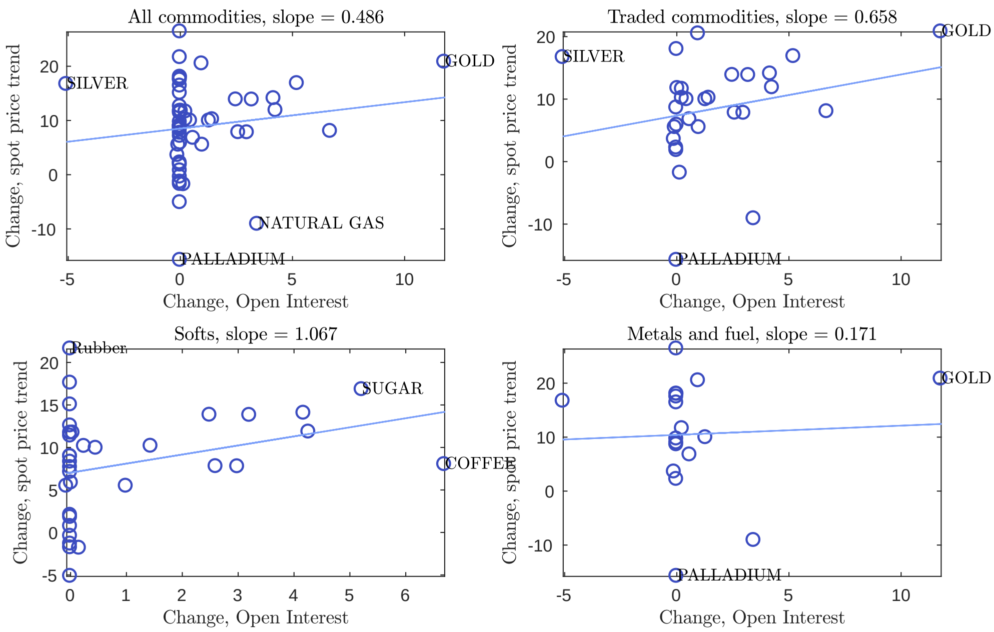


Figure 49 scatters the change in the standard deviation against the change in net financial flows.

Figure 50: Change in Trend against change in Open Interest

(a) Annual (change multiplied by 100)



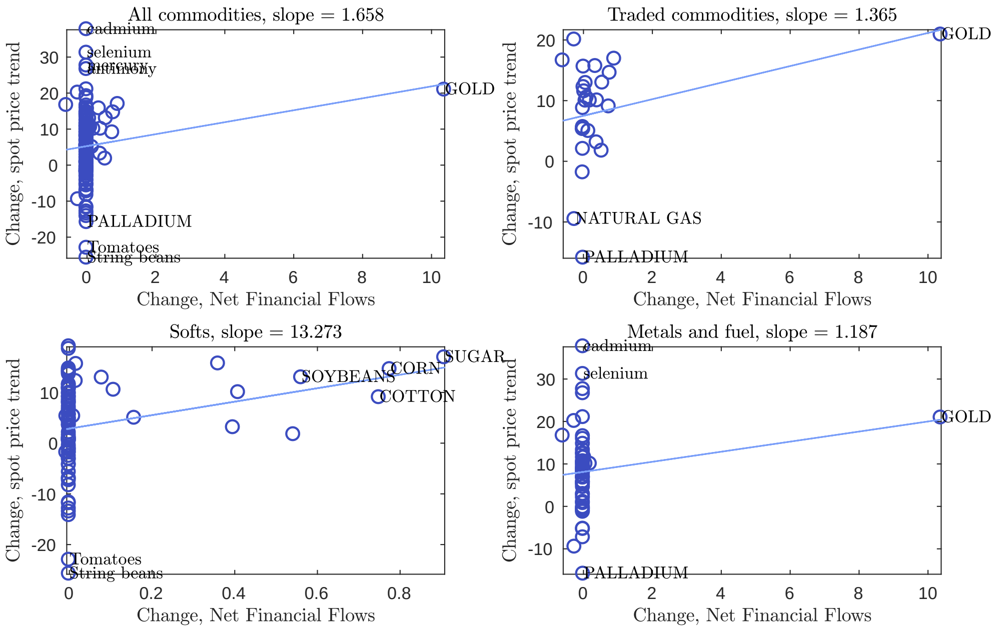


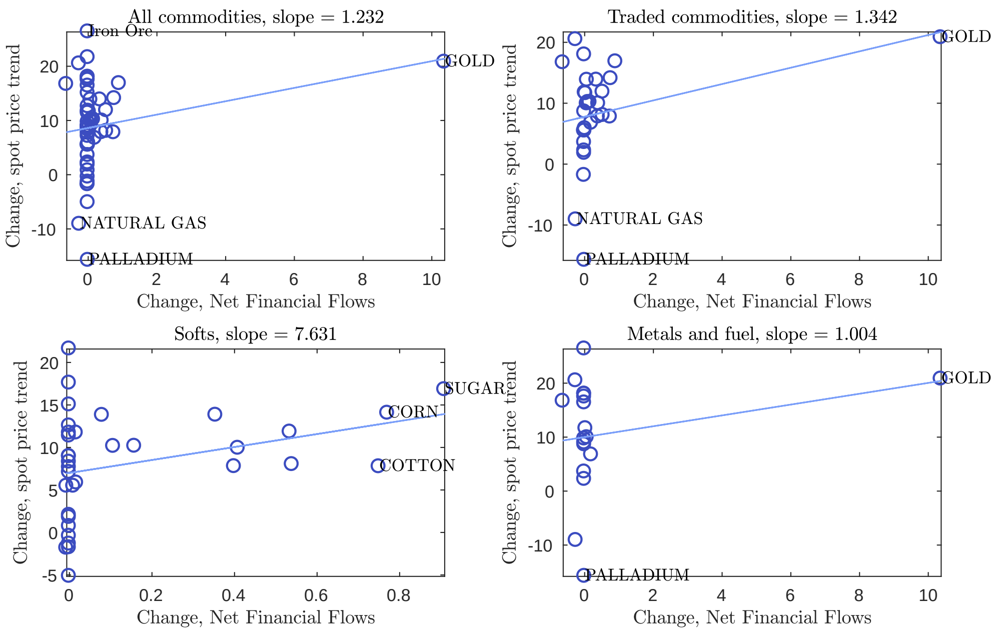
(b) Monthly (change multiplied by 1200)

Figure 50 displays the relationship between the change in the spot price trend in the first and second periods (converted to annual rates) against the change in open interest over the same period.

Figure 51: Change in Trend against change in Net Financial Flows

(a) Annual (change multiplied by 100)



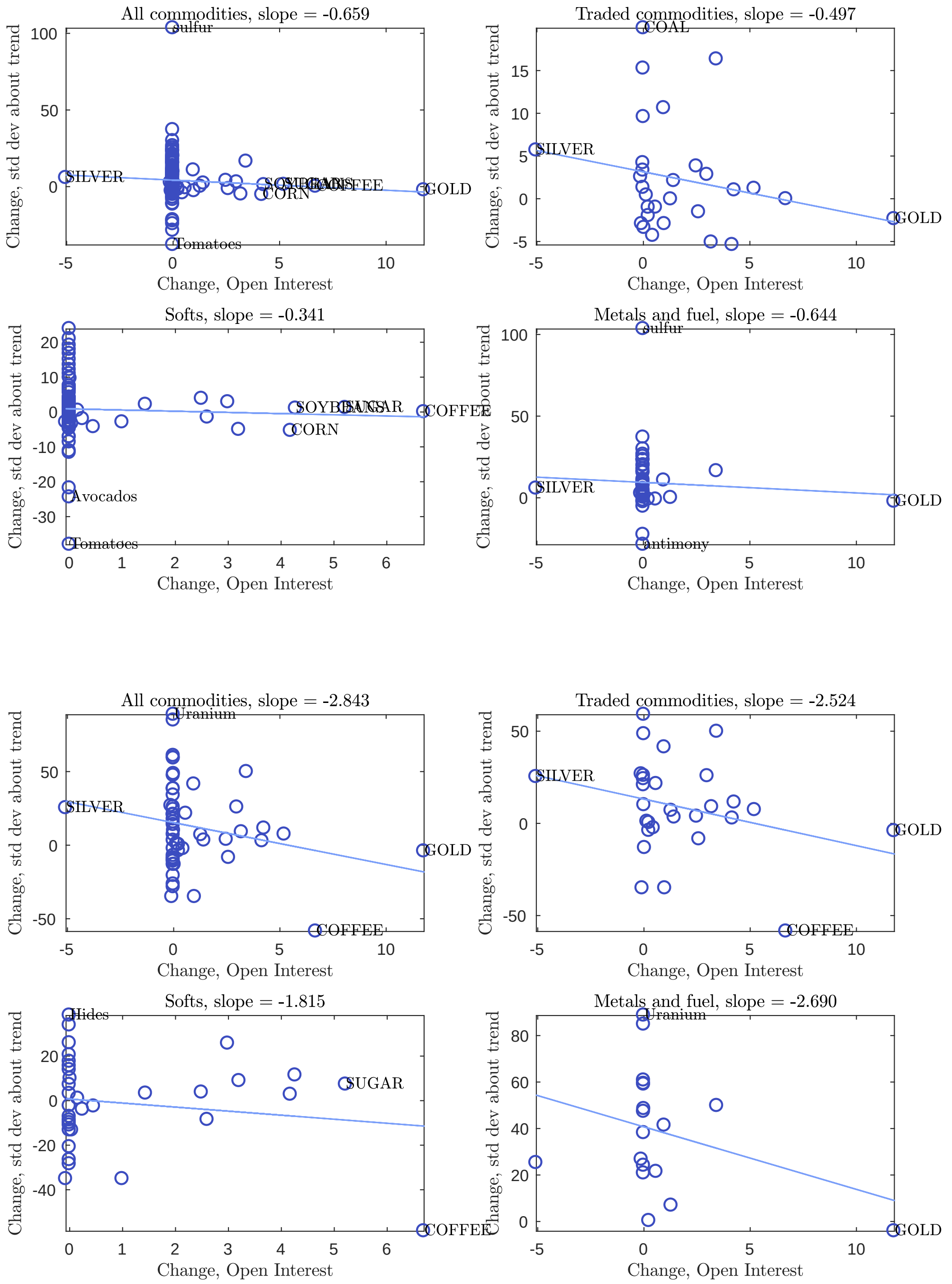


(b) Monthly (change multiplied by 1200)

Figure 51 displays the relationship between the change in the spot price trend in the first and second periods (converted to annual rates) against the change in open interest over the same period.

Figure 52: Change in Std Deviation About Trend Against Change in Open Interest

(a) Annual (change multiplied by 100)

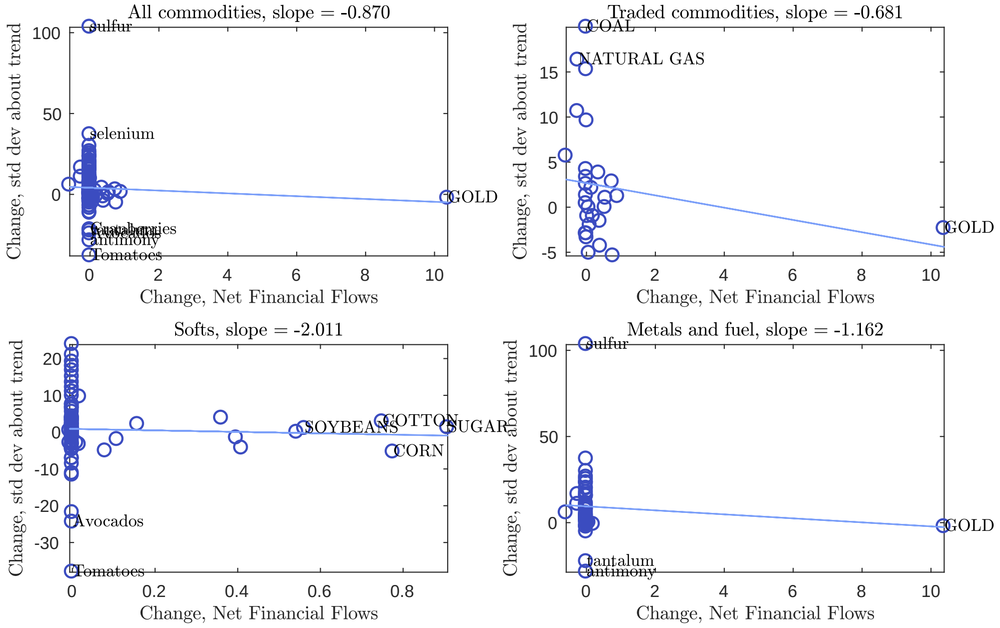


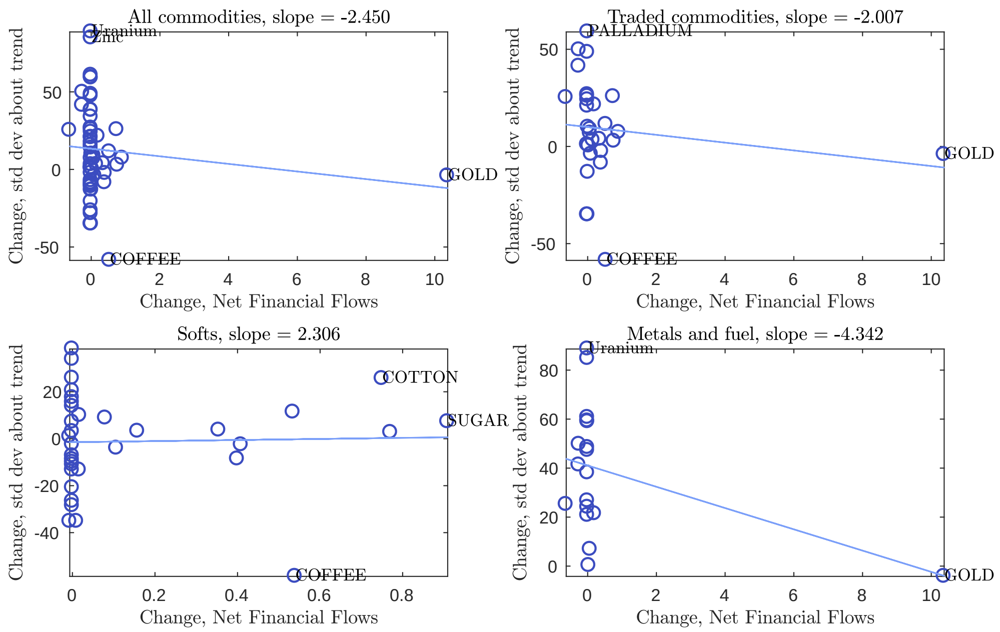
(b) Monthly (change multiplied by 100 times $\sqrt{12}$)

Figure 52 displays the relationship between the change in the spot price trend in the first and second periods (converted to annual rates) against the change in open interest over the same period.

