UNIVERSITÉ DE NEUCHÂTEL

FACULTÉ DES SCIENCES ÉCONOMIQUES

THREE ESSAYS ON COMMODITY MARKETS

Ph.D. thesis submitted to the Faculty of Economics and Business

Institute of Financial Analysis

University of Neuchâtel

For the degree of Ph.D. in Finance

by

Loïc Maréchal

Accepted by the dissertation committee :

Prof. Michel Dubois, University of Neuchâtel, thesis director

Prof. Tim Kroencke, University of Neuchâtel

Prof. Florian Weigert, University of Neuchâtel

Prof. Amit Goyal, University of Lausanne

Prof. Sofia Ramos, ESSEC Business School, France

Dr. Zeno Adams, University of St. Gallen

Defended on July 13, 2021

UNIVERSITÉ DE
NEUCHÂTEL

IMPRIMATUR POUR LA THÈSE

Three Essays on Commodity Markets

Loïc MARÉCHAL

UNIVERSITÉ DE NEUCHÂTEL

FACULTÉ DES SCIENCES ÉCONOMIQUES

La Faculté des sciences économiques,
sur le rapport des membres du jury

Prof. Michel Dubois (directeur de thèse, Université de Neuchâtel)
Prof. Tim A. Kröncke (président du jury, Université de Neuchâtel)
Prof. Florian Weigert (Université de Neuchâtel)
Prof. Amit Goyal (HEC Lausanne)
Prof. Sofia Ramos (ESSEC Business School, Cergy Pontoise, France)
Dr. Zeno Adams (Universität St. Gallen)

Autorise l'impression de la présente thèse.

Neuchâtel, le 14 septembre 2021

Le doyen
Valéry Bezençon

Acknowledgments

First and foremost, I would like to thank my supervisor, Michel Dubois, for his guidance throughout this long journey. Michel is a great mentor from a variety of perspectives. His technical skills, scientific mindset, and frank personality, among many other facets, were essential to develop myself as a researcher. He let me choose my research topics and offered me the necessary freedom to develop them while being always there to ingeniously realign my work with the academic track. In addition to the research competencies, being a Ph.D. student has also been a tremendous opportunity to develop a collection of additional skills such as teaching, public speaking, or time management under pressure. His experience and excellent principles to deal with these aspects considerably helped me to succeed. Finally, I cannot forget to mention how lucky I was to have a supervisor with whom I could do off-track skiing, discover amazing ski areas, or even climb my first 4,000-meter peak in the Alps!

I am grateful to my peers Ph.D. candidates of the Faculty of Business and Economics who made my Ph.D. life not only easier, through their work, moral, and personal support, but also rewarding and fun. I would like to thank in particular, Sidita Hasa and Keven Bluteau, my “office roommates”, Louis Mangeney, Albian Albrahimi, Emanuele Guidotti, and Pierluigi Giosi.

I also wish to thank the Professors at the Institute of Financial Analysis for their guidance on research and teaching and more broadly on all academic topics. I would also like to mention Kira Facchinetti for her great help with any administrative tasks and for guiding me across all the University processes.

A sincere thank goes to all my friends who were always able to change my mind and provide me with a positive spirit over the difficult times.

I also would like to mention some of my life mentors, among many others, who inspired me or guided my interest towards science and research: my grandfathers and Michał Kabata and Christian Merkwirth.

And last but not least, I would like to express my gratitude to my family for their support. Not only did they encourage me to undertake this doctoral process from the beginning, but they assisted me both morally and through countless proofreading and discussions regarding all the above-mentioned subjects, whenever I needed!

Abstract

This dissertation is constituted of three distinct chapters on commodity markets, which cover different aspects of finance. The first chapter focuses on the consequences of the financialization of commodity markets, an important topic from a regulatory perspective.[1] Using a multivariate change-point algorithm, I statistically date when the financialization materializes, with a narrow confidence interval. Using this date as a reference point, I use panel data methods to statistically assess the impact of the financialization on commodity futures prices. I find that it leads to a change in the risk-sharing structure, but without being detrimental neither for a typical investor nor for the functioning of commodity markets.

The second chapter employs economically-motivated variables (time to maturity, inventories, and uncertainty resolution) and shows that they are significant determinants of the contemporaneous volatility.[2] Moreover, for the long-term horizons, I find that these variables improve the out-of-sample forecasts of the most advanced high frequency-based, time-varying parameters, models. In a multi-quantile regression setting, I find that these variables do not yield more robust estimations for the out-of-sample expected shortfall.

The third chapter explores whether the emergence of commodity index traders affects asset pricing properties (factors and characteristics). Using a Bayesian approach, I determine the optimal set of factors pricing commodity futures. I subsequently use this optimal risk adjustment to test how the financialization alters liquidity and insurance premia. Finally, I use a panel data method to test how these characteristics determine commodity returns, controlling for their potential correlation with factors. More broadly, this chapter also questions the existence of risk premia *vs.* pure characteristics (frictions) in the asset class of commodity derivatives, as advocated by *e.g.* Black (1976), and calls for more research in this area.

Keywords: commodity futures, index investment, price pressure, risk premium, liquidity premium, hedging pressure, sunshine trading, predatory trading, realized volatility, term structure, Samuelson effect, supply and demand, risk management, financialization.

[1] In collaboration with Michel Dubois
[2] Published in the Journal of Futures Markets

Résumé

Cette dissertation est composée de trois chapitres distincts, traitant des marchés de matières premières et couvre différents aspects de la recherche en finance. Le premier chapitre analyse les consécquences de la “financiarisation” des marchés de matières premières, un sujet d’importance du point de vue de la régulation des marchés.[1] En utilisant un algorithme de détection de changement structurel sur des séries temporelles multivariées, Je détermine statistiquement la date de materialisation de la finaciarisation en obtenant un intervalle de confiance restreint. En utilisant cette date de référence, j’utilise ensuite des méthodes de données de panel pour évaluer statistiquement l’impact de la financiarisation sur les prix des contrats futures de matières premières. Je documente également un changement dans la structure de transfert de risque entre les agents, mais qui ne se révèle pas dommageable ni pour un investisseur représentatif ni pour le fonctionnement des marchés de matières premières.

Le second chapitre utilise des variables motivées par la théorie économique (temps jusqu’à maturité, niveau d’inventaire et résolution de l’incertitude) et montre qu’elles sont des déterminants significatifs de la volatilité contemporaine.[2] De plus, je trouve que ces variables améliorent les prévisions hors-échantillon, y compris celles générées par les modèles les plus avancés incluant erreurs de mesure et paramètres variant au cours du temps. En utilisant une méthode de régression multi-quantile, je constate cependant que ces variables ne génèrent pas d’estimations plus robustes du déficit attendu.

Le troisième chapitre vise à détecter si l’émergence des investisseurs indiciels affecte les propriétés des matières premières en terme d’évaluation des actifs (facteurs et charactéristiques). En utilisant une approche Bayésienne, je détermine l’ensemble optimal des facteurs de risque des contrats futures de matières premières. J’utilise subséquemment cet ajustement optimal du risque pour tester comment la financiarisation altère les primes de liquidité et d’assurance. Enfin, j’utilise une méthode de données de panel pour déterminer comment ces charactéristiques affectent les rentabilités des matières premières, tout en contrôlant pour la corrélation potentielle avec les facteurs. Plus généralement, ce chapitre questionne également l’existence de primes de risque dans la classe d’actif des dérivés de matières premières, comme évoqué par Black (1976) et ouvre un nouveau champ de recherche dans ce domaine.

[1]En collaboration avec Michel Dubois

[2]Publié dans le Journal of Futures Markets

Mots clefs: contrat futures de matières premières, investissement indiciel, pression sur les prix, prime de risque, prime de liquidité, pression de couverture, sunshine trading, trading prédateur, volatilité réalisée, structure par terme, Samuelson effect, offre et demande, gestion des risques, financiarisation.

Contents

List of Tables

List of Figures

General introduction

Commodity futures constitutes a peculiar asset class. First, the number of different raw materials available is way smaller than the number of stocks or bonds. More importantly, the number of available futures contracts written on the future delivery of these materials is even more restricted. Second, these financial instruments are contingent claims and expire, requesting an investor who holds her position until maturity to receive or deliver the underlying products. Third, the heterogeneity of their production and consumption locations, their sensitivity to logistics costs and inventory constraints, and the numerous independent variables that affect their supply and demand imbalances, such as weather, geopolitics, or inventory decays, makes their integration in a single economic model difficult. Altogether, these specificities represent a challenge for the researcher. It is an issue to explain their performance along with those of other assets, and even more, it is hard to explain their performance within the asset class. This thesis tackles this challenge with both economic and econometric approaches that have proven successful in other assets, in particular, for stocks. It aims to identify the factors affecting their performance, both in terms of returns and risk.

Commodity futures are probably the oldest existing financial instruments.[1] Organized markets to exchange similar contracts began over 300 years ago.[2] Given this longevity, commodity contingent claims are also among the first contracts to be studied.[3] In his "theory of normal backwardation", Keynes (1930) provides a first explanation of commodity futures as an insurance market, in which producers hedge their future output at a discount whereas speculators bear the risk and collect the premiums. A second, non-contradicting, view is the "theory of storage" of Kaldor (1939). In this model, futures prices are function of the spot price, compounded by continuous interest rates and storage costs and discounted by a "convenience yield", which represent the benefits of holding the physical commodity that can accrue to an agent. Both theories build on economic grounds that are common with stocks: in the first theory on the risk premium, in the second, on discounting principles and on the benefit of obtaining a (dividend) yield when holding the actual stock, instead of a claim. However, both theories rationalize the commod-

[1] Clay tablets stating promises for futures quantity of grains were found in Mesopotamia, 5000 years ago.

[2] The Dōjima rice exchange was created in 1697.

[3] See Homma Munehisa https://en.wikipedia.org/wiki/Homma_Munehisa

ity performance with principles common to all assets, but they do not require to integrate them in a global economic context. Black (1976) denies this integration from the capital asset pricing model perspective, since commodity futures are bets with zero-sum positions. Their addition to the market portfolio should not provide any explanation, neither from their returns nor for those of other assets. However, many other features of commodities relate them to the broad economy: (i) because commodities are consumed, they are also natural candidates to be explained by a consumption capital asset pricing model framework, (ii) for the same reason they forecast unexpected inflation, (iii) they condition future states of the economy, and (iv) given the increasing presence of agents aiming to diversify their portfolios or benefit from additional features of commodities, their covariation with other asset classes rises. Additional stylized facts on commodity futures are worth retaining, such as positive skewness, historical positive returns of futures investments, or negative inventory-volatility relationships.

Because of the aforementioned stylized facts of commodity futures, they have become increasingly popular assets within the financial industry, *i.e.* portfolio managers, hedge funds, and mutual funds, among other participants. In the 2000 decade, this attention reached a peak, and new investment vehicles and strategies appeared, a phenomenon coined as the "financialization". These "first generation" index funds had the primary objective to simplify the investment process of any agent willing to obtain exposure on the underlying commodities, without the logistic constraints inherent to the management of these raw materials, and without the need to monitor the timing of contracts expiry. The arrival of these new participants through these vehicles may have changed the risk-sharing structure of the market, and possibly its price process. Similarly, these agents require risk management tools that have been overlooked in commodities. The academic literature is unclear on the consequences of the financialization, despite the needs from a regulatory perspective. In terms of risk management, there is very little evidence that benchmark-realized volatility models apply to commodity futures, and none that integrate specific commodity features to help forecast risk.

This dissertation aims to fill these gaps. The first essay examines the valuation effects of the SP-GSCI roll on commodity contracts. First, we document a surge in investment tracking commodity futures indices in December 2003. Before 2004, the roll period generated average cumulative abnormal price changes amounting to 115 bps for the nearby contract and 146 bps for the first deferred contract. From 2004 to 2010, the average cumulative abnormal price changes of the nearby (first deferred) is equal to -60 bps (-40 bps). However, a strategy that front-runs the roll does not generate abnormal profits at the contract level after (reasonable) transaction costs. A difference-in-differences regression confirms that the financialization has an alleviating effect, with the nearby (first deferred) average cumulative abnormal price changes showing a 158 (166) bps drop, statistically significant at the 1% level. The introduction of electronic trading alone does not affect abnormal price changes during the roll. Finally, the contemporaneous change in hedging pressure is negatively related (statistically significant at the 1% level) to cumulative abnormal price changes, while the contemporaneous change in the commodity index pressure, average hedging pressure, and cross (average) hedging pressure do not affect cumulative abnormal price changes.

The second essay combines recent advances in the realized volatility literature and three economically motivated variables, related to well-known hypotheses of commodity volatility determinants, to improve the volatility forecast of commodity futures contracts. The three forecasting variables are the term structure slope (theory of storage), the time to maturity (Samuelson hypothesis), and a seasonal measure that proxies for the supply and demand uncertainty (uncertainty resolution hypothesis). I first assess the empirical contribution of these variables to explain the realized volatility and find support for the theory of storage and uncertainty resolution. I uncover a positive relationship between time to maturity and volatility, in contradiction with the Samuelson hypothesis. Second, I compare the performance of the HAR (Corsi, 2009), HEXP (Bollerslev, Hood, Huss, and Pedersen, 2018), HARQ (Bollerslev, Patton, and Quaedvlieg, 2016), and time-varying parameter HAR-TV (Chen, Gao, Li, and Silvapulle, 2018) models, with and without the above-mentioned economic variables. I evaluate the in- and out-of-sample performance of these forecasts in econometric tests and find that the inclusion of economic variables is beneficial for long-term horizons only. Finally, I test the validity of these forecasts to improve the expected shortfall modeling. The inclusion of economic variables provides mixed results.

The third essay investigates the effect of the "financialization" of commodity markets in terms of pricing. In this study, I explore whether the emergence of commodity index traders affects weekly returns and turn-over during the roll periods. I split the sample (1994–2017) into the pre-financialization (1994–2003) and the post-financialization (2004–2017). I directly test whether the CIT market share (CIT/Open Interest) contributes to commodity returns, and whether risk adjustments (based on momentum, basis, basis-momentum, open interest, crowding, and average factors) alter the liquidity and insurance premiums documented in Kang, Rouwenhorst, and Tang (2020). I also examine how the financialization affects the liquidity and the insurance premiums. Finally, since previous results are obtained with Fama-MacBeth regressions, I use an alternative method to test how liquidity and insurance premiums determine commodity returns.

Altogether, these essays target three questions related to commodity markets, in terms of pricing: (i) Is there a systematic price effect due to the presence of index investment? (ii) Do economic variables related to commodity futures explain and forecast realized volatility? and (iii) Are there risk factors able to price commodities as for stocks and bonds?

From the first and third papers of this essay, there is evidence that the effect of the financialization is related to liquidity and insurance premiums. The former increased and the latter decreased. Since they go in opposition directions, the global effect is limited. There is however evidence that the arrival of new players through index investments decreased the price of insurance for hedgers. From the second paper, there is also clear evidence that the contribution of classic economic variables to the realized volatility is limited. This question is left for further research. However, the set of characteristics that I document in this monograph merged with recent results in empirical asset pricing should help answer this question, which has puzzled financial economists for half a century.

Chapter I

The valuation effects of index investment in commodity futures

(with Michel Dubois)

1. Introduction

This paper examines whether the increase in market participation of commodity index traders (CITs), known in the literature as the "financialization" of commodity markets, has a material effect on commodity futures prices when commodity indices roll positions from the nearby to the first deferred contracts. To address this question, we focus on the Standard and Poor's Goldman Sachs Commodity Index (SP-GSCI) components.[1] This index is arithmetically weighted based on the relative importance of the commodities in the economy, and the liquidity of the futures contracts. If the financialization has a permanent effect on commodity futures prices, it should be observable when "first generation" products, that replicate the SP-GSCI index, roll their positions from the fifth to the ninth business day of the month preceding maturity (hereafter, the roll). Every day of the roll, the index transfers 20% of its positions in nearby contracts maturing in less than a month to the first deferred contracts. This particular window overlaps with the rebalancing of the BCOM index, and many other U.S. funds invested in commodities.

The emergence of CITs induces a change in the supply/demand of nearby and first deferred contracts, which can be analyzed theoretically through the lens of several models. Brunnermeier and Pedersen (2005) study how predatory traders that know the traders' need to liquidate quickly their position take advantage of them (front trading). Admati and Pfleiderer (1991) propose an alternative view when the trades are known in advance (sunshine trading). In that setting, the impact of the extra demand is mitigated, and trades do not affect prices, everything else

[1]The three major cross-sector indices in terms of tracking OI are respectively, the SP-GSCI (formerly GSCI), the Bloomberg Commodity Index (BCOM, formerly Dow Jones UBS Commodity Index or DJ-UBSCI), and the Deutsche Bank Commodity Index (DBCI). Other diversified indices exist, such as the Reuters-Jefferies CRB Index (RJ-CRB), the Rogers International Commodity Index (RICI), the Chase Physical Commodity Index, Pimco, Oppenheimer, or Bear Sterns. Other non-diversified funds directly track subsectors (energy or agricultural futures) or single futures (crude oil, natural gas, gold).

being equal. Within the Grossman and Miller (1988) framework, the roll creates a temporary order imbalance. The demand (supply) for the immediacy of the first deferred (nearby) contract increases (decreases) its price, which generates positive (negative) price changes. The classic "normal backwardation" theory states that hedgers pay a hedging premium if the demand of hedging is net short (hedging pressure); see Keynes (1930), and Hicks (1946). Roon *de*, Nijman, and Veld (2000) show that price changes of a commodity are theoretically and empirically related to the hedging pressure of commodities that are close substitutes (cross hedging pressure). Kang et al. (2020) demonstrate that speculators provide insurance to hedgers at long-term horizons (over four weeks), whereas hedgers provide liquidity to speculators at weekly horizons. This generates two distinct premiums, *i.e.*, the liquidity premium, and the insurance premium. We conjecture that these premiums are also related to cross hedging pressure since arbitrage capital can easily switch from one commodity to another.[2] Since CITs are structural long investors, it is an empirical issue to see whether changes in CITs pressure, changes in (cross) hedging pressure, and (cross) hedging pressure help explain risk-adjusted price changes during the roll.

Our empirical analysis proceeds in three steps. First, we examine whether a structural break affects simultaneously the CITs' market participation in the 27 SP-GSCI components during the 1999—2010 period. Using the Bai, Lumsdaine, and Stock (1998) algorithm, we find a common break in the proportions of open interest (OI) resulting from index investment in the nearby contracts. This break occurs in December 2003 and is statistically significant at the 1% level. It is identical to that retained in Boons, Roon *de*, and Szymanowska (2014). However, it occurs five years before the one identified by Hamilton and Wu (2015) on WTI crude oil futures, and three years after the one used by Mou (2011). When we restrict the analysis to the 19 SP-GSCI contracts that are under the CFTC supervision, this date is not affected. As an alternative investment index proxy, we use the market share of the long commercial positions (SCL) that include swap dealers and CITs. The break occurs a month before, in November 2003. We also examine six CFTC contracts that are not components of a commodity index. Reassuringly, we find no break in the 2003–2004 period. Finally, we show that the break does not overlap with the staggered introduction of electronically traded contracts that took place from August 2006 to October 2008. Indeed, a Bonferroni test rejects the null hypothesis that the initiation of electronic trading, which is contract-specific, is

[2] Typically, the CME lists intermarket spread with legs on different, but related, commodities. Both legs have the same maturity.

within the confidence interval of the break at the 1% level. This makes us confident that a significant change in market participation is observed by the end of 2003.

Next, we present the results of an event study on the nearby and the first deferred contracts to analyze the valuation effect of the roll. The windows of interest are the pre-roll and the roll periods. We find that the average cumulative abnormal price change (ACAPC) during the roll is equal to 115 bps for the nearby, and 146 bps for the first deferred contracts before 2004, and statistically different from zero at the 1% level. Once we correct for event-induced variance and cross-correlation, the ACAPCs of the nearby and the first deferred remain statistically significant at the 1% level. After 2003, ACAPCs decrease to -60 bps (nearby) and -40 bps (first deferred) respectively, and are no longer statistically significant at the 10% level when t-statistics are corrected for event-induced variance and cross-correlation. The results are not affected when the ACAPCs are value-weighted (SP-GSCI weights) instead of equally-weighted, or when we use raw cumulative price changes instead of cumulative abnormal price changes. We also uncover that cumulative abnormal price changes (CAPCs) do not resist reasonable transaction costs. When we implement a calendar spread strategy based on individual contracts that front runs the index roll by ten days, we find that the residual profits left to arbitragers are lower than the transaction costs, would they enter the trade; see Mou (2011). With a difference-in-differences regression, we analyze the drop in the CAPCs of the rolling contracts. The financialization drop is 158 bps for the nearby and 166 bps for the first deferred over the roll window, both statistically significant at the 1% level, while the introduction of electronic trading does not affect CAPCs. In contradiction with Masters (2008), we find that the financialization is beneficial to the market functioning as the CAPCs during the roll are reduced. We also document a similar pattern in the pre-roll period.

Finally, we demonstrate that the main determinant of CAPCs is the change of net long commercial positions during the roll. Before the financialization, the economic magnitude of this variable is substantial for the nearby (first deferred). A one standard deviation increase is associated with a 1.42% (0.95%) decrease in CAPC. After the financialization, this variable is also statistically significant at the 1% level. Nearby (First Deferred) CAPCs are even more sensitive to changes of the net long commercial positions. A one standard deviation increase translates into a 3.22% (2.15%) decrease in the corresponding CAPC. The net long commercial position aggregates the position of long and short "commercials" over all maturities. To disentangle which maturities generate the CAPCs, we relate them to the turnover of individual contracts. It appears, that CAPCs are positively related to longer-term (second deferred and further) maturities. The change in CITs pressure, average hedging pressure, cross liquidity and cross hedging pressure do not explain CAPCs.

This paper contributes to the literature in several ways. First, we date the financialization using a statistical method instead of relying on anecdotal evidence. Then, we show that the valuation effect of the SP-GSCI "pure" roll is statistically significant at the 1% level before the financialization, even when t-statistics are adjusted appropriately for heteroscedasticity and cross-correlation. The post-financialization does not show statistically significant ACAPCs at the usual levels (after correcting for heteroscedasticity and cross-correlation). This result differs from the valuation effect of the annual index rebalancing ("January roll"). Yan, Irwin, and Sanders (2019) find that price pressure resulting from the weight adjustments is the main driver of the January roll CAPCs, while we conclude that rolls are driven by liquidity needs. Brunetti and Reiffen (2014) study the impact of the roll on the term structure (nearby minus first deferred). We depart from their approach since the roll may have specific effects depending on the contract maturity (selling pressure *vs.* buying pressure from CITs). Our analysis is also related to Kang et al. (2020) who show that the change in net "commercials" position is positively related to futures returns. We demonstrate that the corresponding price of liquidity more than doubles in the aftermath of the financialization. More importantly, our results show that the proportion of CITs investment in OI does not induce a significant price reaction during the roll. This suggests that non informative trades do not affect significantly commodity prices, which gives support to the sunshine hypothesis.

The remainder of the paper is organized as follows. Section 2 presents the hypothesis development in the context of the existing literature. Section 3 provides details on the data and the methodology. Section 4 reports the empirical results, and their robustness is discussed. Section 5 concludes.

2. Prior literature and hypotheses development

Following Master's hearing before the U.S. Senate, several articles examine the economic consequences of the financialization, and when they materialize. However, previous research diverges on its starting date.

2.1. Dating the financialization

A first strand of literature uses ad-hoc, visual, or indirect evidence to date the financialization. Sanders and Irwin (2011), Basak and Pavlova (2016), Boons et al. (2014), and Buyuksahin, Haigh, Harris, Overdahl, and Robe (2009) date the financialization around 2003—2004. Hamilton and Wu (2014) in 2005. Mou (2011) shows that front-running the "SP-GSCI roll" generates abnormal returns as early as 1999. The second strand of literature uncovers a progressive rise in different variables that cause (amount of index investment) or through which the financialization materializes (correlation between commodities and stock returns or excess correlation in indexed commodities). Buyuksahin and Robe (2014) report that the correlation between the SP-GSCI daily price changes and S&P 500 daily returns rises during the 2004-–2010 period, Adams and Gluck (2015) determine statistically the break date (September 2008). Tang and Xiong (2012) demonstrate that prices of non-energy commodity futures have become increasingly correlated with oil prices after 2004. Similarly, Stoll and Whaley (2010) notice that agricultural futures do not converge to their spot prices during the 2006—2009 period. Masters (2008) shows that CITs investment in the total OI changed dramatically from 1998 to 2008. This ratio is a natural candidate to look for a break caused by index investors. As a result, we should observe a permanent change simultaneously affecting CITs market shares of the SP-GSCI components. Conversely, no break should be observed for the remaining futures contracts.

Hypothesis 1. *The shares of index investment in the total OI of SP-GSCI components (SCIT) show a common break during the 1999-–2010 period.*

As an alternative proxy, we use the proportion of the "long commercials" over total OI since swap dealers and CITs are classified as "commercials" in the legacy CFTC reports. If CITs are becoming important players after the financialization, the share of "long commercials" in total OI (SCL) should increase sharply. This aggregate is interesting because it is also observable for contracts that are not in any index. Therefore, we should not find any common break for non-SP-GSCI contracts. Put differently, we use the group of CFTC contracts that are not components of indices as counterfactuals.

2.2. Non-informative trading and commodity prices

2.2.1. Predatory trading

Brunnermeier and Pedersen (2005) study a distressed trader who reveals some information to a predatory trader (*e.g.*, its broker-dealer). Hence, this predatory trader knows the trader's needs to liquidate quickly her position. The model foresees a higher price impact as the predator trades on this information along or before the distressed trader. Thus, the liquidation value of the distressed portfolio decreases. The model also predicts that the more predatory traders compete, the lower the price impact. With an infinity of predatory traders, the effect is similar to price pressure.

Mou (2011) explores the performance of strategies that front-run the SP-GSCI roll from 2000 to 2010. The highest performance is obtained with a calendar spread that takes a long position in the first deferred and a short position in the nearby ten days before the roll. The flows invested in the calendar spread mimics the roll. This strategy delivers an average of 31 bps per roll before transaction costs. Then, he considers the most liquid futures (the WTI crude oil contract) for which he estimates the trading cost at one bp per contract.[3] After adjusting for transaction costs, he finds profitable front trading opportunities. He attributes these findings to the limits of arbitrage capital. As of 2009, he estimates the annual cost at USD 8.4 billion, for a total index investment of USD 211 billion.

2.2.2. Sunshine trading

Some liquidity traders pre-announce the size and timing of their orders to reduce their price impact. The pre-announcement lowers informational asymmetry because trades initiated for liquidity purposes are not related to information that could move prices. Admati and Pfleiderer (1991) show theoretically that transaction costs are reduced when trades are pre-announced ("sunshine trading"). The effect on other traders is ambiguous as far as trading costs and welfare are concerned. Using intraday data from 2008–2009, Bessembinder, Carrion, Tuttle, and Venkataram (2016) fail to identify a systematic use of predatory strategies during the United States Oil Fund rolls, a large fund invested solely in WTI crude oil futures contracts. Consistent with the sunshine trading theory, the authors find more liquidity providers.

Sunshine trading shares many commonalities with the CITs' roll. Typically, public CITs disclose their roll policy, which closely follows that of the tracked index. Since the amount of assets under management is also public information, the daily trades of CITs are known by market participants. However, three differences are noteworthy. First, index fund managers neither explicitly announce their trades,

[3]These transaction costs are extremely low. Hou and Norden (2018) estimate the transaction costs of a calendar spread on the VIX at 15 bps.

nor the intraday timing within the window. Second, CITs also trade to match the incoming flows into their funds. Third, some CITs are private, and their impact is difficult to evaluate. To conclude, the change in CITs position is a noisy proxy of the trades realized during the roll.

2.2.3. Implications

To analyze abnormal price pressure effects, and check whether they are symmetrical on each leg, we compute the ACAPC of the nearby and first deferred contracts. We then state our hypotheses as follows,

Hypothesis 2.

a. *Predatory trading: The ACAPC of the nearby (first deferred) is negative (positive) during the pre-roll.*

b. *Sunshine trading: The ACAPC of the nearby (first deferred) is nil during the roll.*

Raman, Robe, and Yadav (2017) state that the "financialization" of commodities comprises three main players: a) commodity index traders ("CITs"), known as "massive passives", b) managed money traders (hedge funds) that are active in multiple asset classes, and c) institutional financial intraday traders after the inception of electronic trading. They show that electronic trading decreases the variance of the pricing error, narrows the bid-ask spread, and improves market depth for the WTI contract. To make sure that CAPCs are not driven by electronic trading instead of financialization (confounding effect), our hypothesis is as follows,

Hypothesis 3. *CAPCs are affected by the financialization but remain unaffected by the inception of electronic trading.*

Consider a price-taker investor willing to arbitrage the roll. Her investment strategy consists of opening and closing positions on the two legs, which entails paying the bid-ask spread on the nearby and the first deferred contract. The calendar spread can be also negotiated directly.[4] In that case, the bid-ask spread reduces to that of the spread position alone. Since the calendar spread is less risky than

[4]There is an order book for calendar spreads at the CME. However, this dedicated order book is only for convenience, the CME then breaks the order into the two legs of the calendar spread and puts them into the order book of the respective maturities. It seems on the surface that there is a separate order book dedicated to calendar spreads, but it actually feeds through to the order book of the simple long/short orders for any given maturity targeted, and so the bid-ask spread is on each leg of the trade. We thank Jérôme Taillard for providing us with the following explanation.

a one-leg position, one-leg transaction costs overestimate the real cost. Therefore, we consider that a realistic bid-ask spread lies somewhere between a higher bound (bid-ask spread on a single leg) and a lower bound that we set arbitrarily to a quarter of the higher bound. This is by far higher than one bp, in particular for less liquid contracts. Stoll and Whaley (2010) relate the magnitude of the cumulative term-structure changes to the bid-ask spread, with no formal test, however. Hence, our hypothesis is as follows,

Hypothesis 4. *After adjusting CAPCs for transaction costs, the pre-roll and the roll periods do not show trivial arbitrage opportunities, neither before, nor after the financialization.*

2.3. CITs pressure, hedging pressure, and CAPCs

2.3.1. Price pressure

Grossman and Miller (1988) present a three-period model, where two types of agents, market makers, and outside customers, trade a risky asset. Assuming that prices are normally distributed and that investors maximize their expected (exponential) utility, they derive the demand function and analyze the consequences of a liquidity shock that creates a temporary order imbalance. In that setting, they show how market makers are compensated for bearing the risk during the holding period. Specifically, the absolute expected returns increase with the order imbalance and decrease with the number of market makers. In our case, assuming that the market does not absorb instantaneously the CITs positive demand (supply) on the first deferred (nearby) contract, the current price should increase (decrease).[5] Brunetti and Reiffen (2014) derive a two-period model, close to the previous one, where the CITs' demand is exogenous.[6] The empirical part of their study is based on three agricultural contracts with disaggregated data (CFTC Large Traders Reporting System), from 2003 to 2012. Consistent with their model, the slope of the term structure between the two nearest contracts increases as index traders roll their positions. Based on 12 agricultural contracts and a shorter data set (2004—2009), Aulerich, Irwin, and Garcia (2013) report an increase in the slope of the term structure (statistically significant at the 5% level) for five contracts only. Henderson, Pearson, and Wang (2015) examine the impact of financial investors' flow on commodity futures prices through commodity-linked notes (CLNs). This flow, generated by

[5] Hereafter, this positive demand is called "CITs pressure" since we are analyzing the supply/demand of specific groups of traders (CITs, hedgers, and speculators).

[6] Additional assumptions are similar to those of Grossman and Miller (1988).

CLN issuers hedging their liabilities on the commodity futures markets, does not contain new information on futures prices. Nevertheless, commodity futures prices increase under buying pressure and decrease under selling pressure. These results are consistent with Bessembinder et al. (2016) who show that accumulated trading costs resulting from this price pressure amount to 3% per year.

2.3.2. Hedging pressure and speculative demand

The Theory of "Normal Backwardation" assumes that producers hedge their output against future price fluctuations by selling futures contracts. By doing so, they push futures prices below expected spot prices, thereby providing a positive premium, the insurance premium, to long futures investors; see Keynes (1930) and Hicks (1946).[7] Empirical studies show that hedging pressure (short minus long hedgers position scaled by the total OI) is a determinant of the commodity futures premium; see, *e.g.*, Bessembinder (1992).

The role played by hedging pressure can be extended to groups of commodities that are partly substitutable (cross hedging); see Anderson and Danthine (1981). Roon *de* et al. (2000) derive a model within the CAPM framework, where global hedging pressure is split in own-hedging pressure and cross hedging pressure. Over ten commodity contracts, nine show positive own hedging and four cross hedging pressure. Similarly, Kang et al. (2020) extend the hedging pressure framework. On aggregate, speculators are rewarded the insurance premium to act as the hedgers' counterparty (normal backwardation). However, individual speculators have a shorter horizon than hedgers and, at some points, they are in need for liquidity, which generates a second premium. This could explain why some empirical tests on hedging pressure fail to identify an insurance premium when liquidity is not controlled appropriately.[8] More importantly, Kang et al. (2020, p. 388, Table II) report a negative (positive) relation between contemporaneous "commercials" ("non-commercials") position changes and commodity price changes. They conclude that "commercials" trade as contrarians while "non-commercials" are momentum traders. In this paper, our focus is on the reverse relation, *i.e.*, how abnormal price changes are affected by position changes of "commercials", which includes CITs. Assuming that CITs and pure "commercials" positions are fully and simultaneously transferred from the nearby to the first deferred, the impact on

[7] A producer (*e.g.* a farmer) receives an income which depends entirely on its production. Thus, she will be more likely to short futures at a discount than a processor (*e.g.* an airline company), who can pass-through commodity price fluctuations on consumers.

[8] For instance, Gorton, Hayashi, and Rouwenhorst (2012) do not detect a link between the risk premium and hedgers' positions.

commodity prices should be nil. However, market frictions and changes in supply or demand could temporarily affect commodity prices through the liquidity premium. Typically, an increase in the liquidity premium affects negatively the price of both contracts.

Cheng, Kirilenko, and Xiong (2015) examine the response of various groups of traders to external shocks. More specifically, they study the reaction of financial traders to changes in the VIX during the 2008 financial crisis. They show that financial traders reduce their net long position in response to market distress and that hedgers reduce their short positions as prices fall. This "convective" risk flow is the direct consequence of this external shock. Since hedging pressure explains changes in the price of futures commodity contracts, a natural question arises concerning the determinants of hedging demand. Acharya, Lochstoer, and Ramadorai (2013) study a risk-averse producer that faces speculative capital constraints (default risk). The inclusion of speculative limits to arbitrage, and the subsequently higher premium, depress the producer's hedging demand. Instead of hedging, the producer reduces inventories, which in turn decreases the spot price. Their model predicts that the futures risk premium increases with hedging demand proxy by aggregate producers' default risk. The empirical part of their study confirms this result for oil producers. Rampini, Sufi, and Viswanathan (2014) propose an alternative approach based on the trade-off between debt financing and risk management. Because debt and hedging positions need to be collateralized, financially constrained firms hedge less. The empirical part uses airline companies, and shows that the more financially constrained they are, the less they hedge against fuel price changes. These models predict lower hedging for levered firms but the aggregated impact on demand (net "commercials") is unclear.

2.3.3. Implications

Kang et al. (2020) emphasize the role of different groups of investors in determining commodity price changes and the related premiums. Specifically, "non-commercials" speculators create a liquidity premium, while "commercials" hedgers create an insurance premium. Following the same rationale, we make a step further by analyzing the specific impact of CITs. This group of investors is not individualized in Kang et al. (2020) since CITs switching positions from the nearby to the first deferred are not observable in the CFTC legacy reports over the sample period.[9] If

[9] Short/long position in the CFTC legacy report refers to the sum of short/long position on all the open maturities for a given contract. Therefore, CITs position rolling from the nearby to the first deferred are not observable at the aggregate level. Starting in 2006, the DCOT reports CIT positions.

CITs play a role similar to "net commercials", their position change should affect negatively (positively) the nearby (first deferred); see Grossman and Miller (1988).

We test whether the CITs pressure, liquidity, and insurance help explain CAPCs, which are risk-adjusted price changes with a benchmark that does not take these factors into consideration. More broadly, "cross hedging liquidity" as well as "cross hedging insurance" could also affect individual CAPCs since commodity markets are intertwined. Hence, we state the following hypothesis and check whether the financialization of futures markets alters these premiums.

Hypothesis 5. *Changes in CITs pressure, (cross) liquidity, and (cross) insurance demands are determinants of the nearby and first deferred CAPCs.*

3. Data and Methodology

3.1. Dating the financialization

Since 1986, the CFTC provides the weekly "Commitment of Traders" (COT) report that summarizes "commercials" long and short positions, and long, short, and spreading positions of "non-commercials" categories for U.S. futures contracts.[10] In 2006, the CFTC revised these categories. The "supplemental" (SCOT) report was added for contracts having at least 20 active trader positions above a pre-defined threshold. It was followed by the "disaggregated" (DCOT) report because most swap dealers were classified as "commercials" when, in fact, they were CITs. Despite their non-speculative role, they dedicate a large part of their activity to hedge indices-related positions.[11] The disaggregated report splits the original categories of "commercials" and "non-commercials" into hedgers, swap dealers, managed money, and other "reportables". The supplemental report refers to 13 agricultural futures contracts. The positions reported in the aforementioned categories, managed for commodity investment vehicles (*e.g.* ETFs, ETNs, and mutual funds), are aggregated in a specific category (CIT). From August 2006, the index investment is directly observable from the CFTC reports.

To estimate the share of index investment in the total open interest, we start with the total index investment in the Crude oil WTI contract, which is available in the CFTC special report from 2000 to 2010. The index weights are provided by S&P Global.[12] With this information, we are able to reconstruct the volume

[10]See http://www.cftc.gov/MarketReports/CommitmentsofTraders/index.htm

[11]For the same reasons, swap dealers could claim position limits exemptions, as if they were producers or processors in needs of hedging.

[12]We thank S&P Global for providing us with the historical SP-GSCI weights.

invested in each contract, and the total volume invested by CITs in the SP-GSCI. The total OI from the whole term structure is directly available from the CFTC report. We apply Masters and White (2008, p. 49-51) algorithm over the sample period (2000—2010).[13] The algorithm delivers the OI per contract attributable to CITs (VCIT). The share of total index investment (SCIT) is defined as VCIT divided by total OI; see the definition in the Appendix, Table I.A1.

