San Jose State University

From the SelectedWorks of Stoyu I. Ivanov

2011

Cross-Sectional Analysis of Index and Commodity Markets Price Discovery.

Stoyu I. Ivanov, *San Jose State University*

Available at: https://works.bepress.com/stoyu-ivanov/28/

Cross-Sectional Analysis of Index and Commodity Markets Price Discovery

Stoyu I. Ivanov*

This study examines the determinants of relative price discovery between the futures and cash prices in 30 index and commodity markets based on the Gonzalo and Granger (1995) permanent-transitory decomposition methodology. Twenty-eight indexes and commodities have proportional futures market information shares greater than 60%. Two commodities are the only exception: Feeders Cattle and Wheat-Minneapolis have price discovery occurring predominantly in the cash markets with information shares of their futures contracts of 33% and 40%, respectively. The research documents a significant cross-sectional variability of the information shares across the 30 indexes and commodities and finds that the information shares of the futures contract are lower when trading volume of the futures contract is lower, when the contract is on an energy commodity or agricultural commodity, and the commodity or index has a traded ETF.

I. Introduction

On January 13, 2010, the U.S. Department of Agriculture announced that the grains harvest would be better than anticipated (Flood, 2010) which naturally caused significant trading by market participants who were attempting to profit from the new information. Accordingly, significant price adjustments in grains prices occurred. The press was filled at that time with discussions of how fast futures prices dropped and adjusted to the new information. The investing public usually expects that futures prices will adjust quickly, not cash prices. Cash prices in general are expected to adjust with a lag. This is not surprising considering that it is agreed that the well-informed traders and the larger trading volumes are in the futures market, not the cash market. What is surprising is the lack of conclusive evidence with regard to the dominating market, futures or cash market, in the price discovery of the major index and commodities markets. Also, there is no agreement on the factors driving price discovery across these markets. Price discovery refers to the relative contribution in terms of pricing guidance that one market has over another market. A market with the highest price discovery is most likely to trade the fastest, given a common information shock, and thus provide the highest level of pricing guidance to market entities that trade slower and thus get a high proportion of their information from leading markets.

The extant literature in the field has not identified which market provides the most information – the cash market or the futures market. Also, the literature has focused only on a

The author would like to thank the journal editor, Jong Rhim, and two anonymous referees for their insightful comments and suggestions.

***Stoyu I. Ivanov** is an Assistant Professor of Finance at the College of Business, San Jose State University, San Jose, CA, USA.

handful of commodities, and may be because of that, the literature has not identified the major factors affecting price discovery. This study attempts to fill this void in the literature by examining the price discovery in 30 index and commodity markets in light of the recent developments in the price discovery methodology developments. It attempts to identify whether the futures market consistently dominates the cash market in terms of price discovery and whether there is considerable variation in price discovery across the different indexes and commodities. The study also attempts to identify what are the factors affecting this cross-sectional variability. To The appears to be the first study to examine the major commodity markets in a single uniform analysis and incorporate new price discovery identification techniques.

Most indexes and commodities have proportional futures market information shares greater than 60% with the exception of the Feeders Cattle and Wheat-Minneapolis futures contracts. Feeders Cattle and Wheat-Minneapolis futures contracts are led by the cash markets with information shares of the futures contracts of 33% and 40% respectively. A significant cross-sectional variability of the information shares exists across the thirty indexes and commodities. The futures contract information share is lower when trading volume of the futures contract is lower, when the contract is on an energy or agricultural commodity, and when the commodity has a traded ETF.

II. Literature Review

The major study in the field is by Garbade and Silber (1983) and was conducted over two decades ago with focuses on wheat, corn, oats, orange juice, copper, gold, and silver commodities only. They find that the futures commodities market leads the cash commodities market in price discovery and that 75% of the information for wheat, corn, and orange juice is provided by the futures market. In contrast, Quan (1992) does not find that the price formation occurs in the futures market. Quan examines only crude oil commodity. Schwarz and Szakmary (1994) re-examine and criticize Quan's (1992) study and find that the futures market dominates the cash market by looking again at the crude oil commodities market. Schroeder and Goodwin (1991) examine the live hogs market and find that 65% of the price discovery appears in the futures market.

Subrahmanyam (1991) shows that the cost of informational asymmetry is smaller for stock index futures contracts due to the diversification effect across individual stocks. Similarly, interest rate futures are more influenced by macroeconomic public information than by private information. Therefore, we should see variability in information shares of futures contracts depending on whether the underlying commodity is a financial index. Williams and Tse (2009) examine the FTSE-100 Futures Contract on open-outcry and electronic platforms. They find no efficiency differences between the two platforms. Franke and Hess (2000) and Theissen (2002) examine price discovery in floor-traded and electronic platforms of the Bund Future contract. They find conflicting evidence on the dominating market for the price discovery. Franke and Hess (2000) suggest that the price leadership is provided by the electronic market during low information intensity but that the floor market provides the price leadership during high information intensity. Similarly Theissen (2002) finds that electronic markets do not seem to dominate floor markets when transaction prices are used in the analysis. The variability in information shares relates to the electronic or floor traded nature of the futures contract.

Mattos and Garcia (2004) focus only on agricultural futures contracts in their study of price discovery. They focus only on agricultural contracts because of their special characteristics and because of their lower liquidity relative to financial assets. Therefore, agricultural futures contracts should have lower price discovery. Lien and Tse (2002) study the feeder cattle futures contract and the switch from physical delivery to cash settlement of the contract in September 1986. The authors argue that there are several reasons why investors might prefer cash versus physical delivery settlement, such as lower delivery costs, less effective cornering and squeezing of the market, and improvement in convergence of spot and futures prices at maturity. Thus, the settlement of the futures contract might be related to the price discovery of the futures contract. Lien and Tse (2002) find that the switch to cash settlement in the feeders cattle market is beneficial to the futures market; however, they make no prediction for the price discovery.