Figure 53: Change in Std Dev About Trend against change in Net Financial Flows

(a) Annual (change multiplied by 100)



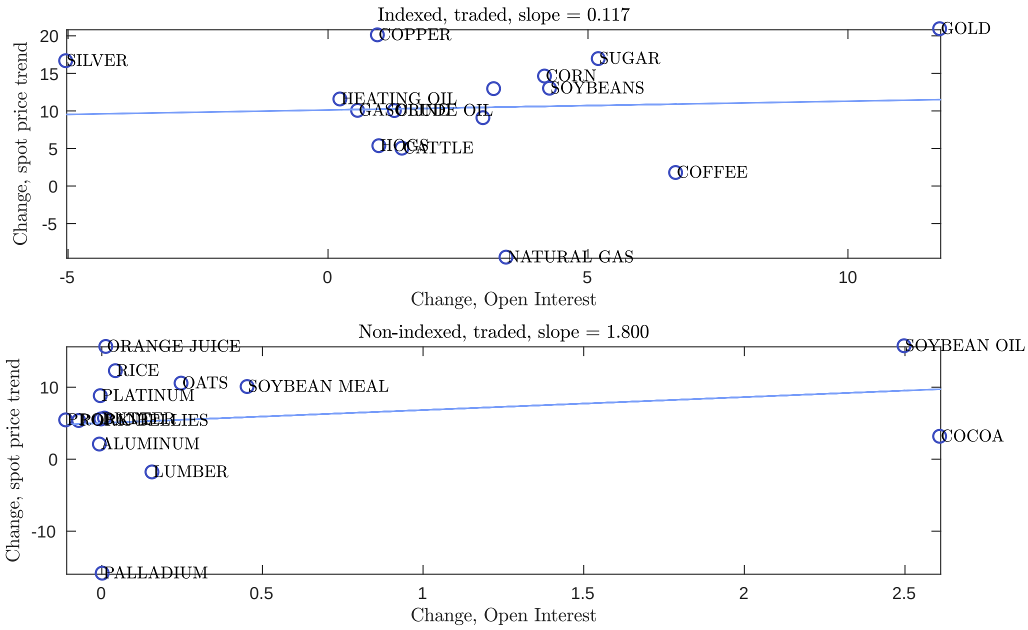


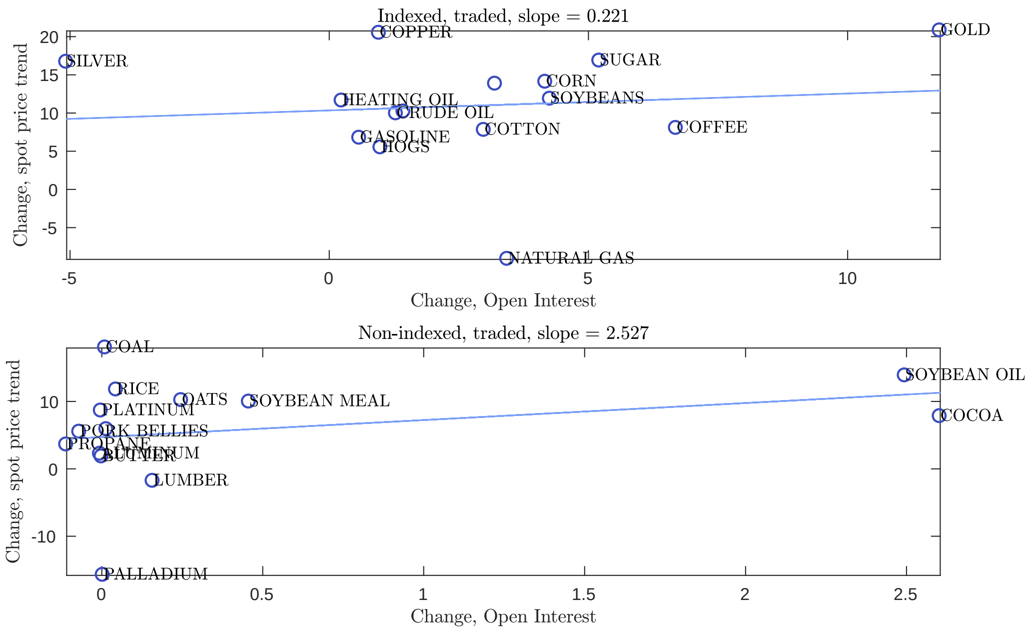
(b) Monthly (change multiplied by 100 times$\sqrt{12}$)

Figure53 displays the relationship between the change in the spot price trend in the first and second periods (converted to annual rates) against the change in net financial flows over the same period.

Figure 54: Change in Trend and Open Interest

(a) Annual (change multiplied by 100)



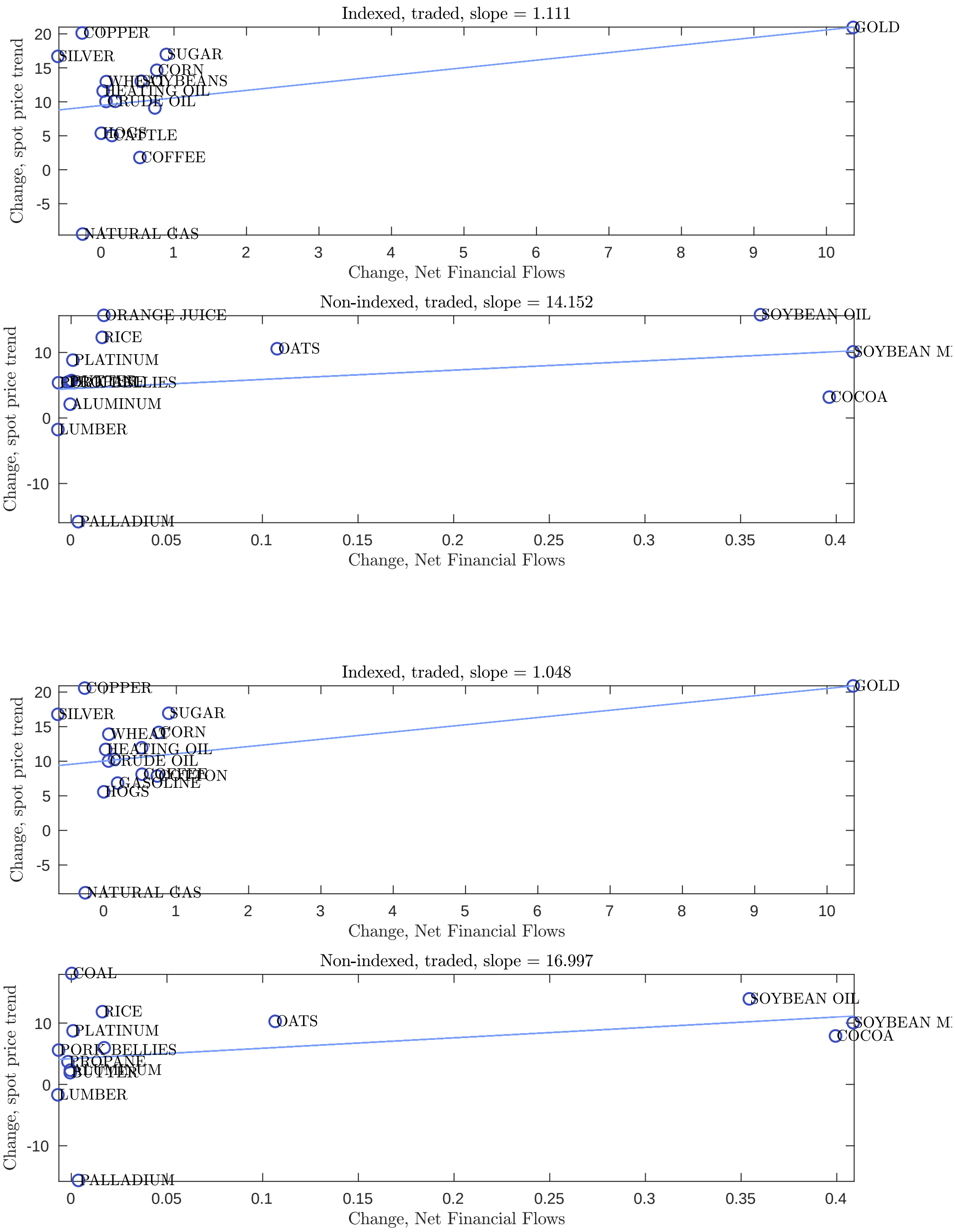


(b) Monthly (change multiplied by 1200)

Figure displays the relationship between XXX add XXX

Figure 55: Change in Trend and Net Financial Flows

(a) Annual (change multiplied by 100)

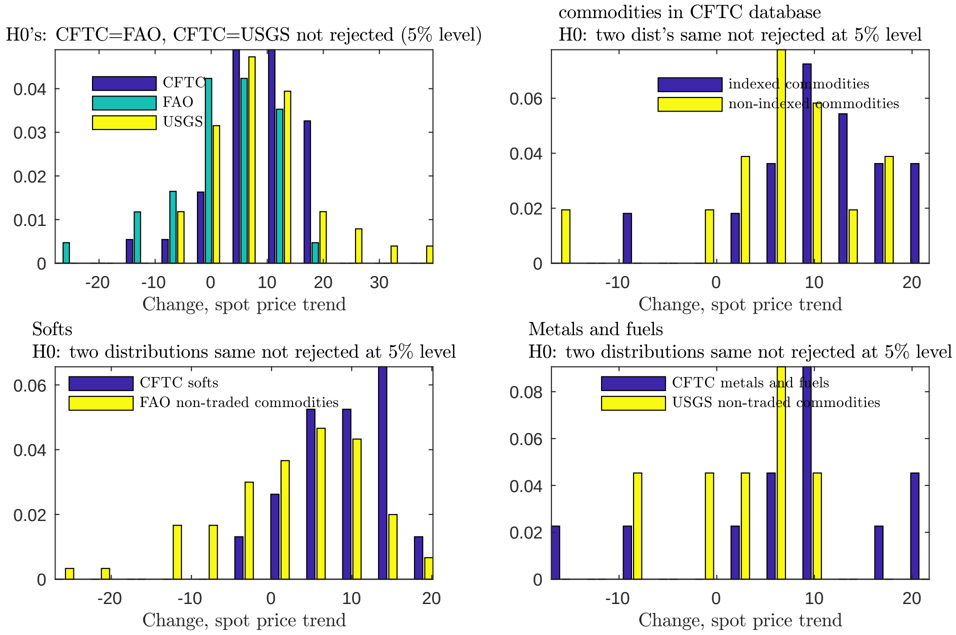


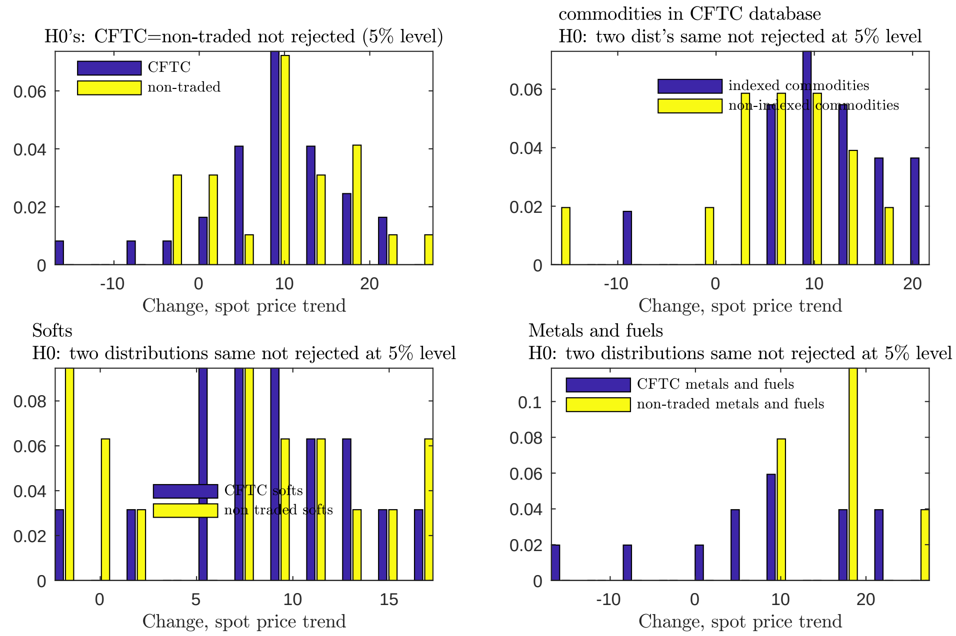
(b) Monthly (change multiplied by 1200)