Sanders and Irwin (2013) argue that the algorithm produces index investment figures that are sensitive to low-weighted contracts. In particular, they find a low, sometimes negative, correlation between $VCIT_{CFTC}$ directly available from the CFTC (after 2006), and $VCIT_{Masters}$ computed with the algorithm. We repeat the exercise with the 13 futures contracts of the CFTC supplemental report. We find that the lowest correlation is 62% (Kansas wheat), and the remaining ones are above 77%. We also run a pooled regression of $VCIT_{Masters}$ on $VCIT_{CFTC}$ to check whether the former is an unbiased estimate of the latter. The global R^2 is equal to 76%, the constant (-0.05) is not statistically different from 0 at the 1% level, and the slope is equal to 0.76 (statistically different from 1 at the 1% level). The bias in the slope may well result from the fact that $VCIT_{Masters}$ accounts for CITs positions in the SP-GSCI and BCOM index, while $VCIT_{CFTC}$ accounts for all index investors. Assuming that market shares of funds related to indices are constant over time or that non-GSCI/BCOM market share is more important in the 2006–2010 period, the break date should not be affected by this measurement issue.

We start the empirical analysis with the 27 SP-GSCI contracts. Following Bai et al. (1998), the dynamics of $\mathbf{SCIT_t}$ is a Vector Auto Regression with q lags, $q = 1, 2, \ldots, 10$, and the number of lags is chosen based on the Bayesian Information Criterion (BIC), computed over the full sample with no breaking covariates. We test for a break on the intercept since we are primarily interested in a structural shift in the level (mean) of the index investment (and not in the trend). We estimate the following equation:

$$\mathbf{SCIT_t} = \left({}^{\mathbf{T}}\mathbf{G_t} \otimes \mathbf{G_t}\right)\theta + d_t\left(k\right)\left({}^{\mathbf{T}}\mathbf{G_t} \otimes \mathbf{I_n}\right){}^{\mathbf{T}}\mathbf{S}\mathbf{S}\delta + \epsilon_{\mathbf{t}}, \tag{1}$$

where ${}^{\mathbf{T}}\mathbf{SCIT_t} = \left[SCIT_{c,t}\right], c = 1, 2, \ldots, 27$, ${}^{\mathbf{T}}\mathbf{G_t}$ is a row vector containing $\mathbf{SCIT_{t-1}}$, $\mathbf{SCIT_{t-2}}, \ldots, \mathbf{SCIT_{t-q}}$, n the number of equations, and $\mathbf{S}$ a selection matrix such that only the intercept is allowed to break with $\mathbf{S} = s \otimes \mathbf{I_n}$ and $s = [1, 0, \ldots, 0]$. Eq. (1) is estimated for every date k, such that $d_t\left(k\right) = 0$ if

[13] For 1999, we start with the January figures in Masters and White (2008) and the first figures available in 2000. Then, we make a linear interpolation and allocate the corresponding amounts to each month.

$t < k$ and 1 otherwise. The potential break date is the value of k that generates the maximum F-statistic, which is statistically significant as soon as this statistic is higher than the limiting χ^2 distribution. Finally, the confidence intervals are constructed for three levels of statistical significance, *i.e.* 90%, 95%, and 99%; see Bai et al. (1998, p. 401).

3.2. The valuation effect of the roll

Daily closing prices of nearby and first deferred SP-GSCI/CFTC contracts are collected from Barchart.[14] We also consider six alternative contracts (ALT) covered by the CFTC, actively traded in the U.S., that are neither components of the SP-GSCI, nor of other popular indices such as BCOM. The sample starts on January 2, 1999, and ends on December 31, 2010. The daily change in commodity price is computed as:

$$r_{c,t}^m = \ln F_{c,t}^m - \ln F_{c,t-1}^m,$$

where $F_{c,t}^m$ is the futures price, c the commodity, t the day and m the maturity of the contract. The weekly "commitment of traders" report is downloaded from the CFTC; see Appendix, Table I.A2.

We estimate the valuation effect of the roll with an event study. As already mentioned, this classic methodology is used in a non-standard context, and several differences are worth mentioning. First, the unit of analysis is a futures contract instead of a firm. Second, each unit is subject to the event (roll) at known intervals, up to 12 times per year for seven contracts. Third, the number of events is high with a minimum of 60 rolls (agricultural) to a maximum of 144 rolls (energy) over the sample period, while the number of contracts is small (19). Fourth, despite the growing literature on commodity asset prices, there is no well-accepted benchmark concerning commodity futures contracts; see, *e.g.*, Bakshi, Gao, and Rossi (2019), and Boons and Porras Prado (2019).

Following Henderson et al. (2015), the benchmark is a multi-factor model that describes price changes of individual futures contracts. The corresponding variables are downloaded from Thomson Reuters. These are the MSCI emerging market index, S&P 500 index, USD index, VIX, T-Bond, Baltic Dry Index, and inflation indices. The period over which CAPCs could materialize is not well identified either (pre-roll or roll). Therefore, we examine the prospectus of various ETFs and check

[14] The SP-GSCI contracts not under CFTC supervision have specificities that make them hard to compare with others. Moreover, CITs and indices providers are vague regarding the methodology they use to roll non-CFTC contracts (ICE-UK and LME).

how closely they follow the roll.[15] We observe that most of them stick to the SP-GSCI definition. In addition, we observe that the trading volume of the futures contract directly written on the SP-GSCI performance concentrates during the five days of the roll; see Appendix, Figure I.A1. To test Hypothesis 2, we estimate the valuation effects of the pre-roll (H2a) and roll (H2b) on the daily log price changes of individual nearby $\left(r_{c,t}^{N}\right)$ and first deferred $\left(r_{c,t}^{FD}\right)$ SP-GSCI/CFTC contracts. The following regression is estimated in a single pass,

$$r_{c,t}^{m} = \alpha_{0,c} + \alpha_{1,c} JAN_c \gamma_{c,t} + \sum_{r=1}^{R} PRECAPC_{c,r}^{m} \delta_{c,t}^{r} + \sum_{r=1}^{R} CAPC_{c,r}^{m} \Delta_{c,t}^{r} + {}^{\mathbf{T}}\mathbf{a_c}\mathbf{HPW_t} + \epsilon_{c,t}, \tag{2}$$

where $m = \{Nearby = N, Firstdeferred = FD\}$, c is the commodity contract, t is the time (day), $\gamma_{c,t}$ is a dummy variable equal to 1 if $\textit{January roll} - 5 \leq t \leq \textit{January roll} + 4$, r is the current roll, $\textit{date}\,(r)$ the date of roll r, $\delta_{c,t}^{r} = 0.20$ if $\textit{date}\,(r) - 5 \leq t \leq \textit{date}\,(r) - 1$, and 0 otherwise, $\Delta_{c,t}^{r} = 0.20$ if $\textit{date}\,(r) \leq t \leq \textit{date}\,(r) + 4$, and 0 otherwise, $PRECAPC_{c,r}^{m}$ and $CAPC_{c,r}^{m}$ capture the cumulative abnormal price change of contract c on pre-roll and roll r respectively, and $\mathbf{HPW_t}$ are the explanatory variables that capture the dynamics of commodities price changes.[16]

Yan et al. (2019) perform an event study on the annual index rebalancing ("January" roll) when the SP-GSCI weights experience their annual revision. Twenty-four SP-GSCI components are changing their respective index weights but seven of them are also rolling from the nearby to the first deferred. To avoid any confounding effect, we discard January CAPCs. We also estimate pre-roll CAPCs simultaneously as we expect them to be "abnormal".

Every month, 16 to 24 futures commodity contracts roll simultaneously. Therefore, the CAPCs are potentially cross-correlated because the error terms are not independent (missing variables in the benchmark for example). Ignoring this cross-correlation shrinks the standard errors of the CAPCs, which in turn leads to an over-rejection of the null hypothesis. To overcome this cross-correlation effect, we estimate the CAPCs every month for all the contracts, and keep the CAPCs of the contracts that are not rolling as counterfactuals; see Section 4.3.2. Eq. (2) is estimated as a Seemingly Unrelated Regression. This approach is also appropriate given the small number of observations in cross-section and the long time series; see

[15] See, *e.g.*, iShares S&P GSCI Commodity-Indexed Trusts (BlackRock), iPath® S&P GSCI® Total Return Index ETN (Barclays), VelocityShares, or Lyxor S&P GSCI (Société Générale).

[16] The benchmark is: $r_{c,t} = \beta_0 + \beta_{c,EM} r_{EM,t} + \beta_{c,EM(1)} r_{EM,t+1} + \beta_{c,S\&P} r_{S\&P,t} + \beta_{c,USD} r_{USD,t} + \beta_{c,TB} r_{TB,t} + \beta_{c,VIX} r_{VIX,t} + \beta_{c,BDI} BDI_t + \beta_{c,INF} INF_t + \beta_{c,lag} r_{c,t-1} + \epsilon_{c,t}$, where the variables are described in Henderson et al. (2015, p. 1293).

Karafiath (1994).

The vector $\mathbf{CAPC_{c,r}} = {}^{\mathrm{T}}[CAPC_{1,r}, \ldots, CAPC_{C,r}]$ gives the abnormal returns during the (pre) roll. We construct an equally (value) weighted portfolio $P = \{\text{equally-weighted, SP-GSCI-weighted}\}$ to obtain $CAPC_{P,r} = [w_{1,r}, \ldots, w_{C,r}] \times {}^{\mathrm{T}}[CAPC_{1,r}, \ldots, CAPC_{C,r}]$ and the average abnormal cumulative price change $ACAPC_P = \frac{1}{R}\sum_{r=1}^{R} CAPC_{P,r}$. Finally, we adjust the standard errors for event-induced variance and cross-correlation; see Boehmer, Musumecci, and Poulsen (1991), and Kolari and Pynnonen (2010). The variance of the (pre) roll is equal to $var\left(CAPC_{P,t}\right) = S^2_{P,t} = {}^{\mathrm{T}}[w_{1,t}, \ldots, w_{C,t}] \times \mathbf{\Omega} \times [w_{1,t}, \ldots, w_{C,t}]$ at time t where $\mathbf{\Omega} = [cov\left(\mathbf{CAPC_i}, \mathbf{CAPC_j}\right)]$ with $(i, j) \in [1 : C] \times [1 : C]$, and $CAPC_i = [CAPC_{i,1}, \ldots, CAPC_{i,R}]$. Under the null hypothesis, $CAPC_{P,r} \sim \mathcal{N}\left(0, \sigma^2_{P,r}\right)$ and $\frac{CAPC_{P,t}}{\sqrt{S^2_{P,t}/C}} \sim t_{C,t}$, where C is the number of contracts at time t. The central limit theorem gives the critical statistics $\frac{mean\left(t_{C,t}\right)}{\sqrt{var\left(t_{C,t}\right)/R}} \sim \mathcal{N}\left(0, 1\right)$ to test the null hypothesis $(ACAPC_P = 0)$, where R is the number of rolls.

3.3. The impact of the financialization on CAPCs

To check the impact of the financialization on CAPCs, we use a difference-in-differences regression. The treated CAPCs are those estimated for SP-GSCI contracts at the time of the roll. The controls are the CAPCs of SP-GSCI contracts that do not roll, *i.e.*, the nearby contracts with an expiry date beyond that of the next roll. We add six alternative contracts (ALT) that are not components of any major commodity index; see Appendix, Table I.A2. We are aware that the choice of non-treated contracts (ALT) may lead to a selection bias. In particular, the choice of the contracts included in the SP-GSCI depends on the liquidity and the importance of the commodity in the economy. These characteristics may differ from those of ALT contracts. However, because we limit our analysis to a short period (roll window), we hope that this bias is negligible with respect to the effect we target. As a robustness check, we use only the SP-GSCI contracts that do not roll on a given month as non-treated. The financialization was not only characterized by an increasing number of commodity index traders but also by a major change in the trading of commodities, *i.e.* the progressive introduction from 2006 to 2008 of electronic trading; see Raman et al. (2017). Non-treated contracts are the ones that are not yet traded electronically. To disentangle the respective effects of "electronification" and "financialization", we write the following baseline model:

$$\begin{aligned} CAPC^m_{c,r} = \mu_c + \beta_1 DROLL_{c,r} + \beta_2 DFIN_{c,r} + \beta_3 DELEC_{c,r} \\ + \beta_4 DROLL_{c,r} \times DFIN_{c,r} + \beta_5 DROLL_{c,r} \times DELEC_{c,r} + \epsilon_{c,r}, \end{aligned} \tag{3}$$

where $CAPC^m_{c,r}$ is a panel of nearby (first deferred) CAPCs at time r, μ_c is a contract fixed effect; $DROLL_{c,t}$ is a dummy equal to "1" for the contract rolling at r and to "0" otherwise; $DFIN_r$ is a dummy equal to "1" if $date(r) \geq$ *Break Date* and to "0" otherwise, with the Break Date estimated as described in Section 3.1.; $DELEC_{c,r}$ is a dummy equal to "1" when contract c is traded electronically, and to "0" otherwise. We estimate Eq. (3) as a Panel with contract fixed effects since the demand/supply of contracts depends on the underlying commodity. Given the small number of observations in cross-section, we do not introduce time fixed effect. The standard errors are clustered at the contract level to account for the contract characteristics (high dispersion of the variance across contracts). If the financialization affects CAPCs, the coefficient β_4 should be significantly different from zero. The coefficient β_4 captures the marginal effect of electronic trading on the rolling contracts. According to Hypothesis 3, β_5 is expected to be nil.

3.4. Front trading the roll

Following Mou (2011), we mimic the roll ten days before it starts. As he put it: [*Calendar spread positions are created on each day . . . , which runs from 10 to 6 business days before the SP-GSCI's first rolling date*]. The spread position is short of the maturing contracts in the SP-GSCI and long the deferred contracts that it will roll into. We assume that both predators and arbitragers are price takers because the timing of the roll is an essential feature for such agents. Transaction costs are decomposed into a fixed component (exchange commission and operational costs) and a variable component, the bid-ask spread. As a proxy for the bid-ask spread, we use the low-frequency estimator proposed by Abdi and Ranaldo (2017), and consider that the fixed component is negligible.

$$TC_{c,t} = \max\left[\sqrt{\frac{4}{N}\sum_{t=1}^{N}(C_{c,t} - M_{c,t})\,(C_{c,t} - M_{c,t+1})}; 0\right],$$

where N is the number of days, $C_{c,t}$ the closing price of contract c, on day t, and $M_{c,t}$ the difference between the "high" and "low" price of contract c, on day t. We compute this estimator for the nearby contracts. To test Hypothesis 4, we subtract transaction costs to the calendar spread for each commodity.

3.5. CAPCs, CIT, liquidity, and insurance demands

We rely on the weekly (from Tuesday to Tuesday) CFTC statistics that are released every Friday and aggregated over the term structure at the commodity level. Note that the beginning of the roll can be any day of the week. To estimate the change in price pressure resulting from the roll, we assume that CITs' positions are concentrated on the nearby contract before the roll, and fully transferred onto the first deferred contract at the end of the roll. Consequently, the change in CITs pressure is equal to minus the CIT position the last Tuesday before the roll.

The liquidity is the change in the "commercial net long position" over the roll, scaled by the total open interest of the last Tuesday before the roll. This variable is not directly affected by the roll if CITs fully report their positions from the nearby to the first deferred. We assume that changes in "commercial net long position" (liquidity demand) are uniformly spread over short periods, and not related to the roll. The average hedging pressure (insurance) is the 52-week moving average of the "commercial net long" position. The cross hedging pressures are computed within a group of commodities. We define four groups, *i.e.*, Agricultural, Energy, Metals, and Softs. On this ground, we use the following baseline panel regression specification:

$$CAPC^{m}_{c,r} = \gamma_1 CIT_{c,r} + \gamma_2 Q_{c,r} + \gamma_3 AHP_{c,r} + \gamma_4 QG_{c,r} + \gamma_5 AHPG_{c,r} + \mu_c + \epsilon_{c,r}, \quad (4)$$

where the subscripts c and r represent the contract, and the roll date, respectively. The dependent variable $CAPC^{m}_{c,r}$ is the abnormal cumulative price change of contract c, maturity m, and roll r; $CIT_{c,r}$ is the demand of CITs for the first deferred (nearby) contract;[17] $Q_{c,r}$ is the change in hedging pressure during the roll; $AHP_{c,r}$ is the 52-week moving average of the hedging pressure; $QG_{c,r}$ is the change in cross hedging pressure computed as the change in hedging pressure within the group and without the current contract; $AHPG_{c,r}$ is the 52-week moving average hedging pressure within the group and without the current contract. These variables are defined in the Appendix, Table I.A1. Eq. (4) is estimated with contract fixed effect, and variance clustering at the contract level.

Under Hypothesis 5, the coefficients γ_1 and γ_3 of Eq. (4) should be positive, γ_2 is expected to be negative, while γ_4 and γ_5 can be either positive or negative. Finally,

[17] With this definition, $CIT_{c,r}$ is negative (supply) for the nearby contract. According to Hypothesis 5, $CAPC^{N}_{c,r}$ should be negative. Therefore, γ_1 is also positive.

we check whether our results are sensitive to the financialization. For that purpose, we split the sample into two sub-samples based on the financialization date. We estimate Eq. (4) on both sub-samples.

4. Empirical results

4.1. Break test

Table I.1 reports the results concerning the existence of a common break affecting CITs' monthly market share of the SP-GSCI components. With the BIC computed over the full sample (and no break), we select a Vector Autoregression with five lags. The Bai et al. (1998) algorithm described in Section 3.1 identifies a break in December 2003.[18] The financialization materializes at a date close or equal to that retained by Sanders and Irwin (2011), Buyuksahin et al. (2009), and Boons et al. (2014) among others. When we restrict the sample to the CFTC contracts, the break date does not change. For robustness purposes, we repeat the test with the share of "commercial" long positions (SCL). A break is detected in November 2003, which is close to the one previously documented. When we turn our attention to six non-indexed CFTC (ALT) contracts, we observe a significant break in April 2008, long after the previous ones. Altogether, these results are in favor of a break that occurs at the end of 2003.

[**Insert Table I.1 here**]

The introduction of electronically traded futures contracts took place from August 2006 to October 2008. We test the null hypothesis (Hypothesis 1) that the inception of electronic trading dates lie inside the confidence interval of the break date. The corresponding Bonferroni test (19 simultaneous hypotheses) rejects the null hypothesis at the 1% level. We conclude that the financialization does not overlap with the electronification of individual commodity contracts. We are not aware of another alternative hypothesis that could explain the break documented here. Altogether, these results make us confident that the date identified as the financialization is not related to any other innovation at the exchange level.

4.2. Descriptive statistics

The composition of the index is remarkably stable over the 1999–2010 period. Two contracts, Orange Juice and Platinum show a decreasing weight in the index,

[18]At the 1% level, this translates into a six-day confidence interval around the break date. The critical values are available in Bekaert, Harvey, and Lumsdaine (2002, p. 244).

and from the 2004 January roll their weight is equal to zero. Table I.2, Panel A provides the descriptive statistics by contract for the 1999–2003 period. Column 1 reports the average of the nearby CAPCs. We observe a wide range of CAPCs from -96.17 bps (Coffee) to 496.61 bps (Natural Gas) with a median of 14.75 bps. In column 2, the first deferred CAPCs exhibit a very similar pattern, the CAPCs ranging from –67.42 bps (Feeder cattle) to 498.81 bps (Natural Gas), and 84.91 bps for the median.

[Insert Table I.2 here]

In column 3, the average change in the market share of CITs during the roll (CIT) varies widely from -33.83% (Lean Hogs) to -1.23% (Silver), and the median (-12.77%). It is interesting to note that, even before the financialization, the change in CITs' position represents an important part of the open interest. Not surprisingly, in column 4, we see that the change in hedging pressure during the roll (Q) has a lower magnitude with a minimum of -3.13% (Platinum), and a maximum of 2.45% (Coffee). An important part of the (uninformative) trading originates from CITs, probably over the two nearest maturities, while only a small part is generated by "commercials", with a non-identified maturity. Column 5 shows that the vast majority of the contracts exhibit positive hedging pressures (minimum = 9.58%, median = 6.58%, maximum = 44.16%). In column 6, the change of cross hedging pressure (QC) presents similar characteristics (minimum = -1.97%, median = -0.30%, maximum = 0.65%) to that of hedging pressure with a lower dispersion, however. Column 7 shows the average hedging pressure. It ranges from 2.30% (Heating Oil) to 0.83% (Orange Juice) with a median equal to 20.08%, and the dispersion is half that of the own average hedging pressure.

Table I.2, Panel B provides the descriptive statistics over the 2004–2010 period. To summarize, the financialization has two major consequences. First, the distribution of nearby and first deferred average CAPCs is shifted to the left, even when raw sugar (SB), an extreme value, is excluded. Second, as expected, the change in CITs pressure is (one and a half to two times) higher after the financialization. The remaining variables (change in hedging pressure, average hedging pressure at the contract level or the group level) are smaller and less dispersed than before.

The last point worth mentioning is the extremely high correlation of nearby and first deferred CAPCs. This correlation is equal to 97% for the 1999–2010 period, and it is stable over sub-periods. A similar result is obtained with the raw price changes.

4.3. Valuation effect

4.3.1. Univariate analysis

Table I.3 reports the results of the event study for the nearby and first deferred contracts. The 1999–2003 pre-roll ACAPC is equal to 102.87 bps (nearby), and 135.42 bps (first deferred). The corresponding t-stats are statistically significant at the 1% level when the t-stat is adjusted for heteroscedasticity and cross-correlation. In 2004–2010, the pre-roll ACAPC is equal to –65.47 (nearby), and -39.94 (first deferred) respectively. ACAPCs are not statistically significant at the 10% levels after adjusting for heteroscedasticity and cross-correlation. This entails a rejection of Hypothesis 2a in both the pre- and post-financialization periods.

[Insert Table I.3 here]

The ACAPC of the roll is 115.25 (146.36) bps for the nearby (first deferred), and the adjusted t stat is statistically significant at the 1% level. During the 2004–2010 period, the ACAPC is equal to -59.73 bps (nearby) and -40.70 bps (first deferred), and not statistically significant at the 10% level with the adjusted t-stat. To summarize, ACAPCs are statistically significant at the 1% level over the 1999–2003 period while there are not significant anymore over the 2004–2010 period (rejection of Hypothesis 2b).

We recompute the ACAPCs by weighting individual CAPCs with their respective index weights. Again, we observe similar qualitative results. We reiterate our analysis with raw (log) price changes, and the results are similar to those previously documented. Finally, we compute the ACAPCs with a non-parametric benchmark (peer contract written on a similar commodity but not a component of an index). We also used Bakshi et al. (2019) model as a benchmark. The results, available upon request, are virtually the same as those presented in Table I.3. Considered altogether, they support the sunshine trading hypothesis (Hypothesis 2b), at least in the post-financialization period. To summarize, the absolute value of the CAPC roll drops substantially after the financialization, which is positive for the market functioning.

4.3.2. Multivariate analysis

Table I.4 reports the results of the difference-in-differences regression. These results provide sharper support for the financialization effects associated with the commodities included in the SP-GSCI index. In columns 1 and 2, we estimate the baseline model with non-SP-GSCI and non-rolling contracts as controls. The difference-in-differences coefficient (β_4) is negative for both the nearby (-1.58) and the first deferred (1.66) contracts, and statistically significant at the 1% level. As in Bessembinder et al. (2016), the alleviating effect of the financialization also supports the sunshine trading hypothesis. The difference-in-differences coefficient related to electronic trading (β_5) is not statistically significant at the usual levels so that there is no evidence that the staggered introduction of electronically traded contracts affects the CAPCs; see Raman et al. (2017), and Martinez, Gupta, Tse, and Kittiakarasakun (2011). Consistent with Hypothesis 3, the financialization is responsible for the CAPCs drop.

[**Insert Table I.4 here**]

In columns 3 and 4, the controls are restricted to non-rolling contracts. We report similar results to the ones previously documented. We also estimate the model for the pre-roll week. The drop is lower in magnitude, but still statistically significant at the 5% level, with non-SP-GSCI and non-rolling contracts as controls. When only non-rolling contracts are considered the magnitude drops again as well as the statistical significance.

4.4. Arbitrage opportunities around the roll

Mou (2011) shows that the calendar spread starting ten days before the roll is the most rewarding strategy before transaction costs. We check whether significant profits remain when we consider reasonable (bid-ask spread on the long leg) or reduced transaction costs (a quarter of the bid-ask spread).

[**Insert Table I.5 here**]

We report the results in Table I.5. In column 1, the raw change in price is statistically significant (at the 5% level or less), for seven over 19 commodities. More specifically, the equally-weighted portfolio with the 19 contracts is rewarded 23 bps per roll, and statistically significant at the 1% level. However, this performance does not resist even mild transaction costs. In the pre-financialization period, only two (five) contracts have a positive performance after full (reduced) transaction costs that are statistically significant at the 10% (5%) level. In columns 8 and 9, the post-financialization period shows no contract with a positive performance after full transaction costs statistically significant at the usual level. Only two contracts have a positive performance with reduced transaction costs statistically significant at the 5% level (and one at the 10% level).

We find that the index investment industry bears an annual (average) economic cost of the roll that went from USD 0.8 billion before financialization to USD 2.66 billion after. However, this surge is largely attributable to the higher contract prices. These results depart from Mou (2011) who uses the crude oil contract minimum tick to estimate transaction costs. Nevertheless, our results are consistent with the "efficient market" hypothesis with frictions, as transaction costs are larger than CAPCs on average. They align with Stoll and Whaley (2010) hypothesis relating the reaction of the term structure to the bid-ask spread. Our results are also consistent with Bessembinder et al. (2016) who document narrower bid-ask spreads and increased liquidity during the roll of the U.S. Oil Fund. Finally, because we omit other components of the transaction cost and execution risks, our conclusion in support of Hypothesis 3 is further reinforced. However, we cannot discard the possibility of intraday predatory trading achieved by speculators ahead of indexed fund managers, penalizing the performance of CITs.

4.5. Explaining the CAPCs: A panel approach

Table 6 displays the results from OLS estimations of Equation (4) with variance clustered at the contract level.

[**Insert Table I.6 here**]

The first (second) column reports the results corresponding to the baseline specification for the nearby (first deferred) in the pre-financialization period. On one hand, the change in the net long position of "commercials" (Q) appears to be the main determinant of cumulative abnormal price changes for the nearby (first deferred) contract. The coefficient is equal to -34.35 (-33.72), and statistically significant at the 1% level. The economic magnitude of this variable is substantial. A one standard deviation increase is associated with a 1.42% (0.95%) decrease in the nearby (first deferred) CAPC. On the other hand, CIT is not significant at the 10% level for both the nearby and the first deferred. This result is not surprising by itself since CITs are not supposed to play a major role before the financialization. Nevertheless, the proportion of open interest that switches from the nearby to the first deferred is above 10%. This result can be interpreted as evidence that CITs' non-informative trades do not move prices while unexpected position changes do. The average hedging pressure, cross liquidity, and cross hedging pressure are not statistically significant at usual levels either. The post-financialization period shows contrasted results. As before, Q remains statistically significant at the 1% level. Its economic magnitude is more than double in the post-financialization. A one standard deviation increase is associated with a 3.22% (2.15%) decrease in the nearby (first deferred) CAPC. As before, CIT is not statistically significant at the 10% level. This reinforce the view that known trades do not generate abnormal price changes. The remaining variables (average hedging pressure, cross liquidity, and cross hedging pressure) are not statistically significant at the usual levels.

The change in net commercial position does not provide information on the group of investors generating CAPCs. The DCOT report is more informative since the variable Q is decomposed in four groups "hedgers" (producer/merchant/processor), "swap dealer", "money managers", and "other reportables". CITs are included in the "swap dealer" group, which allows us to analyze who is providing liquidity to the market. We estimate a simplified version of the previous model,

$$CAPC_{c,r}^{m} = \gamma_1 Q_{c,r}^{MM} + \gamma_2 Q_{c,r}^{SWAP} + \gamma_3 Q_{c,r}^{HEDGE} + \mu_c + \epsilon_{c,r}, \tag{5}$$

Table 7 displays results from OLS estimations of Equation (4) with variance clustered at the contract level.

[Insert Table I.7 here]

The coefficient of Q^{HEDGE} is equal to -54.02, and statistically significant at the 5% level. It captures most of the estimated coefficient of Q in the previous table. A one standard deviation decrease is associated with a 1.80% (1.90%) increase in the nearby (first deferred) CAPC. To summarize, producers reduce their hedging demand to provide liquidity to swap dealers, and possibly money managers. While the net change of "swap dealers" and "money managers" is of the opposite sign, it is not statistically significant at the usual levels. We interpret this result as evidence that the "swap dealers" and "money managers" categories are not homogeneous.

5. Conclusion

In this paper, we examine the consequences of the financialization of commodity futures markets. We identify a common structural break in the SP-GSCI components that occurs in December 2003. We study the abnormal price reaction of commodity contracts during the roll and the week before (pre-roll). We do find supportive results for the sunshine trading hypotheses in the post-financialization period, as the magnitude and the statistical significance of the cumulative abnormal price changes decrease. Hence, the financialization eases the activity of index investors, in the context of predictable trade awareness. We also question the size of the cumulative abnormal price changes, and relate them to the transaction costs bore by a price taker arbitrager. We show that these potential profits disappear as soon as reasonable transaction costs are accounted for. We regress cumulative abnormal price changes on changes in CITs pressure, changes of hedging pressure, average hedging pressure, changes in cross hedging pressure, and cross hedging pressure. We show that change in hedging pressure is an important determinant of abnormal price changes. This study reconciles two contrasting findings of the literature on commodity index investment. On the one hand, we document a significant shift in the risk sharing structure of commodity markets, with the entry of CITs. On the other hand, we show that it is unlikely that CITs have modified the term structure of the contracts that are involved in the roll. A related question, that we leave for future research, is whether the change in hedging pressure materializes at any time as in Kang et al. (2020) or only during the roll periods.

Table I.1: **Break detection in CITs' market share**

This table reports the break dates in the monthly CITs' market share series (SCIT). The break is determined with the Bai et al. (1998) algorithm. The sample covers the 1999—2010 period. Row 1 presents the results obtained with the 27 SP-GSCI contracts. In row 2, we reiterate the procedure with 21 SP-GSCI contracts only. We exclude contracts listed on the LME (Metals) for three main reasons: a) the roll is not well-defined, b) there is an unusual number of maturities for individual contracts, and c) the open interest of the contracts is not reliable. In row 3, only the 19 contracts covered by the CFTC are considered. Row 3 presents the break date for the "commercials" long positions (SCL) of the legacy report as an alternative proxy for CITs' market share series. Row 4 reports the results for six non-SP GSCI/CFTC contracts. The estimation involves vector auto-regression with the number of lags determined with the Bayesian Information Criterion. We test for a break in the intercept. We report the mean breakpoints as well as the 90%, 95%, and 99% confidence intervals for the break. We report the intercept increase (mean change) after the w.r.t. before the breakpoint in % and the Sup-Wald statistics as well as its significance level based on the asymptotic critical values obtained by simulation in Bekaert et al. (2002). *** indicates a significance level of 1%.

Variables	1999–2010							
	# Contracts	BIC	Break date	CI (10%)	CI (5%)	CI (1%)	Intercept ($\times$ 100)	Sup-Wald statistics
SCIT/SP-GSCI	27	5	Dec-03	0.00	0.08	0.25	6.17	204.78
SCIT/SP-GSCI	21	5	Dec-03	0.07	0.10	0.20	6.53	225.60
SCIT/SP-GSCI/CFTC	19	5	Dec-03	0.08	0.11	0.21	3.81	196.88
CL/SP-GSCI/CFTC	19	4	Mar-03	0.53	0.76	1.07	3.40	55.92
CL/non SP-GSCI/CFTC	6	1	Oct-06	1.71	2.45	4.40	0.40	34.54

Table I.2: **Descriptive statistics: CAPCs and pressure**

This table reports the cumulative abnormal prices change of the nearby and the first deferred during the roll, the CITs position of the Tuesday before the roll (CIT), net trading (Q), cross net trading (QG, trading of commodities in the same group without the current contract), the 52-week average hedging pressure (AHP), and the 52-week average hedging pressure (AHPG, hedging pressure of commodities in the same group without the current contract) of 19 SP-GSCI futures contracts under the CFTC supervision. The groups are Agricultural, Energy, Softs, and Metals. The periods of interest are 1999—2003 (before financialization) and 2004—2010 (after financialization). The variables are defined in the Appendix, Table I.A1.

	1999–2003								
Tickers	PRECAPC (N)	PRECAPC (FD)	CAPC (N)	CAPC (FD)	CIT	Q	AHP	QG	AHPG
C	87.15	95.37	82.32	87.50	-10.21	-0.68	-3.12	0.65	11.81
CC	10.16	25.78	24.87	40.00	-2.93	-0.27	-2.98	-0.69	5.11
CL	224.20	232.75	246.28	246.82	-21.89	-0.82	1.44	-1.84	6.01
CT	69.21	101.29	77.10	108.33	-11.99	-0.26	0.83	-0.31	6.14
FC	-63.93	-66.62	-62.70	-67.42	-25.38	0.45	-9.58	-0.13	8.56
GC	-17.83	-19.58	-19.00	-21.05	-4.59	-0.82	3.17	0.39	11.24
HO	206.09	277.61	216.42	287.41	-19.03	-1.31	11.20	-1.68	2.17
KC	-85.06	-56.14	-96.17	-62.60	-5.82	2.45	7.89	-0.28	4.81
KW	27.51	64.67	32.59	67.65	-16.41	-0.09	8.13	-0.07	7.48
LC	70.73	329.42	70.79	332.63	-19.65	-0.35	2.66	-0.26	7.71
LH	87.22	188.97	120.34	212.12	-33.83	-1.08	2.48	-0.25	7.73
NG	481.74	489.22	496.61	498.81	-7.38	-0.37	5.26	-0.97	4.17
OJ	40.66	37.69	33.89	28.74	-19.54	-0.26	14.48	0.25	6.57
PL	121.79	134.54	121.76	131.49	-10.55	-3.13	35.32	-1.67	-1.44
RB	-	-	-	-	-	-	-	-	
S	86.38	83.28	88.46	82.31	-5.31	-0.37	10.81	0.45	7.43
SB	100.36	82.85	97.37	62.40	-13.55	-0.33	10.16	-1.22	3.29
SI	-32.53	-35.25	-6.05	-7.81	-1.23	0.69	44.16	-1.97	5.22
W	72.17	85.43	77.96	93.73	-26.43	0.04	9.82	-0.33	4.04
Min	-85.06	-66.62	-96.17	-67.42	-33.83	-3.13	-9.58	-1.97	-1.44
Median	71.45	84.36	77.53	84.91	-12.77	-0.34	6.58	-0.30	6.07
Max	481.74	489.22	496.61	498.81	-1.23	2.45	44.16	0.65	11.81
	2004–2010								
C	84.40	78.36	30.11	24.38	-19.43	-0.06	4.59	0.35	13.23
CC	-0.96	-1.76	-56.59	-60.28	-10.49	0.29	18.17	0.09	1.98
CL	-125.49	-64.45	-162.55	-99.28	-48.66	0.02	2.40	-0.24	34.02
CT	-43.36	-38.70	-63.20	-57.43	-22.08	0.11	6.54	0.06	2.08
FC	-47.43	-50.33	-45.15	-51.63	-41.45	-0.26	-8.83	0.00	9.10
GC	-23.59	-25.38	-17.84	-19.81	-8.75	-0.22	39.37	0.01	18.03
HO	-90.69	-61.64	-128.31	-97.96	-40.35	0.35	7.95	0.01	2.40
KC	-82.70	-62.66	-132.49	-112.22	-13.83	0.52	14.29	-0.01	0.84
KW	50.54	42.94	15.16	7.13	-33.58	0.25	13.23	0.12	7.45
LC	-124.80	-92.74	-132.42	-102.98	-35.57	-0.33	-0.50	-0.08	7.72
LH	-104.40	-20.60	-103.72	-33.80	-39.20	-0.12	-2.89	0.03	7.54
NG	-30.11	-2.98	-59.82	-31.17	-16.76	0.08	-5.13	0.09	3.95
OJ	-	-	-	-	-	-	-	-	
PL	-	-	-	-	-	-	-	-	
RB	-68.64	-20.50	-89.33	-43.71	-25.96	-0.11	8.65	0.03	1.25
S	201.29	176.40	184.50	161.34	-13.40	-1.05	5.06	-0.10	7.14
SB	-556.03	-554.32	-641.55	-630.14	-19.47	0.28	17.45	-0.45	12.44
SI	-53.61	-53.67	-121.95	-122.64	-4.20	0.09	41.43	-0.56	40.47
W	8.05	9.47	-35.03	-32.84	-39.17	0.18	-6.43	-0.10	2.38
Min	-556.03	-554.32	-641.55	-630.14	-48.66	-1.05	-8.83	-0.56	0.84
Median	-47.43	-25.38	-63.20	-51.63	-22.08	0.08	6.54	0.01	7.45
Max	201.29	176.40	184.50	161.34	-4.20	0.52	41.43	0.35	40.47

Table I.3: **Event study around the roll of nearby SP-GSCI contracts**

This table reports the results of an event study based on the 19 SP-GSCI components covered by the CFTC around the pre-roll and the roll periods. The periods of interest are 1999—2003 (before financialization) and 2004—2010 (after financialization). Individual abnormal cumulative price changes (CAPCs) are computed with respect to the benchmark presented in Henderson et al. (2015, p. 1293). January rolls are excluded. Panel A reports the average cumulative abnormal price changes (ACAPC) of an equally- and a value-weighted portfolio (SP-GSCI weights), and the t-statistics with: a) no adjustments, b) corrected for event-induced variance (BMP), and c) corrected for event-induced variance and cross-correlation (KP). Panel B reports the results with raw (log) price changes. Statistical significance at the 10%, 5%, and 1% levels are indicated by *, **, and *** respectively.