Table 1. Contract and Commodity Descriptions

Commodity	Ticker	Contract/Series Start Date	Trading Platform	Settlement	Exchange
S.BEAN OIL	BO	06/02/1969	OO	PD	CBOT
CORN	C	06/02/1969	OO	PD	CBOT
CRUDE BRENT	CB	08/17/1990	both	Cash	ICE
COCOA	CC	01/02/1986	electronic	PD	ICE
CRUDE Sweet	CL	03/26/2002	both	PD	CME
COTTON	CT	03/26/2002	electronic	PD	NYCE
DOW 30	DJ30	03/26/2002	OO	Cash	CBOT
FEEDERS	FC	07/16/1979	OO	Cash	CME
GOLD	GC	03/26/2002	both	PD	COMEX
COPPER	HG	03/26/2002	OO	PD	COMEX
HEATING OIL	HO	03/26/2002	both	PD	NYMEX
GAS UNLD	HU	03/26/2002	both	PD	NYMEX
COFFEE	KC	03/26/2002	OO	PD	CSC
WHEAT, KANSAS CITY	KW	06/02/1969	OO	PD	KCBOT
CATTLE	LC	06/02/1969	both	PD	CME
HOGS	LH	06/02/1969	both	Cash	CME
WHEAT, MINNEAPOLIS	MW	02/01/1983	OO	PD	MGE
NATURAL GAS	NG	09/14/2000	both	PD	NYMEX
NASDAQ 100 Index	NS	06/28/1996	both	Cash	CME
OATS	O	06/02/1969	OO	PD	CBOT
PALLADIUM	PA	06/09/1987	both	PD	NYMEX
P. BELLIES	PB	06/02/1969	OO	PD	CME
PLATINUM	PL	06/09/1987	OO	PD	NYMEX
RUSSELL 2000	RT	09/10/1987	both	Cash	CME
SOYBEANS	S	06/02/1969	OO	PD	CBOT
SUGAR #11	SB	01/02/1980	both	PD	NYMEX
SILVER	SI	06/02/1969	OO	PD	COMEX
S.BEAN MEAL	SM	06/02/1969	OO	PD	CBOT
S&P 500	SP	01/04/1960	OO	Cash	CME
WHEAT, CHICAGO	W	04/04/1975	OO	PD	CBOT

OO is open outcry trading platform; PD is physical delivery of the contractual commodity. ICE is International Exchange; NYCE is New York Cotton Exchange; CBOT is the Chicago Board of Trade; CME is the Chicago Mercantile Exchange; COMEX is the New York Mercantile Exchange and Commodity Exchange; NYMEX is the New York Mercantile Exchange; MGE is the Minneapolis Grain Exchange; KCBOT is the Kansas City Board of Trade; and CSC is the Coffee, Sugar and Cocoa Exchange.

This is the first study to examine the major commodity markets in a single uniform analysis and incorporate new price discovery identification technique. The question posited in this study is with regard to the different information shares of the futures price relative to the cash price among the 30 different indexes and commodities studied. Perhaps the information shares might be different because of the different nature of the futures contract.

III. Data and Methodology

This study uses daily data similar to those of Garbade and Silber (1983) for the period ending on January 29, 2010; the start dates are different for the different commodities. The cash and futures data are from Glancedata.com. The analysis that follows uses the nearby contract prices. Table 1 provides detailed description of each commodity and futures contract. The table is arranged in alphabetical order of the ticker symbol of the index or commodity. The trading platform category is related to whether the contract trades electronically or on a trading floor. Most commodities have contracts which trade on both platforms.

For example, the Corn Futures contract with ticker symbol C is open outcry/ trading floor traded and is an E-mini contract. It is designed as one fifth of the regular size futures contract.[1] The contract size of the futures contract is with a physical delivery of 5,000 bushels with deliverable grades of the commodity: #2 Yellow corn quality at the contracted price, #1 Yellow grain quality at a 1.5 cent/bushel premium, and #3 Yellow grain quality at a 1.5 cent/bushel discount of the contracted price with the delivery months January, March, May, July, August, September, and November.

Table 2 provides summary statistics of each of the 30 futures contracts examined in this study and the respective cash market index or commodity. The table provides summary information on the log cash price, log futures price, futures contract open interest, and futures contract traded volume. The range of values of the number of observations for the analyzed series are from 1,753 observations or approximately seven years of daily data for the Japanese Yen to 10,238 observations or approximately 40 years of daily data for corn. All data end on January 29, 2010.

There are two major competing information share decomposition methodologies, the Hasbrouck (1995) information share and Gonzalo and Granger (1995) permanent-transitory decomposition methodology. A recent study by Mattos and Garcia (2004) examines the futures cash price discovery in Brazilian agricultural markets by utilizing Hasbrouck's (1995) information share methodology. However, Figuerola-Ferretti and Gonzalo (2010) show that Hasbrouck's (1995) information share methodology produces an upper and lower bound for the information shares. They suggest that the Hasbrouck (1995) information share methodology's problems are diminished when high frequency data are used. Figuerola-Ferretti and Gonzalo (2010) suggest that for low frequency data such as the daily data in this study, the information share methodology developed by Gonzalo and Granger (1995) is more appropriate.

Additionally, Baillie, Booth, Tse and Zabotina (2002) find that the Hasbrouck (1995) and Gonzalo and Granger (1995) models will differ any time there is a high level of cross-

[1] Detailed description of the corn futures contract is available on the Chicago Mercantile Exchange's website: http://www.cmegroup.com/trading/commodities/grain-and-oilseed/corn_contract_specifications.html.