Figure displays the relationship between XXX add XXX

Figure 56: Change in Trend and Net Financial Flows in Histogram

(a) Annual (change multiplied by 100)



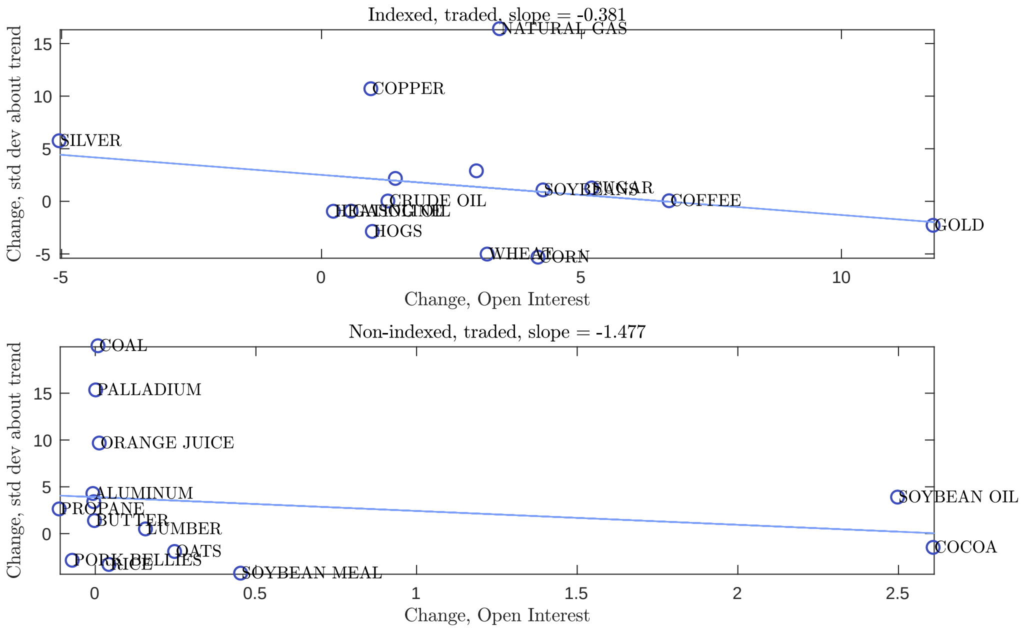


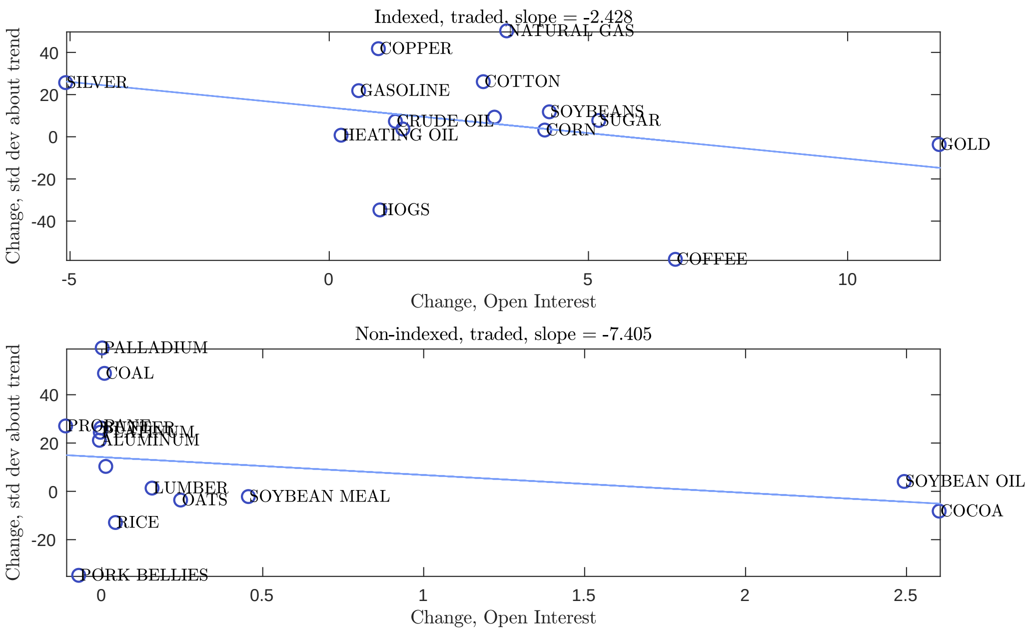
(b) Monthly (change multiplied by 1200)

Figure displays the relationship between XXX add XXX

Figure 57: Change in Std Dev About Trend and Open Interest

(a) Annual (change multiplied by 100)



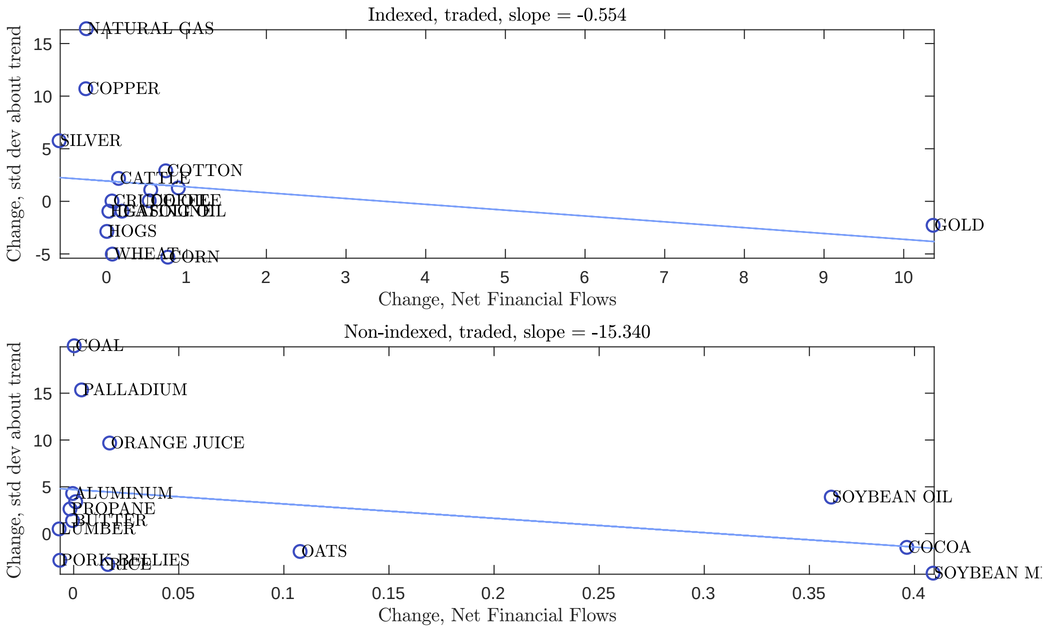


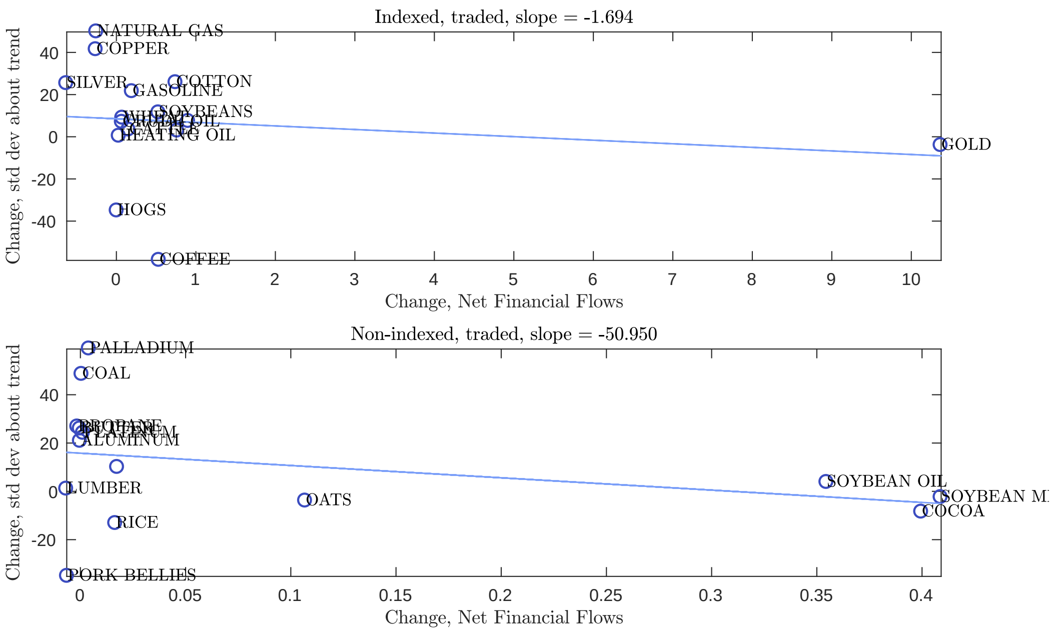
(b) Monthly (change multiplied by 1200)

Figure displays the relationship between XXX add XXX

Figure 58: Change in Std Dev About Trend and Net Financial Flows

(a) Annual (change multiplied by 100)



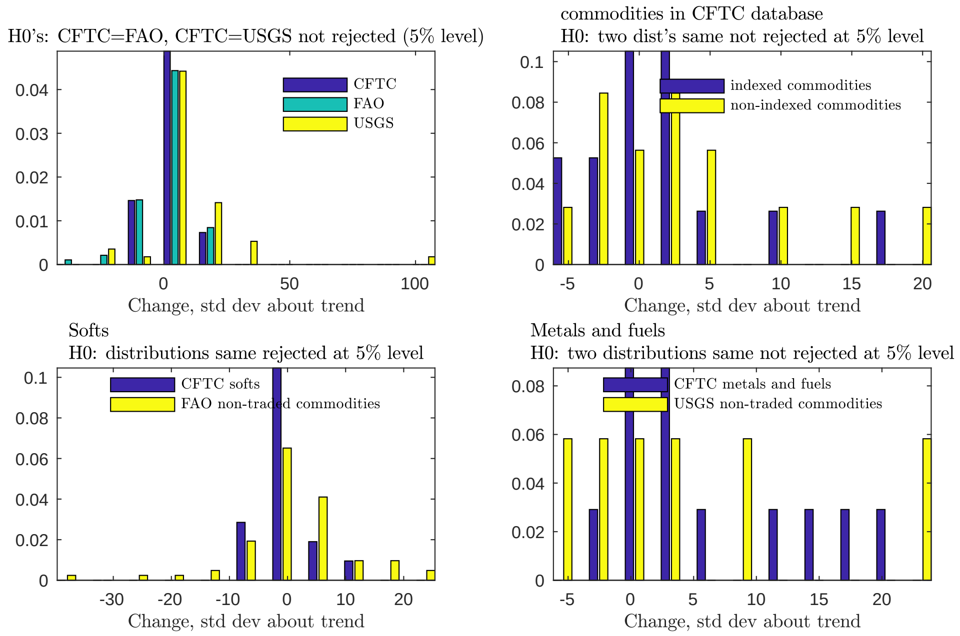


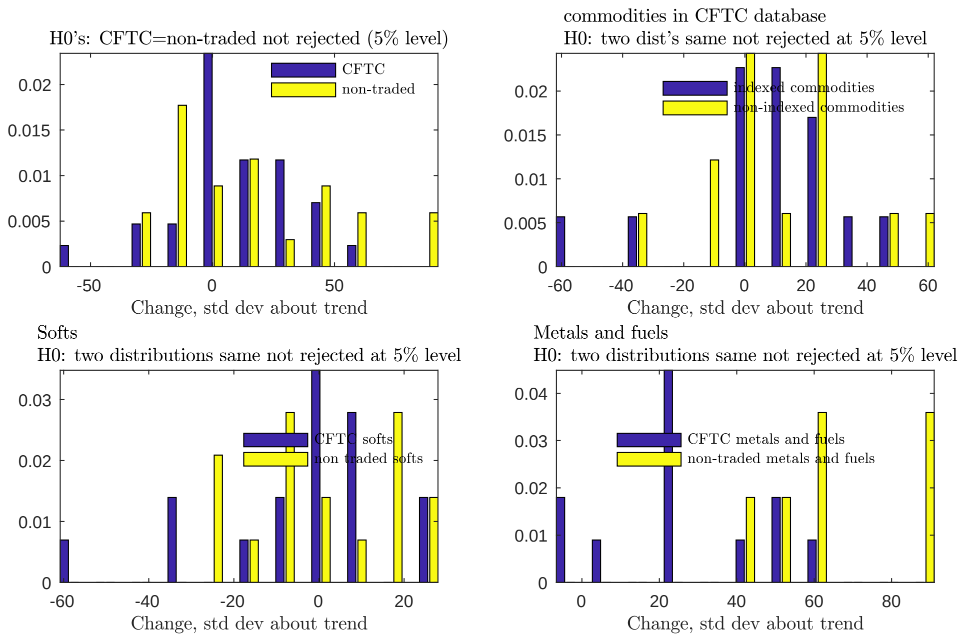
(b) Monthly (change multiplied by 1200)

Figure displays the relationship between XXX add XXX

Figure 59: Change in Std Dev About Trend and Net Financial Flows in Histogram

(a) Annual (change multiplied by 100)





(b) Monthly (change multiplied by 1200)

Figure displays the relationship between XXX add XXX

Figure 60: Time Series of Volatility on Volume: Monthly

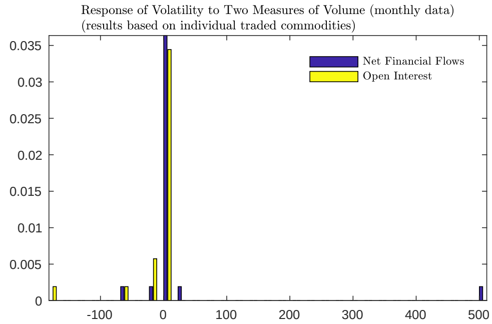


Figure60 is similar to Figure33, except that it applies to monthly data.