	Equally-weighted				Value-weighted			
	1999–2003		2004–2010		1999–2003		2004–2010	
	Pre-roll	Roll	Pre-roll	Roll	Pre-roll	Roll	Pre-roll	Roll
	Panel A: HPW parametric benchmark							
Nearby								
	102.87	115.25	-65.47	-59.73	93.02	119.96	-52.99	-51.16
Unadj. t-stat	4.92	5.49	-3.66	-3.23	3.70	4.98	-2.62	-2.59
HAC	4.39	3.87	-2.58	-1.83	3.23	4.79	-2.20	-1.67
HAC and cross-correlation	3.63	3.84	-1.57	-1.26	3.40	3.81	-2.10	-1.52
First deferred								
	135.42	146.36	-39.94	-40.7	113.38	141.77	-27.27	-23.44
Unadj. t-stat	6.84	7.45	-2.34	-2.34	4.95	6.46	-1.41	-1.34
HAC	6.34	5.96	-1.33	-0.90	4.77	6.43	-0.97	-0.27
HAC and cross-correlation	5.08	5.05	-1.03	-0.92	4.42	4.90	-1.14	-1.03
	Panel B: Raw price changes							
Nearby								
	100.94	123.21	-64.42	-89.44	85.86	128.86	-57.34	-72.42
Unadj. t-stat	4.80	5.83	-3.25	-4.55	3.41	5.30	-2.38	-3.17
HAC	4.21	4.23	-1.92	-3.09	2.82	5.13	-1.69	-2.23
HAC and cross-correlation	3.55	4.06	-1.09	-1.50	3.26	4.00	-1.47	-2.37
First deferred								
	135.03	151.35	-38.43	-70.93	108.40	147.80	-31.04	-44.02
Unadj. t-stat	6.79	7.65	-2.03	-3.81	4.74	6.71	-1.33	-2.16
HAC	6.29	6.18	-0.70	-2.33	4.50	6.68	-0.61	-0.98
HAC and cross-correlation	5.07	5.14	-0.68	-1.26	4.40	4.98	-0.70	-1.95

Table I.4: **The Financialization and CAPCs during the roll**

This table reports the results of the difference-in-differences panel regression analyzing whether the cumulative abnormal price changes observed during the roll are altered after the financialization of commodities markets. The dependent variable, $CAPC$ is estimated with the Henderson et al. (2015). $DROLL_{c,t}$ is a dummy variable set to "1" ("0") when the commodity is (not) part of the SP-GSCI and do (not) roll. , is a dummy variable set to "1" ("0") in the post- (pre-) financialization period. $DFIN_t$, is a dummy variable set to "1" ("0") in the post- (pre-)electronification period. The controls are all CFTC contracts that do not roll (columns 1 and 3) or the SP-GSCI/CFTC contracts that do not roll (columns 2 and 4). The estimations include contract and year fixed effects. The sample period is 1999–2010. The estimations correct for within-contract error clustering. The robust standard errors are in parenthesis, and statistical significance at the 10%, 5%, and 1% levels are indicated by *, **, and *** respectively.

	Panel A: Roll period			
	CFTC N	CFTC FD	GSCI N	GSCI FD
DFIN	-0.379	-0.386	-0.408	-0.367
	(0.406)	(0.297)	(0.507)	(0.422)
DROLL	0.969**	1.309***	1.004*	1.352***
	(0.440)	(0.390)	(0.506)	(0.434)
DELEC	0.421	0.415*	0.576*	0.508*
	(0.246)	(0.207)	(0.297)	(0.270)
DROLL × DFIN	-1.579**	-1.657***	-1.553*	-1.678**
	(0.700)	(0.587)	(0.797)	(0.672)
DROLL × DELEC	-0.157	-0.243	-0.313	-0.336
	(0.376)	(0.374)	(0.394)	(0.396)
Constant	0.122	0.0547	0.00316	-0.0266
	(0.258)	(0.209)	(0.337)	(0.283)
Observations	3,540	3,540	2,676	2,676
R-squared	0.013	0.018	0.017	0.023
Number of contracts	25	25	19	19
	Panel B: Pre-roll period			
DFIN	-0.293	-0.363	-0.661	-0.438
	(0.325)	(0.261)	(0.431)	(0.424)
DROLL	0.927**	1.326***	0.731*	1.265***
	(0.350)	(0.346)	(0.417)	(0.399)
DELEC	-0.00297	0.0633	0.0978	0.0239
	(0.201)	(0.145)	(0.239)	(0.216)
DROLL × DFIN	-1.374**	-1.284**	-1.002	-1.207*
	(0.615)	(0.560)	(0.686)	(0.653)
DROLL × DELEC	0.0617	-0.225	-0.0425	-0.187
	(0.377)	(0.316)	(0.408)	(0.349)
Constant	0.247	0.0453	0.201	0.00423
	(0.262)	(0.222)	(0.369)	(0.346)
Observations	3,540	3,540	2,676	2,676
R-squared	0.011	0.017	0.016	0.022
Number of contracts	25	25	19	19

Table I.5: **Front running the index roll**

This table reports the performance of a calendar spread that front runs the SP-GSCI roll by ten days; see Mou (2011). Column 1 reports the raw performance, Column 2 accounts for the bid-ask spread of the day estimated with Abdi and Ranaldo (2017). Column 5 is the total number of rolls per contract during the corresponding sub-period. The RBOB gasoline contract inception is in November 2006. The number of observations is 645 and 953, respectively and statistical significance at the 10%, 5%, and 1% levels are indicated by *, **, and *** respectively.

	1999–2003					2004–2010				
	Raw returns (%) (1)	TC (%) (2)	Profit (%) (1) - (2) (3)	Profit (%) (1) − (2)/4 (4)	# rolls	Raw returns (%) (1)	TC (%) (2)	Profit (%) (1) - (2) (3)	Profit (%) (1) − (2)/4 (4)	# rolls # rolls
C	0.08	0.37	-0.29	-0.01	25	0.40***	0.51	-0.11	0.27**	35
CC	0.20	0.75	-0.55	0.01	25	0.18***	0.62	-0.44	0.03	35
CL	0.32***	0.68	-0.36	0.15	60	0.54***	1.02	-0.48	0.28*	84
CT	0.40*	0.62	-0.22	0.24	25	0.69***	0.62	0.07	0.54**	35
FC	-0.11	0.18	-0.29	-0.16	40	-0.15	0.25	-0.41	-0.22	56
GC	-0.11	0.02	-0.13	-0.12	30	-0.11	0.21	-0.33	-0.17	42
HO	0.10	0.90	-0.80	-0.13	60	0.41***	0.97	-0.56	0.17	84
KC	0.45***	0.95	-0.49	0.22**	25	0.11**	0.68	-0.57	-0.06	35
KW	0.39***	0.19	0.20*	0.34***	25	0.18***	0.88	-0.7	-0.04	35
LC	0.37	0.31	0.05	0.29	30	0.01	0.4	-0.39	-0.09	42
LH	1.17**	0.50	0.67	1.04**	35	0.33	0.61	-0.28	0.18	49
NG	0.99***	1.33	-0.34	0.66**	60	0.56***	1.36	-0.8	0.22	84
OJ	0.10	0.47	-0.43	-0.03	30	0.08	0.85	-0.67	-0.11	42
PL	-0.68	0.19	-0.87	-0.72	60	3.59	0.54	3.05	3.46	84
RB	N A	NA	N A	N A	60	0.04	0.59	-0.56	-0.11	84
S	-0.01	0.49	-0.50	-0.13	35	-0.33	0.71	-1.05	-0.51	49
SB	0.74	0.82	-0.08	0.53	25	-0.05	0.74	-0.79	-0.24	35
SI	-0.02	0.04	-0.05	-0.02	30	-0.08	0.37	-0.45	-0.17	42
W	0.58***	0.40	0.18*	0.48***	25	0.35***	1.08	-0.73	0.08	35
Agriculture	0.34***	0.35	-0.01	0.26**	215	0.08	0.6	-0.52	-0.07	301
Energy	0.35***	0.73	-0.38	0.17*	240	0.39***	0.98	-0.6	0.14*	336
Metals	-0.37	0.11	-0.48	-0.40	120	1.75	0.42	1.33	1.65	168
Softs	0.38	0.71	-0.35	0.20	130	0.21**	0.71	-0.47	0.04	182
Equally-weighted	0.23***	0.50	-0.28	0.10	705	0.50**	0.72	-0.22	0.32	987

Table I.6: **CAPC, CITs pressure, liquidity and insurance during the roll**

This table reports the results of a panel regression examining whether cumulative abnormal price changes observed during the roll depend on CITs, and short-term changes in price pressure, and the long-term in own (cross) price pressure. The dependent variable, $CAPC$ is an estimate of the change in prices in excess of the Henderson et al. (2015) model. The model is estimated for three sub-periods, *i.e.*, 1999–2003, the pre-financialization period, and 2005–2010, the post-financialization period; 2004 is excluded because two variables (AHP and AHPG) are computed with pre-financialization information. CIT is the position of CITs the Tuesday before the roll divided by the open interest estimated with the Masters and White algorithm (1999–2003, 2005–2010). Q is the change in net long positions of "commercials" divided by the open interest of the contract. AHP is the 52-week average of the (minus) "commercials" net long position divided by the open interest of the contract. QG and AHPG are computed as above over a group of contracts (Agricultural, Metal, Energy, and Softs) to which the current contract belongs without the current contract itself. The variables are defined in the Appendix I.A1. The specifications contain a contract fixed effect, and estimated within-contract error clustering. Robust standard errors are in parenthesis and statistical significance at the 10%, 5%, and 1% levels are indicated by *, **, and *** respectively.

	Nearby		First Deferred	
	1993–2003	2005–2010	1993–2003	2005–2010
	(1)	(2)	(3)	(4)
CIT	-0.866	-1.600	-2.177	-1.765
	(2.678)	(1.879)	(2.556)	(1.811)
Q	-30.04***	-67.99***	-28.59***	-64.84***
	(9.723)	(11.60)	(8.983)	(10.88)
AHP	0.619	3.736	1.388	3.116
	(0.999)	(2.156)	(0.822)	(2.199)
QG	-2.018	-5.538	-1.440	-7.345
	(5.031)	(4.619)	(4.839)	(4.395)
AHPG	0.541	-3.382	0.617	-1.634
	(1.324)	(3.684)	(1.392)	(3.154)
Constant	0.428	-0.345	0.429	-0.398
	(0.357)	(0.725)	(0.328)	(0.703)
Observations	575	676	575	676
R-squared	0.165	0.181	0.185	0.186
Number of contracts	18	17	18	17

Table I.7: **Cumulative abnormal price change during the roll and position changes**

This table reports the results of a panel regression examining whether the cumulative abnormal price changes observed during the roll depends on the change of net long positions of money managers, swap dealers, and hedgers. The model is estimated for the 2006–2010, when the DCOT reports are available. The variables are defined in the Appendix Table I.A1. The specifications contain a contract fixed effect, and are estimated using within-contract error clustering. Robust standard errors are in parenthesis. Statistical significance at the 10%, 5%, and 1% levels are indicated by *, **, and *** respectively.

	2006–2010	
	Nearby	First Deferred
QMM	13.86	7.05
	(22.77)	(18.99)
QSWAP	50.03	35.56
	(35.10)	(32.29)
QHEDGE	-54.02**	-58.00***
	(21.28)	(17.58)
Constant ($\times 10^2$)	-4.18***	8.34***
	(1.35)	(1.27)
R-square		
Within	0.235	0.237
Between	0.230	0.332
Overall	0.238	0.240
Observations	523	523
Number of contracts	17	17

Appendix

Table I.A1: **Variable definition**

We define the variables, their units, frequency, period of availability and additional descriptions if needed. The variables are computed over the 1999–2010 period.

Variable	Computation	Description	Source
Futures price	$F^m_{c,t}$	Price of a futures contract c, at time t, maturing at m.	Barchart
Futures price change	$r^m_{c,t} = \ln(F^m_{c,t}) - \ln(F^m_{c,t-1})$		
Cumulative abnormal price change	$CAPC^m_{c,w}$	CAPC of contract c, maturing at m, over the roll period r.	
Open interest per contract	$OI_{c,t}$	Open interest in USD of contract c at time t	CFTC Legacy report.
Total open interest	$TOI_t = \sum_c OI_{c,t}$		
Total SP-GSCI investment	See below		
Amount tracking SP-GSCI contracts	$VCIT_{c,t}$	See Masters and White (2008, p. 45-51)	
Market share of SP-GSCI contracts	$SCIT_{c,t} = {}^{VCIT_{c,t}}/_{OI_{c,t}}$	Market share of CITs on contract c, at time t.	
Market share of “commercials” long	$SCL_{c,t} = {}^{CL_{c,t}}/_{OI_{c,t}}$	Market share of “commercials” long on contract c, at time t.	
“Commercials” long	$CL_{c,t}$	“Commercials” long on contract c, at time t.	CFTC legacy report.
“Commercials” short	$CS_{c,t}$	“Commercials” short on contract c, at time t.	CFTC legacy report.
“Commercials” net long	$CNL_{c,t} = CL_{c,t} - CS_{c,t}$	“Commercials” net long on contract c, at time t.	
Hedging pressure	$HP_{c,t} = {}^{-CNL_{c,t}}/_{OI_{c,t}}$	Hedging pressure on contract c, at time t.	
Cross hedging pressure	$HPG_{c,t} = -\frac{\sum_{g=\{G\}-c} CNL_{g,t}}{\sum_{g=\{G\}-c} OI_{g,t}}$	Hedging pressure on contract c, at time t, of group g.	
Short term change in hedging pressure	$Q_{c,t} = {}^{(CNL_{c,t}-CNL_{c,t-1})}/_{OI_{c,t-1}}$	Change in hedging pressure on contract c, from $t-1$ to t.	
Average hedging pressure	$AHP_{c,t} = \frac{1}{52} \sum_{\tau=t-51}^{t} HP_{c,\tau}$	52-week average hedging pressure.	
Cross short-term	$QG_{c,t} = \frac{\sum_{g=\{G\}-c} (CNL_{g,t}-CNL_{g,t-1})}{\sum_{g=\{G\}-c} OI_{g,t}}$		
Average cross hedging pressure	$AHPG_{c,t} = \frac{1}{52} \sum_{\tau=t-51}^{t} HPG_{c,t}$	52-week average cross hedging pressure.	

Table I.A2: **Contract characteristics**

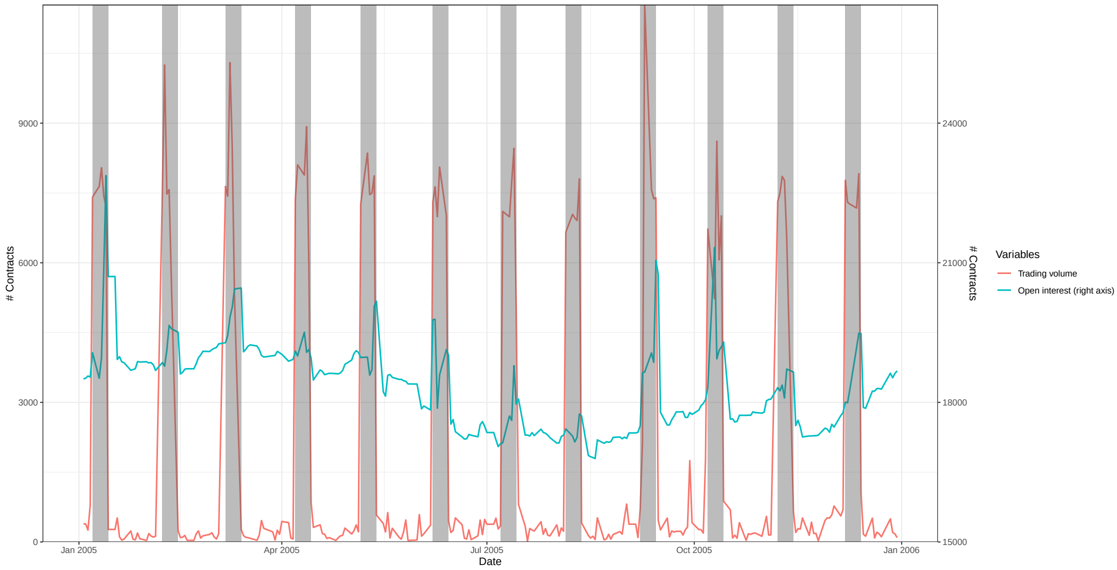
The Table reports the specifications of 26 contracts covered by the CFTC; 19 contracts are included in the SP-GSCI, and 7 are not. The reported characteristics are the trading venues, tickers, underlying commodities, units, maturity months with the appropriate code letter, and the inception date of the corresponding electronically traded contract.

Commodity	Trading venue	Ticker	Unit	Maturity	Electronification
		Panel A: SP-GSCI contracts			
Corn	CBT	C	bu (5,000)	HKNUZ	01/08/2006
Cocoa	ICE-US	CC	MT (10)	HKNUZ	02/02/2007
WTI crude oil	NYMEX/ICE	CL	bbl (1,000)	FGHJKMNQUVXZ	05/09/2006
Cotton	ICE-US	CT	lbs (50,000)	HKNVZ	02/02/2007
Feeder cattle	CME	FC	lbs (50,000)	FHJKQUVX	24/11/2008
Gold	CMX	GC	ozt (100)	GJMQVZ	04/12/2006
Heating oil	NYMEX	HO	gal (42,000)	FGHJKMNQUVXZ	05/09/2006
Coffee	ICE-US	KC	lbs (37,500)	HKNUZ	02/02/2007
Kansas wheat	KBT	KW	bu (5,000)	HKNUZ	13/01/2008
Live cattle	CME	LC	lbs (40,000)	GJMQVZ	04/12/2006
Brent crude oil	ICE-UK	LCO	bbl (1,000)	FGHJKMNQUVXZ	12/06/2006
Gasoil	ICE-UK	LGO	MT (100)	FGHJKMNQUVXZ	12/06/2006
Lean hogs	CME	LH	lbs (40,000)	GJMNQVZ	04/12/2006
Natural gas	NYMEX/ICE	NG	MMBtu (10,000)	FGHJKMNQUVXZ	05/09/2006
Orange juice	ICE	OJ	lbs (15,000)	FHKNUX	02/02/2007
Platinum	NYMEX	PL	ozt (50)	FGHJKMNQUVXZ	04/12/2006
RBOB gasoline	NYMEX	RB	gal (42,000)	FGHJKMNQUVXZ	05/09/2006
Soybeans	CBT	S	bu (5,000)	FHKNQUX	01/08/2006
Raw sugar	ICE-US	SB	lbs (112,000)	HKNVZ	02/02/2007
Silver	CMX	SI	ozt (5,000)	FHKNUZ	04/12/2006
Wheat	CBT	W	bu (5,000)	HKNUZ	01/08/2006
		Panel B: Non-indexed contracts			
Minneapolis Wheat	CME	MWE	bu (50,000)	HKNUZ	15/12/2004
Lumber	CME	LB	m^3 (110,000)	FHKNUX	20/10/2008
Class III Milik	CME	DC	lbs (200,000)	FGHJKMNQUVXZ	17/09/2007
Oats	CBOT	O	bu (5,000)	HKNUZ	01/08/2006
Palladium	NYMEX	PA	ozt (100)	HMUZ	04/12/2006
Rough rice	CBOT	RR	CWT (2,000)	FHKNUX	01/08/2006

Maturity month code: F = January, G = February, H = Mars, J = April, K = May, M = June, N = July, Q = August, U = September, V = October, X = November, Z = December.

Fig. I.A1: **Volume concentration during the roll**

We plot the total daily trading volume and OI of the futures contract directly written on the SP-GSCI performance. The shaded areas correspond to the 12 roll periods of the -randomly selected-year 2005. The frequency is daily.



Chapter II

Do economic variables forecast commodity futures volatility?

1. Introduction

The realized volatility framework introduced by Andersen and Bollerslev (1998a,b) delivers higher forecast accuracy than generalized autoregressive conditional heteroskedasticity (GARCH) or stochastic volatility models for spot and futures prices of stocks, bonds and currencies. Yet, the literature on the realized volatility of commodity futures is scarce. While commodity futures share many commonalities with financial assets, the physical nature of their underlyings introduces specificities such as inventory, consumption, or decay. Hence, I ask whether the introduction of economic variables (EVs), which characterize individual commodity futures contracts, helps to model their realized volatility (RV).

I start the analysis with a regression model that tests whether seasonality (uncertainty resolution), time to maturity (Samuelson conjecture), and the slope of the term structure (a proxy for inventories), explain commodity futures RV. The empirical analysis is based on three groups of commodities (agriculture, energy, and metals). I select the three most liquid commodity contracts in each group, over the 2008–2019 period, and estimate the system of equations with a Seemingly Unrelated Regression (SUR), to account for omitted variables common to commodity contracts.[1] I capture seasonal variations of RV with dummies reflecting critical months. In line with economic predictions, this variable is statistically significant at the 1% level for agricultural products (seasonal crops) and natural gas (higher heating demand in winter). I also study the time to maturity, a controversial variable related to the functioning of futures markets. The time to maturity positively determines the commodity futures RV, with statistical significance at the 1% level (for eight out of the nine commodities). Lastly, consistent with the theory of storage, I show that the slope of the term structure is related to the RV. More specifically,

[1]The agriculture contracts are corn, soybeans, and wheat. The energy contracts are crude oil, heating oil, and natural gas. The metal contracts are gold, copper, and silver.

the absolute value of the slope shows a positive relation with RV that is statistically significant at the 1% level for the three energy contracts and copper.

Given the strong and long-lasting autocorrelation of RV, I introduce several autoregressive specifications in the baseline model. I find that the explanatory power of EVs is robust to this inclusion. I first introduce the heterogeneous autoregressive RV (HAR); see Corsi (2009). The motivation for using the HAR is based on the econometric performance (both in- and out-of-sample), and justified by the general long-term memory that commodity futures RV exhibit. I show that the explanatory power of EVs is not altered since EVs are altogether statistically significant at the 1% level. I repeat this test with two alternative competing models: the HEXP in which the lagging terms are exponentially smoothed, and the HARQ which accounts for measurement errors; see Bollerslev et al. (2018) and Bollerslev et al. (2016) respectively. I find that the explanatory power of EVs remains when these alternative specifications are considered. Finally, I test the baseline model nested with autoregressive and measurement errors variables in time-varying (TV) parameter models (HAR-TV and HARQ-TV see, *e.g.*, Chen et al., 2018; Casas, Ferreira, and Orbe, 2019). The explanatory power of EVs is also robust to these specifications. I carry on with a horse race of all competing models, that nest EVs or not, and compare their in- and out-of-sample forecast accuracy at three different horizons. In-sample, the best performing model for one-day ahead forecasts is the HARQ-TV, whereas the HAR-TV outperforms for the one-week and one-month horizons. Out-of-sample, the HARQ dominates when the forecast horizon is lower than one-week, whereas the EVHARQ-TV produces the best forecasts at the one-month horizon. Comparing one to one the one-day ahead forecasts of each model, I find that EVs yield lower errors at the 1% level, when they are introduced in the HEXP, HAR-TV, and HARQ-TV, compared to their constrained version. I use these out-of-sample forecasts in multi-quantile VaR regressions, and extract expected shortfall for up to four coverage levels. I jointly test the parameters for the forecast bias, and uncover that the simple autoregressive models (HAR and EVHAR) provide the lowest rejection rates.

The contribution of this article is threefold. First, commodity futures are widely overlooked in the RV literature in comparison to other asset classes. To fill this void, I explore the performance of RV models that include theoretically motivated EVs. Second, I show that the addition of EVs to autoregressive models improves their explanatory power, and the accuracy of their forecasts. Since the RV provides a better approximation of the true latent volatility process than the (G)ARCH and stochastic volatility approaches, the reexamination of such EVs in this context is particularly relevant; see Andersen and Bollerslev (1998b). Finally these tests also highlight the contracts peculiarities in terms of RV.

The remainder of the article proceeds as follows. Section 2 develops the literature review on RV and, more specifically, its economic determinants. Section 3 presents the research design, including the data organization and models to test. Section 4 discusses the empirical results. Section 5 compares the performance of out-of-sample forecasts from both autoregressive and EV models in tail risk management. Section 6 concludes.

2. Previous research and hypotheses

The daily realized variance is defined as:

$$RV_{t+1}^{2}(\Delta) \equiv \sum_{j=1}^{1/\Delta} \left(r_{\Delta,t+j\times\Delta}^{2} \right), \tag{1}$$

where $1/\Delta$ is the number of observations in one day and r_{Δ}^{2} represents the squared change in intraday (log) prices; see, *e.g.* Andersen, Bollerslev, and Diebold (2007). Andersen and Bollerslev (1998a) show that the realized variance is the limit of the integrated variance, as the frequency increases to infinity:

$$\lim_{\Delta\to 0} RV_{t+1}^{2}(\Delta) \to \int_{t}^{t+1} \sigma^{2}(s)\, ds + \sum_{t<s\leq t+1} \kappa^{2}(s) \tag{2}$$

The right hand side of Eq. (2) is the integrated variance of the diffusion process σ, and discrete jumps of size κ. Disentangling jumps from the continuous process is an empirical issue; see, *e.g.*, Aït-Sahalia (2002, 2004).

Generally, RV (square root of the realized variance) is computed using constant time-intervals ranging from 1 to 30 minutes; see, *e.g.*, Aït-Sahalia, Mykland, and Zhang (2005); Andersen, Bollerslev, and Meddahi (2011b). Patton and Sheppard (2015) use consecutive transactions. When 5-minute RV is taken as the benchmark, Liu, Patton, and Sheppard (2015) find little evidence that it is outperformed by any other measure. However, when using inference methods that do not require to specify a benchmark, there is some evidence that more sophisticated measures outperform. For example, Andersen et al. (2011b) propose two estimators of particular interest: a) the average estimator and b) the optimal estimator. The linear forecasts obtained by averaging standard sparsely sampled RV measures generally perform on par with the best alternative robust measures. Overall, 5-minute RV is difficult to beat.

2.1. *Stylized facts*

Changes in consecutive (log) prices of financial assets, including stock, bonds and currencies, present common characteristics that are also shared by commodities futures:

a. standard deviation dominates the mean over daily and weekly return horizons
b. daily, weekly, and monthly horizons show excess kurtosis *vs.* a normal distribution
c. squared and absolute returns are strongly autocorrelated
d. there are periods of high volatility (volatility clustering)
e. outliers and jumps are more frequent than they should be (*vs.* a normal distribution).

Two main differences between commodity and stock index futures have been documented. The first noticeable discrepancy is the inverse asymmetric reaction between commodity futures price and volatility, *i.e.* the "inverse leverage effect", arising from shocks on inventories. Typically, when the resources are scarce, the supply on the corresponding market becomes inelastic. A decrease of one unit in inventory leads to a dramatic price upward revision; see Ng and Pirrong (1994); Carpantier (2010); Carpantier and Dufays (2012); Carpantier and Samkharadze (2012).

The second difference with other financial assets has to do with the underlying stochastic process that generates price changes. Commodity futures price changes are positively skewed and, contrary to stock returns, this skewness strongly shows up at the contract level; see, *e.g.*, Gorton and Rouwenhorst (2006).

2.2. *The economic determinants of RV*

Anderson and Danthine (1983) hypothesize that the key determinant of volatility is the time at which the production uncertainty is resolved. The uncertainty resolution is seasonal, for instance at the end of a crop when the supply is publicly known; see also Anderson (1985). This seasonality should be particularly visible for agricultural products whose production are concentrated in a single annual harvest in the northern hemisphere.[2] It should also be present for the natural gas contract since the demand rises every winter in the northern hemisphere. Despite the fact

[2]See the statistics from the US Department of Agriculture: `https://www.nass.usda.gov/`. Although soybeans production is also very large in the southern hemisphere, the underlying product specifications and delivery locations of the contract studied are all in the northern hemisphere.

that these turning points should primarily affect the cash market, Anderson and Danthine (1983) additionally show that the link between the cash and futures markets ensures the volatility diffusion from the former to the latter. The research also shows that intangible commodities like electricity or those whose exchange value is higher than their consumption value, such as gold or silver, behave more like traditional financial assets. Anderson (1985); Khoury and Yourougou (1993); Galloway and Kolb (1996) find a seasonal component in volatility, combined with a time to maturity effect.

Samuelson (1965, 1976) conjectures that the volatility of commodity contracts is higher when the remaining time to maturity is lower. Despite many empirical tests, the results are contradictory. On the one hand, Rutledge (1976) and Grammatikos and Saunders (1986) do not find evidence of any increase in volatility. On the other hand, Milonas (1986), and Galloway and Kolb (1996) find support for all commodities. Consistent with the Samuelson hypothesis, Bessembinder, Coughenour, Seguin, and Monroe-Smoller (1996) develop a model in which the spot price has negative covariance with the slope of the term structure. This implies a temporary price change, which is more likely to occur in real assets than in financial assets. Indeed, recent empirical tests on the NIKKEI (Chen, Duan, and Hung, 2000) and on the S&P 500 futures (Moosa and Bollen, 2001) strongly reject the Samuelson conjecture, whereas Bessembinder et al. (1996) find empirical support mainly for agricultural commodity futures.

The theory of storage states that the relation between the volatility of storable commodities and the level of inventories is convex and negative; see, *e.g.*, Working (1933); Kaldor (1939); Brennan (1958). More recent versions of the theory of storage in equilibrium (*e.g.* Deaton and Laroque, 1992) also predict this link, which is confirmed empirically; see Fama and French (1988); Ng and Pirrong (1994); Geman and Nguyen (2005); Geman and Ohana (2009); Carpantier and Samkharadze (2012). Inventories are difficult to measure at the aggregate level with a daily frequency. At the monthly frequency however, Gorton et al. (2012) find empirical evidence of a negative relation between actual inventories and the spot price volatility. More importantly, they confirm the tight link between the term structure and inventories. Thus, because the term structure is readily measurable at any frequency, this variable allows to test inventory-related hypotheses at the daily frequency. Kogan, Livdan, and Yaron (2009) extend the theory of storage prediction to a non-monotonic and convex relation between volatility and inventories. They first document empirically that the volatility increases when the inventory levels approach their physical limits (empty or filled storage). To explain this "v-shape" pattern, they derive a model that links the volatility to investment constraints through the capacity of firms to absorb demand shocks. They introduce the slope of the term structure conditioned on its sign in GARCH models, and find that the corresponding coefficients are statistically significant at the 1% level; see also Haugom, Langeland, Molnar, and Westgaard (2014). To conclude, both low and high inventories lead to high volatility. Consequently, I state my hypotheses as follows. The volatility of commodities futures:

a. is seasonal for commodities that show seasonality in the supply or the demand
b. increases when the time to maturity decreases
c. is positively related to the intensity of both low and high inventory states.

2.3. Endogenous determinants of RV

The main idea of Corsi (2009) is that the RV at time t depends on past values of the RV at time $t-1$, $t-2$, $\ldots$, $t-p$ where p can be very high (20 or more), suggesting a long-memory process. However, this process is mean reverting toward a long-term component. Therefore, the transitory component of the daily variance relates to the RV at $t-1$ and the introduction of two additional components (weekly and monthly RVs) smooths its dynamics. Altogether, these variables give a parsimonious representation of the typical volatility exponential decay; see, *e.g.* Andersen, Bollerslev, Diebold, and Labys (2003). In the empirical part of the paper, Corsi (2009) estimates the model with the S&P 500 index, the USD/CHF exchange rate, and a 30-year US T-Bond futures. Based on the BIC criterion, the one-day ahead in-sample performance of this model is higher than that of an AR (22), which clearly shows that the HAR (3) model is parsimonious. Out-of-sample, the model steadily outperforms the short-memory models (AR (1) and AR (3)) at the one-day, one-week, and one-month horizons. In addition, it is on par with an (long-memory) ARFIMA model. Lastly, the superior performance of the ARFIMA and HAR (3) increases with the forecasting horizon.

Several versions of the model have been proposed using the RV, its log, and its square. Andersen et al. (2007) show that the log of RV is the closest to normality and that jumps are negligible in terms of RV forecasting. Microstructure effects could introduce measurement errors, which lead to biased coefficients. Nevertheless, the residuals of log RV are still heteroskedastic, and the parameters of the HAR are not constant over time; see Buccheri and Corsi (2019). Using a simple linear process could be insufficient for at least three reasons: a) jumps, b) measurement errors, and c) time-varying parameters. Andersen, Bollerslev, and Huang (2011a) consider that RV has (i) a continuous component for the day that is well described by an HAR-GARCH model, (ii) a jump component for the day, and (iii) a GARCH component for the night, leading to the HAR-CJN model. However, the out-of-sample performance of this model is just slightly higher than that of the HAR.

Given the measurement error that plagues the estimation of RV, Bollerslev et al. (2016) introduce the "realized quarticity" (variance of the RV) in the HAR model. The authors write an extension (HARQ model) where the coefficients are a linear function of the quarticity. The idea is to put less weight on past high values of RV when they are subject to potential mismeasurement. By the same token, this variable is supposed to capture microstructure effects and jumps. However, HARQ also shows signs of misspecification. As an alternative, Corsi and Reno (2012), and Patton and Sheppard (2015) examine whether the RV reacts symmetrically to positive and negative shocks that affect prices, *i.e.*, the so-called "leverage effect". Casas, Mao, and Veiga (2018) nest both models. Cipollini, Gallo, and Otranto (2017) show that HARQ is observationally equivalent to another model where a quadratic term in RV accounts for a faster mean reversion when volatility is high. They argue that the realized quarticity and a time-varying mean seem to play a more important role than measurement errors. In these models, the time-varying coefficients are linear functions of the realized quarticity (parametric specification). Chen et al. (2018) generalize this approach by considering a log HAR model with time-varying coefficients of unspecified functional forms (HAR-TV). These coefficients are approximated with a local linear function of time. Casas et al. (2018) extend Chen et al. (2018) in two directions. First, they consider a potential asymmetric reaction to negative shocks. Second, the coefficients are no longer a local linear function of time but a linear function of the realized quarticity (semi-parametric approach). Unfortunately, the forecasting performance of the RV is not examined specifically since the main purpose of the paper is to forecast the stock market.

Bekierman and Manner (2018) take a different stance. They propose a state-space representation of the HAR model that can be augmented by functions of the realized quarticity. They attribute the higher performance of the state-space HAR models to the fact that the realized quarticity is a noisy proxy for the true measurement error, which is likely to be greater in periods of high volatility. Furthermore, their state-space models are able to capture other sources of time variation in the parameters that are not explained by the measurement error. Buccheri and Corsi (2019) generalize the state-space representation approach in several directions. Their state-space model allows for a time-varying error, and considers that the updated parameters depend on the level of uncertainty that is based on the score function. This model (SHARK) appears to perform very well, both in- and out-of-sample, but there is no straightforward extension for multivariate estimations.[3]

Overall, significant progress has been made with the development of sophisticated specifications. Their main purpose is to clean the data from microstructure effects, and to account for an asymmetric reaction to negative exogenous shocks. Yet, the focus of empirical applications has been on stock indices, currencies, individual stocks, bonds, and less frequently on futures contracts. The following methodology details the implementation of the aforementioned benchmark models to test the contribution of EVs.

3. Methodology

I start with a linear model that incorporates the EVs discussed in Section 2.2. Then, I test whether these economic determinants have explanatory power beyond that of the past realizations of RV. Finally, I test several specifications allowing coefficients to vary over time. All specifications are systems of equations (SUR), estimated with FGLS. Therefore, they account for the contemporaneous correlation of error terms across equations induced by potential omitted variables common to all contracts.

3.1. Baseline model

To check whether the economic determinants of RV have any explanatory power, three EVs observable at the daily frequency are considered. First, I introduce monthly dummies for the corn, soybeans, wheat, and natural gas contracts, to test for seasonal effects (uncertainty resolution hypothesis). The monthly dummies are set in July for the agricultural products, which corresponds to the harvest month of the soft red winter wheat contract in the US, and more generally for winter wheat in the northern hemisphere. It also corresponds to the "filling" month for corn and soybeans in the northern hemisphere, which is more critical than the subsequent harvesting months. For the natural gas, I select January which corresponds to the coldest month, and to the highest consumption month in the US, historically. This choice also matches the unconditional seasonal pattern of RV that is displayed in Figure II.1.[4]

[3] Multivariate RV series modeling is out of the scope of this paper.

[4] For supply-related information of agricultural products see, https://ipad.fas.usda.gov/countrysummary/. For demand-related information of the natu-

[Insert Figure II.1 here]

Second, I introduce the log of the time to maturity (Samuelson hypothesis), crossing the timestamp with the contract maturity information available in the full ticker.[5] I set it on a calendar basis, since I assume that the latent maturity information exists even when futures are not traded.[6] Finally, consistent with the theory of storage, I consider the slope of the nearest term structure. Since the maturity gap differs across contracts, I normalize the slope measured for each contract by the maturity gap.[7] More specifically, I test the "v-shape" hypothesis (Kogan et al., 2009; Haugom et al., 2014) by adding a "backwardation" dummy set to "1" when the slope of the nearest term structure is negative, and "zero" otherwise. Its interaction with the slope allows to capture simultaneously the inventory effects related to either the "v-shape" or the traditional theory of storage. The following model is estimated:

$$\begin{split} RV_{c,t} = \alpha_{0,c} + \alpha_{1,c}M_{c,t} + \alpha_{2,c}TM_{c,t} + \alpha_{3,c}SL_{c,t-1} + \alpha_{4,c}B_{c,t-1} + \\ \alpha_{5,c}B_{c,t-1} \times SL_{c,t-1} + \epsilon_{c,t}, \end{split} \tag{3}$$

where $M_{c,t}$ is a dummy equal to "1" during the critical month of the corresponding contract and to "0" otherwise, $TM_{c,t}$ is the (log) time to maturity, $SL_{c,t-1}$ is the annualized (log) term structure slope between the nearby and first deferred contract, and $B_{c,t-1}$ is a dummy equal to "1" ("0") when this slope is in backwardation (contango). These variables are defined in the Appendix, Table II.A2.

3.2. The autoregressive components of RV

Next, I control for the autoregressive component of the RV by introducing five different specifications, *i.e.*, the HAR (Corsi, 2009), the HEXP (Bollerslev et al., 2018), the HARQ (Bollerslev et al., 2016), the HAR-TV (Chen et al., 2018), and the HARQ-TV. I choose these models because they are parsimonious and competitive in terms of explanatory power and forecasting ability.

- EVHAR

The EVHAR model is,

$$RV_{c,t} = \boldsymbol{\alpha' EV_{c,t}} + \beta_{1,c}RV_{c,t-1} + \beta_{2,c}RV_{c,t-2|t-5} + \beta_{3,c}RV_{c,t-6|t-22} + \epsilon_{c,t}, \tag{4}$$

ral gas see, https://www.eia.gov/outlooks/steo/report/natgas.php. For the unconditional level of historical RV of other contracts see Appendix, Figure II.A1.

[5]Bessembinder et al. (1996) consider the square root of time to maturity instead of the log.

[6]This is also the methodology underlying the VIX computation. In unreported robustness tests I also use business time with virtually no differences in the results.

[7]The contracts maturities are reported in Appendix, Table II.A1.

where $RV_{c,t}$ is the log RV in time t for commodity c, $\boldsymbol{EV_{c,t}}$ is the vector of EVs in Eq. (3) and a constant, $RV_{c,t-n|t-p}$ is the log average RV computed over the days $t-n$ to $t-p$ (previous week and month); see Corsi (2009).