Table 2. Descriptive Statistics

		Cash Log Price		Futures Log Price		Open Interest (Futures)		Volume (Futures)	
	N	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
S.BEAN OIL	10216	3.0798	0.3421	3.0799	0.3405	17253	24254	4900	4874
CORN	10238	5.4698	0.306	5.5050	0.3045	83611	116021	17063	19425
CRUDE BRENT	4863	3.3481	0.5908	3.3495	0.596	57978	36856	32631	30878
COCOA	5980	7.4249	0.2857	7.2564	0.3121	16542	19478	3019	3387
CRUDE Sweet	1926	3.9682	0.4321	3.9685	0.4326	182667	100135	157274	94305
COTTON	1962	3.9365	0.1776	3.9982	0.1726	38200	42262	6196	7001
DOW 30	1766	9.2589	0.1552	9.2594	0.1568	36813	11101	8382	7694
FEEDERS	5532	4.3723	0.1578	4.3063	0.1237	3370	2286	652	521
GOLD	1952	6.3138	0.3916	6.3147	0.3913	70744	90982	34329	49272
COPPER	1931	5.182	0.5821	5.1789	0.5811	28074	26732	7048	6849
HEATING OIL	1933	0.3760	0.4505	0.3825	0.4529	38827	21162	25968	9285
GAS UNLD	1933	0.3859	0.4139	0.3898	0.401	38949	21433	27103	10255
COFFEE	1922	4.4745	0.4040	4.5477	0.3207	32058	31956	6593	6389
WHEAT, KANSAS CITY	8518	5.9553	0.2699	5.9146	0.2665	15106	17025	2974	2949
CATTLE	5894	4.1561	0.1789	4.1705	0.174	19274	12663	6212	3723
HOGS	7950	3.7142	0.2903	3.7888	0.2867	7017	6954	2246	1956
WHEAT, MINNEAPOLIS	6802	6.0937	0.2851	5.9636	0.2912	6665	6617	1461	1275
NATURAL GAS	2311	1.7255	0.4030	1.7477	0.3934	61157	33836	48037	26674
NASDAQ 100 Index	3419	7.3480	0.3723	7.3522	0.3730	40900	24169	8720	6661
OATS	8727	5.1627	0.2625	5.0757	0.2690	3251	3017	533	559
PALLADIUM	5670	5.3399	0.5750	5.3325	0.5755	3545	4383	394	600
P. BELLIES	10209	4.0230	0.3661	4.0891	0.3318	3607	3098	1832	2017
PLATINUM	3589	6.1693	0.3974	6.1684	0.3982	8173	6555	1833	1853
RUSSELL 2000	3936	6.1067	0.3624	6.1098	0.3622	17586	13085	1296	1415
SOYBEANS	10216	6.3802	0.3050	6.3995	0.3085	33097	45447	12856	14187
SUGAR #11	7523	2.2999	0.4266	2.2579	0.4228	76307	92773	13715	15983
SILVER	9617	6.3937	0.5177	6.3965	0.5147	21942	24607	7571	9500
S.BEAN MEAL	8210	5.0879	0.3111	5.1060	0.3069	11016	12088	4230	4130
S&P 500	6797	6.3228	0.7608	6.3262	0.7600	262596	204898	53847	40622
WHEAT, CHICAGO	8769	5.8419	0.2450	5.8787	0.2624	34451	48578	7172	8554

market/asset return correlation (as would be expected when dealing with futures vs. cash returns on the same contract) and over lower data frequencies given that one market is less distinguished from the other market. Therefore, this study uses the Gonzalo and Granger (1995) permanent-transitory decomposition methodology to identify the information shares of the futures and cash markets.

The permanent-transitory component decomposition here employs the following vector error correction model (VECM) specification:

$$\Delta p_t = \alpha\beta' p_{t-1} + \sum_{i=1}^{k} \Gamma_i \Delta p_{t-i} + u_t , \tag{1}$$

where p_t is the 2x1 vector of log futures, and cash prices and u_t is the error term. In the VECM estimation, I use log prices because the difference between the log prices results in rates of return. Additionally, the VECM lag length is set at five to represent up to one week of trading activity. This study follows the Engle-Granger cointegration methodology as discussed in Enders (2004). I test for stationarity in the commodities futures and cash prices first are tested for stationarity. The presence of natural association between the futures and cash commodity prices suggests cointegration based on the Granger representation theorem as discussed in Engle and Granger (1987). The Granger representation theorem states that if two random series are integrated of order one, there is a possible natural combination of them which will not be integrated, hence the estimation of the VECM.

Gonzalo and Granger (1995) methodology identifies the permanent component of the decomposition based on the estimated VECM as W_t

$$W_t = \alpha_{\perp}' p_t \tag{2}$$

as an observable linear combination of the original futures and cash prices and a cointegrating relation Z_t:

$$Z_t = \beta' p_t \tag{3}$$

Based on these observable permanent component and cointegration relation obtained from the VECM, the permanent-transitory decomposition of the vector of futures and cash prices may be derived:

$$p_t = A_1 W_t + A_2 Z_t \tag{4}$$

where

$$A_1 = \beta_{\perp}(\alpha_{\perp}'\beta_{\perp})^{-1} \tag{5}$$

and

$$A_2 = \alpha(\beta'\alpha)^{-1} \tag{6}$$

with the information shares based on this method computed as the proportions of the futures and spot prices in the price discovery function $\alpha_{\perp}$.