6.8. Measures of Volatility Against Volume of Trade

Figure 61: Volatility and Whether or Not a Commodity is Traded

(a) Annual

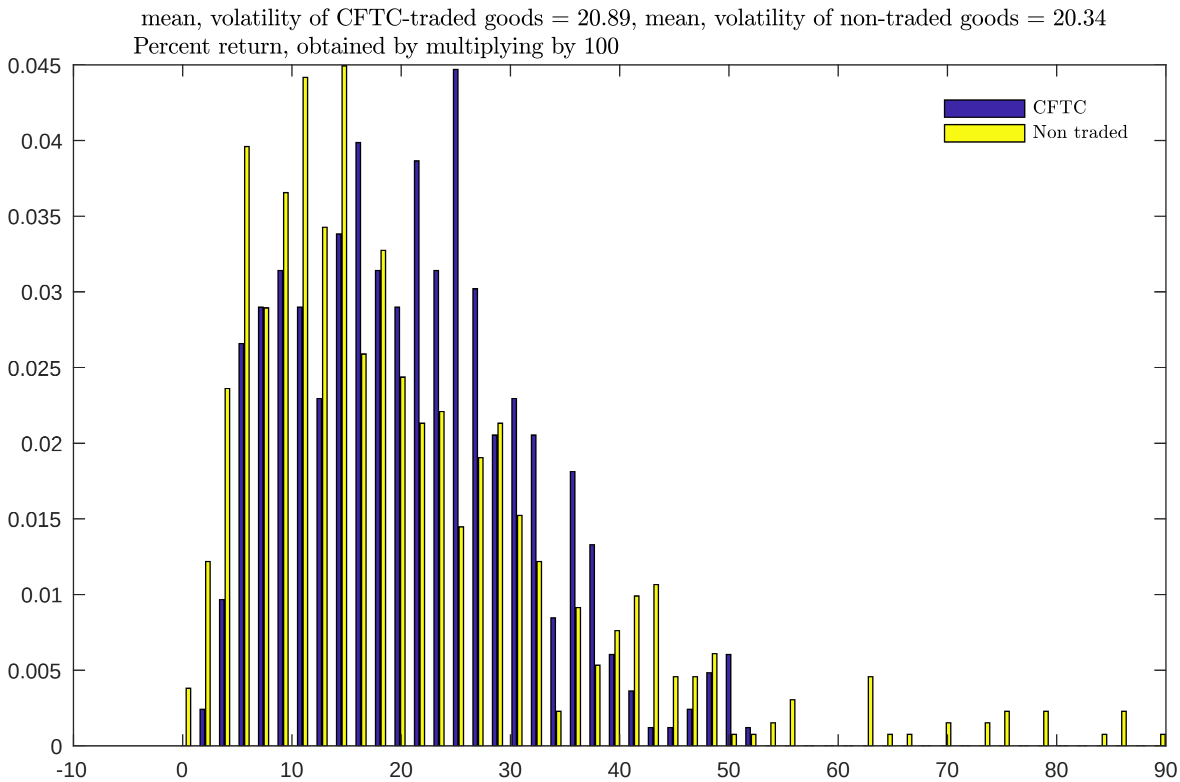
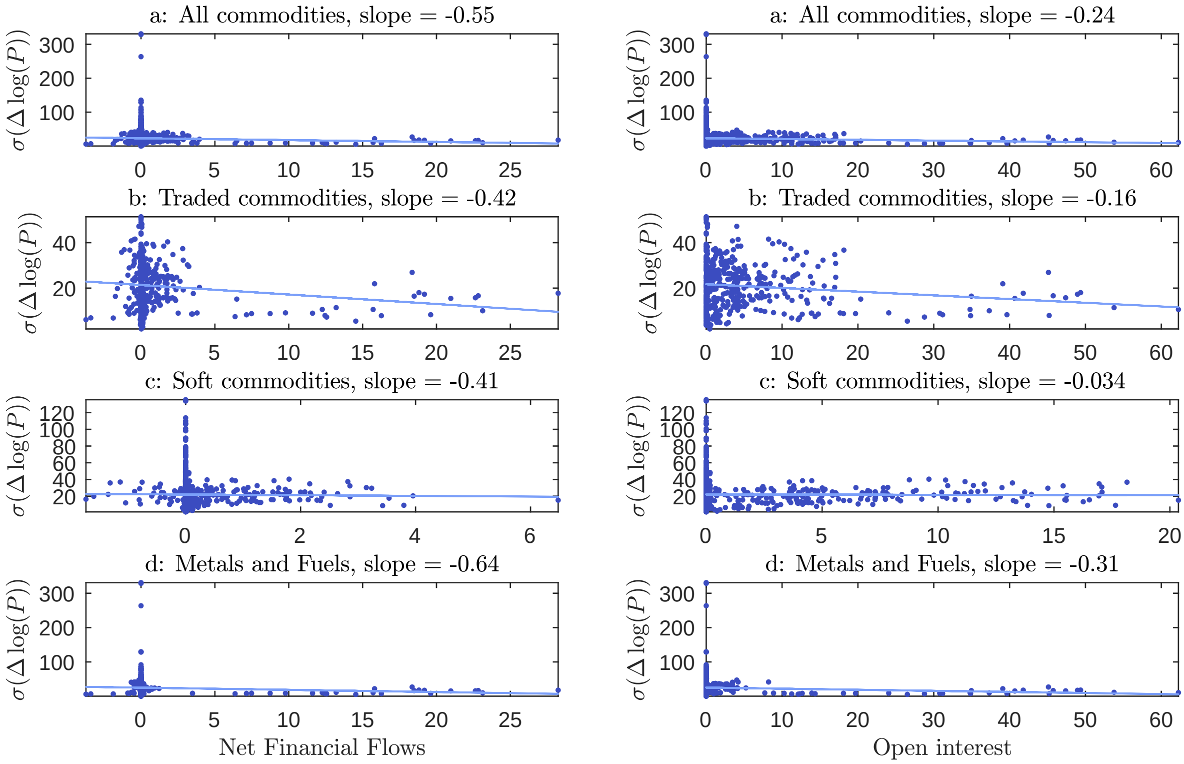


Figure61 displays the empirical density of the centered-window estimates of the standard deviation of the logarithmic first difference of commodities in the CFTC data set versus the non-CFTC data. The first panel contains results for the annual data, in which standard deviations have been converted to percent terms by multiplication by 100. The second panel contains results for monthly data and the results have been presented in annual percent terms by multiplication by$100\sqrt{12}$. Note that, as we have seen in other figures, the volatilities of the non-traded commodity prices are more spread out than the volatility of the CFTC commodities. Still, the mean of the traded volatilities is greater than the mean of the non-traded volatilities. The results are similar for annual and monthly data, although in the annual data, the mean volatility of CFTC-traded goods is only trivially greater than what it is for non-traded goods.

Figure 62 below compares the annual and monthly results for the scatter of price volatility on the level of volume. Price volatility is the standard deviation of log price growth based on a 4 year centered window of the monthly first difference of log, scaled price. The variable on the horizontal axis is the volume of trade, measured by net financial flows in the left column and open interest in the right column. The top panel is based on our annual data set (that figure reproduces Figure Figure 34 below). The bottom panel is based on the monthly data set. The volatilities in the top panel have been multiplied by 100. In the bottom panel they are multiplied by $100 \times \sqrt{12}$ to convert to percent and annual terms. Note that the numbers are broadly similar in magnitude between monthly and annual data. The slopes are also quite similar and quite small.

Figure 62: Volatility and Financialization

(b) Annual Da



Next, we display the scatter of the detrended volatility data against the detrended volume data, where trends were removed at the commodity level. Figure 63 displays the data, where the annual data are copied for convenience from Figure 35. Note that the slope terms are small in magnitude. Again, these results are consistent with the view that the level of financialization has little effect on commodity prices at the high frequency.

Figure 63: Detrended Volatility and Detrended Financialization

(b) Annual

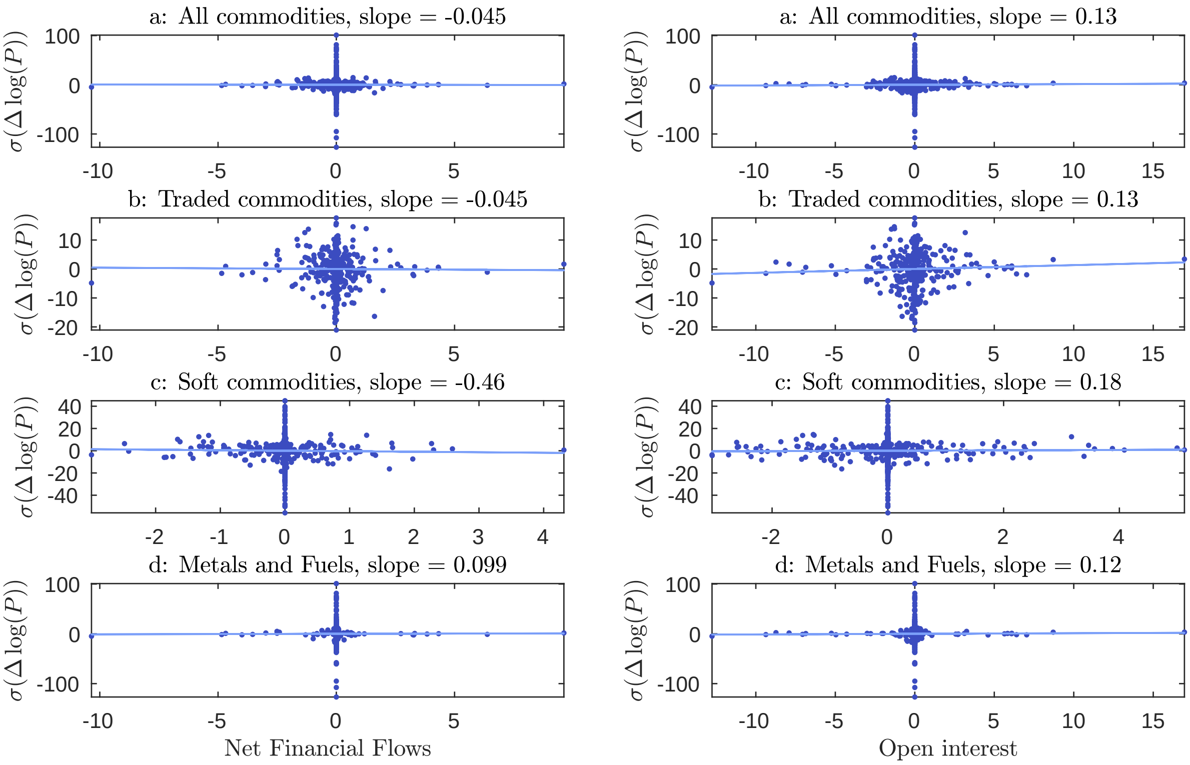


Figure62 corresponds to Figure34, except it uses our monthly data. The regression coefficients are extremely small.