- EVHEXP

The Heterogeneous Exponential of Bollerslev et al. (2018) is similar to the HAR. It uses mixtures of exponentially smoothed past log RV. Each term is computed as, $RV_{c,t}^{CoM(\lambda)} = \sum_{i=1}^{500} \frac{e^{-i\lambda}}{e^{-\lambda}+e^{-2\lambda}+...+e^{-500\lambda}}$, with $\lambda = \ln\left(1+\frac{1}{\text{CoM}}\right)$, for decay rates $\lambda = 0.693, 0.182, 0.039,$ and 0.008 corresponding to centers of mass (CoM) of 1, 5, 25, and 125 days, respectively:

$$RV_{c,t} = \boldsymbol{\alpha' EV_{c,t}} + \gamma_{1,c} RV_{c,t-1}^{CoM_1} + \gamma_{2,c} RV_{c,t-1}^{CoM_5} + \gamma_{3,c} RV_{c,t-1}^{CoM_{25}} + \gamma_{4,c} RV_{c,t-1}^{CoM_{125}} + \epsilon_{c,t}, \quad (5)$$

- EVHARQ

The HARQ uses the realized quarticity to account for measurement errors. Barndorff-Nielsen and Shephard (2002) define the estimator of the log realized quarticity (hereafter, RQ_t) as:

$$RQ_t = \frac{2}{3} \frac{\sum_{i=1}^{1/\Delta} r_{i,t}^4}{\left(\sum_{i=1}^{1/\Delta} r_{i,t}^2\right)^2}$$

I retain a parsimonious version of the model where only the first coefficient of the HAR is penalized for measurement errors,

$$RV_{c,t} = \boldsymbol{\alpha' EV_{c,t}} + (\delta_{1,c} + \delta_{1Q,c} RQ_{c,t-1}) \times RV_{c,t-1} + \delta_{2,c} RV_{c,t-2|t-5} + \delta_{3,c} RV_{c,t-6|t-22} + \epsilon_{c,t}, \quad (6)$$

3.3. Are the parameters time-varying?

- EVHAR-TV

To check whether parameters are time-varying, the semi-parametric (local kernel) estimation approach is employed; see Chen et al. (2018) and Casas et al. (2019). This method also allows for an estimation in system (SUR), which makes the comparison with the competing models consistent. I use the Nadaraya-Watson

(Nadaraya, 1964; Watson, 1964) estimator,

$$\hat{\alpha_h}(x) = \frac{\sum_{i=1}^{n} K_h(x - x_i)\, y_i}{\sum_{j=1}^{n} K_h(x - x_j)},$$

where K is the Epanechnikov kernel for a bandwidth h. The procedure uses "leave-one-out cross-validation" to select the optimal bandwidth. An additional advantage of the kernel regression is that the covariance matrix of errors for the feasible generalized least squares (FGLS) is itself time-varying, with a bandwidth selected similarly. The model is,

$$RV_{c,t} = \boldsymbol{\alpha'(\tau_t)\, EV_{c,t}} + \theta_{1,c}(\tau_t)\, RV_{c,t-1} + \theta_{2,c}(\tau_t)\, RV_{c,t-2|t-5} + \theta_{3,c}(\tau_t)\, RV_{c,t-6|t-22} + \epsilon_{c,t}, \tag{7}$$

where the coefficients $\alpha'(\tau_t)$ are time-varying, $\tau_t = \frac{t}{T}$ the smoothing variable for $t = 1, 2, \ldots, T$, and T the sample size.

- EVHARQ-TV

The following model nests time-varying parameters and measurement errors, penalizing the first term of the HAR-TV with the (log) realized quarticity. The model is,

$$RV_{c,t} = \boldsymbol{\alpha'(\tau_t)\, EV_{c,t}} + (\phi_{1,c}(\tau_t) + \phi_{1Q,c}(\tau_t)\, RQ_{c,t-1}) \times RV_{c,t-1} + \phi_{2,c}(\tau_t)\, RV_{c,t-2|t-5} + \phi_{3,c}(\tau_t)\, RV_{c,t-6|t-22} + \epsilon_{c,t}, \tag{8}$$

3.4. Data and descriptive statistics

3.4.1. Data and variable definition

Nine contracts evenly spread in three main subgroups, *i.e.*, agriculture (wheat, corn, and soybeans), energy (WTI crude oil, natural gas, and heating oil), and metal (copper, gold, and silver) are selected because they have the highest open interest and turnover in their own subgroup.[8] From the Barchart API, I download 5-minute closing prices of the nearby and first deferred commodity futures contracts from May 6, 2008 to January 18, 2019.[9] The data include a timestamp and the maturity date of each contract, and are restricted to the Globex session (exchange market)

[8]Contracts specifications are in the Appendix, Table II.A1, and their liquidity characteristics in Table II.A4.

[9]https://www.barchart.com/

only. The trading hours are reported in Appendix, Table II.A1, along with the corresponding maximum 5-minute observations per day. Finally, estimations of the SUR system requires a perfect time match of the nine contracts, to avoid forward-looking bias entering the residuals covariance matrix in the FGLS. Consequently, I set a common cut-off time at 4 pm, CT, which I define as the end of the trading day.[10]

I compute 5-minute log price changes for each nearby futures contract available as $r_{c,t,j} = f^N_{c,t,j} - f^N_{c,t,j-1}$ where f is the log of the futures price and the subscripts c, t and j stand for commodity, day, and time of the observation, respectively. The arithmetic RV (ARV) is defined as,

$$ARV_{c,t} = \sqrt{\sum_{j=1}^{1/\Delta} r^2_{t,\Delta\times j}}$$

and the log RV (RV) as,

$$RV_{c,t} = \ln\left(ARV_{c,t}\right)$$

where $1/\Delta$ is the number of observations available given the market open hours of each contract. I choose the 5-minute sampling since previous literature document its performance over alternative frequencies (see, *e.g.*, Liu et al. (2015).[11]

[Insert Table II.1 here]

3.4.2. Summary statistics of 5-minute RV and daily market data

Table II.1 compares both distributional and memory properties of the ARV and RV. The daily mean of the ARV ranges from 1.02% (gold) to 2.72% (natural gas). The ARVs of commodity futures exceed those found in previous literature for exchange rates, sovereign bonds, and stock indices. Instead, they are included in the range of typical RVs found for large traded stocks. For instance, the mean of the ARV is 0.5% for the US T-Bond, 0.67% for the DEM/USD exchange rate, and 0.93% for the S&P 500 over the 1986–2002 period; see Andersen et al. (2007). On the other hand, Bollerslev et al. (2016) find that, over the 1997–2013 period, the mean ARV of 27 Dow Jones stocks is in the 1.68%-5.42% range. Buccheri and Corsi (2019) find a slightly larger but similar range (0.95%-11.10%), for 18 NYSE stocks

[10]In unreported robustness tests, I roll the nearby onto the first deferred five business days before maturity, with virtually the same results. Previous research on futures price changes justify this procedure because of possible market squeezes and thinly traded contracts immediately before the maturity.

[11]Summary statistics in the Appendix, Table II.A3 also show that it is a reasonable choice across the nine contracts, when compared to alternative frequencies.

over the 2006–2014 period. Overall, the null hypothesis of normality is rejected for both the ARV and RV. However, in line with Andersen et al. (2003), the RV distribution is closer to normality than that of the ARV. The skewness of the ARV is at least twice that of the RV for eight contracts, and even more for the crude oil contract. The excess kurtosis of the ARV is also much larger, (up to two orders of magnitude for the crude oil contract). Moreover, the persistence increases when using the RV in place of the ARV. The Ljung-Box statistics of the RV are twice those of the ARV for eight contracts. The log-periodogram parameters of the RV are also superior for agriculture and energy, but not for metal products. Overall, the memory properties of the ARV and RV are similar to previous results found for exchange rates (Andersen et al., 2003, 2007), S&P 500, and US T-Bonds (Andersen et al., 2007).

[**Insert Table II.2 here**]

Table II.2 reports the four moments of the distribution for the nine nearby futures contracts. The average log price changes (Panel A) is close to zero over the sample period, for all contracts. Similar to financial assets, their distributions strongly depart from normality, with an average standard deviation that vastly exceeds the mean and an important excess kurtosis, from 1.93 (wheat) to 17.04 (corn). The skewness that has long been perceived as positive in commodity futures, together with positive excess returns, also range from -1.17 (soybeans) to 0.33 (natural gas); see Gorton and Rouwenhorst (2006). I additionally provide the proportion of days during which the contracts are in contango, which stands between 60.16% (soybeans) and 97.55% (wheat). The group means is 81%, 80.12%, and 67.55%, for agriculture, energy, and metal, respectively. These statistics differ from the classical view according to which agricultural products are more subject to contango, given their important storage costs; see, *e.g.*, Keynes (1930); Fama and French (1987). Nonetheless, it does confirm that precious metals have the least contango, given the irrelevance of storage costs with regard to their underlying value; see, *e.g.*, Ng and Pirrong (1994). Finally, I report the optimal sampling frequency, to identify the best trade-off between resolution and market microstructure noise; see Aït-Sahalia et al. (2005). This optimal sampling is roughly related to the market trading volume (Panel B), minutes per day that have at least a transaction (Panel C), and bid-ask spread (Panel D). In brief, these results tend to indicate that the higher the trading activity, the lower the bid-ask spread and the higher the optimal sampling frequency. In the remainder of the article, I define RV as the log RV computed with a 5-minute sampling frequency.[12]

4. Results

4.1. *In-sample estimation*

4.1.1. *EV and EVHAR*

Table II.3 reports the coefficient estimates of the EV (Eq. (3)) and the EVHAR (Eq. (4)) models.

[Insert Table II.3 here]

In the EV specification, the monthly dummies associated with agricultural products (July) and the natural gas contract (January) are positive, and statistically significant at the 1% level. This first result is consistent with the seasonal RV structure displayed in Figure II.1 and the uncertainty resolution hypothesis; see Anderson and Danthine (1983). Second, there is a positive relation between the time to maturity and volatility. For seven out of the nine contracts, the corresponding coefficients are statistically significant at the 1% level, in contradiction with the Samuelson conjecture. Third, in line with Kogan et al. (2009) and Haugom et al. (2014), I find that for the crude oil contract, the magnitude of the slope matters but not its sign. SL (contango) loads positively, the interacted variable $SL \times B$ (backwardation) loads negatively, and both coefficients are statistically significant at the 1% level. Thus, contango and backwardation states are positive predictors of RV, supporting the "v-shape" hypothesis. This pattern is present for the crude oil, heating oil, natural gas, and copper contracts. The contango slope coefficient SL of the wheat contract is also positive and significant at the 1% level, but not the backwardation slope coefficients. This result is likely induced by the fact that the wheat contract is in contango 97.55% of the time over the sample period. The SL coefficient is negative and significant at the 1% level for the corn, gold, and silver contracts, thereby showing a negative contango-RV relation. If the result for the corn contract is difficult to explain (or is spurious), those for the gold and silver contracts may be related to previous empirical evidence. For instance, Ng and Pirrong (1994) document that precious and industrial metals have large differences in their inventory-volatility

[12]RV summary statistics with alternative sampling frequencies of 1-, 5-, 15-, and 60-minute are reported in Appendix, Table II.A3. It verifies the good compromise that the 5-minute sampling delivers, both for distributional and memory properties. The 5-minute frequency remains also superior to the aggregated measure that averages all the aforementioned variables, which is reported at the bottom of the Table; see Andersen et al. (2011b).

relations, explained by the fact that scarcity is irrelevant for the precious metals group. Finally, the explanatory power of the EV model also varies across contracts, with R^2 for individual equations ranging from 2% (copper) to 26% (crude oil).

When I include the HAR terms into the EV specification, the explanatory power of the models is improved, with adjusted R^2 ranging from 33% (soybeans) to 75% (crude oil). The likelihood ratio test (constrained *vs.* unconstrained model) rejects the null hypothesis at the 1% level showing that EVs continue to explain the RV beyond the autoregressive terms.[13] The HAR coefficients are statistically significant at the 1% level and closely aligned with those of Andersen et al. (2007) for the DEM/USD exchange rate, S&P 500, and US T-Bond. In the EVHAR model, the statistical significance levels of the monthly dummies decrease to 5% for the agricultural products. They remain at the 1% level for the natural gas (coefficient size in the EV model thrice larger than in the EVHAR). This points to a more pronounced seasonal structure for natural gas that seems difficult to capture only with autoregressive terms; see Geman and Ohana (2009). Similarly, the time to maturity coefficients only maintain statistical significance at the 1% level in metal products. Switching from the EV to the EVHAR model, the statistical significance of the "v-shape" coefficients is maintained at the 1% level for all energy contracts. Even if the explanatory power of EVs is partly subsumed by the HAR terms, and that memory properties capture some of the EV features, their contribution is robust from both futures markets (time to maturity), and supply and demand (seasonality and term structure slope) perspectives. Therefore, I keep EVs in the forthcoming autoregressive specifications and test further the robustness of their contribution.

4.1.2. Alternative autoregressive specifications

Table II.4 reports the joint estimation of the EVHEXP (Eq. (5)) and EVHARQ (Eq. (6)) models.

[Insert Table II.4 here]

Given their important lag structure (500 lags, and up to 125-day CoM), HEXP variables are likely to impact the cyclical EVs: monthly dummies and time to maturity. For instance, the 125-day CoM coefficient is negative and statistically significant at the 1% level for the soybeans contract, one of the "seasonal" commodities. Indeed, this variable is centered with a half-year lag, its negative coefficient points to a seasonal structure of the RV, in which one season is in a high state, relative to the "opposite" season. Instead, this parameter remains positive and statistically

[13]The constrained model estimations are reported in Appendix, Table II.A5.

significant at the 1% level, for the heating oil and silver contracts, which was not hypothesized to be seasonal. In all contracts, the monthly dummies are robust to this deep lag structure, and remain statistically significant at the 1% level. In addition, the significance of the time to maturity and term structure-related coefficients are comparable to those of the EVHAR. Most of the variance in the EVHEXP is explained by the 1-day CoM coefficient (equivalent to a one day-lag) and the 25-day CoM coefficient, which display statistical significance at the 1% level for nine and seven contracts, respectively. The 5-day CoM coefficient is statistically significant at the 5% level for five contracts only. These results indicate that the (EV)HEXP may still properly explain the RV in a more parsimonious specification. Lastly, the likelihood ratio ${}^{EVHEXP}/_{HEXP}$, which tests the null that the contribution of EVs is nil in the HEXP model is statistically significant at the 1% level, which implies that the EVs' contribution is also robust to this model.[14]

The coefficients related to the quarticity are statistically significant at the 5% level for six contracts, while the inclusion of the HARQ does not affect the coefficient of the seasonal dummy (statistically significant at the 5% and 1% level for the natural gas and agricultural products, respectively), it does affect the other EVs. Seven out of nine time to maturity coefficients in the EVHARQ specification are smaller than those of the EVHAR, suggesting that measurement errors (or faster mean reversion when volatility is high) may relate to contract maturity; see Cipollini et al. (2017). The coefficients associated with the slope decrease further after the inclusion of the realized quarticity but not consistently across contracts. The explanatory power of the EVHARQ is consistently higher than those of the EVHAR and EVHEXP in terms of individual equations and system R^2. Lastly, the likelihood ratio ${}^{EVHARQ}/_{HARQ}$, which tests the null that the contribution of the EVs is nil in the HARQ specification, is also rejected at the 1% level. This supports the robustness of EVs when both autoregressive and measurement error terms are included.

4.1.3. Time-varying coefficients

Table II.5 reports the results of the EVHAR and EVHARQ models with time varying parameters (EVHAR-TV, Eq. (7) and EVHARQ-TV, Eq. (8)). The means of the coefficients time series and their minima and maxima.

[Insert Table II.5 here]

[14]The results of the estimation of the restricted HEXP and HARQ specifications are reported in Appendix, Table II.A5.

The extrema provide information about the extent to which parameters are time-varying. The largest ranges are obtained for the intercepts (RV spread of up to 23 for the corn contract), and EVs.[15] In contrast, all autoregressive parameters are more stable across contracts (minimum-maximum RV range of 4 at most). The larger range of the intercept compared to the autoregressive coefficients aligns with the results of the score-driven HAR (SHAR); see Buccheri and Corsi (2019). The means of the parameters of the EVHAR(Q)-TV lie in the same range as those of the static EVHAR(Q), and the improvement in explanatory power when the specification accounts for measurement errors is similar. The likelihood ratio tests ${}^{EVHAR-TV}/_{HAR-TV}$ and ${}^{EVHARQ-TV}/_{HARQ-TV}$ are statistically significant at the 1% level, implying that EVs significantly improve the explanatory power of the most exhaustive autoregressive specifications, which accounts for both measurement errors and time-varying parameters. Moreover, the extrema ranges of the HAR-TV- and HARQ-TV-related parameters are considerably reduced after the inclusion of the EVs.[16] For instance, the range in RV of the intercepts of the silver contract decreases from 0.93 in the HAR-TV, to 0.12 in the EVHAR-TV, which suggests that EVs play a role in capturing the structural/long-term levels of the RV.

The cyclical components of EVs, time to maturity and monthly dummies, are partly captured by the time-varying intercepts. Yet, in both models, these parameters remain time-varying. This points to a non-linear relation between the EVs and the level of RV; see, *e.g.*, Hong (2000). The term structure-related coefficients remain heterogeneous, with only five contracts that display slope and interaction parameters following the "v-shape" pattern (jointly positive and negative, respectively). Finally, I find that time-varying coefficients are statistically different from zero at the 1% level. It provides evidence that even when the coefficients are small, they are mildly time-varying.

4.1.4. EV and RV: A synthesis

In Table II.6, I summarize the results for the EVs across the nine contracts. For each model and variable, I report how many coefficients are significant at the 5% level (negative or positive).

[Insert Table II.6 here]

First, the coefficients for the monthly dummies are positive and significant at the 1%

[15]See also Appendix, Figure II.A2 for the time-varying pattern of the intercepts. Other coefficients are available upon request.

[16]I report the results of the estimation of the restricted HAR-TV and HARQ-TV in Appendix, Table II.A6.

level for all four contracts and six models. The robustness of this seasonal term, even in autoregressive specifications, is striking. This clearly supports the contribution of EVs, beyond that of the autoregressive components and the uncertainty resolution hypothesis; see Anderson and Danthine (1983).

Second, the results of the time to maturity are partly robust. In the restricted EV model, seven out of nine contracts have positive coefficients, significant at the 5% level. When this variable is nested in the alternative static (time-varying) models, four (eight) coefficients remain positive and significant at the 5% level, although the interpretation of the significance in the dynamic version is not straightforward. Altogether these results strongly reject the Samuelson hypothesis, since the time to maturity coefficients are positive and significant at the 1% level. Moreover, the rejection of the Samuelson hypothesis found in the precious metal, which are supposed to behave more like financial assets, also relates to the results for the NIKKEI and S&P 500 futures; see Chen et al. (2000); Moosa and Bollen (2001). Contrary to Bessembinder et al. (1996), I do not find empirical support for this hypothesis in agricultural commodity futures. Coefficients are not statistically significant at the 1% level.

Third, the results regarding the term structure provide mixed support to the theory of storage and the "v-shape" hypotheses. Four out of nine commodity contracts have statistically significant coefficients at the 1% level. I relate these results to the contract-storage peculiarities detailed in Section 4.1.1. In particular, the contracts for which the "v-shape" hypothesis is supported are those supposed to embed important storage costs (three energy commodities and copper). However, the coefficients of the three agricultural contracts, also viewed as having significant storage costs, do not display the "v-shape" pattern. Lastly, despite the mixed results of the baseline model, they hold across autoregressive specifications, with from three (EVHARQ) to six contracts (EVHAR-TV and EVHARQ-TV) displaying the "v-shape" pattern.

To summarize the aforementioned results, EVs alone do explain the RV. When autoregressive terms are added, the explanatory power of the regressions increases and the size of the EVs coefficients decreases but their sign and statistical significance are maintained. This indicates that the information content of the EVs is captured to some extent by the various lags of the HAR(Q) and HEXP. Yet, the unrestricted versions (with EVs) improve the explanatory power of the models. Moreover, all likelihood ratios reject the null of no improvement (at the 1% level) when EVs are added. Thus, the remainder of the paper analyzes how EVs improve the in- and out-of-sample forecast accuracy of these models.

4.2. In-sample performance

I compute in-sample forecasts from the 11 models at the one-day, one-week, and one-month horizons. I keep the time to maturity and monthly dummies contemporaneous since those are deterministic variables. Therefore, the set of information that determines their future state is fully available ex-ante for every agent, such that this approach does not entail forward-looking bias. Table II.7 reports the results of the Model Confidence Set (MCS) procedure with which the performance of the models is directly compared based on the three loss functions, *i.e.*, mean squared errors (MSE), mean absolute errors (MAE), and QLIKE; see Hansen, Lunde, and Nason (2011); Patton (2011).

[Insert Table II.7 here]

Table II.7 reports the tests related to the one-day ahead forecasts. The HARQ-TV emerges as the best model with the three loss functions. The EVHARQ ranks second, although the MAE losses are excluded from the 90% confidence interval. The one-week and one-month ahead comparisons indicate that the HAR-TV is superior, excluding those of the one-month ahead MSE, for which the HARQ-TV ranks first. These results are in line with those of Bollerslev et al. (2016) for the S&P 500 (MSE) and 27 Dow Jones stocks. However, for longer horizons they find the opposite. Thus, across all models, only a single EV-based model (the EVHARQ) steps up in the 90% confidence set for the MSE and QLIKE losses, at the one-day horizon. Lastly, the losses are consistently smaller (greater) in the static (time-varying) versions of the models, when EVs are included. This points to time-varying specifications appropriately capturing the time variations of EVs through the levels (intercepts).

4.3. Out-of-sample performance

I compute out-of-sample forecasts at the one-day, one-week, and one-month horizons based on a training period from April 22, 2010 to January 30, 2013. I obtain the static model forecasts by computing $\mathbb{E}[RV_t]$. For the time-varying specifications, I use the multi-stage non parametric predictor approach; see Chen, Yang, and Hafner (2004) and Chen et al. (2018). In this procedure, I compute the one-step ahead conditional expectations and re-use them to select the new conditional optimal bandwidth for the next step(s), iteratively. Next, I use the MCS procedure and the modified Diebold-Mariano test (Harvey, Leybourne, and Newbold, 1997) to benchmark these out-of-sample forecasts; see Diebold and Mariano (1995). The results are reported in Table II.8. In this analysis, I add the RiskMetrics model as a generic benchmark, given its wide use in risk management.[17] The RiskMetrics model is a parsimonious version of the HEXP with a single exponentially smoothed lagged variable. The decay rate λ is set at 6%, which corresponds to a 16-day CoM. Although the RiskMetrics is not calibrated for log variances or volatilities, in unreported tests, I find that this decay rate remains a good trade-off for log RVs. The RiskMetrics equation is,

$$RV_{c,t} = \mu_{0,c} + \mu_{1,c} RV_{c,t-1}^{CoM_{16}} + \epsilon_{c,t},$$

[**Insert Table II.8 here**]

Table II.8, Panel A shows that the HARQ strictly dominates all other models at the one-day and one-week horizons. In addition, the benefit of including EVs vanishes over these horizons. However, at the one-month horizon, the best performing model is the EVHARQ-TV, and its MSE (MAE) is almost a third (half) of the HARQ-TV model. The forecast improvement arising from EVs is present in almost all models and for all loss functions. This supports the benefits of including these exogenous variables when the time horizon increases. These forecasting improvements may also come from the fact that the time to maturity and monthly dummies are deterministic variables. As they are readily available for the n-ahead periods, they could improve the forecasting power at longer horizons. Finally, Table II.8, Panel B reports the results of a one-to-one direct comparison of the 12 models using the modified Diebold-Mariano test, at the one-day ahead horizon and for the MSE loss. These results indicate a strict dominance of the HARQ and EVHARQ-TV models. When these two models are compared, the χ^2 statistic is not statistically significant (-1.31), hence supporting their equally superior forecasting ability at this horizon, in line with previous results. Finally, these individual comparisons allow to test the forecasting accuracy improvement of the unconstrained (with EVs) *vs.* their constrained autoregressive counterparts. EVs inclusion is beneficial (statistically significant at the 1% level) in the HEXP, HAR-TV, and HARQ-TV, whereas it is not useful (statistically significant at the 10% level) in the HAR and HARQ-TV specifications.

[17] https://www.msci.com/documents/10199/5915b101-4206-4ba0-aee2-3449d5c7e95a

5. Tail-risk modeling

I now use the out-of-sample forecasts from the 12 models, and compare their ability to forecast the (left) tail risk. A large strand of literature adopts the expected shortfall (ES) methodology, in place of the value at risk (VaR). I adopt the multi-quantile regression approach; see, *e.g.*, Couperier and Leymarie (2020); Bayer and Dimitriadis (2020). It models the left tail of a distribution with a high granularity, since any sequence of coverage levels may be used. Table II.9 reports the p-values testing the coefficients from forecast accuracy panel regressions, $\boldsymbol{RV_t} = \beta_0 + \beta_1 \widehat{\boldsymbol{RV_t}}$, where $\boldsymbol{RV_t} = [RV_{1,t}, RV_{2,t}, \ldots, RV_{9,t}]'$. The null hypotheses are:

- $H_{0,J_1} : \sum_{j=1}^{p} \left(\beta_0\left(\tau_j\right) + \beta_1\left(\tau_j\right)\right) = p$
- $H_{0,J_2} : \sum_{j=1}^{p} \beta_0\left(\tau_j\right) = 0$ and $\sum_{j=1}^{p} \beta_1\left(\tau_j\right) = p$
- $H_{0,I} : \sum_{j=1}^{p} \beta_0\left(\tau_j\right) = 0$
- $H_{0,S} : \sum_{j=1}^{p} \beta_1\left(\tau_j\right) = p$

Where p is the number of quantiles, and τ_j is the corresponding quantile level for $j = 1, 2, \ldots, p$. In other words, J_1 tests the null hypothesis that the sum of the intercepts and slopes sum to p, J_2 that the sum of the intercepts and slopes sum to zero and p, respectively, and I (S) that the sum of the intercepts (slopes) sum to zero (p), individually.

[Insert Table II.9 here]

The J_2 test rejects the null hypothesis for the EV, the time-varying, and the RiskMetrics models when $p = 1$ (equivalent to the 97.5% VaR), and for finer granularities up to $p = 4$.[18] The alternative test, J_1 whose null hypothesis is that the sums of all multi-quantile regressions parameters (both intercepts and coefficients) are equal to p, is only rejected for the HAR-TV and $p = 2$, at the 10% level. Similarly, I and S hypotheses are almost never rejected at the 5% level. Therefore, I focus on the rejection rates of J_2 to compare the relative performance of the models. First, the static parameter version of the HAR delivers the highest p-values (minimum of 0.25). The p-value decreases as the coverage (the granularity of the quantile subdivisions) increases, in line with Couperier and Leymarie (2020). In

[18]The Mincer-Zarnowitz test is a particular case of the J_2, when there is a single coverage level: $p = 1$.

static parameter (time-varying) specifications, the inclusion of EVs decreases (increases) the expected-shortfall forecasting performance.[19] The bottom of the panel presents the J_2 statistics for $p = 4$, at the one-week and one-month horizon. In brief, the one-week horizon does not reveal any dominance from the restricted or unrestricted specifications. Instead, at the one-month horizon, all unrestricted EV models generate p-values that are higher than their restricted versions.

In Table II.10, I present the percentage of violations occurring over the sample for each contract, for a single quantile regression, with a coverage level set at the 97.5% VaR. The horizon of the out-of-sample forecasts is of one day.

[**Insert Table II.10 here**]

On average, the EVHARQ provides the lowest violation percentage at the 97.5% VaR level, but there are important disparities across contracts. There is no systematic benefit arising from the EVs inclusion. Similarly, the most complex time-varying specifications do not deliver better forecasts in terms of VaR violations. Since ES aggregates information from the left tail, the noise arising from the center and the right tail of the distribution may shadow the actual ES signal.[20] Finally, in four contracts, the single EV model yields the lowest violation percentages. Despite the fact that these results stand for a single coverage level, and do not encompass an entire expected shortfall violation, they point to the benefits of including EVs when modeling tail risk of some contracts.

6. Conclusion

This research aims to test whether economic variables, theoretically related to the volatility of commodity futures contracts, add value to autoregressive RV models. Using joint estimations for nine commodities, my results strongly support the uncertainty resolution hypothesis; see Anderson and Danthine (1983) and Anderson (1985). I also find that the RV is positively related to time to maturity, which rejects the Samuelson hypothesis. Lastly, the inclusion of the slope of the term structure yields mixed results. On the one hand, I find support for the theory of storage and “v-shape” hypothesis of Kogan et al. (2009) for the heating oil, natural gas, copper, and crude oil contracts; see also Haugom et al. (2014). On the

[19] Appendix, Table II.A7 reports the coefficients of the predictive regressions for each coverage level τ, individually, when the number of quantiles is set to $p = 4$. It also shows that while all specifications yield a significant forecasting bias, the static HAR and EVHAR models generate the lowest bias, and this for all coverage levels used.

[20] I thank an anonymous referee for this comment

other hand, I do not identify any support for the remaining contracts. However, concerning precious metals, it is likely that the determinants of the term structure are unrelated to supply and demand, and to their inherent storage issues. Finally, the performance of the economic variables considered alone, lies below those of all autoregressive models, including their most parsimonious versions such as Risk-Metrics. However, nesting EVs in the autoregressive specifications always brings a statistically significant improvement in explanatory power, at the 1% level. The inclusion of economic variables also improves the out-of-sample forecast accuracy of three (HEXP, HAR-TV, and HARQ-TV) out of five models, at the one day horizon (significant at the 1% level). It additionally improves the out-of-sample forecasting accuracy for a longer time horizon (one-month ahead), even in specifications accounting for measurement errors and time-varying coefficients. These gains vanish for expected shortfall backtests from multi-quantile regressions. Yet, surprisingly, for four out of nine contracts, the EV model alone generates less 97.5% VaR violations than any other model and delivers good results for the remaining contracts.

Table II.1: **Summary statistics: Daily RV**

This table reports summary statistics for four estimators of daily realized volatility. These estimators are based on 5-minute log price changes of the nearby futures commodity contract. I display these statistics for the arithmetic realized volatility ARV_t and the log realized volatility RV_t. The Table displays the first four moments of the distribution, the Jarque-Bera statistic JB, the Ljung-Box statistic (20[th] order serial correlation) $Q_{(20)}$, and the parameter d of the log-periodogram regression (Geweke and Porter-Hudak, 1983; Robinson, 1995) based on a bandwidth exponent of 4/5 as in Andersen et al. (2003). The sample period is May 5, 2008–April 1, 2019 for nine commodity futures contracts. The number of observations per contract is 2755.

	Agriculture			Energy			Metal		
	Corn (C)	Soybeans (S)	Wheat (W)	WTI crude oil (CL)	Heating oil (HO)	Natural gas (NG)	Gold (GC)	Copper (HG)	Silver (SI)
				Daily arithmetic realized volatility, ARV_t					
Mean%	1.72	1.41	1.94	2.06	1.80	2.72	1.02	1.57	1.83
$\sigma\%$	1.01	0.80	0.84	1.21	0.93	1.36	0.58	0.93	1.06
Skewness	7.50	4.86	2.41	4.13	2.82	4.35	2.94	2.84	2.96
Kurtosis	145.87	52.33	11.81	49.42	21.66	46.60	16.78	11.57	16.52
JB	2,471,897	325,629	18,710	288,677	57,598	258,315	36,358	19,095	35,395
$Q_{(20)}$	4,856	6,723	7,139	26,574	25,471	8,768	18,971	27,365	15,973
d	0.24	0.35	0.32	0.50	0.50	0.41	0.44	0.59	0.49
				Daily logarithmic realized volatility, RV_t					
Mean%	−415.13	−434.90	−400.54	−400.03	−412.02	−369.11	−468.93	−427.66	−412.04
$\sigma\%$	46.48	46.19	41.54	46.69	43.93	39.85	50.87	46.46	52.02
Skewness	2.06	2.37	2.16	0.51	0.42	0.55	1.61	0.65	0.58
Kurtosis	13.73	17.02	18.39	0.61	0.36	1.01	13.22	1.14	7.55
JB	23,627	35,914	41,036	161	95	259	21,286	343	6,706
$Q_{(20)}$	11,283	10,703	8,201	32,443	31,598	18,330	12,106	23,773	11,329
d	0.34	0.40	0.32	0.52	0.53	0.51	0.33	0.51	0.38

Table II.2: **Summary statistics: Daily market-level data**

This table reports summary statistics of daily data for nine nearby commodity futures contracts. I display the statistics for the log price changes (Panel A), the trading volume (Panel B), the number of minutes with at least one transaction (Panel C), and the bid-ask spread estimated on 1-minute data with the Roll (1984) methodology (Panel D). I report the first four moments of the distributions, the minima, and the maxima. Panel A also reports the Ljung-Box statistic (20^{th} order serial correlation for the log price changes) $Q_{(20)}$ and the proportion of days during which the nearest term structure was in contango. Finally, I report the optimal sampling in minutes based on Aït-Sahalia et al. (2005, p. 361), assuming that the microstructure noise is gaussian and only driven by the bid-ask spread. The sample period is May 5, 2008–April 1, 2019. The number of observations per contract is 2755.

	Agriculture			Energy			Metal		
	Corn (C)	Soybeans (S)	Wheat (W)	WTI crude oil (CL)	Heating oil (HO)	Natural gas (NG)	Gold (GC)	Copper (HG)	Silver (SI)
Panel A: Daily log price changes of nearby contracts									
Mean%	−0.02	−0.01	−0.02	−0.02	−0.02	−0.05	0.01	−0.01	−0.004
σ%	1.92	1.60	2.07	2.38	1.93	2.96	1.12	1.73	1.99
Skewness	−1.07	−1.17	0.16	0.05	−0.17	0.33	−0.07	−0.16	−0.99
Kurtosis	17.04	9.30	1.93	4.19	3.10	3.25	7.52	4.01	7.55
$Q_{(20)}$	26.54	31.48	31.37	52.41	30.09	60.09	17.29	74.53	22.54
Contango%	85.30	60.16	97.55	80.50	74.52	85.36	71.62	62.39	68.68
Optimal sampling (min)	43	20	30	21	17	29	14	18	25
Panel B: Trading volume (in million USD)									
Mean	1,466.37	2,151.16	827.84	16,506.63	1,658.83	2,548.02	8,529.10	1,948.14	2,699.74
σ	1,223.27	2,111.44	621.02	9,594.49	1,172.04	1,461.08	11,200.83	2,205.74	3,003.31
Skewness	0.86	1.06	0.60	0.54	0.57	1.31	1.34	1.42	2.13
Kurtosis	0.82	1.81	0.64	0.89	−0.39	5.47	2.01	2.96	11.25
Min	0.03	0.06	0.04	0.13	0.08	0.60	0.08	0.07	0.07
Max	8,277.70	15,145.34	4,330.91	73,000.47	6,605.90	15,031.34	94,868.16	16,387.95	35,581.24
Panel C: Minutes with at least one transaction									
Mean	528.16	518.13	509.41	1,201.31	675.81	898.16	679.75	704.80	744.90
σ	254.11	316.54	264.43	308.60	260.95	258.02	604.09	556.00	552.40
Skewness	−0.47	−0.17	−0.61	−0.99	−0.60	−0.52	0.22	−0.15	−0.31
Kurtosis	−0.70	−1.28	−0.67	2.71	0.49	2.50	−1.48	−1.54	−1.50
Min	1	1	1	1	1	3	1	1	1
Max	1,051	1,054	1,051	1,380	1,353	1,378	1,380	1,375	1,380
Panel D: Bid-ask spread (in bps)									
Mean	2.67	1.27	2.19	1.93	1.33	2.99	0.68	1.26	1.82
σ	1.73	1.31	1.86	1.73	1.46	2.42	0.85	1.43	1.80
Skewness	0.65	2.35	2.54	1.55	1.85	3.91	2.65	2.39	2.00
Kurtosis	3.62	13.64	28.30	4.92	7.16	63.38	14.26	10.30	8.48
Min	0.01	0.00	0.01	0.04	0.09	0.15	0.00	0.00	0.00
Max	16.61	16.04	30.23	14.98	15.65	52.10	10.19	13.37	17.41

Fig. II.1: **Unconditional monthly average RV**

These plots display the monthly average of daily RV for the nearby futures contract written on corn, soybeans, wheat, and natural gas. The red line represents the average and the gray area represents the 90% confidence bands. For a better alignment, the plot of the natural gas contract is centered in January. The sample period is May 5, 2008–April 1, 2019. The number of observations per contract is 2755.