Based on this methodology, computing the respective information shares of the cash and futures markets will help us determine which market provides the price leadership. Based on these computed ranges of information shares for the 30 commodities and indexes, I attempt to identify the factors affecting the cross sectional variation of these information shares. The variation in information shares may be due to the different nature of the futures contract.

Subrahmanyam (1991), Williams and Tse (2009), Franke and Hess (2000), and Theissen (2002) model and examine the behavior of futures contracts based on their different nature. Subrahmanyam (1991) shows that stock index futures contracts and interest rate futures contracts behave differently than futures contracts with a non-financial underlying asset. Williams and Tse (2009), Franke and Hess (2000), and Theissen (2002) examine price discovery in floor traded and electronically traded contracts. Tse, Xiang, and Fung (2006) suggest that price discovery differences may arise due to the impact of behavior and participation of informed and uninformed traders. Specifically, electronic markets dominate the price discovery of open outcry markets given that informed traders prefer electronic markets whereas uninformed traders prefer open outcry markets. The reason is that electronic markets offer informed traders a greater level of order flow anonymity, liquidity, and immediacy. Also, I partition the trading activity variables (volume and open interest) into their expected and unexpected portions to represent informed and uninformed traders trading activity as addressed in Bessembinder and Seguin (1992, 1993).

Therefore, after I compute the information shares for each commodity in the cash and futures markets is computed, we can attempt to determine the driving forces behind the differences in information shares in these markets. Based on the theoretical and empirical studies discussed above I develop the cross sectional analysis model of the information shares of each commodity futures contract. I relate the cross sectional variability of the information share of each commodity futures contract to some observable commodity and futures contract characteristics. The cross sectional regression analysis is based on the following model:

$$\begin{aligned} PT_i &= \alpha_0 + \beta_1(LVOL) + \beta_2(LOI) + \beta_3(RATIO) + \beta_4(CASH) + \beta_5(AGR) \\ &+ \beta_6(ENERGY) + \beta_7(OO) + \beta_8(ETF) + \beta_9(SD) + \beta_{10}(FINANCIAL) \\ &+ \beta_{11}(SDOI) + \beta_{12}(SDVOL) + \varepsilon_i \end{aligned} \tag{7}$$

where PT is the information share of the futures contract; LVOL is the log of futures volume; LOI is futures open interest; RATIO is the ratio of the average trading volume to average open interest; CASH is a dummy variable of one if the futures contract has cash settlement and zero for physical delivery; AGR is a dummy variable of one if the futures contract is agricultural and zero otherwise; ENERGY is a dummy variable of one if energy future contract, zero otherwise; OO is a dummy variable of one if the futures contract is open outcry traded, and zero otherwise; ETF is a dummy variable of one if an ETF exists within the examined period, and zero otherwise; SD is the standard deviation of the futures contract price over the examined period; FINANCIAL is a dummy variable of one if index future contract, zero otherwise; SDOI is the standard deviation of the futures contract open interest over the examined periodl and SDVOL is the standard deviation of the futures contract traded volume over the examined period.

Table 3. Augmented Dickey Fuller Test Results and Johansen Cointegration Test Results (Log Prices)

	Unit Root Test		Cointegration Test			
	Cash	Futures				
	p-value	p-value	H0: R=	H1: R>	Trace	Critical Value
S.BEAN OIL	0.7347	0.7434	1	1	0.1349	4.14
CORN	0.7130	0.7144	1	1	0.1999	4.14
CRUDE BRENT	0.7348	0.7319	1	1	0.2443	4.14
COCOA	0.6921	0.6898	1	1	0.0731	4.14
CRUDE Sweet	0.7327	0.7327	1	1	0.4305	4.14
COTTON	0.7147	0.7094	1	1	0.3902	4.14
DOW 30	0.6758	0.6757	1	1	0.3828	4.14
FEEDERS	0.6799	0.6876	1	1	0.0223	4.14
GOLD	0.7307	0.7303	1	1	4.4706	4.14
COPPER	0.7423	0.7422	1	1	2.0608	4.14
HEATING OIL	0.4210	0.4406	1	1	0.4369	4.14
GAS UNLD	0.2390	0.3363	1	1	0.7742	4.14
COFFEE	0.7542	0.7217	1	1	0.3154	4.14
WHEAT, KANSAS CITY	0.6865	0.6871	1	1	0.0044	4.14
CATTLE	0.7217	0.7205	1	1	0.5657	4.14
HOGS	0.7029	0.7220	1	1	0.1187	4.14
WHEAT, MINNEAPOLIS	0.7012	0.6882	1	1	0.0874	4.14
NATURAL GAS	0.4857	0.5708	1	1	0.2612	4.14
NASDQ 100 Index	0.7083	0.7079	1	1	0.5747	4.14
OATS	0.6742	0.6677	1	1	0.0135	4.14
PALLADIUM	0.7217	0.7208	1	1	0.3704	4.14
P. BELLIES	0.6146	0.6763	1	1	0.0014	4.14
PLATINUM	0.7151	0.7138	1	1	0.2215	4.14
RUSSELL 2000	0.7250	0.7244	1	1	1.9677	4.14
SOYBEANS	0.7165	0.7180	1	1	0.2621	4.14
SUGAR #11	0.6894	0.6006	1	1	0.0002	4.14
SILVER	0.7551	0.7531	1	1	0.8150	4.14
S.BEAN MEAL	0.6927	0.6971	1	1	0.0294	4.14
S&P 500	0.7469	0.7461	1	1	3.6878	4.14
WHEAT, CHICAGO	0.6732	0.6840	1	1	0.0001	4.14