Part VI

Tables

Table 12: Mean and Volatility Properties of Commodity Price Growth

(a) Annual

Commodity	$n^{(i)}$	Mean, spot price growth ($\times 100$)			Standard deviation ($\times 100$)		
		1992-2000	2001-2012	diff (p-values$^{(ii)}$: theory$^{(iii)}$, boot)	1992-2000	2001-2012	diff (p-values : theory, boot)
All	136	-0.66	3.8	4.4 (10,10)	22	25	3.2 (8.6,22)
Indexed	15	-0.51	6.2	6.7 (12,12)	19	21	2.1 (26,38)
Non indexed	14	0.28	3.9	3.6 (20,20)	19	22	3.3 (7,16)
Traded	29	-0.13	5.1	5.2 (14,14)	19	22	2.6 (12,26)
Non trade	107	-0.8	3.4	4.2 (12,12)	22	26	3.4 (9.1,21)

(b) Monthly

Commodity	$n^{(iii)}$	Mean$^{(i)}$			Standard Deviation$^{(ii)}$		
		1992-2000	2001-2012	diff (p-values$^{(iv)}$: theory, boot)	1992-2000	2001-2012	diff (p-values$^{(iv)}$: theory, boot)
All	52	-1.5	5.1	6.6 (2.9,2.9)	22	25	3.2 (7.5,9)
Indexed	15	-0.35	5.6	5.9 (10,10)	24	26	2.9 (15,18)
Non indexed	14	-0.22	3.8	4 (17,17)	22	27	4.1 (3.1,5.3)
Traded	29	-0.29	4.7	5 (11,11)	23	26	3.5 (5.3,7.5)
Non traded	23	-3	5.6	8.7 (0.8,0.83)	20	23	2.8 (27,28)

Note for (a): (i) n denotes the number of commodities in the specified group. (ii) p-value indicates the probability, under the null hypothesis of no difference, of getting an even higher difference than was realized in the data. (iii) p-values are reported using a particular sampling theory and by a bootstrap procedure for robustness (see appendix for details).
Note for (b): (i) Monthly log difference of spot prices, converted to annual percent by multiplying by 1200; (ii) Standard deviation converted to annual percent by multiplying by $100\times\sqrt{12}$; statistics apply to standard standard; (iii) n denotes the number of commodities in the specified group; (iv) p-value indicates the probability, under the null hypothesis of no difference, of obtaining an even higher difference than was realized in the data; the sampling theory underlying our p-values is based on asymptotic theory and on a bootstrap, as reported in the appendix.

Table 13: Decadal Change in Spot Price Behavior Versus Decadal Change in Financialization

Annual

$z_i = \alpha + \gamma q_i + u_i$

left-hand variable, z_i	$\Delta std\left(log\frac{P_i}{P_{i,-1}}\right)$		$\Delta\beta_i$		$\Delta\sigma_i$	
right-hand variable, q_i	Δnff_i	Δoi_i	Δnff_i	Δoi_i	Δnff_i	Δoi_i
Variable group	γ ($p-$value)	γ ($p-$value)	γ ($p-$value)	γ ($p-$value)	γ ($p-$value)	γ ($p-$value)
All	-0.26 (45)	-0.34 (38)	1.7 (15)	0.83 (7.8)	-0.87 (24)	-0.66 (17)
Indexed	-0.054 (48)	-0.15 (37)	1.1 (16)	0.12 (42)	-0.55 (21)	-0.38 (18)
Non-indexed	-21 (13)	-2.4 (25)	14 (15)	1.8 (25)	-15 (11)	-1.5 (29)
Traded	-0.19 (44)	-0.31 (29)	1.4 (14)	0.51 (17)	-0.68 (19)	-0.5 (12)
Softs	7.2 (22)	0.99 (24)	13 (0.93)	1.7 (1.9)	-2 (33)	-0.34 (31)
Metals	-0.97 (28)	-0.75 (28)	1.1 (24)	0.7 (23)	-1.2 (22)	-0.74 (26)
Fuels	-26 (24)	-1.5 (39)	48 (7)	-4.7 (11)	-43 (12)	2.4 (26)

Monthly

All	0.015 (45)	-0.33 (25)	1.2 (16)	0.49 (16)	-2.4 (25)	-2.8 (6.8)
Indexed	0.023 (45)	-0.48 (19)	1 (16)	0.22 (33)	-1.7 (27)	-2.4 (12)
Non-indexed	6.8 (31)	0.76 (40)	17 (11)	2.5 (17)	-51 (13)	-7.4 (19)
Traded	-0.023 (43)	-0.43 (18)	1.3 (14)	0.66 (11)	-2 (25)	-2.5 (8)
Softs	6.9 (12)	0.29 (39)	7.6 (3.7)	1.1 (4.5)	2.3 (44)	-1.8 (19)
Metals	-0.57 (17)	-0.41 (16)	0.81 (26)	0.52 (29)	-5 (17)	-3 (15)
Fuels	23 (17)	-3 (13)	40 (13)	-5.5 (10)	-85 (14)	5.6 (26)

NOTE : (i) table reports γ and $p-$value for null hypothesis, $\gamma = 0$, obtained by 1,000,000 bootstrap simulations; for each bootstrap simulation we construct a data set having the same length as the actual data by randomly drawing with replacement from the empirical data; the left and right hand variables were drawn independently to ensure consistency with the null hypothesis; (ii) Δ denotes value of variable in second half of data minus value in first half; (iii)σ_i denotes standard deviation of error in regression of log, i^{th} commodity price, on constant and time trend; (iv) β_i denotes coefficient on time of the regression of log, i^{th} commodity price, on constant and time trend; (v) $std\left(\Delta log P_i\right)$ denotes standard deviation of log, first difference of i^{th} commodity price; (v) non-traded goods are not included because in this case $q_i = 0$ for all i; (vi) because the number of traded metals is relatively small, a tiny fraction of bootstrap-simulated regressions produced non-invertible $X'X$ matrices and these were ignored in computing the $p-$values.

7. Regression of Volatility on Volume

Tables **??** and 15 report the results of regressing volatility on volume for all commodities, as well as for subgroups of commodities. Two types of regressions were performed. In one, (non-adjusted) the regression involves a constant and slope term that is common across all velocities. In the other, (adjusted) there is a distinct constant and slope for each commodity.

Two approaches were taken to doing the sampling theory under the null hypothesis that there is no relationship between volatility and volume. Method 1 is reported in **??** and Method 2 is reported in 15. Loosely, method 1 samples the underlying data and then computes the volatilities and method 2 samples the volatilities directly. The advantage of method 1 is that by sampling the volatilities the method allows for the serial correlation of the volatilities induced by the moving window method for computing the standard deviations.

7.1. Method 1.

That involves randomly sampling from the actual price growth data and volume data for each commodity. With each such random dataset we then compute the time series of volatilities and then perform the adjusted and unadjusted regressions. We then see how often the bootstrap-simulated slope terms exceed the estimated slope term. That is the $p-$value.

7.1.1. Unadjusted regression

In the unadjusted regression, we regress the volatilities on measures of volume for all dates and commodities ('all') as well as for subsets of those commodities ('soft', 'metals and fuels'). In this regression, the constant terms and coefficient on volume are held fixed across commodities. We focus on the sampling distribution of the coefficient on volume.

7.1.2. Adjust regression

This case is similar to 'unadjusted' with one exception. The regression allows for a different constant and time trend for each commodity. Again, the slope on volume is fixed across all commodities. Again, it is this slope term that we focus on in the analysis. Table 3 reports the estimated slope term, as well as the centered 95% confidence interval for the simulated slope terms.

Table 14: Regression, Volatility of Spot Price Growth on Volume of Futures Trades[1]

(a) Annual

$$volatility_t = control + \gamma \times volume_t^{2,3}$$

	Net Financial Flows		Open Interest	
Variables	Non Adjusted	Adjusted	Non Adjusted	Adjusted
All	-0.55	-0.045	-0.24	0.13
	(-0.71,-0.072)	(-0.26,0.25)	(-0.33,-0.07)	(-0.17,0.18)
Traded	-0.42	-0.045	-0.16	0.13
	(-0.55,0.097)	(-0.26,0.25)	(-0.24,0.043)	(-0.17,0.18)
Softs	-0.41	-0.46	-0.034	0.18
	(-1.7,0.73)	(-0.88,0.85)	(-0.42,0.15)	(-0.32,0.34)
Metals & Fuels	-0.64	0.099	-0.31	0.12
	(-0.85,-0.14)	(-0.27,0.26)	(-0.4,-0.088)	(-0.22,0.22)

(b) Monthly

$$volatility_t = control + \gamma \times volume_t^{2,3}$$

	Net Financial Flows		Open Interest	
Variables	Non Adjusted	Adjusted	Non Adjusted	Adjusted
All	-0.13	0.071	-0.058	0.18
	(-0.25,-0.052)	(-0.023,0.022)	(-0.12,-0.02)	(-0.018,0.018)
Traded	-0.26	0.071	-0.17	0.18
	(-0.36,-0.15)	(-0.023,0.022)	(-0.22,-0.11)	(-0.018,0.018)
Softs	0.28	0.0061	0.079	0.19
	(-0.25,0.35)	(-0.07,0.075)	(-0.11,0.14)	(-0.045,0.042)
Metals & Fuels	-0.18	0.084	-0.099	0.18
	(-0.3,-0.082)	(-0.025,0.023)	(-0.15,-0.043)	(-0.02,0.02)

Notes: (1) the table reports our least squares estimates of γ, the (common) slope coefficient in a regression of volatility (a two-year moving, centered standard deviation of one-period real spot price growth) on our two measures of volume (net financial flows and open interest); in the case of monthly data, reported γ's and boundaries of probability intervals are multiplied by $\sqrt{12}$ to make results comparable with results based on annual data (see text for discussion), (2) "non-adjusted" means that "control" is a constant that is common across all commodities, (3) "adjusted" means that "control" is a separate constant and time trend for each commodity. In the case of (3), estimation is done in two steps. In the first step, the volatility and volume data are replaced by the error in their (commodity-specific) regression on a constant term and a time trend. In the second step the error in price volatility (that is, "adjusted" price volatility) is regressed on adjusted volume and γ is the common slope coefficient across all commodities. It is easy to verify that results for "all" variables and "traded" variables are mathematically identical. (4) "All" means analysis is done using all commodities, "traded" means only commodities in our CFTC data included in the analysis; "softs" and "metals & fuels" means only commodities classified as softs and metals and fuels included in the analysis (see text for further discussion).

7.2. Method 2

In this method the sampling is done directly on the volatilities (i.e., not the underlying data, as in method 1) and on the measures of volume (like before). The unadjusted regression is unchanged apart from this one difference. To visualize what is involved, consider the scatter plot of volatilities and measures of volume in Figure 62. The slope term is the coefficient of the regression of (annualized) volatility on volume. Thus, if open interest goes up by one times world production, then volatility goes down by 0.058 percentage points, when all commodities are considered.

Table 15: Regression, Volatility of Spot Price Growth on Volume of Futures Trades[1]

(a) Annual

$$volatility_t = control + \gamma \times volume_t^{2,3}$$

	Net Financial Flows		Open Interest	
Variables	Non Adjusted	Adjusted	Non Adjusted	Adjusted
All	-0.55	-0.045	-0.24	0.13
	(-0.73,-0.39)	(-0.21,0.21)	(-0.31,-0.19)	(-0.14,0.15)
Traded	-0.42	-0.045	-0.16	0.13
	(-0.6,-0.27)	(-0.22,0.21)	(-0.24,-0.12)	(-0.14,0.15)
Softs	-0.41	-0.46	-0.034	0.18
	(-1.2,0.42)	(-0.91,0.9)	(-0.26,0.045)	(-0.54,0.55)
Metals & Fuels	-0.64	0.099	-0.31	0.12
	(-0.85,-0.45)	(-0.16,0.17)	(-0.38,-0.23)	(-0.11,0.12)

(b) Monthly

$$volatility_t = control + \gamma \times volume_t^{2,3}$$

	Net Financial Flows		Open Interest	
Variables	Non Adjusted	Adjusted	Non Adjusted	Adjusted
All	-0.13	0.071	-0.058	0.18
	(-0.25,-0.16)	(-0.047,0.046)	(-0.093,-0.059)	(-0.034,0.034)
Traded	-0.26	0.071	-0.17	0.18
	(-0.39,-0.28)	(-0.047,0.046)	(-0.21,-0.17)	(-0.034,0.034)
Softs	0.28	0.0061	0.079	0.19
	(-0.06,0.21)	(-0.093,0.094)	(-0.0072,0.055)	(-0.064,0.065)
Metals & Fuels	-0.18	0.084	-0.099	0.18
	(-0.3,-0.2)	(-0.053,0.051)	(-0.13,-0.089)	(-0.039,0.039)

Notes: (1) the table reports our least squares estimates of γ, the (common) slope coefficient in a regression of volatility (a two-year moving, centered standard deviation of one-period real spot price growth) on our two measures of volume (net financial flows and open interest); in the case of monthly data, reported γ's and boundaries of probability intervals are multiplied by $\sqrt{12}$ to make results comparable with results based on annual data (see text for discussion), (2) "non-adjusted" means that "control" is a constant that is common across all commodities, (3) "adjusted" means that "control" is a separate constant and time trend for each commodity. In the case of (3), estimation is done in two steps. In the first step, the volatility and volume data are replaced by the error in their (commodity-specific) regression on a constant term and a time trend. In the second step the error in price volatility (that is, "adjusted" price volatility) is regressed on adjusted volume and γ is the common slope coefficient across all commodities. It is easy to verify that results for "all" variables and "traded" variables are mathematically identical. (4) "All" means analysis is done using all commodities, "traded" means only commodities in our CFTC data included in the analysis; "softs" and "metals & fuels" means only commodities classified as softs and metals and fuels included in the analysis (see text for further discussion).