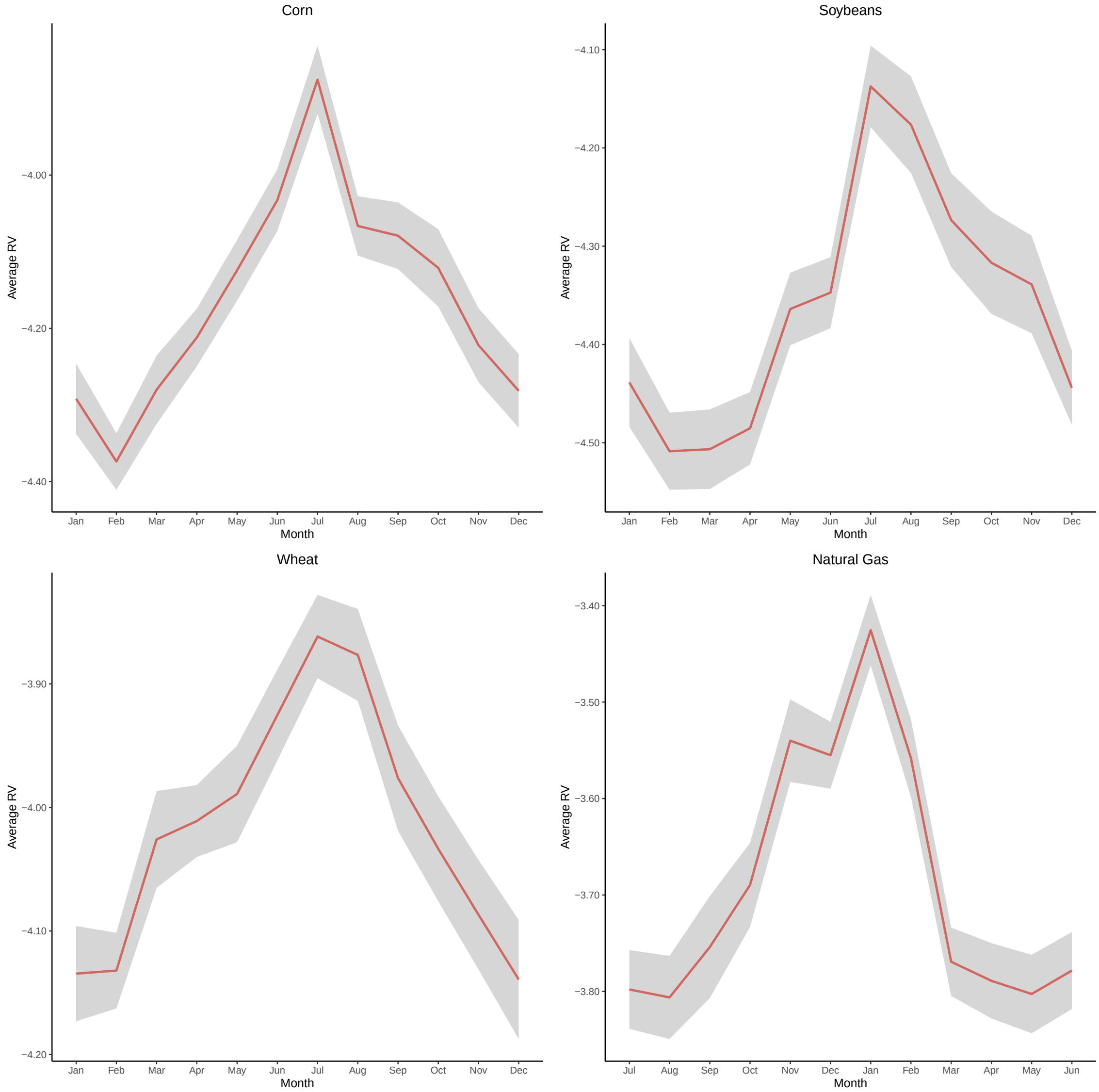


Table II.3: **EV and EVHAR estimation of RV_t**

This table reports the coefficients of the EV and EVHAR. $RV_{c,t}$ is the 5-minute log realized volatility. $M_{c,t}$ is one monthly dummy set to "1" during the month of uncertainty resolution and to "0" otherwise (July for the three agriculture products and January for the natural gas contract). $SL_{c,t-1}$ is the log of the nearest term structure slope. $B_{c,t-1}$ is a dummy set to "1" ("0") when the slope of the nearest term structure is negative (positive). $TM_{c,t}$ is the (log) time to maturity. The three lagged variables are the daily $RV_{c,t-1}$, weekly $RV_{c,t-2|t-5}$ *i.e.* the average of the RV from $t-2$ to $t-5$, and monthly $RV_{c,t-6|t-22}$ *i.e.* the average of the RV from $t-6$ to $t-22$. I report Newey and West (1994) standard errors with automatic lag selection in parenthesis. I report the adjusted R^2 for the individual equations of the system, the overall OLS R^2, McElroy R^2, the log likelihood, and the likelihood ratio test (LR) that compares the unrestricted model EVHAR with the restricted HAR. The sample period is May 5, 2008–April 1, 2019. The number of observations per equation is 2733.

	Agriculture						Energy						Metal					
	C		S		W		CL		HO		NG		GC		HG		SI	
Constant	−4.34***	−1.25***	−4.58***	−1.51***	−4.22***	−1.09***	−4.17***	−0.55***	−4.34***	−0.72***	−3.96***	−0.80***	−4.83***	−1.08***	−4.57***	−0.75***	−4.34***	−1.11***
	(0.02)	(0.09)	(0.02)	(0.10)	(0.02)	(0.11)	(0.02)	(0.05)	(0.02)	(0.06)	(0.02)	(0.08)	(0.02)	(0.12)	(0.02)	(0.09)	(0.02)	(0.10)
$M_{c,t}$	0.31***	0.05**	0.18***	0.06**	0.21***	0.05**					0.29***	0.11***						
	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)					(0.03)	(0.02)						
$TM_{c,t}(10^2)$	4.81***	0.27	4.46***	2.08**	0.95	0.93	0.31	−0.44	3.52***	0.44	5.98***	1.23	11.01***	5.85***	7.79***	6.42***	18.57***	11.89***
	(1.08)	(1.03)	(1.09)	(1.05)	(1.17)	(1.07)	(1.03)	(0.69)	(1.00)	(0.70)	(1.48)	(1.22)	(1.23)	(1.16)	(1.27)	(1.09)	(1.36)	(1.31)
$SL_{c,t-1}(10^{-2})$	−1.63***	−0.40	0.36	0.74	3.07***	0.61**	3.38***	0.63***	5.44***	1.53***	1.22***	0.46***	−55.21***	−12.25***	11.91***	2.84	−43.72***	−9.92***
	(0.55)	(0.50)	(1.12)	(1.03)	(0.29)	(0.29)	(0.15)	(0.12)	(0.30)	(0.21)	(0.10)	(0.08)	(3.17)	(3.24)	(2.47)	(2.00)	(2.81)	(2.95)
$B_{c,t-1}$	0.07***	0.00	0.05**	0.02	0.16***	0.07	−0.08***	−0.02	−0.05***	−0.04***	0.03	−0.02	−0.06	−0.05	−0.08***	−0.06***	0.10**	0.05
	(0.02)	(0.02)	(0.02)	(0.02)	(0.06)	(0.05)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.04)	(0.03)	(0.02)	(0.02)	(0.04)	(0.04)
$SL_{c,t-1} \times B_{c,t-1}(10^{-2})$	0.63	−0.16	−1.41	−1.32	3.92	1.47	−5.05***	−1.80***	−8.70***	−3.23***	−3.27***	−1.21***	−25.92	−47.00	−20.17***	−11.96***	52.10***	23.50**
	(0.58)	(0.52)	(1.13)	(1.03)	(4.91)	(4.43)	(0.61)	(0.39)	(0.38)	(0.26)	(0.18)	(0.15)	(37.01)	(33.19)	(3.29)	(2.70)	(11.77)	(10.68)
$RV_{c,t-1}$		0.32***		0.26***		0.21***		0.42***		0.37***		0.21***		0.20***		0.20***		0.16***
		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)
$RV_{c,t-2\|t-5}$		0.18***		0.21***		0.26***		0.25***		0.24***		0.37***		0.28***		0.38***		0.32***
		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)		(0.03)		(0.03)		(0.03)		(0.03)
$RV_{c,t-6\|t-22}$		0.21***		0.22***		0.27***		0.20***		0.23***		0.23***		0.32***		0.28***		0.30***
		(0.02)		(0.02)		(0.03)		(0.02)		(0.02)		(0.03)		(0.03)		(0.03)		(0.03)
Adj. R^2	0.11	0.41	0.06	0.33	0.09	0.38	0.26	0.75	0.20	0.71	0.19	0.55	0.16	0.42	0.02	0.48	0.17	0.42
OLS R^2	0.15	0.50																
McElroy R^2	0.11	0.47																
log likelihood	−3916.70	249.93																
LR EVHAR > HAR	559.47***																	

$^{***}p < 0.01; ^{**}p < 0.05; ^{*}p < 0.1$

Table II.4: **EVHEXP and EVHARQ estimation of RV_t**

This table reports the results of the EVHEXP and EVHARQ models. Realized volatility $RV_{c,t}$ is computed with 5-minute (log) prices. I compute the HEXP with four terms exponentially smoothed over a backward looking 500-day window for four lagged variables, *i.e.*, centers of mass (CoM) of 1, 5, 25, and 125 days. I compute the HARQ with three lagged variables, the daily $RV_{c,t-1}$, weekly $RV_{c,t-2|t-5}$ *i.e.* the average of the RV from $t-2$ to $t-5$, and monthly $RV_{c,t-6|t-22}$ *i.e.* the average of the RV from $t-6$ to $t-22$. $M_{c,t}$ is a monthly dummy set to "1" during the months of uncertainty resolution and to "0" otherwise (July for the three agriculture products and January for the natural gas contract). $TM_{c,t}$ is the log time to maturity. $SL_{c,t-1}$ is the log difference of the nearest term structure slope, $B_{c,t-1}$ a dummy set to "1" ("0") when the slope of the nearest term structure is negative (positive). I report Newey and West (1994) corrected standard errors with automatic lag selection in parenthesis. I report the adjusted R^2 for each individual equation of the system, the overall OLS R^2, McElroy R^2, the log likelihood, and the likelihood ratio tests (LR) that compares the unrestricted EVHARQ and EVHEXP models with the HARQ and HEXP, respectively. The sample period is April 22, 2010–April 1, 2019. The number of observations per equation is 2255 for the EVHEXP and 2733 for the EVHARQ.

	Agriculture						Energy						Metal					
	C		S		W		CL		HO		NG		GC		HG		SI	
Constant	−1.50***	−1.21***	−2.33***	−1.40***	−1.23***	−1.07***	−0.66***	−0.50***	−0.66***	−0.66***	−1.14***	−0.66***	−0.87***	−1.08***	−0.60***	−0.70***	−0.82***	−1.08***
	(0.12)	(0.09)	(0.19)	(0.10)	(0.14)	(0.11)	(0.08)	(0.05)	(0.08)	(0.06)	(0.16)	(0.07)	(0.18)	(0.12)	(0.12)	(0.09)	(0.14)	(0.10)
$M_{c,t}$	0.10***	0.05**	0.08***	0.05**	0.08***	0.05**					0.12***	0.09***						
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)					(0.02)	(0.02)						
$TM_{c,t}(10^2)$	0.50	0.11	2.22**	1.01	0.88	0.99	−0.35	−0.53	−0.17	1.19*	0.05	2.48**	6.32***	3.61***	7.21***	5.08***	13.36***	10.53***
	(1.07)	(1.04)	(1.07)	(1.06)	(1.08)	(1.10)	(0.72)	(0.66)	(0.75)	(0.68)	(1.25)	(1.18)	(1.16)	(1.26)	(1.07)	(1.29)	(1.30)	(1.52)
$SL_{c,t-1}(10^{-2})$	0.13	−0.51	−0.78	0.99	0.45	0.56*	0.75***	0.65***	1.67***	1.43***	0.48***	0.40***	−8.97**	−14.65***	5.09**	2.31	−3.20	−10.77***
	(0.59)	(0.50)	(1.17)	(1.02)	(0.30)	(0.29)	(0.13)	(0.11)	(0.23)	(0.20)	(0.08)	(0.08)	(3.82)	(3.26)	(2.03)	(2.01)	(3.25)	(2.97)
$B_{c,t-1}$	0.02	0.00	0.02	0.02	0.09*	0.07	−0.01	−0.02	−0.03***	−0.04***	−0.03	−0.03	−0.06*	−0.05	−0.05**	−0.06***	0.05	0.05
	(0.02)	(0.02)	(0.02)	(0.02)	(0.05)	(0.05)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.04)	(0.04)
$SL_{c,t-1} \times B_{c,t-1}(10^{-2})$	−0.81	−0.04	0.20	−1.56	3.21	1.39	−2.02***	−2.10***	−3.51***	−3.19***	−1.27***	−1.04***	−64.56*	−47.00	−14.25***	−11.39***	16.34	23.98**
	(0.61)	(0.52)	(1.18)	(1.03)	(4.46)	(4.45)	(0.41)	(0.38)	(0.29)	(0.26)	(0.15)	(0.14)	(33.11)	(33.22)	(2.74)	(2.70)	(10.69)	(10.75)
$RV_{c,t-1}^{CoM1}$	0.47***		0.37***		0.31***		0.64***		0.58***		0.22***		0.18***		0.23***		0.12**	
	(0.04)		(0.04)		(0.05)		(0.04)		(0.04)		(0.05)		(0.05)		(0.05)		(0.06)	
$RV_{c,t-1}^{CoM5}$	−0.17**		−0.10		0.05		−0.05		−0.06		0.47***		0.26**		0.39***		0.44***	
	(0.07)		(0.07)		(0.09)		(0.06)		(0.06)		(0.08)		(0.10)		(0.09)		(0.11)	
$RV_{c,t-1}^{CoM25}$	0.30***		0.44***		0.37***		0.25***		0.21***		0.10		0.41***		0.25***		0.16	
	(0.06)		(0.07)		(0.09)		(0.05)		(0.06)		(0.07)		(0.10)		(0.08)		(0.10)	
$RV_{c,t-1}^{CoM125}$	0.08		−0.22***		−0.01		−0.00		0.14***		−0.07		0.00		0.04		0.16**	
	(0.06)		(0.07)		(0.07)		(0.04)		(0.04)		(0.07)		(0.07)		(0.05)		(0.07)	
$RV_{c,t-1}$		0.34***		0.31***		0.22***		0.54***		0.48***		0.41***		0.20***		0.22***		0.16***
		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)
$RV_{c,t-1} \times RQ_{c,t-1}$		0.07		0.19***		−0.01		0.77***		0.67***		0.72***		0.11***		0.07**		0.04
		(0.04)		(0.03)		(0.03)		(0.05)		(0.05)		(0.05)		(0.02)		(0.03)		(0.03)
$RV_{c,t-2\|t-5}$		0.17***		0.19***		0.27***		0.17***		0.17***		0.24***		0.26***		0.37***		0.32***
		(0.02)		(0.02)		(0.02)		(0.02)		(0.02)		(0.03)		(0.03)		(0.03)		(0.03)
$RV_{c,t-6\|t-22}$		0.21***		0.20***		0.27***		0.16***		0.20***		0.18***		0.32***		0.28***		0.30***
		(0.02)		(0.02)		(0.03)		(0.02)		(0.02)		(0.02)		(0.03)		(0.03)		(0.03)
Adj. R^2	0.36	0.41	0.29	0.33	0.36	0.38	0.74	0.77	0.68	0.72	0.53	0.58	0.43	0.43	0.50	0.49	0.43	0.42
OLS R^2	0.49	0.51																
McElroy R^2	0.45	0.48																
log likelihood	−40.03	492.69																
LR EVHEXP > HEXP	638.69***																	
LR EVHARQ > HARQ	459.90***																	

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table II.5: **EVHAR-TV and EVHARQ-TV estimation of RV_t**

This table reports the results of the EVHAR-TV and EVHARQ-TV model estimated with three lagged variables, the daily $RV_{c,t-1}$, weekly $RV_{c,t-2|t-5}$ *i.e.* the average of the RV from $t-2$ to $t-5$, and monthly $RV_{c,t-6|t-22}$ *i.e.* the average of the RV from $t-6$ to $t-22$. $M_{c,t}$ is a dummy set to "1" during the months of uncertainty resolution and to "0" otherwise (July for the three agriculture products and January for the natural gas contract). $TM_{c,t}$ is the log time to maturity. $SL_{c,t-1}$ is the log difference of the nearest term structure slope, $B_{c,t-1}$ a dummy set to "1" ("0") when the slope of the nearest term structure is negative (positive). The reported statistics are the means of the time-varying coefficients and *** indicates the significance of the t-test that the time-varying coefficients are not equal to zero, at the 1% level. $^{Min}/_{Max}$ indicate the minima and maxima of the time-varying coefficients. I report the optimal bandwidth selected by "leave-one-out cross-validation", the pseudo R^2 for each individual equation, the log likelihood, and the likelihood ratio tests (LR) that compares the unrestricted EVHAR-TV and EVHARQ-TV models with the HAR-TV and HARQ-TV, respectively. The sample period is April 22, 2010–April 1, 2019. The number of observations per equation is 2733.

	Agriculture						Energy						Metal					
	C		S		W		CL		HO		NG		GC		HG		SI	
Constant	−1.27***	−1.23***	−1.53***	−1.42***	−1.10***	−1.09***	−0.58***	−0.52***	−0.75***	−0.68***	−0.80***	−0.66***	−1.09***	−1.09***	−0.76***	−0.70***	−1.13***	−1.09***
	−1.42/−1.19	−1.38/−1.15	−1.61/−1.49	−1.49/−1.38	−1.23/−1.02	−1.22/−1.02	−0.66/−0.50	−0.57/−0.47	−0.84/−0.67	−0.75/−0.62	−0.82/−0.79	−0.68/−0.66	−1.15/−1.06	−1.13/−1.06	−0.80/−0.73	−0.74/−0.68	−1.20/−1.06	−1.16/−1.04
$M_{c,t}$	0.05***	0.05***	0.06***	0.06***	0.05***	0.05***					0.11***	0.09***						
	0.04/0.07	0.04/0.07	0.05/0.06	0.05/0.06	0.05/0.06	0.05/0.06					0.10/0.12	0.08/0.09						
$TM_{c,t}(10^2)$	0.33***	0.23***	2.17***	1.09***	0.97***	1.05***	−0.45***	−0.54***	0.34***	1.11***	1.20***	2.44***	5.84***	3.65***	6.42***	5.13***	11.97***	10.62***
	−0.44/1.16	−0.61/1.13	1.83/2.62	0.75/1.54	0.74/1.27	0.84/1.31	−0.69/−0.24	−0.86/−0.22	−0.17/0.72	0.78/1.53	0.95/1.78	2.16/2.85	5.74/5.92	3.38/4.05	6.30/6.49	4.92/5.39	11.75/12.33	9.93/11.30
$SL_{c,t-1}(10^{-2})$	−0.39***	0.41	0.84***	0.34	0.59***	0.38	0.62***	0.77	1.53***	0.72	0.46***	0.58	−12.35***	0.43	2.88***	0.49	−9.76***	0.42
	−0.52/−0.14	−0.63/−0.30	0.56/1.06	0.86/1.31	0.52/0.64	0.46/0.60	0.61/0.64	0.63/0.66	1.50/1.61	1.41/1.49	0.44/0.47	0.39/0.41	−14.40/−11.11	−16.11/−13.42	2.59/3.56	2.02/2.99	−10.19/−9.38	−11.08/−10.20
$B_{c,t-1}$	0.00***	0.00***	0.02***	0.02***	0.07***	0.07***	−0.02***	−0.02***	−0.04***	−0.04***	−0.02***	−0.03***	−0.04***	−0.05***	−0.06***	−0.06***	0.05***	0.05***
	0.00/0.01	0.00/0.01	0.02/0.03	0.02/0.03	0.06/0.08	0.06/0.08	−0.02/−0.01	−0.02/−0.01	−0.05/−0.04	−0.04/−0.04	−0.03/−0.02	−0.03/−0.02	−0.05/−0.03	−0.06/−0.03	−0.07/−0.06	−0.07/−0.06	0.03/0.05	0.04/0.05
$SL_{c,t-1} \times B_{c,t-1}(10^{-2})$	−0.18***	−0.06***	−1.43***	−1.66***	1.50***	1.39***	−1.83***	−2.11***	−3.24***	−3.20***	−1.21***	−1.03***	−44.90***	−45.53***	−12.07***	−11.48***	23.59***	24.24***
	−0.41/−0.04	−0.24/0.07	−1.66/−1.15	−1.89/−1.44	0.46/2.40	0.67/2.10	−2.00/−1.72	−2.26/−2.03	−3.35/−3.16	−3.29/−3.14	−1.23/−1.20	−1.06/−1.03	−58.14/−34.02	−63.04/−30.51	−13.37/−11.63	−12.56/−11.04	19.72/25.79	20.56/26.38
$RV_{c,t-1}$	0.32***	0.34***	0.26***	0.31***	0.21***	0.21***	0.42***	0.54***	0.36***	0.47***	0.21***	0.41***	0.20***	0.20***	0.20***	0.21***	0.16***	0.16***
	0.31/0.33	0.32/0.35	0.25/0.26	0.30/0.31	0.20/0.22	0.20/0.22	0.40/0.43	0.52/0.55	0.34/0.38	0.46/0.48	0.21/0.21	0.41/0.41	0.19/0.21	0.19/0.21	0.19/0.21	0.21/0.22	0.15/0.17	0.15/0.17
$RV_{c,t-1} \times RQ_{c,t-1}$		0.06***		0.19***		−0.01***		0.77***		0.67***		0.72***		0.11***		0.06***		0.04***
		0.04/0.09		0.18/0.20		−0.02/0.00		0.75/0.79		0.66/0.67		0.72/0.73		0.10/0.12		0.05/0.07		0.02/0.06
$RV_{c,t-2\|t-5}$	0.18***	0.17***	0.20***	0.18***	0.26***	0.27***	0.25***	0.17***	0.24***	0.17***	0.37***	0.24***	0.28***	0.26***	0.38***	0.37***	0.32***	0.32***
	0.17/0.18	0.16/0.17	0.20/0.21	0.18/0.19	0.25/0.27	0.25/0.28	0.23/0.27	0.14/0.19	0.23/0.25	0.16/0.18	0.37/0.37	0.24/0.25	0.26/0.29	0.25/0.28	0.37/0.38	0.36/0.38	0.30/0.34	0.30/0.33
$RV_{c,t-6\|t-22}$	0.21***	0.21***	0.22***	0.20***	0.27***	0.27***	0.20***	0.16***	0.23***	0.20***	0.23***	0.18***	0.32***	0.32***	0.28***	0.28***	0.30***	0.30***
	0.20/0.22	0.20/0.21	0.21/0.22	0.20/0.20	0.27/0.28	0.27/0.27	0.19/0.20	0.16/0.17	0.22/0.23	0.20/0.20	0.22/0.23	0.17/0.18	0.31/0.33	0.31/0.33	0.27/0.29	0.27/0.29	0.29/0.31	0.29/0.31
Bandwidth	0.56	0.40	0.75	0.36	0.65	0.58	6.50	0.67	20.00	20.00	1.16	0.95	0.90	0.61	0.59	0.88	0.62	0.45
Pseudo R^2	0.41	0.41	0.33	0.34	0.38	0.38	0.75	0.77	0.71	0.72	0.55	0.58	0.42	0.43	0.49	0.49	0.42	0.42
log likelihood	231.50	477.56																
LR EVHAR-TV>HAR-TV	313.65***																	
LR EVHARQ-TV>HARQ-TV	231.59***																	

*** $p < 0.01$

Table II.6: **Summary of EV performance**

This table summarizes the performance of the EV across the nine contracts, in the restricted and in all autoregressive specifications. For each model, I report the number of occurrence that each EV is positive at the 5% level (X^+), or negative at the 5% level (X^-). M is the monthly dummy set to "1" during the month of uncertainty resolution and to "0" otherwise, and is included in only four out of nine contracts, TM is the (log) time to maturity, SL is the log of the nearest term structure slope, B is a dummy set to "1", ("0") when the slope of the nearest term structure is negative (positive). The sample period is April 22, 2010–April 1, 2019.

	EV	EVHAR	EVHEXP	EVHARQ	EVHAR-TV	EVHARQ-TV
M	4^+	4^+	4^+	4^+	4^+	4^+
TM	7^+	4^+	4^+	4^+	8^+ / 1^-	8^+ / 1^-
SL	5^+ / 3^-	4^+ / 2^-	4^+ / 1^-	3^+ / 2^-	6^+ / 3^-	6^+ / 3^-
B	4^+ / 3^-	2^-	2^-	2^-	4^+ / 5^-	4^+ / 5^-
$SL \times B$	1^+ / 4^-	1^+ / 4^-	4^-	1^+ / 4^-	2^+ / 7^-	2^+ / 7^-

Table II.7: **In-sample model comparison**

This table reports the results of the model confidence set procedure; see Hansen et al. (2011). The reported values for each model are the mean of the losses of the three functions: (i) squared errors $MSE = \left(\hat{\sigma} - \sqrt{h}\right)^2$, (ii), absolute errors $MAE = |\hat{\sigma} - \sqrt{h}|$, and (iii) $QLIKE = \ln h + \frac{\hat{\sigma}^2}{h}$. When applicable, the superscripts 1, 2, ..., n indicate the ranking of the models that are included in the confidence interval $\widehat{\mathcal{M}}_{90\%}$. I display the results for one day, one week, and one month ahead. The sample period is April 22, 2010–April 1, 2019. The number of observations is $2255 \times 9 = 20295$.

	EV	HAR	EVHAR	HEXP	EVHEXP	HARQ	EVHARQ	HAR-TV	EVHAR-TV	HARQ-TV	EVHARQ-TV
One day ahead											
MSE	0.1316	0.0785	0.0773	0.0806	0.0792	0.0768	^{2}*0.0759*	0.0774	0.0774	^{1}**0.0756**	0.0760
MAE	0.2764	0.1984	0.1981	0.2018	0.2014	0.1962	0.1959	0.1971	0.1982	^{1}**0.1948**	0.1960
QLIKE	3.8969	3.8909	3.8908	3.8913	3.8911	3.8907	^{2}*3.8906*	3.8908	3.8908	^{1}**3.8906**	3.8906
One week ahead											
MSE	0.1331	0.0925	0.0902	0.0941	0.0919	0.0916	0.0895	^{1}**0.0746**	0.0906	0.0792	0.0900
MAE	0.2783	0.2194	0.2180	0.2217	0.2204	0.2187	0.2173	^{1}**0.1931**	0.2183	0.2018	0.2178
QLIKE	3.8971	3.8926	3.8923	3.8928	3.8925	3.8924	3.8922	^{1}**3.8905**	3.8924	3.8910	3.8923
One month ahead											
MSE	0.1361	0.1131	0.1078	0.1138	0.1086	0.1127	0.1075	^{2}*0.0823*	0.0950	^{1}**0.0823**	0.0942
MAE	0.2808	0.2486	0.2430	0.2497	0.2442	0.2485	0.2427	^{1}**0.2073**	0.2243	^{2}*0.2074*	0.2234
QLIKE	3.8975	3.8949	3.8943	3.8950	3.8944	3.8949	3.8943	^{1}**3.8915**	3.8929	^{2}*3.8915*	3.8928

Table II.8: **Out-of-sample model comparison**

Panel A reports the results of the model confidence set procedure; see Hansen et al. (2011). The reported values for each model are the mean of the losses of the three functions: (i) squared errors $MSE = \left(\hat{\sigma} - \sqrt{h}\right)^2$, (ii), absolute errors $MAE = |\hat{\sigma} - \sqrt{h}|$, and (iii) $QLIKE = \ln h + \frac{\hat{\sigma}^2}{h}$. When applicable, the superscripts 1, 2, …, n indicate the ranking of the models that are included in the confidence interval $\widehat{\mathcal{M}}_{90\%}$. I report the results for the one day, week, and month forecasts. Panel B reports the χ^2 statistics for the modified Diebold-Mariano test; see Diebold and Mariano (1995) and Harvey et al. (1997). The null hypothesis is that the model in the row-entry is equal to the one of the column-entry. I report the statistics for the one-day ahead forecasts MSE. The in-sample period is April 22, 2010–January 30, 2013 and the out-of-sample period is January 31, 2013–April 1, 2019. The number of observation in the out-of-sample period is $1555 \times 9 = 13995$.

	EV	HAR	EVHAR	HEXP	EVHEXP	HARQ	EVHARQ	HAR-TV	EVHAR-TV	HARQ-TV	EVHARQ-TV	RiskMetrics
Panel A: Model confidence set												
One day ahead												
MSE	0.2220	0.0810	0.0808	0.0927	0.0905	^{1}**0.0788**	^{2}*0.0791*	0.0863	0.0819	0.0815	^{3}0.0793	0.0930
MAE	0.3800	^{5}0.2038	0.2059	0.2228	0.2217	^{1}**0.2012**	^{3}0.2032	0.2113	0.2075	^{4}0.2043	^{2}*0.2027*	0.2230
QLIKE	3.9293	3.9124	3.9124	3.9138	3.9136	^{1}**3.9121**	^{2}*3.9121*	3.9130	3.9125	3.9124	^{3}3.9122	3.9139
One week ahead												
MSE	0.2209	^{4}0.0963	^{3}0.0953	0.1140	0.1097	^{2}*0.0950*	^{1}**0.0943**	0.9771	0.1000	0.1593	^{5}0.0973	0.1049
MAE	0.3798	^{2}*0.2284*	^{4}0.2301	0.2515	0.2481	^{1}**0.2277**	^{3}0.2291	0.5166	0.2351	0.3022	^{5}0.2316	0.2395
QLIKE	3.9292	3.9143	3.9141	3.9163	3.9158	^{2}*3.9141*	^{1}**3.9140**	3.9588	3.9147	3.9209	3.9143	3.9152
One month ahead												
MSE	0.2245	0.1302	0.1278	0.1639	0.1543	0.1300	0.1274	0.3099	0.1050	0.2908	^{1}**0.1038**	0.1340
MAE	0.3833	0.2782	0.2776	0.3121	0.3035	0.2783	0.2771	0.4125	0.2421	0.4095	^{1}**0.2406**	0.2789
QLIKE	3.9297	3.9183	3.9180	3.9222	3.9210	3.9182	3.9179	3.9451	3.9152	3.9452	^{1}**3.9151**	3.9186
Panel B: Modified Diebold-Mariano test												
One day ahead												
EV												
HAR	59.35***											
EVHAR	60.11***	0.83										
HEXP	57.79***	−18.95***	−17.28***									
EVHEXP	58.98***	−14.50***	−17.16***	5.51***								
HARQ	58.08***	6.88***	5.14***	18.37***	15.74***							
EVHARQ	59.52***	5.12***	7.85***	17.79***	17.73***	−0.87						
HAR-TV	55.17***	−7.00***	−7.03***	8.51***	5.14***	−8.99***	−8.63***					
EVHAR-TV	60.01***	−2.26*	−3.90***	16.35***	16.76***	−6.98***	−8.05***	6.15***				
HARQ-TV	57.46***	−1.26	−1.53	16.57***	12.41***	−7.76***	−5.47***	7.54***	0.79			
EVHARQ-TV	59.43***	3.97***	4.20***	18.42***	19.04***	−1.31	−0.87	9.14***	11.95***	5.83***		
RiskMetrics	56.51***	−16.65***	−15.61***	−0.47	−3.45***	−17.05***	−16.34***	−9.65***	−15.65***	−16.25***	−17.36***	

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

Table II.9: **Forecast bias of multi-quantile regressions**

This table reports the p-values of the Wald tests achieved on the parameters from multi-quantile predictive panel regressions estimated with the daily log price changes and one-day, one-week, and one-month ahead out-of-sample RV forecasts, and for four number of quantiles $p = 1, 2, 4$, as in Couperier and Leymarie (2020). The quantile levels always start at $\tau_1 = 0.975$, which is the Basel Committee on Banking Supervision (Basel III) regulatory level, and each level u_i within is determined as: $\tau + (u_i - 1) \times {}^{(1-\tau)}/_{p}$, for each $u_i = 1, 2, \ldots, p$. The tests are for the following null hypotheses: (i) $H_{0,J_1} : \sum_{j=1}^{p} (\beta_0(\tau_j)) + (\beta_1(\tau_j)) = p$, (ii) $H_{0,J_2} : \sum_{j=1}^{p} \beta_0(\tau_j) = 0$ and $\sum_{j=1}^{p} \beta_1(\tau_j) = p$, (iii) $H_{0,I} : \sum_{j=1}^{p} \beta_0(\tau_j) = 0$, and (iv) $H_{0,S} : \sum_{j=1}^{p} \beta_1(\tau_j) = p$. The p-values are obtained with a bootstrap of 1,000 replications. The sample period is January 31, 2013–April 1, 2019. The number of observations is $1555 \times 9 = 13995$.

	EV	HAR	EVHAR	HEXP	EVHEXP	HARQ	EVHARQ	HAR-TV	EVHAR-TV	HARQ-TV	EVHARQ-TV	RiskMetrics
One-day ahead												
$p = 1$												
J_1	0.82	0.61	0.82	0.68	0.39	0.29	0.67	0.18	0.27	0.55	0.68	0.23
J_2	0.01	0.42	0.31	0.15	0.05	0.22	0.26	0.00	0.04	0.02	0.04	0.02
I	0.23	0.29	0.59	0.74	0.78	0.07	0.24	0.87	0.87	0.46	0.50	0.93
S	0.80	0.60	0.85	0.69	0.40	0.28	0.65	0.19	0.27	0.57	0.70	0.24
$p = 2$												
J_1	0.21	0.74	0.35	0.61	0.83	0.32	0.20	0.09	0.74	0.41	0.93	0.18
J_2	0.04	0.35	0.19	0.08	0.04	0.18	0.13	0.00	0.03	0.01	0.04	0.01
I	0.05	0.31	0.06	0.82	0.38	0.08	0.03	0.96	0.51	0.67	0.20	0.98
S	0.20	0.72	0.33	0.62	0.86	0.31	0.19	0.10	0.76	0.43	0.90	0.19
$p = 4$												
J_1	0.17	0.63	0.23	0.74	0.98	0.31	0.18	0.17	0.75	0.47	0.87	0.21
J_2	0.04	0.33	0.16	0.12	0.04	0.17	0.16	0.00	0.05	0.01	0.04	0.01
I	0.05	0.26	0.06	0.70	0.28	0.08	0.06	0.92	0.45	0.70	0.20	0.98
S	0.16	0.62	0.22	0.76	0.95	0.30	0.17	0.18	0.77	0.49	0.85	0.22
One-week ahead												
$p = 4$												
J_2	0.04	0.55	0.48	0.04	0.07	0.42	0.19	0.00	0.00	0.00	0.01	0.01
One-month ahead												
$p = 4$												
J_2	0.07	0.18	0.30	0.00	0.01	0.18	0.36	0.00	0.02	0.00	0.03	0.00

Table II.10: **Percentage of 97.5% VaR violations**

This table reports the violation percentage of the 97.5% VaR which corresponds to a single quantile regression of coverage $\tau = 0.975$ from the one-day ahead forecasts and the average violation percentage for the nine commodities. The sample period is January, 2013–April 1, 2019. The number of observations is $1555 \times 9 = 13995$.

	EV	HAR	EVHAR	HEXP	EVHEXP	HARQ	EVHARQ	HAR-TV	EVHAR-TV	HARQ-TV	EVHARQ-TV	RiskMetrics
C	1.35	1.04	1.04	1.04	0.96	1.12	1.04	1.04	1.04	1.20	1.12	1.20
S	1.20	1.35	1.20	1.20	1.27	1.12	1.12	1.35	1.20	1.12	1.27	1.04
W	1.04	1.20	1.12	1.20	1.12	1.12	1.12	1.20	1.20	1.20	1.12	1.27
CL	1.35	1.20	1.20	1.12	1.20	1.20	1.20	1.20	1.20	1.20	1.20	1.35
HO	0.96	1.59	1.27	1.27	1.20	1.51	1.20	1.51	1.27	1.27	1.20	1.51
NG	1.12	1.35	1.51	1.20	1.20	1.27	1.12	1.27	1.20	1.35	1.20	1.27
GC	1.27	1.12	1.12	1.12	1.27	1.12	1.04	1.12	1.27	1.04	1.27	1.43
HG	1.20	1.35	1.35	1.35	1.20	1.43	1.43	1.35	1.35	1.43	1.43	1.43
SI	1.27	0.96	1.20	1.27	1.43	1.12	1.12	1.35	1.12	0.96	1.12	1.20
Average	1.20	1.24	1.22	1.20	1.20	1.22	1.15	1.27	1.20	1.20	1.21	1.30

Appendix

Table II.A1: **Description of futures contracts**

This table reports the specifications of the futures contracts written on the nine selected commodities. The specifications include the ticker, underlying commodity, unit, maturity, the trading hours used for the sample (Globex) and the maximum number of 5-min trade observations per day.

Ticker	Trading venue	Underlying	Unit	Maturity	Trading hours (CT)	# 5-min obs.
				Agriculture		
C	CBT	Corn	bu (5,000)	HKNUZ	Sunday-Friday, 7:00 p.m.-7:45 a.m. and Monday-Friday, 8:30 a.m.-1:20 p.m.	211
S	CBT	Soybeans	bu (5,000)	FHKNQUX	Sunday-Friday, 7:00 p.m.-7:45 a.m. and Monday-Friday, 8:30 a.m.-1:20 p.m.	211
W	CBT	Chicago wheat	bu (5,000)	HKNUZ	Sunday-Friday, 7:00 p.m.-7:45 a.m. and Monday-Friday, 8:30 a.m.-1:20 p.m.	211
				Energy		
CL	NYMEX/ICE	WTI crude oil	bbl (1,000)	FGHJKMNQUVXZ	Sunday-Friday 5:00 p.m.-4:00 p.m. 60-min break at 4:00 p.m.	276
HO	NYMEX	Heating oil	gal (42,000)	FGHJKMNQUVXZ	Sunday-Friday 5:00 p.m.-4:00 p.m. 60-min break at 4:00 p.m.	276
NG	NYMEX/ICE	Natural gas	MMBtu (10,000)	FGHJKMNQUVXZ	Sunday-Friday 5:00 p.m.-4:00 p.m. 60-min break at 4:00 p.m.	276
				Metal		
GC	CMX	Gold	oz (100)	GJMQVZ	Sunday-Friday 5:00 p.m.-4:00 p.m. 60-min break at 4:00 p.m.	276
HG	COMEX	Copper	lb (25,000)	FGHJKMNQUVXZ	Sunday-Friday 5:00 p.m.-4:00 p.m. 60-min break at 4:00 p.m.	276
SI	CMX	Silver	oz (5,000)	FHKNUZ	Sunday-Friday 5:00 p.m.-4:00 p.m. 60-min break at 4:00 p.m.	276

Letter code: F = January, G = February, H = Mars, J = April, K = May, M = June, N = July, Q = August, U = September, V = October, X = November, Z = December.

Table II.A2: **Variable definition**

This table defines all variables used throughout the article, with their symbol, plain definition, computation, unit, frequency, and additional descriptions when applicable. For all variables, the sample period is: May 5, 2008–April 1, 2019.