The expected and unexpected trading activity of Bessembinder and Seguin (1992, 1993) is reflected by modifying the model described in equation (7) to account for the fact that the sum of expected and unexpected volume equals actual volume and that the sum of expected and unexpected open interest equals actual open interest. Therefore, this becomes the estimate of the alternative model specification:

$$\begin{aligned} PT_i = \alpha_0 &+ \beta_1(LFVOL) + \beta_2(LFOI) + \beta_3(RESVOL) + \beta_4(RESOI) + \beta_5(RATIOF) \\ &+ \beta_6(RATIORES) + \beta_7(CASH) + \beta_8(AGR) + \beta_9(ENERGY) + \beta_{10}(OO) + \beta_{11}(ETF) \\ &+ \beta_{12}(SD) + \beta_{13}(FINANCIAL) + \beta_{14}(SDOI) + \beta_{15}(SDVOL) + \varepsilon_i \end{aligned} \tag{8}$$

where all parameters are as defined in equation (7) with the exception of Lfvol - the logarithm of expected volume, Lfoi - the logarithm of expected open interest, Resvol - the unexpected volume, Resoi - the unexpected open interest. Ratiof is the ratio of the average expected trading volume to average expected open interest, and Ratiores is the ratio of the average unexpected trading volume to average unexpected open interest.

Table 4. Information Shares

	Proportional Contributions	
	Futures	Cash
S.BEAN OIL	0.9271	0.0729
CORN	0.7861	0.2139
CRUDE BRENT	0.9982	0.0018
COCOA	0.6515	0.3485
CRUDE Sweet	0.9991	0.0009
COTTON	0.9035	0.0965
DOW 30	0.9646	0.0354
FEEDERS	0.3321	0.6679
GOLD	0.9999	0.0001
COPPER	0.9867	0.0133
HEATING OIL	0.9984	0.0016
GAS UNLD	0.9999	0.0001
COFFEE	0.9144	0.0856
WHEAT, KANSAS CITY	0.7916	0.2084
CATTLE	0.9992	0.0008
HOGS	0.6537	0.3463
WHEAT, MINNEAPOLIS	0.4042	0.5958
NATURAL GAS	0.9982	0.0018
NASDAQ 100 Index	0.9956	0.0044
OATS	0.6079	0.3921
PALLADIUM	0.9974	0.0026
P. BELLIES	0.9258	0.0742
PLATINUM	0.9978	0.0022
RUSSELL 2000	0.9987	0.0013
SOYBEANS	0.9463	0.0537
SUGAR #11	0.6322	0.3678
SILVER	0.9934	0.0066
S.BEAN MEAL	0.9234	0.0766
S&P 500	0.9487	0.0513
WHEAT, CHICAGO	0.6920	0.3080
Average	0.8656	0.1344
Median	0.9475	0.0525
Standard Deviation	0.1870	0.1870
Minimum	0.3321	0.0001
Maximum	0.9999	0.6679

IV. Price Discovery Results

Stationarity test results based on Augmented Dickey Fuller Test with a null hypothesis of unit root series and Johansen Cointegration Test results are presented in Table 3. Results for no cointegrating relation are not reported in the table because the null hypothesis of no cointegration is rejected for all commodities; thus in the interest of brevity, these results are not reported but are available upon request.

Schroeder and Goodwin (1991), Quan (1992), Schwarz and Szakmary (1994), and Mattos and Garcia (2004) provide detailed description of the unit root test. Schroeder and Goodwin (1991) and Mattos and Garcia (2004) provide detailed description of the cointegration test. Philips-Perron Stationarity Test results are not reported but are identical to the Augmented Dickey Fuller test results and are available upon request. The unit root tests fail to reject the null hypothesis of non-stationarity. This suggests that based on the Granger representation theorem as discussed in Engle and Granger (1987) and the presence of a natural association between the futures and cash commodity prices the futures and cash prices might be cointegrated. The Granger representation theorem states that if two random series are integrated of order one, there is a possible natural combination of them which will not be integrated. The Johansen Cointegration Test results suggest the presence of at least one cointegrating vector among the futures and cash prices with the exception of the gold futures prices and cash prices.

Proportional information shares of the futures and cash prices are reported in Table 4. The information shares are computed as in equation (4) to satisfy the cointegration relation between the futures and cash prices based on the Vector Error Correction Model as discussed in equation (1). Futures prices lead cash prices with an average proportional information share of 86.56%. Most indexes and commodities have proportional information shares greater than 60% with the exception of the Feeders Cattle and Wheat-Minneapolis futures contracts which are lead by the cash markets with information shares of the futures contracts of 33% and 40%, respectively.

Thus, I document a range of price discovery among the futures contracts with Feeders and Wheat-Minneapolis futures contracts having the minimum information shares of 33% and 40% whereas the Gas Unleaded and Gold futures contracts have the maximum information shares of 99.99%. This might be explained with the fact that the Feeders Cattle contract has the third lowest average trading volume of 652 contracts and the second lowest average open interest of 3,370.

However, the low information share is a bit surprising for the Wheat-Minneapolis futures contract which does not have the lowest trading volume and open interest of the futures contracts. Palladium has the lowest average trading volume of 394, and oats has the lowest average open interest of 3,251.