Table 16: Futures Returns and Financialization

$$correlation_t = c + \beta \times volume_t$$

	Correlation of Futures Returns with:			
	Equity Returns		3 month T-bill	
Measure of Volume	Monthly	Daily	Monthly	Daily
Net financial flows	3.6	4	0.88	0.17
	(-2.7,2.8)	(-4.6,11)	(-1.9,2.5)	(-0.3,0.54)
Open interest	1	1.4	1.1	0.25
	(-1.7,2.1)	(-5,5.7)	(-0.94,1.3)	(-0.19,0.28)

$$volatility_t = c + \beta \times volume_t$$

	Volatility of Futures Returns	
Measure of Volume	Monthly	Daily
Net financial flows	-0.065	0.0014
	(-0.29,0.38)	(-0.049,0.058)
Open interest	-0.11	-0.004
	(-0.15,0.28)	(-0.02,0.049)

Notes: see text for a discussion of estimates of β, which are the number not in parentheses. Numbers in parentheses are 95 confidence intervals under the null hypothesis, $\beta = 0$.

Part VII

Tang-Xiong Commodity Return Correlations

Tang and Xiong (2012) studied the effects of financialization by comparing traded commodities that are included in major indexes ('indexed commodities') and those that are not ('non-indexed'). For our purposes, a commodity is indexed if it belongs to both of the two major indexes, the Standard and Poor GSCI (Goldman-Sachs Commodity Index) and Dow Jones UBSCI (UBS Commodity Index). Tang and Xiong report that the average pairwise correlation between indexed returns is substantially higher than the pairwise correlation among non-indexed returns. They hypothesize that this reflects the effects of financialization. We show that the differences depend on the correlations being computed on daily returns. When computed on monthly returns, the differences are very small. We infer that the effects uncovered by Tang and Xiong may only be transitory in nature, and may not be evidence of an important impact from financialization onto resource allocations.

The Tang and Xiong results may be seen in Figure 64. The 1,1 panel of that figure reproduces the Tang and Xiong findings. To understand the findings, let F_{t+1} denote the price of a futures contract on day $t+1$. Then, $r_{t+1} = F_{t+1}/F_t - 1$ denotes the one month (fully collateralized) rate of return on the contract. (Technically, a fully collateralized contract requires that the nominal rate of interest should be included in the formula for r_{t+1}, but this makes little quantitative difference.) In computing r_{t+1}, F_t and F_{t+1} pertain to the same futures contract, i.e., ones with the same maturity date. For each r_{t+1}, F_{t+1} is the price of the nearest maturing contract. Note that r_t need not pertain to the same contract for different t. We see from the 1,1 panel that the pairwise correlations between indexed commodities is much bigger than it is for non-indexed commodities, as we get into the 2000's.

The 2,2 panel of the figure displays the analogous results for monthly returns. Now the subscript, t, pertains to quarters. Also, if $r_{t+1} = F_{t+1}/F_t - 1$, then F_{t+1} is the price on the first day of month $t+1$ of the nearest maturity contract on that day and F_t is the price on the first day of month t of the same contract. Note that the pairwise correlations of the indexed returns does not change much as we move to monthly, but the pairwise correlations of the non-indexed returns increases substantially, essentially wiping out the difference between the average correlation among indexed and non-indexed returns.

To better understand the reason for the above result, we used the daily data to compute the τ day rate of return on a commodity futures, $r_t^\tau = F_t/F_{t-\tau} - 1$. Here, F_t represents the price, on day t, of the nearest maturing contract on that date and $F_{t-\tau}$ represents the price of the same contract on date $t-\tau$. The 1,2 and 2,1 panels in Figure 64 displays the average pairwise correlations of r_t^τ for τ =10 and 20, respectively. Note that the discrepancy between the pairwise correlations is reduced for $\tau = 10$ and virtually eliminated for $\tau = 20$. This is accomplished by an increase in the pairwise correlations among non-indexed commodity returns.

The above findings suggest a hypothesis about what is driving the results in panel 1,1. The hypothesis is consistent with Tang and Xiong's conclusion that financialization has affected the returns on commodities, but suggests the possibility that the effects are transient and benign. The idea is that purchases or sales by the major index funds acts as a correlated demand or supply shock across all the commodities in the index. The impact of these shocks on commodity prices are transitory, and their effects are eliminated in a few days or weeks as the market absorbs them. If we suppose that the effects are present in indexed commodities, but not in non-indexed commodities, then pairwise correlations should become equated across the two categories of commodities as we consider weekly and monthly returns. The data suggest that as these correlations become equal, it is because the non-indexed commodities become more highly correlated in weekly and monthly returns. This suggests the presence of idiosyncratic, transitory shocks to the price of all commodities. In what follows, we formalize these observations with an example.

The rate of return over a month is roughly the sum of the rates of return over each day of the month. To see this, note that

$$r_t^\tau = F_t/F_{t-\tau} - 1 \approx log\left(F_t/F_{t-\tau}\right)$$

Then,

$$r_t^\tau = r_t^1 + r_{t-1}^1 + \cdots + r_{t-\tau+1}^1$$

That is, the τ period return is the sum of τ one period returns. Let x_t^i denote the one-day return on the i^{th} indexed commodity and z_t^j denote the one-period return on the j^{th} indexed commodity, for $i = 1, ..., N_x$ and $j = 1, ..., N_z$. Suppose

$$x_t^i = \varepsilon_t^x - \varepsilon_{t-1}^x + y_t + \varepsilon_{i,t}^x - \varepsilon_{i,t-1}^x,$$

for all i. All variables to the right of the equality are zero-mean, normally distributed random variables, independent over time and with each other. The variable, ε_t^x, is a transitory shock to the price of commodity futures, which is common across indexed commodities. The variable, y_t, is a shock to the return on commodities, and has a random-walk effect on the price of commodities. It is common across commodities and is the source their common, long-run stochastic trend. The variable, $\varepsilon_{i,t}^x$, is a transitory shock to the price of a commodity and it is idiosyncratic across commodities. Note that

$$X_t^i \equiv \frac{1}{\tau}\left[x_t^i + x_{t-1}^i + \cdots + x_{t-\tau+1}^i\right] = \frac{1}{\tau}\left[\varepsilon_t^x - \varepsilon_{t-\tau}^x + \varepsilon_{i,t}^x - \varepsilon_{i,t-\tau}^x\right] + \frac{1}{\tau}\left[y_t + \cdots + y_{t-\tau}\right]$$

so that

$$var\left(X_t^i\right) = \frac{2}{\tau}\left(\sigma_x^2 + \sigma_{i,x}^2\right) + \tau\sigma_y^2.$$

The fraction of the variance in X_t^i due to the transitory shocks to the price of commodities is

$$\frac{\sigma_x^2 + \sigma_{i,x}^2}{\sigma_x^2 + \sigma_{i,x}^2 + \left(\tau^2/2\right)\sigma_y^2} \rightarrow_{\tau\rightarrow\infty} = 0.$$

In higher order returns, the impact of temporary shocks to futures prices disappears. Thus, in the case of indexed commodities, as we look at higher order returns a source of positive correlation is reduced (i.e., the common transitory shocks to futures prices) while sources of non-correlation (i.e., the idiosyncratic transitory shocks to futures prices) are reduced too.

Now consider the non-indexed commodities. We assume these have the same time series structure as the indexed commodities, with the exception that the common transitory shock to the commodity price level is not present. Thus,

$$z_t^j = y_t + \varepsilon_{j,t}^z - \varepsilon_{j,t-1}^z.$$

In the case of non-indexed commodities, we expect pairwise correlations between different returns to increase as we examine higher order returns. The reason is that now idiosyncratic shocks to futures prices are only sources of non-correlation between returns, sources that become less important in higher order returns.

We computed the following numerical example. We set $N_x = N_z = 2$, and we set the variance of y_t to unity. The variance of ε_t^x and $\varepsilon_{i,t}^x, i = 1, 2$ were all set to 10. The variance of $\varepsilon_{j,t}^z, j = 1, 2$ were each set to 3. We generated 20 million observations on x_t^i and z_t^j for $i, j = 1, 2$. Then,

$$corr\left(x_t^1, x_t^2\right) = 0.5, corr\left(z_t^1, z_t^2\right) = 0.05.$$

Evidently, the pairwise correlation between daily indexed returns, 0.5, is substantially higher than the pairwise correlation between non-indexed returns, 0.05. We then converted the data into 'monthly' averages computing averages over 20 observations and then sampling every 20th observation. We denote an averaged and sampled variable by a capitalized letter. Then,

$$corr\left(X_t^1, X_t^2\right) = 0.5, corr\left(Z_t^1, Z_t^2\right) = 0.5.$$

Note that the correlation between the indexed variables have not changed (to one significant digit). But, the correlation of the non-indexed variable has jumped to the point where both correlations are identical.

Figure 64: Impact of Temporal Aggregation on Pairwise Correlations Among Commmodity Futures

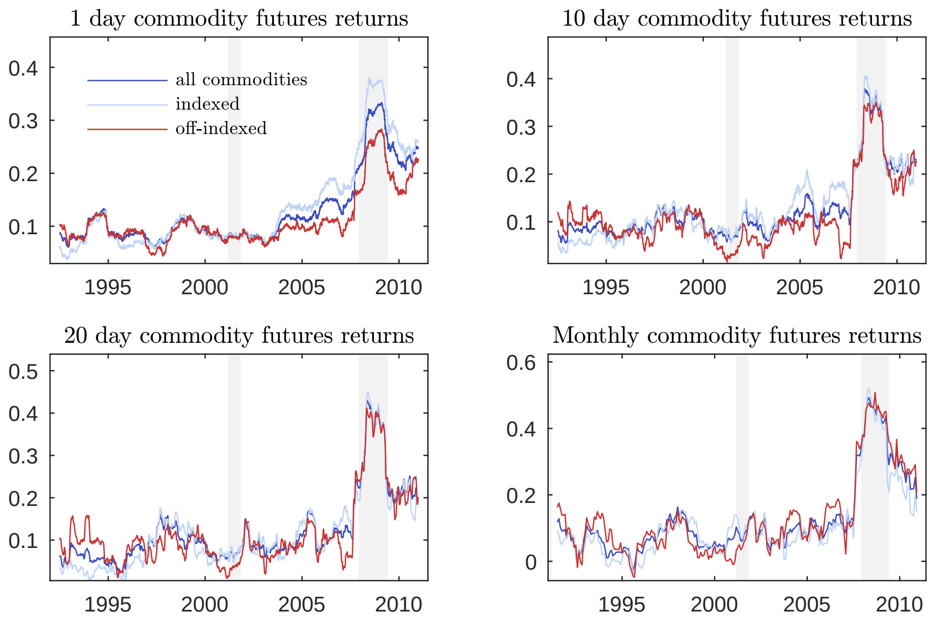
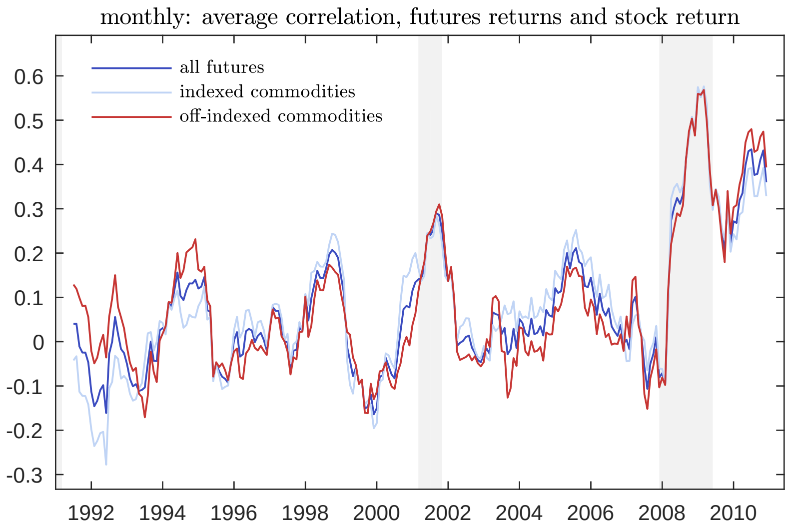
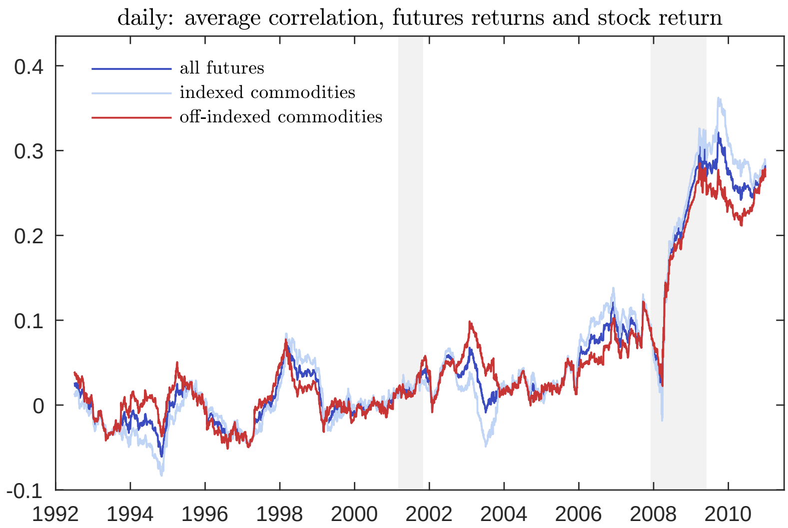