Symbol	Title	Computation	Unit	Frequency	Description
$F^m_{c,t}$	Futures price	–	USD	5 min	m: maturity, c: commodity, t: time
$f^m_{c,t}$	Log futures price	$f^m_{c,t} = \ln F^m_{c,t}$	–	5 min	m: maturity, c: commodity, t: time
$r_{c,t}$	Log price change	$r_{c,t} = f_{c,t} - f_{c,t-1}$	%	5 min	m: maturity, c: commodity, t: time
$M_{c,t}$	Monthly dummy	$M_{c,t} = \begin{cases} 1 & \text{if } t \in \text{critical month} \\ 0 & \text{if } t \notin \text{critical month} \end{cases}$	–	D	"Critical month": July (corn, soybeans, and wheat) and January (natural gas).
$TM_{c,t}$	(Log) time to maturity	$TM_{c,t} = \ln(T - t)$	s	D	T and t are the time at maturity and current time in POSIX time format.
$SL_{c,t}$	(Annualized log) term structure	$SL_{c,t} = \frac{(f^N_{c,t}) - (f^{FD}_{c,t})}{\delta} \times 250$	–	D	Where N is the nearby contract, FD is the first deferred contract, and δ is the time difference in days between two consecutive maturities. I annualize the variable for ease of interpretation.
$B_{c,t}$	Backwardation dummy	$B_{c,t} = \begin{cases} 1 & \text{if } SL_{c,t} < 0 \\ 0 & \text{if } SL_{c,t} >= 0 \end{cases}$	–	D	–
Δ	The temporal period = 300 s	Reciprocal of $^{10}/_{3}$ mHz in SI	s	s	With a day composed of 211 and 276 5-min (300 s) observations for agriculture contracts and energy and metal contracts, respectively, so that $^{1}/_{\Delta} = [211; 276]$.
$ARV_{c,t}$	Arithmetic realized volatility	$ARV_{c,t} = \sqrt{\sum_{j=1}^{1/\Delta} r^2_{t,\Delta\times j}}$	–	D	
$RV_{c,t}$	Log realized volatility	$RV_{c,t} = \frac{1}{2} \times \ln \sum_{j=1}^{1/\Delta} r^2_{t,\Delta\times j}$	–	D	
$RQ_{c,t}$	(Log realized quarticity)	$RQ_t = \frac{2}{3} \frac{\sum_{i=1}^{1/\Delta} r^4_{i,t}}{\left(\sum_{i=1}^{1/\Delta} r^2_{i,t}\right)^2}$	–	D	In amount of 5-minute periods elapsed each day, *i.e.*, $j = 1, 2, \ldots, n$.
$RV_{c,t-k\|t-n}$	Average of the RV from $t-k$ to $t-n$	$RV_{c,t-k\|t-n} = \frac{1}{n-k+1} \sum_{i=k}^{n} RV_{c,t-i}$	–	D	–
$RV^{CoM(\lambda)}_{c,t}$	Exponential average of the RV	$RV^{CoM(\lambda)}_{c,t} = \sum_{i=1}^{500} \frac{e^{-i\lambda}}{e^{-\lambda}+e^{-2\lambda}+\ldots+e^{-500\lambda}}$	–	D	With $\lambda = \ln\left(1 + \frac{1}{\text{CoM}}\right)$

Table II.A3: **Summary statistics: Daily RV with alternative sampling frequency**

This table reports statistics on daily log realized volatility sampled at 1, 5, 15, and 60 minute-intervals in panels A to D, respectively. In panel E, I add the statistics for the average of these four measures, as in Andersen et al. (2011b). For each panel, I report the four moments of the distribution, the Jarque-Bera statistic (JB χ^2), the Ljung-Box statistic for the 20$^{\text{th}}$ order serial correlation ($Q(20)$), and the parameter of the log-periodogram regression based on a bandwidth exponent of $4/5$ (d), as in Andersen et al. (2003). The sample period is May 5, 2008–April 1, 2019. The number of observations per contract is 2755.

	Agriculture			Energy			Metal		
	Corn (C)	Soybeans (S)	Wheat (W)	WTI crude oil (CL)	Heating oil (HO)	Natural gas (NG)	Gold (GC)	Copper (HG)	Silver (SI)
					Panel A: RV_t 1-minute sampling				
Mean%	−402.74	−430.24	−393.64	−396.43	−409.49	−363.35	−466.62	−424.40	−406.93
σ%	42.72	46.16	40.80	46.28	44.28	38.46	50.95	46.48	51.43
Skewness	3.24	3.15	3.08	0.29	0.51	0.30	1.83	0.89	0.67
Kurtosis	25.50	24.11	27.48	1.96	3.95	3.41	15.20	3.52	9.34
JB	79,586	71,432	91,174	481	1,916	1,379	28,121	1,784	10,245
$Q(20)$	7,995	9,645	7,340	32,279	29,356	17,552	11,641	23,501	10,770
d	0.35	0.38	0.30	0.53	0.45	0.48	0.34	0.50	0.40
					Panel B: RV_t 5-minute sampling				
Mean%	−415.13	−434.90	−400.54	−400.03	−412.02	−369.11	−468.93	−427.66	−412.04
σ%	46.48	46.19	41.54	46.69	43.93	39.85	50.87	46.46	52.02
Skewness	2.06	2.37	2.16	0.51	0.42	0.55	1.61	0.65	0.58
Kurtosis	13.73	17.02	18.39	0.61	0.36	1.01	13.22	1.14	7.55
JB	23,626	35,914	41,036	161	95	259	21,286	343	6,706
$Q(20)$	11,283	10,703	8,201	32,443	31,598	18,330	12,106	23,773	11,329
d	0.34	0.40	0.32	0.52	0.53	0.51	0.33	0.51	0.38
					Panel C: RV_t 15-minute sampling				
Mean%	−421.47	−437.96	−404.39	−402.19	−414.20	−373.01	−470.73	−429.22	−414.93
σ%	50.33	47.83	43.27	47.97	45.23	41.54	52.02	47.27	52.95
Skewness	1.64	2.17	1.93	0.47	0.37	0.47	1.55	0.55	0.67
Kurtosis	10.42	15.17	16.09	0.53	0.37	1.02	12.22	0.91	6.59
JB	13,738	28,629	31,470	132	77	223	18,283	236	5,197
$Q(20)$	10,848	9,618	7,178	29,330	28,077	15,463	11,447	21,689	10,863
d	0.35	0.39	0.31	0.51	0.48	0.44	0.31	0.49	0.37
					Panel D: RV_t 60-minute sampling				
Mean%	−427.89	−441.96	−409.07	−406.34	−418.90	−378.32	−475.29	−432.69	−419.19
σ%	56.01	52.20	48.10	51.65	49.39	47.09	55.49	50.68	56.73
Skewness	1.23	1.67	1.45	0.34	0.23	0.32	1.34	0.39	0.81
Kurtosis	7.13	11.01	10.89	0.39	0.26	0.58	9.67	0.74	5.56
JB	6,549	15,229	14,596	72	33	86	11,567	132	3,854
$Q(20)$	8,606	7,318	5,406	21,352	20,053	10,619	8,700	16,626	8,127
d	0.31	0.35	0.31	0.42	0.40	0.42	0.30	0.41	0.32
					Panel E: RV_t average of 1-, 5-, 15-, and 60-minute sampling				
Mean%	−414.66	−435.20	−400.61	−400.22	−412.68	−369.68	−469.49	−427.72	−412.20
σ%	45.93	46.20	41.28	46.88	44.33	40.14	51.17	46.66	51.57
Skewness	2.14	2.38	2.20	0.50	0.40	0.55	1.62	0.60	0.77
Kurtosis	14.40	17.07	18.86	0.56	0.34	1.05	12.90	1.04	6.87
JB	25,942	36,111	43,102	154	85	268	20,329	289	5,695
$Q(20)$	10,643	10,475	8,160	31,440	30,408	17,361	11,800	22,889	11,387
d	0.35	0.39	0.32	0.52	0.52	0.50	0.33	0.50	0.39

Table II.A4: **Average daily turnover and open interest over five years prior to the sample**

This table reports the average daily turnover and open interest in millions USD for 20 contracts components of the SP-GSCI/BCOM, prior to the sample period selection. The pre-sample period is January 1, 2003–April 30, 2008, and the contracts selected for the study are in bold font.

Energy	**WTI crude oil (CL)**	**Heating oil (HO)**	Brent crude oil (LCO)	Gasoil (LGO)	**Natural gas (NG)**	RBOB gasoline (RB)
Turnover	15,126.82	3,476.32	7,860.83	2,804.78	5,093.58	3,969.98
Open interest	28,209.41	8,962.40	15,468.90	7,704.32	16,167.50	3,528.27

Agriculture	**Corn (C)**	Feeder cattle (FC)	Kansas wheat (KW)	Live cattle (LC)	Lean hogs (LH)	**Soybeans (S)**	**Wheat (W)**
Turnover	1,891.39	203.31	329.13	955.63	478.49	3,056.77	1,111.43
Open interest	10,752.81	1,192.12	2,208.12	6,597.15	2,756.31	10,155.23	5,678.65

Metal	**Gold (GC)**	**Copper (HG)**	Platinum (PL)	**Silver (SI)**
Turnover	4,252.86	1227.43	227.20	1,300.46
Open interest	13,488.65	3792.76	625.33	4,339.29

Soft	Cocoa (CC)	Cotton (CT)	Coffee (KC)	Orange juice (OJ)	Raw sugar (SB)
Turnover	177.56	467.97	602.04	58.25	626.54
Open interest	1,644.18	3,395.23	4,724.07	433.67	4,592.96

Table II.A5: **HAR, HEXP, and HARQ estimation of RV_t**

This table reports the results of the restricted HAR, HEXP, and HARQ estimated with 5-minute log realized volatility $RV_{c,t}$. The HAR and HARQ have three lagged variables, the daily $RV_{c,t-1}$, weekly $RV_{c,t-2|t-5}$ *i.e.* the average of the RV from $t-2$ to $t-5$, and monthly $RV_{c,t-6|t-22}$ *i.e.* the average of the RV from $t-6$ to $t-22$. I implement the parsimonious version of the HARQ, adding the product of the first autoregressive term $RV_{c,t-1}$ with the log realized quarticity measure $RQ_{c,t-1}$ (Barndorff-Nielsen and Shephard, 2002). The HEXP has four terms exponentially smoothed over a backward looking 500-day window for four lagged variables, *i.e.* centers of mass (CoM) of 1, 5, 25, and 125 days. I report Newey and West (1994) corrected standard errors with automatic lag selection in parenthesis. I report the adjusted R^2 for each individual equation of the system, the overall OLS R^2, McElroy R^2, and the log likelihood. The sample period is June 3, 2008–April 1, 2019. The number of observations per equation is 2733 for the HAR and HARQ and 2255 for the HEXP.

	Agriculture								
	C			S			W		
Constant	−1.09***	−1.05***	−1.36***	−1.20***	−1.11***	−1.83***	−0.86***	−0.85***	−1.04***
	(0.08)	(0.08)	(0.10)	(0.09)	(0.09)	(0.17)	(0.09)	(0.09)	(0.13)
$RV_{c,t-1}$	0.33***	0.35***		0.27***	0.32***		0.22***	0.22***	
	(0.02)	(0.02)		(0.02)	(0.02)		(0.02)	(0.02)	
$RV_{c,t-1} \times RQ_{c,t-1}$		0.07*			0.20***			−0.00	
		(0.04)			(0.03)			(0.03)	
$RV_{c,t-2\|t-5}$	0.20***	0.19***		0.23***	0.20***		0.28***	0.28***	
	(0.02)	(0.02)		(0.02)	(0.02)		(0.02)	(0.02)	
$RV_{c,t-6\|t-22}$	0.22***	0.22***		0.24***	0.23***		0.30***	0.30***	
	(0.02)	(0.02)		(0.02)	(0.02)		(0.03)	(0.03)	
$RV_{c,t-1}^{CoM1}$			0.47***			0.33***			0.31***
			(0.04)			(0.04)			(0.05)
$RV_{c,t-1}^{CoM5}$			−0.12**			0.06			0.08
			(0.06)			(0.07)			(0.09)
$RV_{c,t-1}^{CoM25}$			0.32***			0.35***			0.42***
			(0.06)			(0.07)			(0.09)
$RV_{c,t-1}^{CoM125}$			0.02			−0.14**			−0.05
			(0.05)			(0.06)			(0.07)
Adj. R^2	0.41	0.41	0.36	0.32	0.33	0.29	0.38	0.38	0.36
OLS R^2	0.49	0.50	0.48						
McElroy R^2	0.45	0.47	0.43						
log likelihood	−29.81	244.74	−359.38						

	Energy								
	CL			HO			NG		
Constant	−0.36***	−0.31***	−0.44***	−0.43***	−0.37***	−0.36***	−0.46***	−0.36***	−0.95***
	(0.04)	(0.04)	(0.06)	(0.05)	(0.05)	(0.07)	(0.07)	(0.06)	(0.16)
$RV_{c,t-1}$	0.44***	0.55***		0.40***	0.51***		0.25***	0.46***	
	(0.02)	(0.02)		(0.02)	(0.02)		(0.02)	(0.02)	
$RV_{c,t-1} \times RQ_{c,t-1}$		0.76***			0.68***			0.77***	
		(0.05)			(0.06)			(0.05)	
$RV_{c,t-2\|t-5}$	0.27***	0.19***		0.27***	0.20***		0.41***	0.27***	
	(0.02)	(0.02)		(0.02)	(0.02)		(0.03)	(0.03)	
$RV_{c,t-6\|t-22}$	0.20***	0.17***		0.23***	0.20***		0.22***	0.16***	
	(0.02)	(0.02)		(0.02)	(0.02)		(0.03)	(0.02)	
$RV_{c,t-1}^{CoM1}$			0.67***			0.61***			0.28***
			(0.04)			(0.04)			(0.05)
$RV_{c,t-1}^{CoM5}$			−0.03			0.00			0.49***
			(0.06)			(0.06)			(0.08)
$RV_{c,t-1}^{CoM25}$			0.26***			0.15***			0.11
			(0.05)			(0.05)			(0.07)
$RV_{c,t-1}^{CoM125}$			0.01			0.16***			−0.13*
			(0.03)			(0.04)			(0.07)
Adj. R^2	0.75	0.76	0.74	0.70	0.71	0.67	0.53	0.57	0.51

	Metal								
	GC			HG			SI		
Constant	−0.71***	−0.74***	−0.42***	−0.58***	−0.56***	−0.42***	−0.70***	−0.69***	−0.46***
	(0.10)	(0.10)	(0.13)	(0.08)	(0.08)	(0.11)	(0.09)	(0.09)	(0.11)
$RV_{c,t-1}$	0.22***	0.22***		0.23***	0.23***		0.20***	0.17***	
	(0.02)	(0.02)		(0.02)	(0.02)		(0.02)	(0.02)	
$RV_{c,t-1} \times RQ_{c,t-1}$		0.12***			0.13***			0.11***	
		(0.02)			(0.03)			(0.02)	
$RV_{c,t-2\|t-5}$	0.29***	0.28***		0.40***	0.39***		0.34***	0.33***	
	(0.03)	(0.03)		(0.03)	(0.03)		(0.03)	(0.03)	
$RV_{c,t-6\|t-22}$	0.34***	0.36***		0.25***	0.26***		0.32***	0.34***	
	(0.03)	(0.03)		(0.03)	(0.03)		(0.03)	(0.03)	
$RV_{c,t-1}^{CoM1}$			0.21***			0.27***			0.21***
			(0.05)			(0.06)			(0.06)
$RV_{c,t-1}^{CoM5}$			0.29***			0.39***			0.33***
			(0.10)			(0.09)			(0.11)
$RV_{c,t-1}^{CoM25}$			0.34***			0.20**			0.20*
			(0.10)			(0.08)			(0.10)
$RV_{c,t-1}^{CoM125}$			0.09			0.06			0.18***
			(0.06)			(0.05)			(0.06)
Adj. R^2	0.42	0.43	0.42	0.47	0.48	0.48	0.41	0.42	0.42

$^{***}p < 0.01; ^{**}p < 0.05; ^{*}p < 0.1$

Table II.A6: **HAR-TV and HARQ-TV estimation of RV_t**

This table reports the results of the joint estimation (SUR) of the restricted HAR-TV (Eq. (7)) and HARQ-TV (Eq. (8)) models, computed with 5-minute log realized volatility $RV_{c,t}$ and with three lagged variables, the daily $RV_{c,t-1}$, weekly $RV_{c,t-2|t-5}$ *i.e.* the average of the RV from $t-2$ to $t-5$, and monthly $RV_{c,t-6|t-22}$ *i.e.* the average of the RV from $t-6$ to $t-22$. I implement the parsimonious version of the model, adding the product of the first autoregressive term $RV_{c,t-1}$ with the log realized quarticity measure $RQ_{c,t-1}$ (Barndorff-Nielsen and Shephard, 2002). The reported statistics are the means of the time-varying coefficients and *** indicates the significance of the t-test that the time-varying coefficients are not equal to zero, at the 1% level. $^{Min}/_{Max}$ indicate the minima and maxima of the time-varying coefficients. I report the optimal bandwidth selected by "leave-one-out cross-validation", the pseudo R^2 for each individual equation, and the log likelihood of the system. The sample period is June 3, 2008–April 1, 2019. The number of observations per equation is 2733.

	Agriculture						Energy						Metal					
	C		S		W		CL		HO		NG		GC		HG		SI	
Constant	-1.29^{***}	-1.25^{***}	-1.27^{***}	-1.16^{***}	-0.78^{***}	-0.79^{***}	-0.39^{***}	-0.34^{***}	-0.45^{***}	-0.37^{***}	-0.48^{***}	-0.38^{***}	-0.82^{***}	-0.84^{***}	-0.71^{***}	-0.59^{***}	-1.14^{***}	-1.11^{***}
	$^{-1.77}/_{-0.88}$	$^{-1.62}/_{-0.93}$	$^{-1.67}/_{-0.81}$	$^{-1.43}/_{-0.79}$	$^{-1.05}/_{-0.57}$	$^{-1.07}/_{-0.55}$	$^{-0.46}/_{-0.31}$	$^{-0.44}/_{-0.26}$	$^{-0.52}/_{-0.38}$	$^{-0.41}/_{-0.33}$	$^{-0.52}/_{-0.45}$	$^{-0.47}/_{-0.31}$	$^{-1.01}/_{-0.72}$	$^{-1.06}/_{-0.73}$	$^{-1.03}/_{-0.44}$	$^{-0.86}/_{-0.45}$	$^{-1.69}/_{-0.76}$	$^{-1.70}/_{-0.77}$
$RV_{c,t-1}$	0.31^{***}	0.32^{***}	0.26^{***}	0.31^{***}	0.23^{***}	0.23^{***}	0.45^{***}	0.56^{***}	0.40^{***}	0.51^{***}	0.24^{***}	0.45^{***}	0.23^{***}	0.22^{***}	0.23^{***}	0.24^{***}	0.18^{***}	0.16^{***}
	$^{0.22}/_{0.38}$	$^{0.20}/_{0.41}$	$^{0.23}/_{0.30}$	$^{0.24}/_{0.38}$	$^{0.20}/_{0.26}$	$^{0.19}/_{0.27}$	$^{0.44}/_{0.46}$	$^{0.53}/_{0.59}$	$^{0.39}/_{0.41}$	$^{0.49}/_{0.52}$	$^{0.21}/_{0.29}$	$^{0.39}/_{0.52}$	$^{0.19}/_{0.28}$	$^{0.17}/_{0.27}$	$^{0.18}/_{0.29}$	$^{0.21}/_{0.27}$	$^{0.12}/_{0.29}$	$^{0.08}/_{0.29}$
$RV_{c,t-1} \times RQ_{c,t-1}$		0.03^{***}		0.20^{***}		-0.02^{***}		0.75^{***}		0.65^{***}		0.77^{***}		0.14^{***}		0.15^{***}		0.12^{***}
		$^{-0.15}/_{0.20}$		$^{0.05}/_{0.34}$		$^{-0.12}/_{0.07}$		$^{0.68}/_{0.82}$		$^{0.62}/_{0.67}$		$^{0.70}/_{0.84}$		$^{0.12}/_{0.17}$		$^{0.13}/_{0.15}$		$^{0.07}/_{0.16}$
$RV_{c,t-2\|t-5}$	0.18^{***}	0.18^{***}	0.22^{***}	0.20^{***}	0.29^{***}	0.29^{***}	0.27^{***}	0.19^{***}	0.27^{***}	0.21^{***}	0.41^{***}	0.27^{***}	0.29^{***}	0.28^{***}	0.39^{***}	0.38^{***}	0.29^{***}	0.28^{***}
	$^{0.13}/_{0.21}$	$^{0.13}/_{0.22}$	$^{0.14}/_{0.31}$	$^{0.13}/_{0.30}$	$^{0.27}/_{0.32}$	$^{0.27}/_{0.31}$	$^{0.25}/_{0.28}$	$^{0.16}/_{0.21}$	$^{0.26}/_{0.28}$	$^{0.19}/_{0.23}$	$^{0.39}/_{0.43}$	$^{0.24}/_{0.29}$	$^{0.22}/_{0.37}$	$^{0.21}/_{0.34}$	$^{0.36}/_{0.43}$	$^{0.36}/_{0.41}$	$^{0.21}/_{0.37}$	$^{0.20}/_{0.36}$
$RV_{c,t-6\|t-22}$	0.21^{***}	0.21^{***}	0.24^{***}	0.22^{***}	0.30^{***}	0.30^{***}	0.20^{***}	0.16^{***}	0.22^{***}	0.19^{***}	0.23^{***}	0.17^{***}	0.32^{***}	0.33^{***}	0.23^{***}	0.24^{***}	0.27^{***}	0.30^{***}
	$^{0.12}/_{0.32}$	$^{0.12}/_{0.33}$	$^{0.18}/_{0.31}$	$^{0.16}/_{0.31}$	$^{0.23}/_{0.34}$	$^{0.25}/_{0.33}$	$^{0.19}/_{0.20}$	$^{0.14}/_{0.17}$	$^{0.22}/_{0.23}$	$^{0.18}/_{0.19}$	$^{0.20}/_{0.25}$	$^{0.14}/_{0.19}$	$^{0.21}/_{0.42}$	$^{0.23}/_{0.43}$	$^{0.20}/_{0.24}$	$^{0.22}/_{0.25}$	$^{0.15}/_{0.35}$	$^{0.16}/_{0.39}$
Bandwidth	0.23	0.25	0.29	0.30	0.63	0.63	4.03	0.61	20.00	20.00	0.74	0.72	0.53	0.57	0.40	0.81	0.25	0.25
Pseudo R^2	0.42	0.42	0.34	0.34	0.38	0.39	0.75	0.77	0.70	0.71	0.53	0.57	0.42	0.43	0.48	0.49	0.43	0.44
log likelihood	74.80	244.74																

$^{***}p < 0.01$

Table II.A7: **Forecast bias of individual quantile regressions**

This table reports the coefficients β_0 and β_1 of individual predictive quantile regressions, for the nine commodities and for four levels $u_i = 1, 2, \ldots, 4$, when $p = 4$. I also report the χ^2 of the joint test that $\beta_0 = 0\%$ and $\beta_1 = 100\%$, as in Couperier and Leymarie (2020). The data are the negative daily log price changes of the nine commodities and the one-day-ahead daily RV forecasts from all models tested. The sample period is January, 2013–April 1, 2019. The number of observations is $1555 \times 9 = 13995$.

	EV	HAR	EVHAR	HEXP	EVHEXP	HARQ	EVHARQ	HAR-TV	EVHAR-TV	HARQ-TV	EVHARQ-TV	RiskMetrics
$\underline{u_1}$												
β_0	−0.34	0.15	0.08	0.05	0.04	0.23	0.17	0.03	−0.02	0.12	0.11	0.01
	(0.23)	(0.14)	(0.15)	(0.14)	(0.16)	(0.14)	(0.14)	(0.16)	(0.15)	(0.16)	(0.16)	(0.16)
β_1	103.14	97.41	100.65	102.09	103.58	95.25	97.82	107.88	105.76	103.16	101.69	105.93
	(7.47)	(5.31)	(5.54)	(5.45)	(5.67)	(5.31)	(5.18)	(6.43)	(5.21)	(6.22)	(5.88)	(6.01)
χ^2	203.30***	20.85***	34.88***	55.32***	85.17***	41.65***	40.38***	273.62***	106.60***	171.90***	100.48***	153.49***
$\underline{u_2}$												
β_0	−0.43	0.20	0.19	0.11	0.12	0.27	0.27	0.05	0.13	0.12	0.21	0.10
	(0.26)	(0.16)	(0.14)	(0.16)	(0.16)	(0.15)	(0.13)	(0.16)	(0.16)	(0.16)	(0.15)	(0.18)
β_1	105.54	96.26	96.35	100.78	101.50	94.59	93.90	106.94	101.29	103.60	99.22	103.16
	(8.63)	(5.79)	(4.84)	(5.99)	(5.74)	(5.58)	(4.67)	(6.74)	(5.39)	(6.42)	(5.12)	(6.38)
χ^2	205.10***	33.51***	29.46***	65.73***	108.98***	53.73***	45.76***	283.63***	114.44***	207.34***	132.52***	162.51***
$\underline{u_3}$												
β_0	−0.50	0.26	0.27	0.12	0.10	0.31	0.28	−0.02	0.15	0.08	0.24	0.07
	(0.33)	(0.19)	(0.17)	(0.19)	(0.16)	(0.18)	(0.17)	(0.18)	(0.17)	(0.16)	(0.14)	(0.19)
β_1	106.93	94.83	92.91	101.00	101.83	93.54	93.21	110.34	100.59	105.74	97.21	103.99
	(9.75)	(6.56)	(5.13)	(6.94)	(5.43)	(6.49)	(5.35)	(7.21)	(5.97)	(6.17)	(5.27)	(7.53)
χ^2	202.61***	48.07***	40.65***	94.36***	106.41***	61.90***	44.49***	425.66***	105.65***	297.59***	88.79***	164.69***
$\underline{u_4}$												
β_0	−0.88	0.22	0.48	0.04	0.21	0.34	0.45	−0.04	0.22	0.02	0.26	−0.01
	(0.35)	(0.20)	(0.20)	(0.20)	(0.20)	(0.20)	(0.18)	(0.22)	(0.19)	(0.19)	(0.18)	(0.20)
β_1	118.61	97.70	88.44	105.12	97.91	93.85	89.37	110.85	98.52	106.92	96.95	109.73
	(9.92)	(7.09)	(6.69)	(6.82)	(6.22)	(7.01)	(6.25)	(7.52)	(6.22)	(6.66)	(6.24)	(7.73)
χ^2	184.76***	72.15***	128.22***	206.99***	72.35***	86.68***	115.86***	503.51***	103.06***	305.72***	91.30***	497.78***

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

Fig. II.A1: **Unconditional monthly average RV**

These plots display the monthly average of daily RV for the nearby futures contract written on crude oil, heating oil, gold, copper, and silver. The red line represents the average and the gray area represents the 90% confidence bands. The sample period is May 5, 2008–April 1, 2019. The number of observations per contract is 2755.

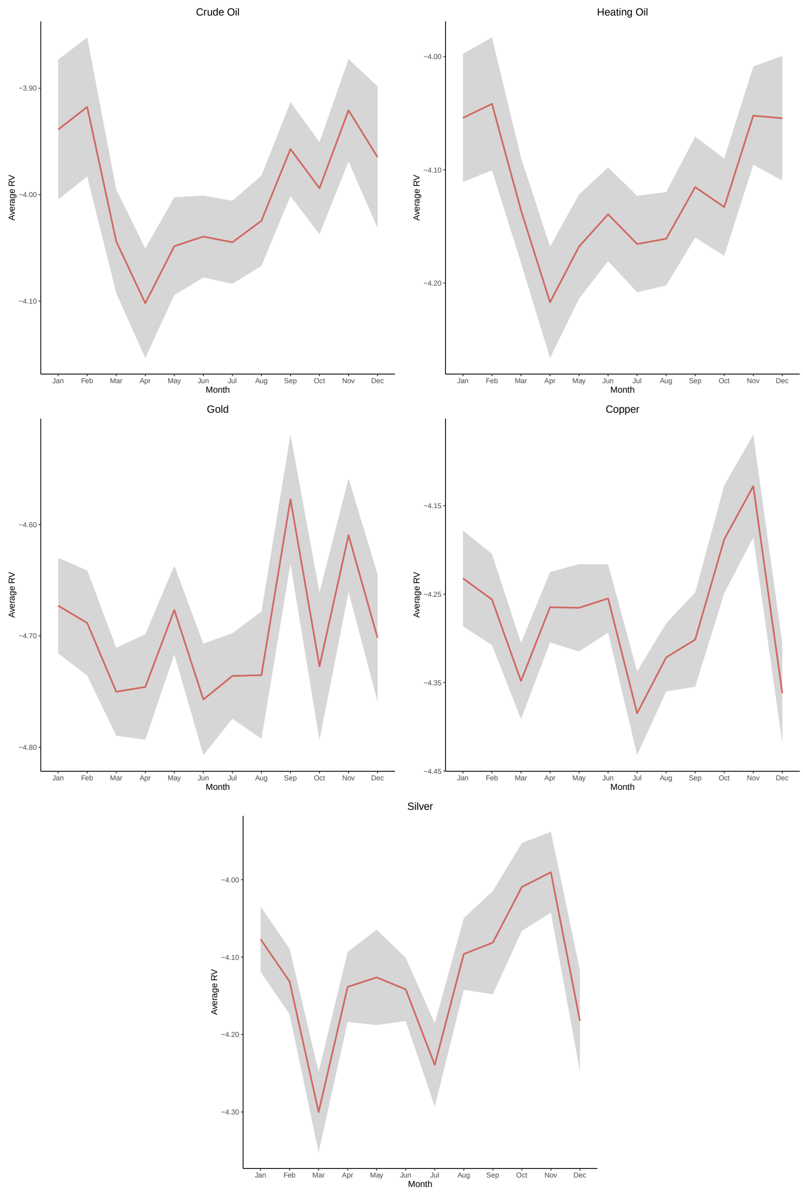
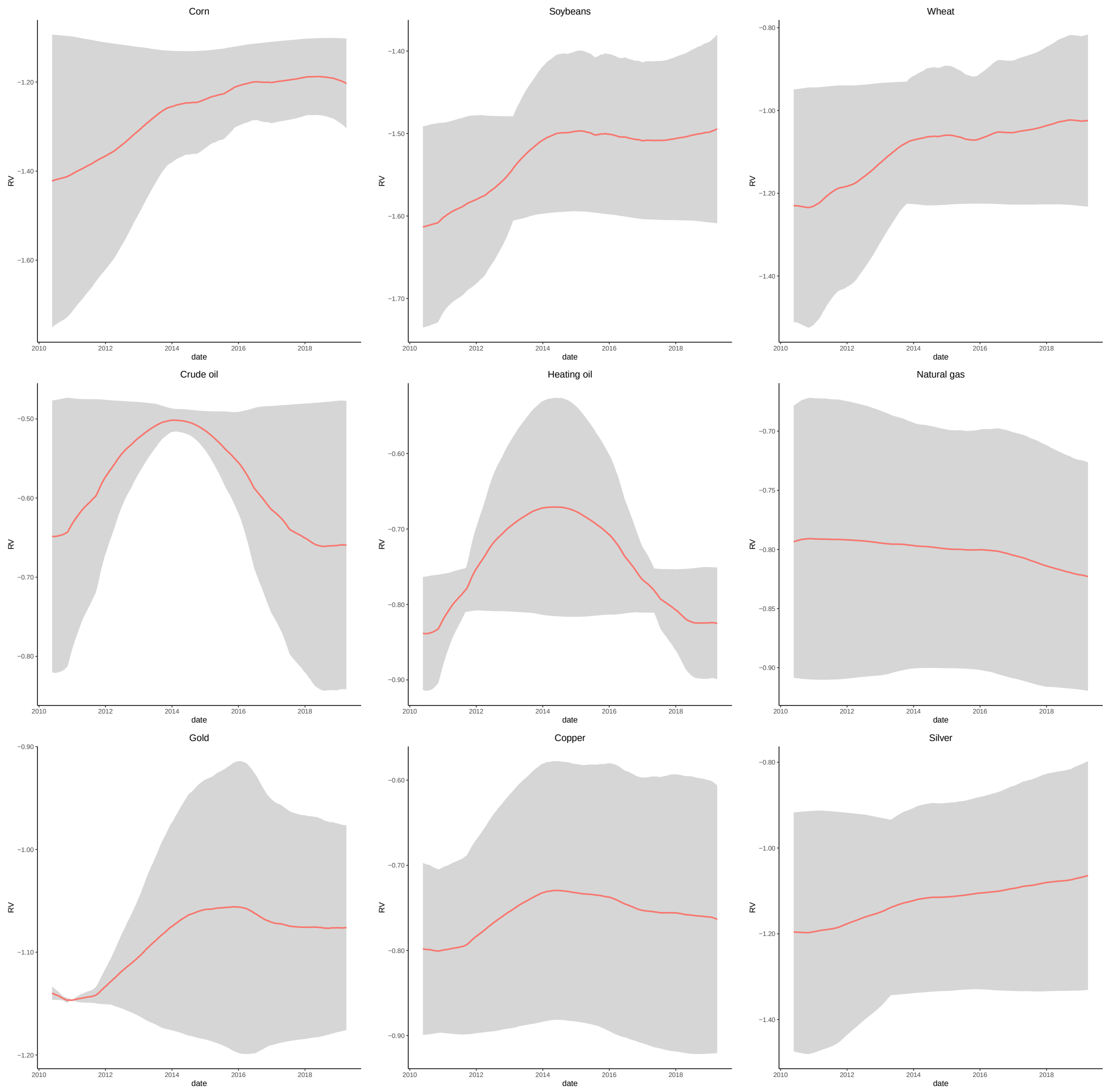


Fig. II.A2: **Time-varying intercepts in the EVHAR-TV model**

This plot displays the pattern of time-varying intercepts in the joint estimation of the EVHAR-TV model (SUR). The intercept unit is RV for the nine nearby futures contracts. The sample period is May 5, 2010–April 1, 2019. The 90% confidence intervals are displayed with the shaded area and the number of observations per contract is 2755.



Chapter III

A tale of two premiums revisited

1. Introduction

In a recent paper, Kang et al. (2020) (hereafter, KRT) uncover two risk premiums in commodity returns, *i.e.*, the long-term horizon insurance premium paid by producers through hedging demand, and the short-term horizon liquidity premium earned by these producers for easing the trading activity of speculators. The goal of this paper is to check the robustness of these results under different scenarios. I conjecture that the presence of a third category of investors, the commodity index traders (CITs), affects these premiums since their objective is to get exposure to the spot price of commodities rather than just grabbing the insurance premium. These investors are net "buyers" of commodity contracts. On one hand, this should decrease the insurance premium and, on the other hand, CITs rolling their position from the nearby to higher maturity contracts should increase the liquidity premium on the nearby contract, at least during the period starting after the roll to the maturity of the contracts. The net effect of CITs on contract demand mirrors that of processors and users. However, the causes of this demand are different. While the former are including commodities in their portfolio for diversification purpose (wealth management), the latter are hedging their position (risk management).

Testing for the existence of (risk) premiums in commodity futures markets is a long-lasting challenge. The reasons include, among others, the small number of contracts available (cross-section), their heterogeneity in terms of demand, their lack of integration with traditional asset classes, and non-available spot prices. From a theoretical perspective, Black (1976) denies the existence of a commodity risk premium resulting from the exposure to a global factor. A large strand of the empirical literature supports this view when additional commodity market peculiarities (*e.g.* hedging pressure) are not controlled appropriately; see, *e.g.*, Dusak (1973), Bodie and Rosansky (1980), or Carter, Rausser, and Schmitz (1983). Moreover, whereas it seems natural to relate commodities returns to changes in aggregate consumption, and despite the existence of theoretical models (consumption-based CAPM) designed for commodity pricing, there is weak empirical evidence concerning the existence of a significant risk premium; see, *e.g.*, Hazuka (1984) and Jagannathan (1985). Empirical tests built on characteristics derived from the theory of "Normal Backwardation" or the theory of "Storage" do not provide more convincing results; see Keynes (1930), Hicks (1946), Kaldor (1939), Working (1949), and Brennan (1958). Similarly, characteristics derived from market frictions common to many asset markets, such as transaction costs, liquidity, or limits to arbitrage, also work well to explain the cross-section of commodity futures returns. This research uses these characteristics, and traded factors built upon these characteristics, to provide risk adjustments.

To identify under which conditions the results of KRT hold after the inclusion of variables related directly or indirectly to the financialization, I determine the set of factors that allows for an optimal risk adjustment. Beyond insurance and risk premiums, I use six additional characteristics, which are known to predict commodity futures returns. Next, from these characteristics, I construct factors (long-short portfolios sorted on characteristics, and rebalanced weekly). As previously documented for monthly or bi-monthly frequencies, I document a superior performance for portfolios sorted on (i) basis (annualized average weekly returns of 15.75%), (ii) momentum (14.46%), and (iii) basis-momentum (14.04%). I also confirm their performance at weekly frequency for portfolios sorted on (i) average hedging pressure (9.99%), (ii) net trading of "commercials" (17.99%), and (iii) crowding (17.85%); see Kang et al. (2020), and Kang, Rouwenhorst, and Tang (2021). In contrast, the portfolios sorted on the β with an average portfolio (Bakshi et al., 2019) or on the growth of open interest (Hong and Yogo, 2012) deliver poor annualized average weekly returns (-0.58% and -1.51%, respectively).

Next, I select the optimal subset of factors with the Bayesian asset pricing test derived by Barillas and Shanken (2018). The initial set of factors does not include the characteristics under scrutiny (average hedging pressure and net trading). I find that the highest probability (61%, for a prior on the maximum attainable Sharpe ratio of 1.25) is obtained with the four-factor model that includes the basis, momentum, basis-momentum, and the crowding factor. This result holds when the test-assets are the remaining portfolios or the remaining portfolios plus the 26 individual commodities of the sample.

First-generation indices have a systematic methodology to carry their long position from the nearby onto the first deferred futures contract called the "roll". The amount rolled every day over the window is known.[1] Using panel data and time series regression, I perform preliminary analyses concerning the impact of the roll days

[1]From the fifth to the ninth workable days for the SP-GSCI.

on commodities returns and turnover panels, and on the factor returns, respectively. I do not find any statistically significant systematic effect of the roll-days on individual commodity returns, and on the factors, in both pre- and post-financialization periods. This result supports the view that the financialization does not affect prices through the rolling of SP-GSCI contracts. Conversely, I find that the roll is associated with a significant increase in turnover, in particular for the nearby contract. Moreover, this effect is the most visible in the post-financialization period, with an average cumulative abnormal turnover of 17.56%.

Next, I replicate the analysis of KRT (Fama-MacBeth regressions, Fama and MacBeth, 1973). I also estimate a related model where I introduce the optimal set of factors selected previously. I find that the insurance (average hedging pressure) and liquidity (net trading) premiums are robust to this adjustment, despite a drop in the economic magnitude (from 0.43 to 0.34), and the statistical significance (from the 1% to the 10%) of the insurance premium. A one standard deviation change in AHP impacts the returns by 8.2 and 6.5 bps, respectively.

I pursue these robustness tests, and control for the financialization with three different approaches. First, I directly add a measure of index funds pressure to the regression. I find that KRT results are not affected by this variable and that the coefficient of this variable does not show any statistical significance at the usual levels. Second, I restrict the Fama-MacBeth regressions to the cross-sections of the weeks that have at least a three-day overlap with the roll of the SP-GSCI. This aims to control whether KRT results are not driven by the particular roll weeks. The significance levels of the insurance and liquidity coefficients drops at the 5% and 10% level, respectively. In addition, while the economic magnitude of the insurance coefficient is unchanged, that of Q is halved at 2.32%. Third, I split the sample in a pre- and post-financialization period to capture more generally a change in the risk-sharing structure based on the study period. I find that the liquidity coefficient is not affected by the period, but that of the insurance is not significant at the usual level (significant at the 10% level) in the pre- (post-)financialization period.