V. Cross-Sectional Analysis Results

It is natural to ask the question why I observe such a large range of information shares across commodities and index products. To address this issue I attempt to relate the information discovery by the futures market to the nature of the underlying commodity and the nature of the futures contract as specified in equations (7) and (8). The hypothesis is that the information

Table 5. Correlogram

	lvol	loi	ratio	cash	agr	metals	energy	oo	etf	sd	fin	sdoi	sdvol	lfvol	lfoi	resvol	resoi	ratiof	rres
lvol	1	0.92	-0.59	-0.04	-0.4	-0.14	0.61	-0.28	0.24	0.37	0.07	0.65	0.76	0.98	0.92	0.53	0.62	0.67	0.31
loi	0.92	1	-0.27	0.06	-0.38	-0.14	0.39	-0.26	0.36	0.39	0.27	0.79	0.7	0.92	0.99	0.45	0.64	0.36	0.32
ratio	-0.59	-0.27	1	0.26	0.05	0.05	-0.5	-0.02	0.12	-0.07	0.42	-0.1	-0.39	-0.59	-0.27	-0.29	-0.18	-0.77	-0.1
cash	-0.04	0.06	0.26	1	-0.27	-0.25	-0.04	-0.12	0.01	0.07	0.71	0.08	-0.01	-0.04	0.06	-0.05	-0.04	-0.2	0.2
agr	-0.4	-0.38	0.05	-0.27	1	-0.48	-0.48	0.33	-0.41	-0.63	-0.42	-0.16	-0.39	-0.4	-0.38	-0.28	-0.2	-0.37	-0.14
metals	-0.14	-0.14	0.05	-0.25	-0.48	1	-0.2	0.06	0.21	0.42	-0.18	-0.06	0.01	-0.14	-0.14	-0.1	-0.16	-0.1	-0.19
energy	0.61	0.39	-0.5	-0.04	-0.48	-0.2	1	-0.48	0.03	0.3	-0.18	0.06	0.49	0.61	0.39	0.53	0.4	0.86	0.24
oo	-0.28	-0.26	-0.02	-0.12	0.33	0.06	-0.48	1	-0.01	-0.13	-0.03	-0.01	-0.27	-0.28	-0.26	-0.25	-0.12	-0.34	-0.03
etf	0.24	0.36	0.12	0.01	-0.41	0.21	0.03	-0.01	1	0.17	0.34	0.14	-0.04	0.24	0.36	-0.18	-0.04	-0.07	-0.14
sd	0.37	0.39	-0.07	0.07	-0.63	0.42	0.3	-0.13	0.17	1	0.14	0.48	0.37	0.37	0.39	0.21	0.37	0.21	0.09
fin	0.07	0.27	0.42	0.71	-0.42	-0.18	-0.18	-0.03	0.34	0.14	1	0.24	0.02	0.07	0.27	-0.05	0.03	-0.29	0.15
sdoi	0.65	0.79	-0.1	0.08	-0.16	-0.06	0.06	-0.01	0.14	0.48	0.24	1	0.67	0.65	0.79	0.36	0.62	0.09	0.17
sdvol	0.76	0.7	-0.39	-0.01	-0.39	0.01	0.49	-0.27	-0.04	0.37	0.02	0.67	1	0.76	0.7	0.87	0.72	0.59	0.49
lfvol	0.98	0.92	-0.59	-0.04	-0.4	-0.14	0.61	-0.28	0.24	0.37	0.07	0.65	0.76	1	0.92	0.53	0.62	0.67	0.31
lfoi	0.92	0.99	-0.27	0.06	-0.38	-0.14	0.39	-0.26	0.36	0.39	0.27	0.79	0.7	0.92	1	0.45	0.64	0.36	0.32
resvol	0.53	0.45	-0.29	-0.05	-0.28	-0.1	0.53	-0.25	-0.18	0.21	-0.05	0.36	0.87	0.53	0.45	1	0.75	0.57	0.68
resoi	0.62	0.64	-0.18	-0.04	-0.2	-0.16	0.4	-0.12	-0.04	0.37	0.03	0.62	0.72	0.62	0.64	0.75	1	0.37	0.59
ratiof	0.67	0.36	-0.77	-0.2	-0.37	-0.1	0.86	-0.34	-0.07	0.21	-0.29	0.09	0.59	0.67	0.36	0.57	0.37	1	0.21
ratiores	0.31	0.32	-0.1	0.2	-0.14	-0.19	0.24	-0.03	-0.14	0.09	0.15	0.17	0.49	0.31	0.32	0.68	0.59	0.21	1

Lvol is the log of futures volume; Lfvol is the log of futures expected volume; Loi is futures open interest; Lfoi is futures expected open interest; Resvol is futures contract unexpected volume; Resoi is the futures contract unexpected open interest; Ratio is the ratio of the average trading volume to average open interest; Ratiof is the ratio of the average expected trading volume to average expected open interest; Rres is the ratio of the average unexpected trading volume to average unexpected open interest. Cash is a dummy variable of one if the futures contract has cash settlement, and zero for physical delivery; Agr is a dummy variable of one if the futures contract is agricultural, and zero otherwise; Energy is a dummy variable of one if energy future contract, zero otherwise; OO is a dummy variable of one if the futures contract is open outcry traded, and zero otherwise; Etf is a dummy variable of one if an ETF exists within the examined period, and zero otherwise; and Sd is the standard deviation of the futures contract price over the examined period. Fin is a dummy variable of one if index future contract, zero otherwise; Sdoi is the standard deviation of the futures contract open interest over the examined period; and Sdvol is the standard deviation of the futures contract traded volume over the examined period.

shares might be different because of the different nature of the futures contract, the contract liquidity, in terms of agricultural versus other type of underlying asset (metal, energy or financial asset), in terms of open outcry or electronic trading platform and cash or physical delivery settlement of the futures contract and involvement of informed and uninformed investors. I also test whether the existence of an ETF on the commodity might have an impact on the commodity price discovery. Cotton, gold, copper, heating oil unleaded gas, coffee, wheat, natural gas, platinum, sugar, silver and the four indexes studied had U.S. listed ETFs prior to January 29, 2010.