Figure 65: Correlation, Futures Returns and Stock Returns

(a) Monthly Data





(b) Daily Data

Figure 66: Average Pairwise Correlations, Different Subsets of Commodities

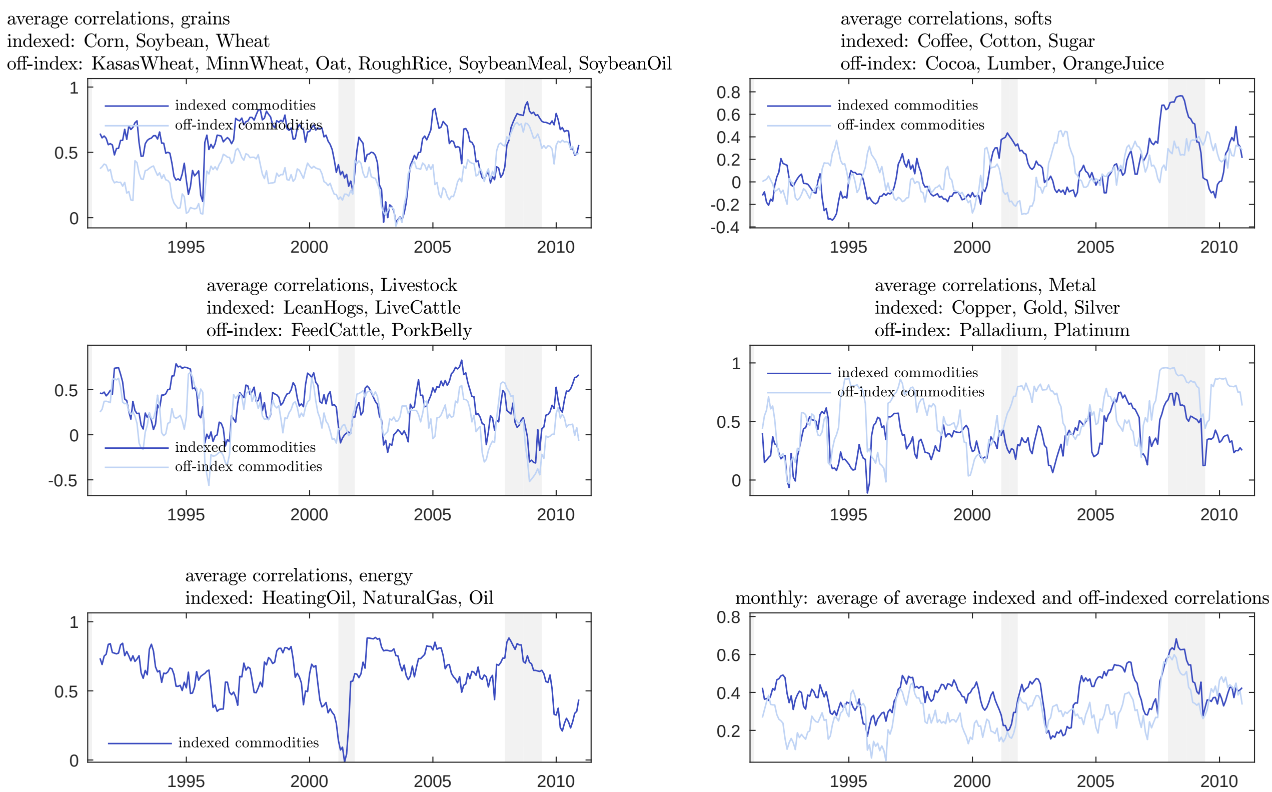


Figure 67: Standard Deviations Monthly Futures Returns

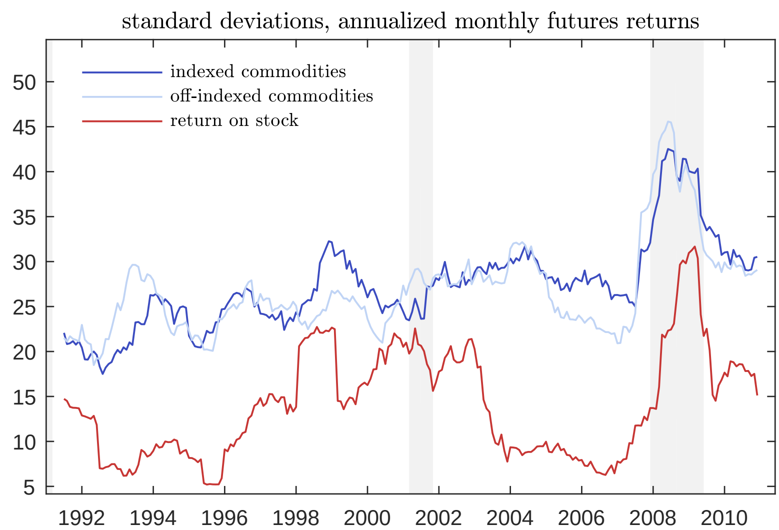


Figure 68

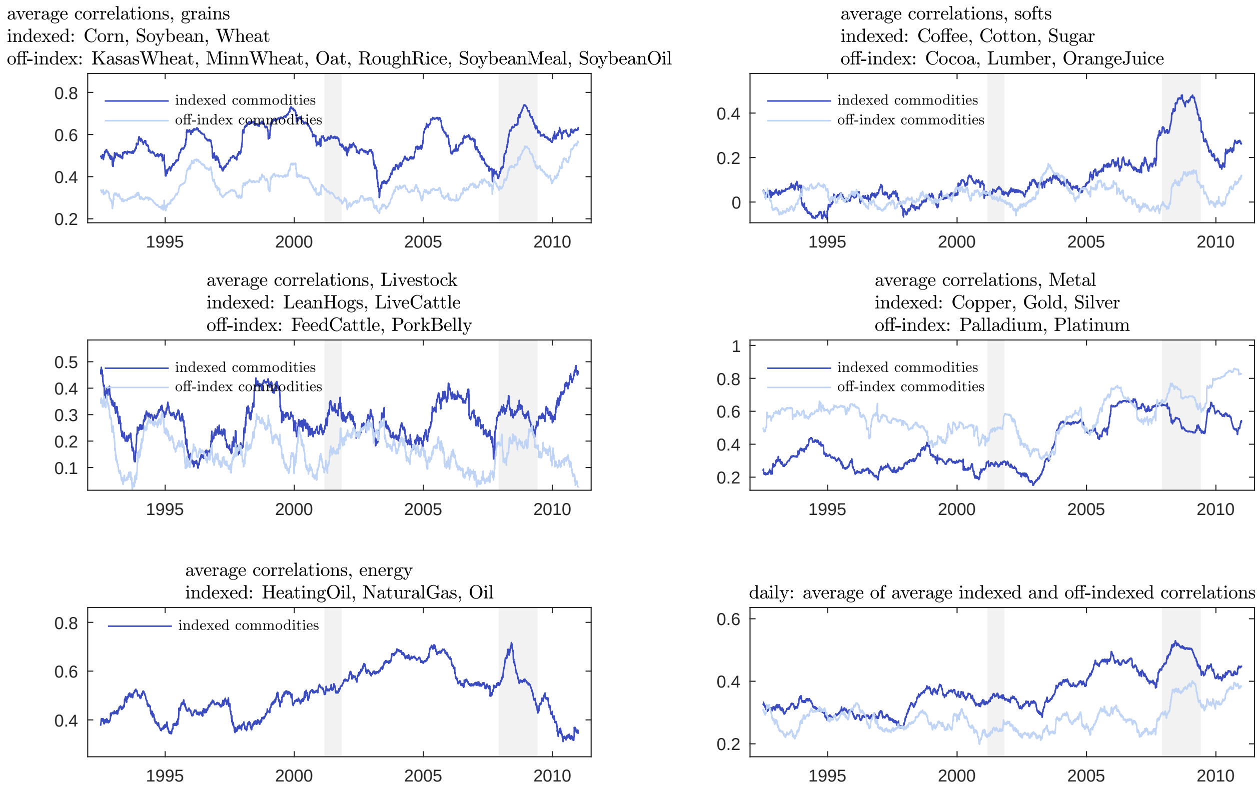
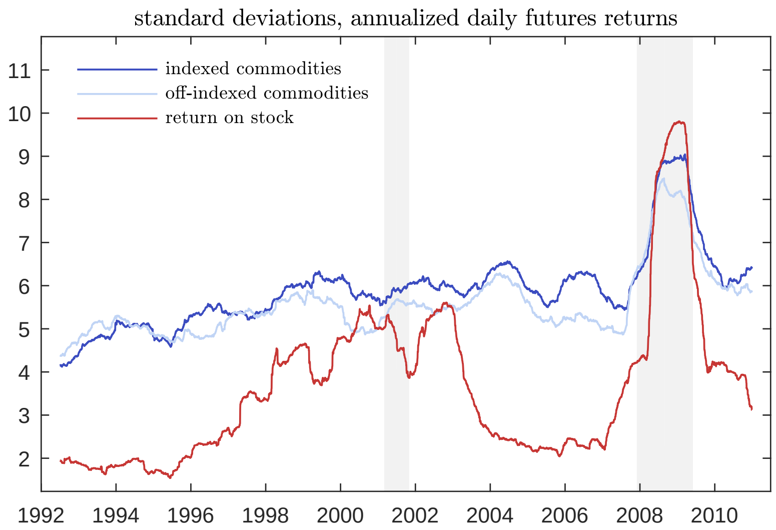


Figure 69



Part VIII

The Return on Commodity Futures

We begin by describing how we construct returns on futures contracts.

We now consider the statistical properties of commodity futures returns. For this, we use the daily data on 27 commodities analyzed in Tang and Xiong (2012).[6] We compute a time series of returns for each commodity as follows. Let $F(t,T)$ denote the price on day t of a contract for delivery at date T. Tang and Xiong kindly provided us with their daily observations on $F(t,T)$ for the 27 commodities. We compute the day t realized one day return on a futures contract by

$$\frac{F(t,T)}{F(t-1,T)}, \tag{4}$$

where T denotes the nearest settlement date from the perspective of time t. (See Appendix A for a discussion of (4).) We also use the daily data to construct and study the monthly returns on futures contracts. We compute the monthly realized return using the appropriate analog of (4). The realized monthly return for a particular month is the price in the first day of the next month on a contract for delivery at the nearest settlement date, divided by the price of the same contract on the first day of the month.

Our equity return data are equally weighted daily equity returns taken from the Center for Research on Securities Prices (CRSP) database. These returns were aggregated into monthly returns by Ferreira (2013), who kindly shared his data with us. We obtained daily and monthly returns on 3 month US government treasury bills ('Tbills') from the online data base, FRED, maintained by the Federal Reserve Bank of St. Louis.

In this appendix we describe a measure of the return on a futures contract, which can be compared with returns on other assets. Specifically, we require a trading strategy which involves an immediate outlay of money, with a payoff that depends on the future realization of futures prices. The measure of return that we describe here is standard we include it only for completeness. For additional discussion, see Jagannathan (1986) and Gorton and Rouwenhorst (2006). Let $F(t,T)$ denote the dollar price at date t of a commodity for delivery at time T. Let r_{t+1} denote the realized return at time $t+1$ on a one period nominally risk free bond (i.e., r_{t+1} is known at t). Consider the following trading strategy.

- Period t. Purchase a quantity, $F(t,T)$, of the risk free bond so $F(t,T)\,r_{t+1}$ is received in $t+1$. In addition, acquire r_{t+1} futures contracts at t. These contracts imply a commitment to purchase r_{t+1} commodities at price $F(t,T)$ in T.

- Period $t+1$. Under the terms of the futures contract, in period $t+1$ the trader receives $[F(t+1,T)-F(t,T)]\,r_{t+1}$ in cash from the exchange. That is, if the futures price goes up relative to $F(t,T)$ the trader receives a payment from the exchange for the difference. If the price goes down, the trader receives a negative payment, i.e., the trader must make a payment to the exchange. In principle, the trader could now walk away, being under no further obligation to the exchange. Adding the cash from the futures transaction to the cash, $F(t,T)\,r_{t+1}$, earned on the bond, the total cash that the trader has as a result of this strategy is $F(t+1,T)\,r_{t+1}$. The cost of the strategy was the purchase of $F(t,T)$ units of the risk free bond in t, so that the one period rate of return on the futures contract is
$$\frac{F(t+1,T)\,r_{t+1}}{F(t,T)}.$$

We can define a $j>1$ period return on a futures contract by continuing the above strategy for $j-1$ more periods.