Several papers have documented the econometric limits of the Fama-MacBeth approach; see Petersen (2009), and Gow, Ormazabal, and Taylor (2010). Therefore, I use an alternative approach developed by Hoechle, Schmid, and Zimmermann (2020).[2] It permits to include contract fixed-effects, a feature that seems necessary, given the important heterogeneity of commodities. It additionally allows to control for the dependence between the construction of risk-adjustment factors, and the

[2]An obvious limitation is the limited cross-sectional dimension, of 26 commodity contracts, in this context.

characteristics of interest (or unobservable characteristics). In this setting, I confirm that contract fixed-effects contributes significantly to explain the returns. The Haussman (1978) test rejects the random effect hypothesis at the 1% level in all specifications. However, I find that the liquidity premium is robust to econometric and risk-adjustments, but not the insurance premium. When fixed effects are added, the coefficient is not significant at the usual levels. Moreover, this latter coefficient becomes negative (and not statistically significant) in the post-financialization period. Finally, I conclude the analysis with an economic robustness specification.The disaggregated CFTC dataset, available since 2007, allows the computation of more accurate values of CITs pressure. I do not find any statistically significant effect related to CITs pressure, and confirm that in the post-financialization period, additionally extended until 2020, the price of insurance is not statistically significant either.

My contribution is threefold. First, I identify a set of factors that optimally price commodity futures returns. This has implications in research, to *e.g.* improve the counterfactual of event-studies or benchmark the performance of future factors. Contrary to Boons and Porras Prado (2019), who compare the performance of the basis-momentum in turn, against the basis and Bakshi et al. (2019) factors (momentum, basis, and average factor), I find that, at the weekly frequency, basis-momentum does not suffice to generate the optimal asset pricing model. This result is also useful for portfolio management in practice. Second, I show that the roll of index funds neither affects the returns, nor the factors. However, I find that the roll days do increase the turnover, implying that CITs modify the functioning of commodity markets. Finally, using an appropriate methodology for longitudinal panels, I find that the financialization does not affect the liquidity premium earned by producers, whereas the price of the insurance premium decreases, and eventually vanishes. This goes against the argument that index funds activity is detrimental to the commodity economic activity, since it benefits commodity producers.

The remainder of the paper is organized as follows. Section 2 provides a literature review and develops hypotheses on previously identified characteristics, risk premiums, and integration of commodity markets in the broader economy. Section 3 describes the data and presents the summary statistics. Section 4 presents the empirical results, and Section 5 provides the econometric robustness tests. Section 6 concludes.

2. Literature review and hypotheses development

2.1. *Theoretical models of commodity prices*

2.1.1. Normal backwardation

The theory of Normal Backwardation states that producers depress the futures price below the expected spot price because they sell future output through futures contracts to hedge against price fluctuations; see Keynes (1930) and Hicks (1946). Hence, a long futures investor obtains a positive premium from the producer that insures herself. Cootner (1960) and Gray (1961) extend the theory to buyers of physical commodities who generate net long hedging pressure, and potentially a sign reversal of the insurance premium. Testing this theory is a threefold question: (i) Is the futures price a downward biased estimate of the expected spot price? (ii) Are speculators' profits positive on average? (iii) Do speculators' profits arise from the insurance premium or superior forecasting ability? A subsequent strand of literature builds on the real or latent substitution effects across commodities, leading to cross-hedging. Many physical commodities do not have corresponding futures markets. Therefore, commodity hedgers may use a resembling or a linear combination of resembling futures contracts to hedge their position; see Anderson and Danthine (1981). If the cross-hedging hypothesis holds, the hedging pressure in one futures contract is also a function of other cash markets; see, *e.g.*, Roon *de* et al. (2000) and Chng (2009).

2.1.2. Normal backwardation and integration with asset markets

To test whether commodity futures contracts embed a risk premium, Dusak (1973) develops a model, similar to the CAPM, where commodity returns covariations with an aggregated wealth index are positive. She tests her model on three agricultural futures contracts and finds that commodity futures betas are not significantly different from zero. Black (1976) argues that, even if (commodity) futures are included in the market portfolio, their aggregation sums up to zero. Thus, there is no reason to detect a risk premium in commodity futures arising from covariation with the market. In relaxing the assumption that speculators are net long, and adding the conjecture that the market portfolio does include commodity futures, Carter et al. (1983) derive a commodity CAPM that includes the market index and the net position of speculators. They argue that commodity futures premiums depend on commodity futures prices covariation with the market, and on the extent to which speculators are incentivized to trade. The empirical findings confirm that speculators do earn excess returns, and that the level of systematic risk is a function of the speculators' position. Hirshleifer (1988, 1990) derives an equilibrium model where physical commodities may be hardly marketable (Mayers, 1973). In this setting, market frictions materialize through trading setup costs (Merton, 1987); see also Stoll (1979) and Carter et al. (1983). Bessembinder (1992, 1993) find empirical support for this model; see also Roon *de* et al. (2000).

2.1.3. Consumption, industrial production, and inflation

Given the nature of commodities, *i.e.*, goods that are produced and consumed, it is natural to relate their price to consumption growth or production growth as well as unexpected inflation. Breeden (1980) derives a commodity consumption-based CAPM, in which the consumption betas gather individual commodity futures exposure to consumption and specific income elasticity of demand.[3] The income elasticity of demand may theoretically drive the consumption beta below zero, depending on the nature of the commodity. He confirms empirically that grain products have negative consumption betas. Interestingly, the derivation of the model through elasticities also allows to switch the model from a consumption to a production-based CAPM. However, Breeden (1980) does not test the relation between consumption betas and returns. Hazuka (1984) regresses a single cross-section of 14 commodity futures monthly returns, defined as the exposures of price changes over the basis, on consumption betas. The author finds that the price of risk is statistically significant at the 1% level.[4] Jagannathan (1985) derives a multi-period commodity consumption model. Based on three commodities (corn, soybeans, and wheat contracts), he shows that the risk aversion and the stochastic discount factor are similar to those obtained for stocks. Nevertheless, the model is rejected at the 1% level. Grauer and Litzenberger (1979) derive a two-period model where commodity futures prices are biased estimates of expected spot prices due to inflation risk borne by investors. The model collapses with the CAPM when the market risk premium is adjusted appropriately (deflated terminal wealth scaled by initial market portfolio value).[5]

[3]This approach considers that constant time to maturity provides the same distribution for all periods (debatable for commodities whose supply is seasonal such as agricultural products). It departs from Dusak (1973), who considers that each maturity has its own (stable) distribution.

[4]Note that the author provides little details on the methodology.

[5]For empirical evidence of commodities (futures) as a hedge against (unexpected) inflation, also see Greer (2000), and Erb and Harvey (2006, 2016).

2.1.4. Theory of Storage

Kaldor (1939), Working (1949), and Brennan (1958) view the futures price as the spot price compounded by interest rate and storage costs, and discounted by a convenience yield, similar to a continuous dividend yield in stock index futures. This convenience yield represents the benefit to physically own the product instead of receiving it later. It depends on the risk aversion of the buyer and the levels of inventory. The theory of Storage is extended by Deaton and Laroque (1992, 1996) who consider the non-negativity constraint of inventories. In particular, their model explains the high skewness of commodity futures returns induced by the non-linear price-inventory relation. They attribute the high level of autocorrelation in commodity futures returns to the presence of speculators who smooth futures prices. They conclude that autocorrelation is likely induced by supply and demand processes that are not identically, and independently distributed (*e.g.* serial dependence in harvests). Other models of storage analyze the physical commodity price as a call option ; see respectively Routledge, Seppi, and Spatt (2000), and Schwartz (1997).

2.2. Market integration and financialization

2.2.1. Common factors

There is weak evidence of commodity markets being integrated with stock and bond markets. Empirical tests confirm that the systematic risk premium is zero; see Dusak (1973) and Bodie and Rosansky (1980).[6] In an exercise similar to Fama and French (1993), Daskalaki, Kostakis, and Skiadopoulos (2014) use various stock factors (market portfolio, size, value) and macro-economic factors (liquidity, consumption, and foreign exchange) to test whether commodities are integrated.[7] They find that none of the classic factors price the cross-section of commodity futures returns. They also show that commodities are heterogeneous within their asset class since the first five principal components of commodity futures returns explain only 60% of the total variance. Moreover, the principal components do not price the cross-section of commodity returns, as opposed to equities; see Roll and Ross (1980).

[6]The scope of this review concerns commodity futures prices in relation with the broader economy. It does not treat commodities futures as pricing factors of other assets; see, *e.g.*, Hong and Sarkar (2008) and Hong and Yogo (2012).

[7]Note that the bond-stock market segmentation is now revised down, and a number of empirical studies find that corporate bonds are exposed to the same risk factors as stocks; see, *e.g.*, Bai, Bali, and Wen (2019).

Despite the segmentation of commodity futures markets relative to other asset classes, academic studies find strong evidence for an integration of the commodity term structure in the broader economy. In particular, Bailey and Chan (1993) directly relate the basis (the price difference between two contracts on the same commodity but with different maturities), to macro-economic variables such as recession, liquidity, and volatility. Similarly, Yang (2013) relates theoretically and empirically the basis to investment shocks in the economy. A plausible explanation is that the basis filters out the heterogeneous and specific commodity effects, leaving the core free to react to economic innovations; see also Hirshleifer (1988, 1989). Finally, the integration of the basis/term structure is also consistent with the interest rate component of the futures prices that spot prices do not carry; see Schwartz (1997).

2.2.2. Limits to arbitrage

In two mirror studies, Acharya et al. (2013) and Rampini et al. (2014) analyze the relation between firm distress, approximated by the expected default frequency (EDF), and hedging demand.[8] Acharya et al. (2013) study this relation from the producer side. If arbitragers (speculators) of commodity futures are capital constrained, they will increase the (aggregate) insurance premium, turning producers to hedge less and to sell their output outright, thereby also decreasing spot prices. To test this model empirically, they use the EDF and financial disclosures of natural gas and crude oil-producing firms to estimate the hedging demand. They confirm that the cross-section of hedging demand is positively related to the firm default risk. Additional tests show that (i) the futures risk premium increases with the producer's hedging demand, (ii) spot prices respond negatively to aggregated measures of default risk, and (iii) the two previous relations are exacerbated when speculators have a low-risk capacity; see Etula (2013). Rampini et al. (2014) consider the buyer's standpoint. Their model shows that hedging is positively related to income (net worth) and inversely related to default risk, which is confirmed empirically with data from airline companies. These findings hold both in cross-section and time series. At the first glance, the results of Acharya et al. (2013) and Rampini et al. (2014) are contradictory, and contribute to an existing debate on the firm's propensity to hedge when they are financially constrained. In brief, theoretical models predict a positive relation between default risk and hedging, whereas empirical evidence point to the opposite; see, *e.g.*, Tufano (1996). However, the peculiarities of the two industries they study matter. First, in Rampini et al. (2014), the

[8] See also Cheng et al. (2015).

model incorporates collateralization for the buyer, which in turns imply a trade-off between hedging demand and financing. This feature is absent from Acharya et al. (2013) and, most probably, absent from the true supplier's decision model. Second, the buyer's profit function is convex and decreasing in input prices. Therefore, the buyer's incentive to hedge, which is costly, decreases in the high price scenarios. Instead, from the producer perspective, the profit function is increasing and linear in the output price. Moreover, the producer may optimize her hedging demand in managing inventories.[9]

2.2.3. Commodity pricing: The consequences of the financialization

Gorton and Rouwenhorst (2006), and Erb and Harvey (2006) document the long-run properties of the investments in commodity futures, either as a diversification asset class or in terms of absolute performance. These papers are concomitant to new and large inflows from investors into commodity markets, a phenomenon coined as the "financialization" of commodities. Subsequently, given the size of the inflows, the novelty of these investments, and the commodity prices increase before the 2008 crisis, both academic research and newspapers suspected these inflows to cause distortions in the functioning of commodity markets. In particular, Masters and White (2008), and Masters (2008) in a public hearing before the US Senate, argue that the sharp increase in the share of index investing in the total open interest of commodities was responsible for market distortions and prices misaligned with their fundamental values. Typically, the financialization generates excess co-movements between commodity futures prices (Tang and Xiong, 2012), increases the correlation between commodities and stock markets (Buyuksahin and Robe, 2014), induces price pressure effects during the roll of the larges commodity indices (Mou, 2011), and changes in the risk-sharing structure of commodity markets; see Brunetti and Reiffen (2014), and Dubois and Maréchal (2021). The date attributed to the materialization of "financialization" varies across studies from 1999 (Mou, 2011) to 2008 (Adams and Gluck, 2015) but the consensus and the statistical evidence point to December 2003; see Dubois and Maréchal (2021). Brunetti and Reiffen (2014) derive a model in which uninformed CITs mitigate hedging costs. Using a non-public data-set of individual traders' positions, they find evidence that CITs supply insurance to producers through their activity. If this risk-sharing in the commodity

[9] The different profit functions of producers and buyers also relate to Keynes (1930) view, that a buyer/processor can report an increase in input prices on output prices (pass-through), whereas the commodity producer's profit only depends on her output price. This also justifies, in the theory of Normal Backwardation, that the futures price should be a downward biased estimate of the expected spot price, even with a negative or nil net hedging pressure.

market changes, it also induces a modification of the liquidity premium.

2.3. Commodity characteristics, risk factors, and returns: Empirical evidence.

2.3.1. Hedging pressure

Chang (1985) disentangles the speculators' forecasting ability and the Normal Backwardation hypothesis for three agricultural commodity contracts split by maturity, over the period 1951-1980. The results vary greatly over time, maturity months, and commodities, but the existence of a risk premium in the Keynes-Hicks sense cannot be discarded. In empirical tests, Bessembinder (1992, 1993) shows that the residual risk explains agricultural futures returns, conditional on hedging pressure. Roon *de* et al. (2000) study the hedging pressure aggregated at the sector level (cross-hedging pressure). Futures premiums depend on the market, own-hedging pressure, and cross-hedging pressure. Additionally, they show that this effect is not conditional on price pressure defined as the period-to-period change of hedging pressure.[10] KRT extend the hedging pressure framework. They uncover a second premium arising from the speculators' liquidity demand. The existence of this premium explains why empirical tests on hedging pressure fail to identify risk premium when liquidity is not controlled appropriately. Speculators have no superior forecasting ability but, on average, they are rewarded the positive insurance premium; see Buyuksahin et al. (2009), Ederington and Lee (2002), Buyuksahin and Harris (2011), and Dewally, Ederington, and Fernando (2013). Basu and Miffre (2013) construct portfolios of commodity futures sorted on hedging pressure. The cross-sectional returns of these portfolios is priced with a long-short inter-quartile factor built on the same characteristics. This premium is orthogonal to momentum and carry premiums; see also Szymanowska, Roon *de*, Nijman, and Goorbergh *van den* (2014).

2.3.2. Test of the theory of Storage and the basis

There are three approaches to test the theory of Storage. The first one uses actual data on inventories and relate them to the level of prices and basis. However, inventory data are scarce, difficult to aggregate across countries, and often do not include governmental strategic reserves; see Pindyck (1994, 2004), Gorton et al. (2012), and Dincerler, Khokher, and Simin (2020). Combining Deaton and Laroque

[10] For cross-characteristics and commodity relations see Casassus, Liu, and Tang (2013).

(1992) with Routledge et al. (2000), Gorton et al. (2012) find support for non-linear relations between bases and inventories, and between inventories and risk premiums. Since they provide evidence for a close, convex, relation between convenience yield (basis) and inventories, the basis is a noisy proxy for inventories. Further evidence is provided by Fama and French (1987, 1988).

The second approach consists of extracting inventory levels through filtering the convenience yield with a state-space representation. However, the process of the convenience yield (Brownian or Ornstein-Uhlenbeck motion) must be specified ahead; see, *e.g.*, Gibson and Schwartz (1990), Schwartz (1997), and Casassus and Collin-Dufresne (2005).

Third, the nearby futures contract is used as a proxy for the spot price; see, *e.g.*, Fama and French (1987, 1988), Bailey and Chan (1993), Yang (2013), Szymanowska et al. (2014), and Koijen, Moskowitz, Pedersen, and Vrugt (2018). Using 21 nearby monthly returns over the period 1966-1984, Fama and French (1987) find strong support for the theory of Storage, whereas the insurance premium (Normal Backwardation) is present in only five contracts. They uncover further support by examining the basis of LME forward contracts for which spot prices are observable; see Fama and French (1988). Given the fact that commodity bases are readily measurable at any frequency, portfolio returns sorted on basis are used in asset pricing tests. For instance, Yang (2013) combines a basis factor with an average commodity factor to price commodity futures returns in cross-section. He finds that the basis is related to investment shocks affecting the broader economy. This is consistent with the basis being correlated with other macroeconomic risks (stock index dividend yield and corporate bond quality spread); see Bailey and Chan (1993). Szymanowska et al. (2014) also document the superior performance of the basis. They find insignificant alphas for a variety of portfolios (commodity portfolios sorted on basis and other characteristics such as momentum, volatility, hedging pressure, inflation, and liquidity). Finally, Koijen et al. (2018) find that the "carry" factor predicts returns in time series, and in cross-section, for various asset classes such as equities, bonds, commodities, and options.

2.3.3. Momentum

Financial research struggles to produce a risk-based explanation of the stock momentum anomaly; see, *e.g.*, Jegadeesh and Titman (1993) and Asness, Moskowitz, and Pedersen (2013). The rationale of momentum in commodity futures prices is more readily available. As Deaton and Laroque (1992, 1996) emphasize, if the underlying process of supply (*e.g.*, the harvest) is not independent, and identically distributed, the commodity returns are serially correlated.[11] Miffre and Rallis (2007) find that short-term momentum (up to 12-month ranking period, 1-month holding period) is associated with long-term reversals (value effect). Asness et al. (2013) test the value and momentum factors in several asset classes including commodities, and find statistically significant risk premiums induced by exposure to these factors. If the identification of the momentum is straightforward for any asset class, that of the value factor is ambiguous for commodities (longer-term sum of returns). Moskowitz, Ooi, and Pedersen (2012) confirm the momentum effect in various asset classes, including commodities. This predictability is unrelated to other risk factors. It is beneficial to speculators, and detrimental to hedgers. Bakshi et al. (2019) include momentum as one of the three factors describing commodity futures premiums along with the basis and the commodity average portfolio. Taken individually, these factors perform poorly.

2.3.4. Basis-momentum

As for bonds, the convexity of the term structure also matters for commodity futures. Boons and Porras Prado (2019) use the difference between two momentum signals in the nearby and deferred commodity futures contracts to obtain a variable that aggregates both the slope and the curvature of the commodity term structure. In the longest monthly time series available (1959-2014), and broadest cross-section (32 commodities), they find that a portfolio sorted on this signal has a superior pricing ability over other known commodity factors. They additionally document that this factor is not subsumed by momentum and basis factors, and that the curvature is the most important component of the signal. Moreover, in pooled regressions, they find that this factor predicts nearby and first deferred contracts, as well as spreading returns. Finally, they observe that the commodity futures exposure to this factor alone prices the cross-section of individual futures returns and proxies for the volatility risk.[12]

2.3.5. Open interest, commodity liquidity, and crowding

If commodities forecast the economy through their hedging demand, prices may cancel out this demand as it represents the sum of long and short hedging (*e.g.* a commodity producer *vs.* a processor). Following this principle, Hong and Yogo (2012) use the growth in total open interest, averaged over commodity sectors, which

[11] This might explain why the trend-following strategy is so popular in the CTAs industry; see Fung and Hsieh (1997).

[12] See also Groot *de*, Karstanje, and Zhou (2014).

represent the average absolute hedging demand. They find that commodity open interest forecasts commodity futures, stock, and bond returns, even after controlling for the usual conditioning variables including short-term interest rate, yield spread, dividend yield, and stock returns. In their commodity returns predictability test, they also include popular commodity forecasting variables (cross-hedging pressure, basis, and lagged returns). In opposition to the traditional view that commodity markets are segmented from the equity market, their results indicate that open interest of commodity futures is a leading indicator of economic activity, including commodity futures returns. Similarly, Marshall, Nguyen, and Visaltanachoti (2012, 2013) and Beckmann, Belke, and Czudaj (2014) notice that commodity market liquidity and global liquidity, respectively, drive commodity prices. Finally, Kang et al. (2021) uncovers a new "crowding" factor, built as the deviation of the net-long position of speculators from their one-year trailing average. This factor alone predicts the cross-section of commodity futures returns, and also helps to predict the returns of popular commodity factors such as momentum, value, and basis. They do not test whether it improves the performance of the basis-momentum factor.[13][14]

2.4. Hypotheses development

An important result that emerges from the literature is that the "factor zoo" that populates the equity market has contaminated the commodity market; see Cochrane (2011). While only 30 commodities are traded, corresponding to less than 10,000 monthly returns over the 1994–2020 period, the literature identifies at least six factors constructed on contract characteristics. Note that only five factors suffice to price reasonably the cross-section of more than 200,000 stock monthly returns from 1962 to 2014, and four factors perform well to price the cross section of 400,000 corporate bond monthly returns from 1999 to 2020; see Fama and French (2015), Kelly, Pruitt, and Su (2019), and Kelly, Palhares, and Pruitt (2020). Given the aforementioned results, I first check whether the factor dimensionality can be reduced :

Hypothesis 1. *There is a (parsimonious) subset of previously identified factors, which optimally adjust commodity futures weekly returns for risk.*

Beyond risk adjustments, commodity markets have experienced the emergence of a new class of investors (CITs) in 2000-2003, whose potential impact on returns,

[13] Since the earliest version of Boons and Porras Prado (2019) was circulating only in 2016, it is possible that this factor is not yet "crowded".

[14] These findings confirm previous results derived from the equity market. The literature documents that the less crowded the strategy is, the higher its returns are; see, *e.g.*, Baltas (2019).

turnover, and factor returns, deserve a special attention, I specify my hypothesis as follows:

Hypothesis 2. *The days of the roll of the SP-GSCI affect:*

a. *Commodity futures returns*

b. *Turnover*

c. *Factor returns*

Among all documented effects of the financialization developed in Section 2.2.3, the price-pressure generated by CITs around the roll could affect KRT findings. In particular, Brunetti and Reiffen (2014) document that CITs act as insurance suppliers. They modify the risk-sharing structure in the post-financialization, and potentially the prices of insurance. By the same token, liquidity needs are increased during the roll because CITs roll positions, which should affect the price of liquidity during this specific week. Therefore, I specify the last hypothesis:

Hypothesis 3. *The liquidity and insurance premiums are not altered after:*

a. *The risk adjustment*

b. *The financialization (roll days, measure of the CITs' activity, and the pre- and post-financialization periods).*

3. Data and methodology

3.1. Data

From Thomson-Reuters, I download the daily closing prices of 26 commodity futures nearby and first deferred contracts, among which, 18 are SP-GSCI components. The sample starts on January, 1994 and ends on January, 2020. I compute the Tuesday to Tuesday, weekly arithmetic excess returns for each maturity m as: $R^m_{c,t} = \frac{F^m_{c,t} - F^m_{c,t-1}}{F^m_{c,t-1}}$ where $F^m_{c,t}$ is the futures price of commodity c, on day t. To mitigate the effect of thinly traded contracts as they approach maturity, I follow KRT. I define the nearby contract as the actual nearby for the weeks ending with a Tuesday prior to the seventh calendar day of the month. I define the nearby contract as the first deferred if it matures in the current month and if the week ends on a Tuesday that is on or after the seventh day of the month. All contracts are under the CFTC supervision. This institution releases the commitment of traders weekly report (COT) from Tuesday to Tuesday. The COT includes the long and short positions of the "commercials" and the long, short, and spreading positions of the "non-commercials" categories for every U.S. futures contract. In 2006, the CFTC revised the compositions of these groups and added the "disaggregated" COT report (DCOT), that splits "commercials" and "non-commercials" trades into "Producer/Merchant/Processor/User", "Money managers", "Swap dealers", and "Other reportables".[15]. From the CFTC, the following three contract characteristics are computed, (i) the net trading of the "commercials" category (intensity with which commercial provide short-term liquidity), (ii) the average hedging pressure (demand for insurance), (iii) the "crowding" (intensity with which speculators are involved). The latter variable is used by Kang et al. (2021).

Next, from the commodity futures prices, I compute the following four characteristics, (i) the basis *i.e.*, (ii) the momentum, (iii) the β, and (iv) basis-momentum. I summarize the characteristics and their definition in the table below,

[15]See http://www.cftc.gov/MarketReports/CommitmentsofTraders/index.htm

Name	Formula	Variables	Meaning	Authors
Net trading	$Q_{c,t} = \frac{(CL_{c,t}-CS_{c,t})-(CL_{c,t-1}-CS_{c,t-1})}{OI_{c,t-1}}$	CL (CS) is the long (short) position of "commercials" and OI the total open interest.	The intensity at which the "commercials" category ease the short-term speculative activity of "non-commercials".	Kang et al. (2020, 2021); Dubois and Maréchal (2021)
Average hedging pressure	$AHP_{c,t} = \frac{\frac{1}{52}\sum_{j=0}^{51}(CS_{c,t-j}-CL_{c,t-j})}{OI_{c,t}}$	see Net trading	The net short position of "commercials" *i.e.* the intensity at which producers hedge their cash position, smoothed over the previous 52 weeks.	Cootner (1960); Gray (1961); Roon *de* et al. (2000); Chng (2009); Gorton et al. (2012); Basu and Miffre (2013); Kang et al. (2020)
Basis	$B_{c,t} = \frac{\ln F^2_{c,t}-\ln F^1_{c,t}}{T_2-T_1}$	T_m is the time to maturity of the contract of maturity m.	Proxies for the levels of inventory, as scarcity implies backwardation in the theory of storage.	Bailey and Chan (1993); Gorton et al. (2012); Yang (2013); Szymanowska et al. (2014); Koijen et al. (2018); Bakshi et al. (2019); Kang et al. (2021)
Momentum	$M_{c,t} = \prod_{j=0}^{51}\left(1+R^1_{c,t-j}\right)$		The past-year performance predicts future performance.	Jegadeesh and Titman (1993); Miffre and Rallis (2007); Moskowitz et al. (2012); Asness et al. (2013); Bakshi et al. (2019)
Basis-momentum	$BM_{c,t} = \prod_{j=0}^{51}\left(1+R^1_{c,t-j}\right) - \prod_{j=0}^{51}\left(1+R^2_{c,t-j}\right)$		Difference in two momentum signals of contracts of difference maturities, relates to the term-structure convexity.	Boons and Porras Prado (2019)
Average factor β	$\beta_{c,t} = \frac{Cov\left(R^1_{c,t-51\|t};AVG_{t-51\|t}\right)}{Var\left(AVG_{t-51\|t}\right)}$	AVG is the returns on an equally-weighted portfolio of the 26 commodities.	Relates to the CAPM principles.	Bakshi et al. (2019)
Crowding	$CR_{c,t} = \frac{NCL_{c,t}-NCS_{c,t}}{OI_{c,t}} - \frac{1}{52}\sum_{j=0}^{51}\frac{NCL_{c,t-j}-NCS_{c,t-j}}{OI_{c,t-j}}$	NCL (NCS) is the long (short) position of "non-commercials" and OI the total open interest.	The intensity at which speculators enter systematic strategies, thereby diminishing future returns.	Kang et al. (2021)
Open interest growth rate	$\Delta OI_{c,t} = \prod_{j=0}^{51}\left(\frac{OI^{USD}_{c,t-j}}{OI^{USD}_{c,t-j-1}}\right)^{1/52} - 1$	$OI^{USD}_{c,t} = OI_{c,t} \times F^1_{c,t} \times DQ_c$. $OI_{c,t}$ is the open interest in contracts, $F^1_{c,t}$ the price of the nearby contract and DQ_c, its deliverable quantity; see Appendix III.A1, column "underlying".	Considers commodity futures as predictors of the general economy, through the quantity of open interest (relates to hedging).	Hong and Yogo (2012)

3.2. Methodology

3.2.1. Factors construction and selection

To test hypothesis 1, I use the eight characteristics defined above, and compute the weekly returns of the corresponding portfolios. I use the characteristic observed on week t, and sort five quantile portfolios based on the characteristic. Each quintile contains five contracts, except for the median quintile which includes six. To select the optimal set of factors, I use the Bayesian approach of Barillas and Shanken (2018). This approach builds upon Gibbons, Ross, and Shanken (1989), and the "Bayes factor". The likelihood ratio of the unrestricted time series regressions (with intercept) with that of the restricted (without intercept) indicates the marginal contribution of the α.

$$ML_z = |X'X|^{N-2}|S_z|^{-\frac{T-K}{2}} \times H_z, \tag{1}$$

where X is the (sub)set of factors considered, N is the number of test assets, T the number of time observations and K the number of factors. S_z is the $N \times N$ cross-product of the time series regressions residuals for the unrestricted $z = 1$ and restricted $z = 0$ cases. In addition,

$$H_z = \begin{cases} 1 & \text{if } z = 0 \\ \left(1 + \frac{a}{a+k}\left(\frac{W}{T}\right)\right)^{-(T-K)/2}\left(1 + \frac{k}{a}\right)^{-N/2} & \text{if } z = 1 \end{cases}$$

where W is the Gibbons et al. (1989) F-statistic, $a = \frac{1+\text{Sharpe}_X^2}{T}$ and k is a prior multiple given the prior chosen for the maximum Sharpe ratio, such that $k = \frac{\text{Sharpe}_{max}^2 - \text{Sharpe}_X^2}{N}$, where Sharpe_X is the Sharpe ratio of the set of factors X and Sharpe_{max} the maximum prior Sharpe ratio, that I select in turn as 1.25, 1.5, 1.75, and 2; see Barillas and Shanken (2018, Proposition 1). Finally, Barillas and Shanken (2017) demonstrate that the optimal factor selection does not depend on additional test assets and that using factors out of the subset as dependent variables suffices to test whether the α depart from zero (relative test). In the alternative (absolute test), both remaining factors and test assets are used as dependent variables. Thus, I use relative and absolute tests. Since the cross-section of commodities is limited, the additional assets in the absolute tests are simply the individual commodity returns. Because the purpose of this study is to optimally adjust for the risk, I leave out the two factors built upon the liquidity (Q) and insurance (AHP) characteristics from the selection, and report the optimal factor selection from the full set in the Appendix III.A4.

3.2.2. Day of the roll effect

To test hypothesis 2, I capture the individual effects of each five SP-GSCI roll days on the 18 contracts that are concerned with dummy variables.The following panel data-regression is estimated for the 26 contracts simultaneously,

$$R^m_{c,t} = \mu_c + \sum_{i=1}^{5} \delta_i d_{c,t,i} + \epsilon_{c,t}, \tag{2}$$

where μ_c is a contract fixed effect, and $d_{c,t,i}$ is a dummy variable set to "1" when the contract c, in a week t, includes the i^{th} day of contract roll, and to "0" otherwise. I cluster the variance at the contract level.

To capture the effect of the roll-days on the factors individually, I regress each factors time series on five dummy variables set to "1" when the factor f in week t includes the i^{th} day of the SP-GSCI, *i.e.* roll.

$$R_{f,t} = \alpha_f + \sum_{i=1}^{5} \delta_i d_{t,i} + \epsilon_t, \tag{3}$$

where $R_{f,t}$ is the return on f, $f = 1, 2, \ldots, 8$. These regressions allows to control for systematic roll-day effects on the factors, with no assumptions on whether the individual contracts, experience a roll or not.

3.2.3. Fama-MacBeth regressions

To make the results easy to compare with KRT, I use Fama-MacBeth regressions to replicate and extend the tests on the forecasting power of the liquidity and insurance characteristics. The cross-sectional estimation on week t is,

$$R^1_{c,t} = \beta_{0,t} + \beta_{1,t} AHP_{c,t-1} + \beta_{2,t} Q_{c,t-1} + \beta_{3,t} CIT_{c,t-1} + {}^{\mathbf{T}}\mathbf{b_t} \mathbf{RISK}_{c,t-1} + \epsilon_{c,t}, \tag{4}$$

where $R^1_{c,t}$ is the returns on the nearby contract in week t, AHP and Q are the insurance and liquidity characteristics of the contract, and CIT is the pressure from index investment weighted by the number of roll-days of the contracts during the week,

$$CIT_{c,t} = \frac{OI_{CIT,c,t} \times d_{c,t}}{OI_{c,t}}, \tag{5}$$

where $OI_{c,t}$ is the total open interest, $OI_{CIT,c,t}$ is the SP-GSCI investment computed using the Masters and White (2008) procedure, and $d_{c,t} = 1/5, 2/5, \ldots 1$, when the contract c in week t includes $1, 2, \ldots, 5$ days. **RISK** includes the three controls of Kang et al. (2020), (i) the basis, (ii) the annualized standard deviations of the residuals from 52-weeks trailing-window regressions of the commodity returns on the S&P 500, multiplied by an indicator variable equals to "1" when "non-commercials" are net long and to "-1" otherwise, and (iii) the lagged returns $R^1_{c,t-1}$. Alternatively, **RISK** includes the factors selected as in Section 3.2.3.[16]

4. Empirical results

4.1. Descriptive statistics

4.1.1. Futures returns

Table III.1 reports the summary statistics of the 26 nearby futures contracts returns, over the 1994–2017 period (KRT), and the 1994–2003 and 2004–2017 subperiods (pre- and post-financialization).[17]

[Insert Table III.1 here]

Over the pre-financialization, annualized mean weekly returns range between -0.81% (livestock) and 15.98% (energy). In the post-financialization period, the mean returns of the nearby contract range between -9.62% (energy) and 11.41% (metal). Table III.1 also reports the skewness, which is positive for 20 out of the 26 contracts, for both nearby and first deferred positions, in line with the results of Gorton and Rouwenhorst (2006). These commodity characteristics underlie the popularity of commodity index investment; see Gorton and Rouwenhorst (2006); Erb and Harvey (2006). However, when the returns are averaged in an equally-weighted sector (or all commodities) portfolio, the skewness vanishes; a result similar to stocks, see, Albuquerque (2012). Commodity futures returns have excess kurtosis, and this at the individual, sector, and average market levels. Finally, the higher moments of commodity returns do not show any sizeable change between the pre- and the post-financialization period, and this at the individual and portfolio levels.

4.1.2. Commodity futures characteristics

Table III.2 reports the summary statistics of the eight commodity futures characteristics and the β detailed in Section 3.1 and for the variable CIT of Eq. (5)

[16] Estimations with CIT defined as the non-weighted ratio: $\frac{OI_{CIT,c,t}}{OI_{c,t}}$ yield similar results. They are available upon request.

[17] I report the statistics for the first deferred contracts in the Appendix III.A2.

for the whole KRT period (1994–2017, “All”), and the pre- (1994–2003, “Pre”) and post-financialization (2004–2017, “Post”) sub-periods.

[Insert Table III.2 here]

In Columns (1-3), the first characteristic AHP shows that commercial traders are structurally net short. Only two contracts, natural gas, and feeder cattle display a net long “commercials” position over the full sample. The metal (livestock) group reports the highest (lowest) hedging pressure, during both sub-periods. The equally-weighted hedging pressure increases by almost three percentage points in the post-financialization period, the largest increase being for the palladium contract (from 11.63 to 53.45%). Given that Q is a week-to-week difference, I report the mean absolute value in Columns (4-6). The highest (lowest) absolute net trading is present in the metal (energy) group over the full sample and the group ranking remains steady across sub-periods.

Columns (7-9) report β computed as the mean of 52-weeks trailing window regressions of each contract on an equally-weighted average factor AVG; see Bakshi et al. (2019). The energy (livestock) group displays the highest (lowest) average β, at 1.35 (0.80). No sizeable difference exist between the two sub-periods, except for the energy (decrease from 1.44 to 1.29) and metal groups (increase from 0.84 to 1.05). Overall, the betas of commodities are steady across periods, excepted for natural gas (from 1.67 to 1.16), silver (from 0.78 to 1.41), and lean hogs (0.84 to 0.29). Columns (10-12) report the mean basis, which (B) indicates how backwardated (negative), or contangoed (positive figure), was the nearby term structure over the period. The contract with the highest contango on average (natural gas, 7.05 bps) is also the one with the lowest annualized mean returns (-14.18%), whereas the most backwardated (soybean meal, -3.16 bps) delivers the highest returns (14.80%). This well-documented cross-sectional relation holds as shown further in the paper. Only five contracts are in backwardation on average, and none in the energy group. The basis of the equally-weighted portfolio increases from the pre- (0.85 bps) to the post-financialization period (1.55), a change mostly driven by the energy group (from 0.28 to 4.57 bps), confirming documented results about commodities becoming more backwardated along with the financialization; see *e.g.*, Adams and Gluck (2015). In columns (13-15), only six contracts display an average negative momentum (M) over the 1994–2017 period, with two sizeable changes in the sub-periods for the energy and metal groups, decreasing from 16.24 to -6.92% and increasing from 3.21 to 11.10%, respectively. The basis-momentum characteristic (BM, columns 16-18) is more homogeneous across contracts, groups and period. The largest period-to-period change occurs for the palladium contract (from -10.94 to 5.69%), while its annualized mean rises from 8.32 to 20.29%). Whereas the growth in open interest (ΔOI, columns 19-21) is steady across contracts and sub-periods, the crowding factor (CR, columns 22-24) increases from -0.44 to 0.07% on average across periods, the largest increase in the metal (from -3.30 to -0.08%), followed by the soft group (from -0.69 to 0.42%). This supports the view of an increase usage of factor investing in commodities; see Miffre (2016) and Kang et al. (2021).[18] Columns (25-27) report the statistics for the CIT, for the 18 contracts of the SP-GSCI. Over the full sample, the share of CITs is the highest for the feeder cattle (9.03%), wheat (8.39%), lean hogs (8.35%), and crude oil (8.16%) contracts. It is the lowest for the silver (0.78%) and gold (1.64%) contracts. These figures depart from the SP-GSCI average weights over the periods, *e.g.*, 26% for the crude oil and less than 0.4% for the feeder cattle. The average share of CITs is 2.53% in the pre-financialization period and 6.25% after. This confirms the sharp increase of CIT investment and the identified break date; see Dubois and Maréchal (2021). The energy group experiences the largest increase (*e.g.* the crude oil contract from 3.73% to 10.63%), and the CIT of all contracts increase over the period.[19]

4.2. Factor performance

Table III.3 reports the annualized weekly returns of the quantile (equally-weighted) portfolios constructed from the eight characteristics detailed in Section 3.1.