Before the regression analysis, however, I examine the cross correlation among the proxy variables. The correlogram is provided in Table 5. Based on this table, special attention is required when examining some of the variables because of the high correlation among them. Lvol, Loi, ratio, ETF, and Sd are highly correlated suggesting that some of these variables need to be excluded because of potential multicollinearity issues in the regression. Therefore, as robustness exercise this study examines several model specifications to ensure stability of results; however, complete results for all model specifications are not reported in the interest of brevity but are available upon request.

The cross sectional regression results are presented in Table 6 based on four model specifications to ensure robustness of results. Model 1 is the unrestricted model as defined in equation 7 based on variables aggregated over entire sample period, which is different for each commodity. The cross section results suggest that liquidity does play an important role in the variability of information shares indicated by the significance of the liquidity proxy used here - log of volume. The agricultural, energy, and cash settled contracts consistently seem to have lower information shares of the futures contract. This is indicated by the consistently significant negative coefficients of the respective variables. Also, commodities and indexes with ETFs tend to have consistently statistically significant lower information shares of the futures contract. This is most likely due to the ETF market attracting informed traders away from the futures market.

The cross sectional regression results separating expected and unexpected volume and open interest as defined in equation 8 are presented in Table 6, Model 2. Again variables are aggregated over entire sample periods. The model is the unrestricted model as defined in equation 8. Again, agricultural, energy, and cash settled contracts and commodities with ETFs consistently have statistically significant negative interaction with the information share.

Models (3) and (4) are estimated based on annual information shares and annual independent variables and pooled cross-section time series analysis. The results presented in Table 6, Model (3) reflect the pooled cross-section time series restricted model as defined in equation (7); but the data are aggregated annually rather than over entire sample periods this time. The restriction is imposed on the intercept and assumes equal effects across commodities. Volume is again positively and significantly related with the information share of the futures contract whereas the agricultural, energy, cash-settled contracts and commodities with ETFs are significantly negatively related to the information share. The open outcry nature plays a significant role when the annual information shares are used and indicates lower information shares of open outcry contracts reinforcing the notion established in the literature that informed investors congregate in the electronic markets. The statistically significant standard deviation of the volume proxy of liquidity indicates that the higher the volatility of the futures contract volume, the lower the information share of the futures contracts. This suggests that the volatility or uncertainty regarding liquidity conditions within the futures marketplace reduces its relative information contribution. This is especially true considering that this is in line, for instance, with trader being

Table 6. Regression Results

	Model 1		Model 2		Model 3		Model 4	
Variable	Estimate	Pr > |t|	Estimate	Pr > |t|	Estimate	Pr > |t|	Estimate	Pr > |t|
Intercept	-0.1289	0.8022	-1.492210	0.2412	0.9638***	<.0001	0.975240***	<.0001
lvol	0.4060**	0.0430			0.1127***	<.0001		
lfvol			-0.334760	0.2427			0.067110*	0.0766
loi	-0.248	0.1921			-0.0867***	0.0026		
lfoi			0.506880	0.1369			-0.046150	0.2816
resvol			-0.000072	0.8543			-0.000011	0.2570
resoi			-0.002140	0.3922			-0.000017*	0.0921
ratio	0.061	0.1035			0.0066***	0.0054		
ratiof			1.791320	0.1084			-0.006050	0.9609
ratiores			0.014240	0.1796			-0.000094	0.3774
cash	-0.1713*	0.0583	-0.146680	0.2099	-0.0693**	0.0301	-0.075120**	0.0206
agr	-0.2338**	0.0326	-0.260600**	0.0442	-0.2329***	<.0001	-0.229320***	<.0001
energy	-0.3153**	0.0263	-0.462300*	0.0529	-0.1515***	0.0021	-0.102350**	0.0447
oo	0.0068	0.9060	0.005850	0.9325	-0.0691***	0.0013	-0.065990***	0.0022
etf	-0.1535**	0.0256	-0.166900**	0.0262	-0.0676***	0.0019	-0.071540***	0.0011
sd	0.2187	0.4507	0.243960	0.4596	-0.1338	0.3289	-0.160980	0.2494
financial	0.1211	0.4047	0.042140	0.8213	0.0698	0.1821	0.090640	0.1020
sdoi	-0.002	0.1534	-0.001070	0.5940	-0.0002	0.7668	-0.000003	0.9963
sdvol	-0.0039	0.1359	-0.006560	0.3450	-0.0022*	0.0786	-0.001660	0.1844
R-sq	0.7789		0.7996		0.1944		0.1934	
N	30		30		709		709	

Models (1) and (3) are OLS regressions based on equation (7) and models (2) and (4) are based on equation (8). Lvol is the log of futures volume; Lfvol is the log of futures expected volume; Loi is futures open interest; Lfoi is futures expected open interest; Resvol is futures contract unexpected volume; Resoi is the futures contract unexpected open interest; Ratio is the ratio of the average trading volume to average open interest; Ratiof is the ratio of the average expected trading volume to average expected open interest; Ratiores is the ratio of the average unexpected trading volume to average unexpected open interest; Cash is a dummy variable of one if the futures contract has cash settlement, and zero for physical delivery; Agr is a dummy variable of one if the futures contract is agricultural, and zero otherwise; Energy is a dummy variable of one if energy future contract zero otherwise; OO is a dummy variable of one if the futures contract is open outcry traded, and zero otherwise; Etf is a dummy variable of one if an ETF exists within the examined period, and zero otherwise. Sd is the standard deviation of the futures contract price over the examined period; Financial is a dummy variable of one if index future contract, zero otherwise; Sdoi is the standard deviation of the futures contract open interest over the examined period; and Sdvol is the standard deviation of the futures contract traded volume over the examined period. Statistical significance at the 1%, 5%, and 10% levels are denoted with ***, **, and * respectively.

less likely to incorporate information into futures prices when they feel that doing so may expose them to positions that are costly or not easily reversible.