- Period $t+1$. Use the cash earned in period $t+1$ to purchase the one-period safe bond, so that $F(t+1,T)\,r_{t+1}r_{t+2}$

[6] The 27 commodities are, in terms of Tang and Xiong's (self-explanatory) mnemonics: Cocoa, Coffee, Copper, Corn, Cotton, FeedCattle, Gold, HeatingOil, KasasWheat, LeanHogs, LiveCattle, Lumber, MinnWheat, NaturalGas, Oat, Oil, OrangeJuice, Palladium, Platinum, PorkBelly, RoughRice, Silver, Soybean, SoybeanMeal, SoybeanOil, Sugar, and Wheat. We obtain time series on intensity of financialization for each of these commodities by associating them with commodities in our CFTC dataset. In all but three cases, the commodities in the Tang-Xiong data base map into a single commodity in our CFTC data. In one of the exceptional cases we associate each of KasasWheat, MinnWheat and Wheat in the Tang-Xiong data based into what we call Wheat in the CFTC data. Also, FeedCattle and LiveCattle are both associated with what we call Cattle. Finally, we associate each of Soybean and SoybeanMeal with Soybeans in the CFTC data. Two of the commodities in our CFTC database do not appear in the Tang and Xiong data, Butter and Propane. These two commodities are left out of the analysis in this section of the paper. In Tang and Xiong (2012)'s Table 1 there are 28 commodities. One of them, RBOB gasoline, is not included in our analysis.

in cash is received in period $t+2$. Simultaneously, purchase $r_{t+1}r_{t+2}$ futures contracts on the commodity for delivery at T. Notice that no money is used or received in period $t+1$.

- Period $t+2$: In period $t+2$, $[F(t+2,T)-F(t+1,T)]\, r_{t+1}r_{t+2}$ units of cash is received (or, paid in case the expression is negative) from the exchange. In this way, the total cash received by this strategy in period $t+2$ is

  $$[F(t+2,T)-F(t+1,T)]\, r_{t+1}r_{t+2}+F(t+1,T)\, r_{t+1}r_{t+2}=F(t+2,T)\, r_{t+1}r_{t+2}.$$

  The two-period return on the future's contract is total cash received in period $t+2$ divided by the cash outlay in period t :

  $$\frac{F(t+2,T)\, r_{t+1}r_{t+2}}{F(t,T)}.$$

Another way to think about the return on a long contract is as follows. The investor buys one contract or $F(t,T)$ and puts up $F(t,T)/r_{t+1}$ in collateral. The exchange credits the investor with $F(t+1,T)-F(t,T)$ at the end of period t. Meanwhile, the collateral is worth $F(t,T)$ at the end of the period (because, it earns interest, r_{t+1}) and so the investor has a net amount at the end of the period that is equal to $F(t+1,T)-F(t,T)+F(t,T)=F(t+1,T)$. This is a 'fully collateralized' long contract because even in the worst case scenario, when $F(t+1,T)=0$, the exchange is completely covered. We can define the return on this operation as the payoff, $F(t+1,T)$, divided by the current outlay, $F(t,T)/r_{t+1}$, or

$$\frac{F(t+1,T)\, r_{t+1}}{F(t,T)}.$$

$H^L+S^L=H^s+S^s$, $H^s-H^L=S^L-S^s$

Now, consider a short contract. The investor buys one contract and receives $F(t,T)-F(t+1,T)$ from the exchange at the end of t.

In practice, we found that it makes no noticeable different in the future's calculation whether or not we included the risk free interest rate. This is because that interest rate fluctuates relatively little by comparison to the fluctuations in futures prices. This is why the interest rate does not appear in (4).

Now consider going short in a future's contract. The trader agrees, at time t, to sell 1 unit of the commodity at time T at price $F(t,T)$. In time $t+1$ the trader receives a payment, $F(t,T)-F(t+1,T)$, from the exchange. A difference with the short contract is that

Part IX

Sampling Theory for Statistics in Table 12

Here we describe the sampling theory used to interpret our mean and variance statistics in Table 12.

7.3. Comparison of Means

Table 12 compare the means of some of our data. This subsection reports how we compute the $p-$values reported in the table. Let $X_{i,t}$ denote some variable (say, the log first difference of the real commodity price) of interest, for commodity i at date t. We compare statistics computed using the 1990s and the 2000s samples. We denote the average, across commodities, of the time series mean by $\bar{X}$:

$$\bar{X}=\frac{1}{N}\sum_{i=1}^{N}\bar{X}_i,$$

where N denotes the number of commodities and

$$\bar{X}_i = \frac{1}{T}\sum_{t=1}^{T} X_{i,t}.$$

Here, T denotes the number of observations in the sample. In our application, T is relatively small while N is relatively large. To motivate the distribution theory that we use to interpret $\bar{X}$, we make use of an insight of Ibragimov and Müller, (2010, 2011).[7] In particular, notice that $\bar{X}$ can also be written

$$\begin{aligned} \bar{X} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} X_{i,t} \\ &= \frac{1}{T}\sum_{t=1}^{T}\overbrace{\left(\frac{1}{N}\sum_{i=1}^{N} X_{i,t}\right)}^{\equiv \bar{X}_t} \\ &= \frac{1}{T}\sum_{t=1}^{T}\bar{X}_t. \end{aligned}$$

If N is sufficiently large and the $X_{i,t}$'s are not too correlated across i, then we can appeal to a Central Limit Theorem, to suppose that $\bar{X}_t$ is, approximately, a realization from a Normal distribution:

$$\bar{X}_t \sim N\left(\mu_t, \sigma_t^2\right).$$

Suppose $\sigma_t^2 = \sigma^2$, for now and we adopt the maintained hypothesis that $\bar{X}_t$ is independently distributed across t. Let

$$\mu = \frac{1}{T}\sum_{t=1}^{T}\mu_t.$$

Then,

$$\begin{aligned} E\left[\bar{X} - \mu\right]^2 &= E\left[\frac{1}{T}\sum_{t=1}^{T}\left(\bar{X}_t - \mu\right)\right]^2 \\ &= E\frac{1}{T^2}\sum_{t=1}^{T}\left(\bar{X}_t - \mu\right)^2 \\ &= \frac{\sigma^2}{T}, \end{aligned}$$

by independence. An unbiased estimator of σ^2 is given by:

$$s^2 = \frac{1}{T-1}\sum_{t=1}^{T}\left(\bar{X}_t - \mu\right)^2.$$

Then,

$$\frac{\bar{X} - \mu}{\sqrt{s^2/T}} \sim t_{T-1},$$

where t_{T-1} denotes a t distribution with $T-1$ degrees of freedom.

We actually wish to compare sample means based on our two data sets. Let $\bar{X}$ and $\bar{Y}$ denote the sample

[7] We thank Mark Watson for drawing our attention to this.

mean based on the first and second datasets, respectively. (The two means, $\bar{X}$ and $\bar{Y}$, are averages over different dates, but to keep the notation simple, our notation does not reflect that.) An implication of our assumption that observations are independent over time is that $\bar{X}$ and $\bar{Y}$ are independently distributed. The variance of the difference between the two means, under the null hypothesis (and, maintained hypothesis of independence), is:

$$\begin{aligned} E\left[\bar{X}-\bar{Y}\right]^2 &= E\left[\frac{1}{T_1}\sum_{t=1}^{T_1}\bar{X}_t - \frac{1}{T_2}\sum_{t=1}^{T_2}\bar{Y}_t\right]^2. \\ &= E\left[\frac{1}{T_1}\sum_{t=1}^{T_1}\left(\bar{X}_t-\mu\right) - \frac{1}{T_2}\sum_{t=1}^{T_2}\left(\bar{Y}_t-\mu\right)\right]^2 \\ &= \frac{1}{T_1^2}E\sum_{t=1}^{T_1}\left(\bar{X}_t-\mu\right)^2 + \frac{1}{T_2^2}E\sum_{t=1}^{T_2}\left(\bar{Y}_t-\mu\right)^2 \\ &= \frac{\sigma_1^2}{T_1}+\frac{\sigma_2^2}{T_2}, \end{aligned}$$

where T_1 and T_2 are the lengths of the first and second dataset, respectively. Also, $\bar{X}_t$ and $\bar{Y}_t$ denote the period t cross section averages of $X_{i,t}$ and $Y_{i,t}$. Estimators of σ_i^2, $i=1,2$ are given by:

$$\begin{aligned} s_1^2 &= \frac{1}{T_1-1}\sum_{t=1}^{T_1}\left(\bar{X}_t-\bar{X}\right)^2 \\ s_2^2 &= \frac{1}{T_2-1}\sum_{t=1}^{T_2}\left(\bar{Y}_t-\bar{Y}\right)^2. \end{aligned}$$

Let

$$s_{\bar{X}-\bar{Y}} = \sqrt{\frac{s_1^2}{T_1}+\frac{s_2^2}{T_2}} \tag{5}$$

Then, we treat the following object as a realization from a t distribution with $r=T_1+T_2-2$ degrees of freedom:

$$\frac{\bar{X}-\bar{Y}}{s_{\bar{X}-\bar{Y}}}\tilde{\ }t_r. \tag{6}$$

This approach assumes variances in the two samples are the same. We also considered an approach that is commonly used when the variances in the two samples are unequal. In this we used the sampling theory, (6), with degrees of freedom:

$$r=\frac{\left(\frac{s_X^2}{n}+\frac{s_Y^2}{m}\right)^2}{\frac{(s_X^2/n)^2}{n-1}+\frac{(s_Y^2/m)^2}{m-1}}.$$

In practice, we found that the two ways of computing degrees of freedom produced roughly the same results.[8]

As a check on the analysis, we also evaluate our statistic, (6), using a bootstrap procedure. In particular, for each i, we draw random samples of length T_1 from the data, $X_{i,t}$, $t=1,...,T_1$ and random samples of length T_2 from the data, $Y_{i,t}$, $t=1,...,T_2$. To ensure that the null hypothesis of equal means is true in the data generating mechanism, we first adjust the data from which we draw, so that they have the same mean. Subject to this adjustment, the samples are drawn, with replacement, with equal probability being assigned to each of the empirical observations. In this way, we compute 10,000 data sets and obtain 10,000 values of the test statistic in (6). The reported $p-$value is fraction of simulated test statistics which take on values that exceed its empirical value.

[8]The degree of freedom adjustment is suggested in, for example, https://onlinecourses.science.psu.edu/stat414/node/275

7.4. Comparison of Variances

We compute the average variance, across commodities:

$$S^X = \frac{1}{N}\sum_{i=1}^{N} s_i^2,$$

where

$$s_i^2 = \frac{1}{T-1}\sum_{t=1}^{T}\left[X_{i,t} - \bar{X}_i\right]^2,\ \bar{X}_i = \frac{1}{N}\sum_{t=1}^{N} X_{i,t}.$$

Then,

$$\begin{aligned} S^X &= \frac{1}{N(T-1)}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[X_{i,t} - \bar{X}_i\right]^2 \\ &= \frac{1}{T-1}\sum_{t=1}^{T}\overbrace{\left(\frac{1}{N}\sum_{i=1}^{N}\left[X_{i,t} - \bar{X}_i\right]^2\right)}^{S_t^X}, \end{aligned}$$

where S_t^X is the variance, in the cross section of commodities at date t, of $X_{i,t}$. These cross sectional variances are centered around the sample mean, $\bar{X}_i$, for each commodity. We take an approach analogous to the one in the previous subsection, by appealing to the Central Limit Theorem to justify the notion that S_t^X is a realization from a Normally distributed random variable:

$$S_t^X \sim N\left(\sigma_t^2, \gamma_t\right).$$

where σ_t^2 denotes the mean of S_t^X and γ_t its variance.

Let S^X and S^Y denote the average variances in first and second samples and consider the following statistic for testing equality of the variances in the two samples by computing

$$\frac{S^X - S^Y}{\sqrt{\frac{s_1^2}{T_1} + \frac{s_2^2}{T_2}}}, \tag{7}$$

where

$$s_1^2 = \frac{1}{T_1 - 1}\sum_{t=1}^{T_1}\left[S_t^X - S^X\right]^2,\ S^X = \frac{1}{T_1}\sum_{t=1}^{T_1} S_t^X.$$

We define s_2^2 in the same way, for the second data set. Under the null hypothesis that the variances in the two samples are the same, we interpret (7) as a realization of a t distribution with $T_1 + T_2 - 2$ degrees of freedom.

We also compute a bootstrap version of the above statistic. We implement the same bootstrap procedure described in the previous section, with one obvious adjustment. The null hypothesis is now that the two variances being compared are the same. As a result, we must adjust the data from which the random samples are drawn, to ensure that the variances are the same. We do this by adjusting one of the two series. Thus, let $x_1, ..., x_T$ denote the series that we adjust and suppose that they have a sample mean of $\bar{x}$ and a sample variance σ^2. Then, we replace x_t with $\tilde{x}_t$, where $\tilde{x}_t$ has mean $\bar{x}$ and variance, $\widetilde{\sigma}^2$, by adopting the following transformation:

$$\tilde{x}_t = (x_t - \bar{x})\frac{\tilde{\sigma}^2}{\sigma^2} + \bar{x}.$$

There are of course many ways to impose equality of variances. We can adjust one series or the other, or both. We found that the results did not change significantly when we considered the two extreme cases. Note that our procedure does not impose equality of means.