The results presented in Table 6, Model (4) reflect the pooled cross-section time series restricted model based on the model as defined in equation (8) based on data aggregated annually. The restriction is imposed on the intercept and assumes equal effects across commodities. Again, the expected volume is significant and positively related to the futures contract information share. Also, the agricultural, energy, cash settled, open outcry contracts, and commodities with ETFs havea negative relationship to the futures contract information share. In this model the unexpected open interest has a negative statistically significant relation to the information share of the futures contract whereas the expected volume has a statistically significant positive relation with the futures contract information share. This suggests that both informed and uninformed traders play a role in the formation of the futures contract information share – informed traders influence price discovery through open interest, and uninformed traders influence price discovery through trading volume.

Additionally, robustness analysis based on annual information shares and annual independent variables pooled cross-section time series analysis allowing for fixed effects in the model is performed. Results are similar to the other estimation methods and thus in the interest of brevity are not reported but are available upon request.

VII. Conclusion

The extant literature in the field of lead lag relationship of cash and futures markets has not identified which market provides the most information – the cash market or the futures market. Also, the literature in this area has focused only on a handful of commodities. The major study in the field is by Garbade and Silber (1983) conducted over two decades ago and focuses on wheat, corn, oats, orange juice, copper, gold, and silver commodities only. This study attempts to fill this void in the literature by examining the price discovery in 30 index and commodity markets and attempts to identify factors causing the difference in information shares across these markets.

Most indexes and commodities have proportional futures market information shares greater than 60% with the exception of the Feeders Cattle and Wheat-Minneapolis futures contracts. Feeders Cattle and Wheat-Minneapolis futures contracts are lead by their cash markets with information shares of the futures contracts of 33% and 40% respectively. The analysis documents a significant cross-sectional variation in the information shares across the 30 indexes and commodities and finds that the information shares of the futures contract are lower when the trading volume of the futures contract is lower and when the contract is on an energy commodity or agricultural commodity, and the commodity has a traded ETF.

References

Baillie, R., Booth, G., Tse, Y., & Zabotina, T. (2002). Price discovery and common factor models. *Journal of Financial Markets*, *5*(3), 309–321.

Bessembinder, H., & Seguin, P. (1992). Futures-trading activity and stock price volatility. *Journal of Finance*, *47*(5), 2015-2034.

Bessembinder, H., & Seguin, P. (1993). Price volatility, trading volume, and market depth: Evidence from futures markets. *Journal of Financial and Quantitative Analysis*, *28*(1), 21-39.

Brockman, P., & Tse, Y. (1995). Information shares in Canadian agricultural cash and futures markets. *Applied Economics Letters*, *2*(10), 335-338.

Enders, W. (2004). *Applied Econometric Time Series* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Engle, R., & C. W. Granger, C. W. (1987). Co-Integration and error correction: Representation, estimation, and resting." *Econometrica*, *55*(2), 251-276.

Figuerola-Ferretti, I., & Gonzalo, J. (2010). Modeling and measuring price discovery in commodity markets. *Journal of Econometrics*, *158*(1), 95-107.

Flood, C. (2010, January 13). Grain prices dip after forecast for record crops. *Financial Times*, p. 22.

Franke, G., & Hess, D. (2000). Information diffusion in electronic and floor trading. *Journal of Empirical Finance*, *7*(5), 455-478.

Garbade, K. D. & Silber, W. L. (1983). Price movements and price discovery in futures and cash markets. *The Review of Economics and Statistics*, *65*(2), 289-297.

Gonzalo, J., & Granger, C. (1995). Estimation of common long-memory components in cointegrated systems. *Journal of Business & Economic Statistics, 13*(1), 27-35.

Hasbrouck, J. (1995). "One security, many markets: Determining the contributions to price discovery. *The Journal of Finance*, *50*(4), 1175-1199.

Lien, D., & Tse, Y. (2002). Physical delivery versus cash settlement: An empirical study on the feeder cattle contract. *Journal of Empirical Finance*, *9*(4), 361-371.

Mattos, F., & Garcia, P. (2004). Price discovery in thinly traded markets: Cash and futures relationships in Brazilian agricultural futures markets. *2004 Proceedings of the NCR-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management*. Retrieved on January 15, 2010 from: http://www.farmdoc.uiuc.edu/nccc134

Quan, J. (1992). Two-step testing procedure for price discovery role of futures prices. *Journal of Futures Markets, 12*(2), 139-149.

Schroeder, T., & Goodwin, B. K. (1991). Price discovery and cointegration for live hogs. *Journal of Futures Markets*, *11*(6), 685-696.

Schwarz, T., & Szakmary, A. (1994). Price discovery in petroleum markets: Price discovery, cointegration, and the time interval analysis. *Journal of Futures Markets*, *14*(2), 147-167.

Subrahmanyam, A. (1991). A theory of trading in stock index futures. *The Review of Financial Studies*, *4*(1), 17-51.

Theissen, E. (2002). Price discovery in floor and screen trading systems. *Journal of Empirical Finance*, *9*(4), 455-474.

Tse, Y., Xiang, J., & Fung, J. (2006). Price discovery in the foreign exchange futures markets. *Journal of Futures Markets*, *26*(11), 1131-1143.

Williams, M., & Tse, Y. (2009). Increased efficiency in electronic markets: Liquidity vs. informed trading. *Global Business and Finance Review, 14*(1), 37-50.