[Insert Table III.3 here]

Table III.3, Panel A reports the annualized weekly returns for the full sample (1994–2017). Five out of eight portfolio sorts (on Q, B, M, BM, and CR) report a monotonic increase of the returns across quintiles (annualized means weekly returns, and probability that week t delivers positive return). As previously documented by Bakshi et al. (2019), the long-short portfolio sorted on the rolling β alone delivers a performance that is not statistically different from zero. The next three columns report the performance of the long-short strategy (P5-P1), for the full period, the pre- and the post-financialization sub-periods. Factor Q delivers the highest mean weekly returns (17.99%), and is statistically different from 0 (t-statistic = 4.80), followed by CR (17.85%, and t-statistic = 3.68) and B (15.75%, and t-statistics = 3.84). Note that the average annualized weekly returns of the Q, B, M, BM,

[18] In commodities factor investing is coined as “third-generation” index investment.

[19] Note that in the post-financialization periods, there is no value for the platinum and orange juice contracts since the SP-GSCI delisted these contracts in 2003–2004.

and CR factors are statistically different from 0 at the 1% level. However, the momentum and crowding factors display noticeable changes across sub-periods. The crowding (momentum) strategy is not significant at the usual levels in the first (second) period, and significant at the 1% level in the second (first) period. This supports the view that factor strategies have been developed along with the financialization (in the form of "third-generation index investing"), the momentum strategy being the most popular; see Miffre and Rallis (2007) and Kang et al. (2021). This increased popularity may have altered the performance due to "crowding", an effect also documented on the stock markets; see, *e.g.*, McLean and Pontiff (2016). Therefore, the parallel surge of the crowding factor performance reinforces Kang et al. (2021) view.

Panel B reports the correlation of the factors over the full-sample period. The highest correlation is reported for ΔOI and M factors.[20] This correlation, however, does not relate to the factor ability, given the poor performance of ΔOI. In contrast, and as documented by Boons and Porras Prado (2019), the correlation of BM with M and B remains below 30%, despite the fact it encompasses them. Overall, the lowest correlation coefficients are between factors constructed from market-level data (B, M, BM, and β) with those constructed from CFTC or trader-level data (AHP, Q, CR). Finally, the CR factor is negatively correlated with four out of seven strategies, an additional indication of its ability to predict the (poor) performance of other factors; see Kang et al. (2021).

4.3. Optimal subset of factors

Table III.4 reports the results of the Bayesian asset-pricing test detailed in Eq. (1); see Barillas and Shanken (2018). I use a set of six factors excluding those constructed on the two characteristics of interest (AHP and Q) out of the eight long-short factors presented in Table III.3.[21] There are $2^n - 1$ arrangements, that is 63 (255) in the six (eight) factors scenario. I report the best results for each of the subset size (from one to six) for each prior on the maximum attainable Sharpe ratio $1.25, 1.50, 1.75, 2.00$.

[Insert Table III.4 here]

The left columns ("Relative") report the statistics for the relative test in which only factors are included in the set of dependent variables of the time series regressions. The right columns ("Absolute"), reports the statistic for which both remaining

[20]Both factors are constructed using some forms of price trends.

[21]I report the full comparison of the eight factors in the Appendix III.A4.

factors and the 26 commodity returns are in the dependent variables. The selection is robust to the choice of the test (relative or absolute) and to the prior chosen. In the six potential subsets, the basis factor is selected, and each restricted set of factors of size n is nested in the $n+1$ selection. This approach also allows to compare the probability of obtaining the optimal set of factors, across the subset size. In both relative and absolute specifications, and for all priors, the highest probability is reached with a four factor model (B-M-BM-CR), with 61% and 36% in the relative and absolute test, respectively. Moreover, this four-factor model also yields the lowest (second to lowest) GRS statistic in the absolute (relative) specification, at 1.41 (0.34). Given the fact that the lowest GRS statistic is only obtained with the inclusion of an additional factor in the relative specification, it further supports this choice from the "frequentist" inference perspective; see Gibbons et al. (1989) and Barillas and Shanken (2017). Thus, I select these factors to optimally adjust commodity returns for risk.

4.4. Effects of the day of the roll

Table III.5 reports the results of the panel and time series regressions of Eq. (2) and Eq. (3). Panel A reports the results of the panel regression that includes the 26 contracts, and for the 18 contracts, components of the SP-GSCI. I also present the results for the first deferred contracts since they might be affected differently during the roll-days.

[Insert Table III.5 here]

Over the full period, no roll-day coefficient is significant at the 10% level. Moreover, no systematic pattern emerges in any set of contracts. For instance, the coefficient for the first roll-day is 16.28 bps, when that of the second is -10.31 bps. Moreover, the period-to-period comparison is inconclusive. Some coefficients reach the traditional significance levels in the post-financialization period, but they are of opposite signs across days. These results confirm the view that no systematic effect on the prices can be attributed to the roll of the major index, thus supporting the "Sunshine trading" view of Admati and Pfleiderer (1991). In addition, the low (negative) adjusted R^2 regressions constitutes further evidence of the little predictability of the roll on returns. Hence I cannot reject the hypothesis 2a.

Table III.5, Panel B, presents the results concerning the turnover of the contracts. I define turnover as as $\frac{VOL^m_{c,t}}{OI_{c,t}}$ where VOL is the trading volume of the contract of maturity $m = 1; 2$, c and t the contract and week, respectively, and OI the total open interest. These results show a different picture. In the first specification (26 nearby contracts), the pre-financialization period does not show any roll-day coefficient statistically significant at the 10% level or less. During the post-financialization period, however, all coefficients are positive, four of which are significant at the 1% level. Zooming into the 18 SP-GSCI contracts confirms this pattern. In addition, the adjusted R^2 is 5.08%. These results point to an effect on the market functioning and risk-sharing structure, with CITs trading significantly more than traditional traders during the roll. The coefficients for the first deferred contracts of the SP-GSCI also increase across periods, but less. Only one coefficient is significant at the 1% level for the third day (10% for the first) and one is negative (fifth day). This post-period difference may be explained by the concurrent entry of second-generation index funds, which are optimizing the "roll yield", investing in contracts of further-deferred maturities.

Finally, Panel C reports the results of time series regressions of the eight factors (see Table III.3) onto SP-GSCI roll-day dummies (one roll each month). The factors computed on the two characteristics of interest (AHP and Q) are not systematically affected by the roll, except for the first day that shows statistical significance at the 1% level. In addition, there is no sizeable change across periods, which reinforces the view that the financialization has not affected these factors. However, except for the β factor, the first day of the roll is the single day which is systematically positive across strategies and periods. It is statistically significant at the 5% level for the factors B, M, BM, and CR. Given its construction, and if speculators use strategies that front-run the index roll, the crowding factor should be affected. Instead, I do not find any coefficient statistically significant at the 5% level, and their sign varies across roll-days. Overall, I reject the hypothesis 2b, when the turnover is considered. There is no evidence for an impact of the roll days on the returns and on the factors, and I do not reject hypotheses 2a and 2c.

4.5. Performance of the liquidity and insurance premiums

Table III.6 reports the results of the Fama-MacBeth regression Eq. (4). I first replicate KRT (see Table VI, 2^{nd} and 3^{rd} columns) results in columns (1) and (2).[22]

[Insert Table III.6 here]

In column (3), I report the results with the optimal set of factors obtained in Section 4.3 (B-M-BM-CR). In this specification, the size of the coefficients of AHP and Q

[22] I obtain very close results for the coefficients of AHP and Q, using the same Newey-West adjustment with four lags, albeit my t-statistics are slightly smaller. I cannot compare the coefficients for risk-adjustment since they do not report them, however, the (unadjusted) R^2 are also identical.

decreases slightly and the coefficient of AHP is statistically significant at the 10% level, while Q is robust to this risk adjustment and remains statistically significant at the 1% level. Thus, I do not reject hypothesis 3a. The risk adjustment increases the average (adjusted) R^2 by about five (two) percentage points. Column (4) restricts the estimations for weeks with at least three roll-days. In this case, the magnitude of the AHP coefficient is unchanged, and its statistical significance reaches the 5% level. In contrast, in this specification, the size of the Q coefficient is halved and is only significant at the 10% level. This indicates that the roll affects the predictability of Q, but not through the CIT inclusion. Column (5) reports the results for the 18 SP-GSCI contracts, including the variable CIT. The AHP coefficient decreases economically and is not statistically significant at the usual levels, whereas the Q coefficient is robust to this specification. Column (6) also includes CIT for the 26 commodity contracts (the CIT value is zero for the eight non-SP-GSCI contracts). This aims to disentangle the CIT from the cross-sectional dimension effect. In this specification, the AHP coefficient is statistically significant at the 5% level. Note that the coefficient of CIT is not statistically significant at usual levels. Columns (7) and (8) split the sample pre- and post-financialization. Whereas the average explanatory power of both estimations is unchanged across periods, the size of the coefficients of AHP and Q increases twofold post-financialization, and the coefficient of AHP is significant (at the 10% level) post-financialization only. However, a t-test that compares the series of coefficients before and after does not reject equality neither for AHP, nor for Q, at the 10% level. Finally, Column (9) extends the sample until December 2020, for which the results hold. Altogether, the results of KRT are valid under this extension, although the statistical significance, and the magnitude of the coefficients, decreases under this optimal risk-adjustment.

Altogether, these results point to a modification of the risk-sharing structure in commodity markets, with major changes related to the insurance (AHP), but not to the liquidity component (Q). Hence, hypothesis 3b is rejected. In line with previous results, the financialization has eased the activity of the "commercials" category, and decreased the magnitude of the insurance premium. There is no evidence that the financialization (through the roll days, CIT pressure, or the change in periods) has affected the liquidity premium paid by speculators. However, the Fama-MacBeth approach has limited power in the context of a small cross-section (26 contracts). The next section intends to overcome these limitations.

5. Robustness tests

5.1. Econometric robustness

I use the "Generalized Portfolio Sorts" (GPS) method, which controls for the correlation between contract characteristics and factors built upon other characteristics; see Hoechle et al. (2020). If characteristics are observable, the inclusion of the interaction terms, and their time series average by contract, controls for the dependence. If they are unobservable, contract fixed effects are introduced to avoid the bias. In particular, a Haussman test (Haussman, 1978) is achieved on either type of fixed effects (FE). The panel estimation is,

$$R^1_{c,t} = \mu_c + {}^{\mathbf{T}}\big({}^{\mathbf{T}}\mathbf{z}_{c,t} \odot {}^{\mathbf{T}}\mathbf{x}_t\big)\mathbf{\Theta} + \epsilon_{c,t}, \tag{6}$$

where $\mathbf{z}_{c,t} = [1, z_{2,c,t}, \ldots, z_{M,c,t}]$ is a vector of constant and contract characteristics $m = 2, \ldots, M$, and $\mathbf{x}_{c,t} = [1, x_{2,t}, \ldots, x_{K,t}]$ is a vector of constant and factors, $\odot$ denotes the Khatri-Rao product. When transposed, it yields the column-wise multiplication of the matrices of characteristics and factors, and hence the fully integrated model.[23] Thus, the dependent variables are (i) a constant, (ii), the characteristics, (iii) the factors, and (iv) the factor-characteristic interactions. I introduce characteristics, and their factor-interactions FE in the specifications that adjust for risk with the optimal set of factors, and apply the corresponding Haussman test, using standard errors based on the Driscoll-Kraay (Driscoll and Kraay, 1998) covariance matrix. To introduce characteristics/interactions FE, I compute the time series average of each characteristics and interactions, pooled over the time dimension; see Hoechle et al. (2020, Eq. 18). I report the results of the estimations in Table III.7, using the FE specifications when the Haussman test rejects the RE assumptions at the 5% level.

[Insert Table III.7 here]

Panel A, reports the results for the full sample without (with) risk adjustment column (1)(column (2)). The coefficient of AHP is equal to 0.33% (*vs.* 0.47% in the Fama-MacBeth regressions), and statistically significant at the 5% level. Q is equal to 2.78/2.80 and statistically significant at the 1% level. However, the coefficient is smaller than that of the Fama-MacBeth regression (3.80).

[23] The original notation of Hoechle et al. (2020) uses a Kronecker product, which provides the correct dimension only if the characteristics are constant over time (in a vector). This notation generalizes it in the case of time-varying characteristics; For transpose of Kathri-Rao products and "face-splitting" products see also Slyusar (1999).

In the two specifications, the Haussman test does not reject the RE assumption for the characteristics and factor-interactions. Again, the CIT coefficient is not statistically significant at the 10% level. Moreover, none of the factors show statistical significance at the usual levels. Columns (3-4) and (5-6) report the results for the pre- and post-financialization periods. In contradiction with the results of Table III.6, the size of the AHP coefficient is equal to 0.50%, and significant at the 5% level in the first period, but of 0.25% and not significant at the 10% level in the second period. Instead, the coefficient of Q increases from 3.19% to 4.79% across periods, and is significant at 1%. III.6. This further supports the view that the financialization has modified the risk-sharing structure from the "commercials" (producer) perspective and decreased, in particular, the price of insurance. For instance when computing the average annualized premiums, I find that of insurance decreases by 43 bps, from 1.11 to 0.68% and that of liquidity increases by 2.92 percentage points. A one standard deviation change in AHP during the pre- (post-) financialization translates into a 8.33 bps (5.12) bps change in weekly returns. For Q, a one standard deviation change implies a 15.91 and 13.05 bps change in returns for the pre- and post-financialization, respectively. Finally, the RE assumption for the characteristics and factor-interactions is rejected at the 5% level, only in the post-period and for the risk-adjusted setting.

5.2. Economic robustness

Since mid-2006, the CFTC adds the disaggregated commitment of traders report (DCOT), in which the categories of the "legacy" COT are refined. The commercials category is split in the "producers" and "swap dealers" category. Swap dealers' commercial activity consists in hedging CIT positions on behalf of index traders. Thus, DCOT allows to obtain a more timely and accurate measure of the CIT pressure. It also improves the measurement of the "commercials", and should lead to less noisy variables AHP and Q. Table III.7, Panel B, reports the results for the 2007–2017 period in column (1), *i.e.*, the overlap between the KRT sample and the DCOT, and for the 2007–2020 period in column (2). I construct Q^{PROD}, AHP^{PROD}, and CIT^{SWAP} with the "producers" and "swap dealers" DCOT positions, with the same methodology as above.[24] I report the results of the restricted estimations (no risk-adjustment) with characteristics FE as implied by the rejection of the RE assumption at the 1% level. The results are very similar to those of the COT in the post-financialization period. The coefficient of AHP^{PROD} is not

[24] I also compute CIT^{SWAP} with no weight on the roll-days overlap and I find no difference in the results. These results are available upon request.

significant at the 10% level and turns negative (about -0.35%), thereby confirming the results of the post-financialization period of Panel A. In contrast, the coefficient of Q^{PROD} remains statistically significant at the 1% level, and is of similar size as the COT/post-financialization setting. Again, the coefficient of CIT^{SWAP} does not show any economic and statistical significance at the usual levels. From an investor perspective, the robustness of Q, in the light of these econometric, risk, and economic adjustments also confirms the validity of the net trading or liquidity demand characteristic up to today. On the other hand, a Chow test does not reject the null hypothesis of stable coefficients over the pre- and post-financialization periods (Panel A) (in particular for AHP). When the coefficients of the post-financialization period are replaced by those of the 2007–2020 period in the DCOT setting, the test does reject the null for AHP. Since the covariance matrices are built on different sets of variables, this is not a statistical test. However, it points to a drastic reduction of the insurance premium paid by hedgers.

6. Conclusion

This paper studies the robustness of the risk premiums uncovered by Kang et al. (2020), in the light of the financialization. From the extant literature, I first select the characteristics that are identified as good predictors of commodity futures returns. I build factors from these characteristics, and using the Bayesian factor selection approach, I find an optimal asset-pricing model consisting of four factors: (i) the basis, (ii) basis-momentum, (iii) momentum, and (iii) the "crowding". This model stands out all other factor combinations, both in terms of alpha and probability. Beyond its usage in this research, this selection already contributes to the literature, reducing slightly the dimension of a rising "factor zoo". Next, I find that the roll-days of popular indices are significant predictors of the turnover, and that this predictability further increases in the post-financialization period. In contrast, I do not find any evidence that the roll affects returns on commodities or factors, in the pre- and post-financialization periods, and the whole sample period. This contributes to the view that the financialization has not influenced prices, but plays a role in the reorganization of futures markets as an insurance market. To find support for this hypothesis, I use the aforementioned uncovered four-factor model to optimally adjust the specification of Kang et al. (2020) for risk, and provide the results of additional tests aiming to disentangle the financialization effects on this market organization. First, I find that Kang et al. (2020) findings are robust to risk-adjustment. Second, I do not find any influence for a direct measure of index funds pressure. Third, when I select only the weeks overlapping the roll, the results are not affected. Fourth, the insurance premium varies across periods, but not that of liquidity, which supports the view of a market reorganization, in which the index traders supply insurance to producers, thereby decreasing its price; see also Brunetti and Reiffen (2014). Conversely, this market reorganization does not affect the premium captured by producers in providing liquidity to speculators. Given these results, the financialization would benefit the producers both ways. Finally, I implement the panel data approach of Hoechle et al. (2020) and include disaggregated data from the CFTC to obtain robustness inferences from both econometric and economic perspectives. I confirm the aforementioned results. The poor performance of the risk-adjustment with common, optimal factors, also confirm the strong heterogeneity of commodities, and possibly the necessity to identify a more restricted set of more efficient factors. A second line of research would be to consider monthly returns. Finally, a possible research extensions includes the estimation of a latent risk factors and time-varying loading with Instrumented Principal Component Analysis; see Kelly et al. (2019), and Kelly et al. (2020). I leave these methodologies for future research.

Table III.1: **Descriptive statistics: Futures returns**

This table reports the annualized means and standard deviations, and the skewness and kurtosis of the weekly returns on the 26 nearby commodity futures contracts. The period of interest is 1994–2017, and two sub-periods before (1994–2003) and after (2004–2010) the financialization. The table reports the statistics for the equally-weighted portfolios for each group (Energy, Metal, Agriculture, Soft and Livestock) and for the 26 contracts equally-weighted portfolio. The tickers are defined in the Appendix Table A.1.

	1994–2017				1994–2003				2004–2017			
Ticker	Mean (%)	S.d. (%)	Skewness	Kurtosis	Mean (%)	S.d. (%)	Skewness	Kurtosis	Mean (%)	S.d. (%)	Skewness	Kurtosis
CL	8.64	33.67	-0.13	4.60	22.56	32.41	-0.34	4.32	-1.31	34.49	0.02	4.79
HO	8.72	31.69	0.12	4.32	14.89	31.08	-0.13	3.50	4.30	32.13	0.29	4.86
NG	-14.18	45.91	0.24	3.93	10.50	50.57	0.17	3.77	-31.84	42.13	0.25	3.87
PL	6.00	24.81	2.12	6.80	14.64	28.59	3.09	83.77	-1.07	21.20	-0.13	4.54
PA	15.06	36.84	-2.67	4.35	8.32	43.67	-3.55	48.98	20.29	30.52	-0.21	3.88
SI	6.51	28.84	0.00	6.26	-0.14	20.46	0.53	4.96	11.23	33.54	-0.12	5.39
HG	10.34	24.91	0.03	5.53	4.66	21.12	0.06	3.61	14.37	27.28	-0.01	5.65
GC	4.02	16.52	0.48	9.27	-1.98	13.34	2.04	21.27	8.30	18.45	0.00	6.18
W	-8.29	29.16	0.45	4.48	-9.17	23.72	0.46	4.34	-7.66	32.52	0.42	4.09
KW	-0.76	27.47	0.38	4.42	1.18	23.35	0.47	5.52	-2.14	30.08	0.35	3.85
MWE	6.02	26.05	0.31	6.82	2.75	21.28	0.72	6.17	8.36	29.01	0.17	6.35
C	-3.15	26.86	0.32	5.60	-7.27	21.80	0.30	4.68	-0.21	29.96	0.30	5.26
O	7.32	34.15	0.54	7.80	5.46	30.40	0.21	4.37	8.62	36.57	0.67	8.63
S	7.45	23.50	-0.01	4.15	3.54	19.95	0.14	4.02	10.23	25.74	-0.07	3.92
BO	-0.64	23.57	0.18	4.06	-1.86	20.80	0.21	3.77	0.23	25.38	0.16	3.97
SM	14.80	27.35	0.23	4.36	12.17	22.92	0.41	4.14	16.68	30.12	0.16	4.09
RR	-7.36	25.47	-0.10	5.21	-12.98	26.92	-0.22	6.65	-3.35	24.38	0.04	3.55
CT	0.53	28.53	0.06	5.02	-3.44	25.98	0.26	4.21	3.37	30.23	-0.05	5.21
OJ	5.26	32.76	0.36	5.18	-5.96	27.24	0.22	5.37	13.20	36.13	0.35	4.72
LB	-8.58	31.12	0.36	3.59	-9.91	31.45	0.13	2.61	-7.64	30.90	0.52	4.34
CC	3.96	29.50	0.68	6.58	2.51	30.43	0.97	7.65	5.00	28.83	0.43	5.61
SB	4.67	30.93	0.05	4.16	7.66	28.96	0.23	3.88	2.53	32.28	-0.04	4.21
KC	3.81	37.45	0.67	7.12	6.69	44.06	0.83	7.44	1.75	31.91	0.26	3.79
LH	-0.69	28.85	0.18	6.72	-4.59	28.72	0.33	7.78	2.06	28.96	0.07	6.02
LC	2.77	15.81	-0.13	5.81	2.70	15.62	-0.42	8.57	2.82	15.95	0.06	3.98
FC	2.70	15.26	-0.39	6.38	-0.39	13.53	-0.90	11.74	4.91	16.39	-0.20	4.31
Energy	1.06	29.81	0.01	3.80	15.98	30.24	-0.09	3.58	-9.62	29.44	0.08	4.02
Metal	8.86	19.31	-0.17	5.97	5.29	15.37	-0.06	9.43	11.41	21.70	-0.22	4.75
Agriculture	1.67	19.02	0.26	4.59	-0.80	15.82	0.45	4.35	3.43	21.02	0.18	4.29
Soft	1.58	16.03	0.04	3.60	-0.45	14.70	0.31	3.32	3.03	16.92	-0.10	3.64
Livestock	1.63	15.11	-0.05	4.71	-0.81	14.95	-0.04	6.18	3.37	15.24	-0.05	3.74
Average	2.71	12.56	-0.06	5.01	2.33	9.64	0.12	3.04	2.98	14.29	-0.10	4.68

Table III.2: **Descriptive statistics: Beta and Futures characteristics**

This table reports the mean and standard deviations of nine commodity characteristics. AHP is the 52 weeks rolling average of the hedging pressure computed as the net-short position of "commercials" traders scaled over total open interest. Q is the liquidity measure of Kang et al. (2020) and is the change in "commercials" net long position in week t from week $t-1$. The beta is obtained from 52 weeks rolling regressions of the nearby returns with the equally-weighted average returns of the 26 commodities. B, the basis is the difference of the first deferred and nearby log prices, scaled by their difference in time to maturity. M is the momentum, geometric sum of the 52 weeks' previous nearby returns. BM, basis-momentum is the difference in the momentum of the first deferred and nearby returns (see Boons and Porras Prado, 2019). CR is the "crowding" factor of Kang et al. (2021), defined as the difference between the net long position of the "non-commercials" with its 52-weeks average. OI is the geometric mean of the growth rate in the open-interest (in USD); see Hong and Yogo (2012). Finally, CIT is the index pressure during the roll, weighted by the number of roll-days overlapping the week; see Eq. (5). The period of interest is 1994–2017. The tickers are defined in the Appendix Table III.A1.

	AHP (%)			$\lvert Q\rvert$(%)			β			*B* (bps)			*M* (%)			*BM* (%)			$\lvert\Delta OI\rvert$ (%)			*CR* (%)			*CIT* (%)		
Ticker	All (1)	Pre (2)	Post (3)	All (4)	Pre (5)	Post (6)	All (7)	Pre (8)	Post (9)	All (10)	Pre (11)	Post (12)	All (13)	Pre (14)	Post (15)	All (16)	Pre (17)	Post (18)	All (19)	Pre (20)	Post (21)	All (22)	Pre (23)	Post (24)	All (25)	Pre (26)	Post (27)
CL	6.25	2.01	9.28	1.72	2.68	1.03	1.37	1.32	1.40	0.73	-2.62	3.13	9.45	22.51	0.10	-2.02	0.65	-3.93	0.64	0.47	0.75	0.48	0.27	0.63	8.16	3.73	10.63
HO	9.15	12.21	6.96	2.49	3.08	2.06	1.31	1.32	1.30	0.60	-0.97	1.73	8.66	13.91	4.90	-0.51	0.82	-1.47	0.51	0.41	0.59	0.20	0.35	0.09	6.57	3.48	8.29
NG	-0.65	7.75	-6.67	1.63	2.50	1.01	1.37	1.67	1.16	7.05	4.42	8.95	-9.88	12.29	-25.75	-8.02	-9.48	-6.99	0.72	0.68	0.74	-0.12	-0.39	0.07	3.05	1.74	3.78
PL	49.49	41.72	55.05	5.93	7.28	4.96	0.96	0.96	0.96	0.05	-0.21	0.27	4.50	10.56	0.16	-1.75	0.82	-3.58	0.81	1.00	0.68	-2.55	-5.34	-0.85	1.67	1.67	
PA	36.01	11.63	53.45	4.43	4.90	4.09	1.24	1.36	1.16	0.64	1.05	0.32	11.20	7.36	13.95	-1.25	-10.94	5.69	0.81	0.62	0.94	0.74	0.04	1.25			
SI	39.71	42.04	38.03	3.67	4.82	2.85	1.14	0.78	1.41	0.45	0.66	0.30	6.69	-1.34	12.43	0.48	0.84	0.23	1.03	1.60	0.69	-1.89	-4.99	-0.24	0.78	0.25	1.07
HG	7.76	17.09	1.08	3.99	5.69	2.78	0.85	0.65	1.00	-0.76	-0.59	-0.89	12.59	2.19	20.03	0.39	-1.03	1.41	1.08	1.74	0.67	-0.32	0.23	-0.61			
GC	23.80	2.49	39.04	4.93	6.77	3.61	0.59	0.43	0.71	0.40	0.61	0.25	4.06	-2.72	8.92	0.32	0.39	0.28	0.80	1.17	0.54	0.70	1.76	0.08	1.64	0.85	2.09
W	1.34	13.01	-7.02	3.03	4.28	2.13	1.46	1.46	1.46	3.25	2.59	3.73	-7.04	-7.20	-6.93	-2.54	-3.55	-1.82	0.62	0.50	0.70	-0.68	-0.29	-0.95	8.39	4.96	10.26
KW	8.51	5.80	10.44	2.82	3.35	2.44	1.35	1.41	1.31	1.53	0.48	2.28	2.02	4.30	0.39	1.40	3.46	-0.06	0.60	0.43	0.72	-0.22	-0.47	-0.04	6.35	3.84	7.19
MWE	8.49	3.86	11.80	2.83	3.44	2.40	1.16	1.22	1.11	0.20	0.39	0.07	9.39	6.01	11.81	3.77	3.63	3.88	0.51	0.11	0.76	0.56	4.18	0.04			
C	2.91	-0.51	5.36	2.37	3.16	1.81	1.39	1.38	1.39	2.49	2.41	2.55	-1.92	-4.89	0.21	0.60	0.42	0.74	0.68	0.62	0.72	-1.15	-2.43	-0.27	3.93	1.82	5.08
O	31.63	38.70	26.58	4.02	4.16	3.91	1.46	1.54	1.40	0.63	0.01	1.05	8.49	8.40	8.55	5.89	6.74	5.27	0.66	0.71	0.63	-0.62	-2.30	0.45			
S	10.75	14.10	8.35	2.78	3.28	2.42	1.31	1.28	1.33	-1.16	-0.74	-1.45	7.85	1.92	12.09	1.17	0.18	1.87	0.63	0.57	0.68	0.97	2.98	-0.35	2.35	0.92	3.15
BO	12.99	14.48	11.92	3.89	4.85	3.20	1.12	0.99	1.21	1.24	1.14	1.31	0.65	-1.60	2.25	0.77	1.66	0.14	0.56	0.47	0.62	-0.85	-1.72	-0.22			
SM	19.00	17.98	19.73	3.46	4.00	3.08	1.30	1.29	1.31	-3.16	-3.10	-3.20	15.26	10.04	19.00	3.58	3.40	3.71	0.54	0.47	0.59	-0.08	0.19	-0.27			
RR	9.73	3.57	14.15	3.75	3.90	3.64	0.73	0.85	0.64	2.88	3.06	2.76	-4.62	-7.77	-2.36	-1.42	-1.98	-1.01	0.93	1.07	0.84	0.41	1.71	-0.43			
CT	8.97	1.75	14.14	4.42	5.62	3.56	0.79	0.68	0.87	1.63	2.21	1.21	1.21	-3.77	4.78	-0.51	-3.60	1.70	0.75	0.70	0.78	0.71	1.65	0.04	4.75	2.06	6.25
OJ	26.48	17.60	32.83	4.85	5.35	4.49	0.77	0.78	0.77	1.51	2.79	0.61	4.45	-4.94	11.17	5.56	6.36	4.99	0.93	0.47	1.26	-0.05	-1.39	0.98	4.67	4.67	
LB	10.15	12.86	8.21	4.59	4.93	4.35	0.53	0.55	0.51	3.47	2.58	4.12	-8.02	-7.00	-8.76	-2.95	-5.35	-1.24	0.67	0.71	0.64	0.01	-0.89	0.66			
CC	13.70	9.55	16.67	2.77	3.21	2.46	0.74	0.83	0.67	1.18	2.03	0.57	3.94	4.50	3.53	1.71	1.46	1.89	0.45	0.50	0.42	-0.53	-1.45	0.12	1.94	0.52	2.71
SB	16.40	16.86	16.07	3.64	5.49	2.31	0.89	0.83	0.93	-0.59	-0.87	-0.39	5.34	9.69	2.23	-1.13	3.81	-4.66	0.65	0.65	0.65	-0.21	-0.72	0.16	4.00	2.25	4.94
KC	13.44	15.78	11.77	3.80	5.29	2.74	1.36	1.45	1.30	1.91	0.82	2.69	6.60	13.27	1.83	1.63	1.46	1.76	0.68	0.69	0.68	-0.50	-1.29	0.07	2.67	1.21	3.46
LH	2.54	1.85	3.04	2.54	3.43	1.89	0.52	0.84	0.29	5.62	4.82	6.19	-2.17	-4.14	-0.75	-0.12	-4.29	2.87	0.69	0.68	0.70	0.28	-0.02	0.49	8.35	5.82	9.77
LC	5.11	4.11	5.83	1.79	2.14	1.55	0.30	0.29	0.30	0.39	-0.34	0.93	2.45	0.75	3.66	1.31	-0.71	2.76	0.43	0.38	0.46	0.64	1.07	0.33	6.80	3.50	8.65
FC	-7.65	-5.77	-8.99	2.13	2.63	1.76	0.09	-0.02	0.17	-0.16	-0.55	0.13	2.66	-0.77	5.11	0.90	0.91	0.90	0.59	0.58	0.60	0.47	0.51	0.44	9.03	6.02	9.47
Energy	4.91	7.32	3.19	1.94	2.75	1.36	1.35	1.44	1.29	2.78	0.28	4.57	2.74	16.24	-6.92	-3.52	-2.67	-4.13	0.62	0.52	0.70	0.19	0.08	0.26	5.93	2.98	7.57
Metal	31.35	22.99	37.33	4.59	5.89	3.66	0.96	0.84	1.05	0.13	0.26	0.04	7.81	3.21	11.10	-0.36	-1.99	0.81	0.92	1.21	0.70	-1.43	-3.30	-0.08	1.58	1.09	1.85
Agriculture	11.70	12.33	11.26	3.22	3.82	2.78	1.25	1.27	1.24	0.88	0.69	1.01	3.34	1.02	5.00	1.47	1.55	1.41	0.63	0.54	0.69	-0.24	-0.26	-0.23	4.16	2.09	5.34
Soft	14.86	12.40	16.62	4.01	4.98	3.32	0.85	0.85	0.84	1.52	1.60	1.47	2.25	1.96	2.46	0.72	0.69	0.74	0.69	0.62	0.74	-0.04	-0.69	0.42	3.70	2.16	4.61
Livestock	0.00	0.06	-0.04	2.15	2.74	1.73	0.30	0.37	0.25	1.95	1.29	2.42	0.98	-1.39	2.67	0.70	-1.36	2.18	0.57	0.55	0.59	0.46	0.52	0.42	8.10	4.87	9.41
Average	14.08	12.40	15.27	3.39	4.24	2.79	1.00	1.01	1.00	1.26	0.85	1.55	3.61	3.14	3.95	0.28	-0.15	0.59	0.69	0.68	0.69	-0.14	-0.44	0.07	4.92	2.53	6.25

Table III.3: **Factor returns**

This table reports the annualized weekly portfolio and factor returns built upon the eight commodity characteristics of Table III.2. Panel A reports the returns on portfolios sorted as follows: 5–5–6–5–5, and the performance of the strategy that is long (short) in the top (bottom) quantile. The table reports the annualized means, standard deviations, and Sharpe ratios, the percentage of weeks during which the portfolio returns were positive and the t-statistic of the mean of the long-short strategy. Panel B reports the correlation between the long-short strategies. The period of interest is 1994–2017 and the pre- (1994–2003) and post-financialization (2004–2017) sub-periods.

Panel A: Factor performance

	1994–2017						Pre	Post	1994–2017						Pre	Post
	AHP								*Q*							
	P1 (low)	P2	P3	P4	P5 (high)	P5-P1	P5-P1	P5-P1	P1 (low)	P2	P3	P4	P5 (high)	P5-P1	P5-P1	P5-P1
Ann. mean (%)	-0.33	-4.67	3.57	6.44	9.66	9.99	6.78	10.45	-7.05	-0.70	2.19	8.23	10.94	17.99	21.37	17.36
Ann. σ (%)	14.92	17.19	17.29	16.89	18.20	19.27	17.02	20.86	17.23	16.39	16.53	16.86	16.03	18.31	18.42	18.35
Ann. Sharpe ratio	-0.02	-0.27	0.21	0.38	0.53	0.52	0.40	0.50	-0.41	-0.04	0.13	0.49	0.68	0.98	1.16	0.95
$P(>0)$	48.03	46.41	50.52	51.41	52.05	53.18	50.68	54.21	46.98	50.52	49.48	52.30	55.12	56.16	56.09	56.15
t-statistic						2.53	1.26	2.06						4.80	3.66	3.89
	β								*B*							
	P1 (low)	P2	P3	P4	P5 (high)	P5-P1	P5-P1	P5-P1	P1 (high)	P2	P3	P4	P5 (low)	P5-P1	P5-P1	P5-P1
Ann. mean (%)	-1.30	5.13	7.69	3.86	-1.88	-0.58	1.88	-2.57	-4.21	-2.75	2.39	7.72	11.54	15.75	17.77	15.13
Ann. σ (%)	11.97	14.97	17.03	19.43	22.32	23.02	22.53	23.75	17.26	17.58	16.17	16.78	17.75	20.05	20.02	21.17
Ann. Sharpe ratio	-0.11	0.34	0.45	0.20	-0.08	-0.02	0.08	-0.11	-0.24	-0.16	0.15	0.46	0.65	0.79	0.89	0.71
$P(>0)$	49.96	52.62	54.71	51.41	49.40	49.15	49.90	48.86	47.06	49.72	49.88	53.10	53.02	55.28	56.29	55.01
t-statistic						-0.12	0.26	-0.44						3.84	2.80	2.94
	M								*BM*							
	P1 (low)	P2	P3	P4	P5 (high)	P5-P1	P5-P1	P5-P1	P1 (low)	P2	P3	P4	P5 (high)	P5-P1	P5-P1	P5-P1
Ann. mean (%)	-3.92	-0.39	2.98	4.70	10.54	14.46	22.79	5.71	-4.79	-3.37	6.21	5.78	9.25	14.04	13.85	10.31
Ann. σ (%)	18.71	16.13	14.84	16.32	20.19	23.37	23.78	23.53	17.42	16.85	15.65	16.18	17.20	19.39	19.72	19.68
Ann. Sharpe ratio	-0.21	-0.02	0.20	0.29	0.52	0.62	0.96	0.24	-0.27	-0.20	0.40	0.36	0.54	0.72	0.70	0.52
$P(>0)$	47.38	49.48	51.73	52.54	53.99	54.07	56.67	51.82	47.86	47.86	51.89	52.22	53.34	54.79	53.58	54.44
t-statistic						3.02	3.02	1.00						3.54	2.21	2.15
	ΔOI								*CR*							
	P1 (low)	P2	P3	P4	P5 (high)	P5-P1	P5-P1	P5-P1	P1 (high)	P2	P3	P4	P5 (low)	P5-P1	P5-P1	P5-P1
Ann. mean (%)	0.37	4.07	1.91	9.33	-1.14	-1.51	12.77	-10.80	-4.82	-1.41	6.52	8.90	13.03	17.85	16.55	18.43
Ann. σ (%)	17.74	16.50	15.82	17.47	17.90	20.32	19.15	21.27	17.81	18.58	17.79	17.17	18.00	19.36	17.93	19.46
Ann. Sharpe ratio	0.02	0.25	0.12	0.53	-0.06	-0.07	0.67	-0.51	-0.27	-0.08	0.37	0.52	0.72	0.92	0.92	0.95
$P(>0)$	49.14	50.43	50.69	53.01	48.97	51.03	56.85	46.92	46.31	51.87	52.48	52.36	54.53	57.07	53.40	57.29
t-statistic						-0.35	1.94	-2.09						3.68	1.30	3.89

Panel B: Factor correlation

	β	*B*	*M*	*BM*	*AHP*	*Q*	*OI*
B	-0.01						
M	0.08	0.44					
BM	-0.07	0.29	0.23				
AHP	0.13	0.25	0.34	0.08			
Q	-0.02	-0.01	-0.02	0.01	0.00		
ΔOI	-0.01	0.11	0.52	-0.02	0.13	-0.02	
CR	-0.01	0.00	-0.17	0.02	-0.02	0.19	-0.23

Table III.4: **Optimal subset of factors**

This table reports the optimal factor selection based on the Bayesian procedure of Barillas and Shanken (2018). The left columns ("Relative") reports the statistics for the relative test of Barillas and Shanken (2017), in which only factors are included in the set of dependent variables of the time series regressions. The right columns ("Absolute"), reports the statistic for which both remaining factors and the 26 commodity returns are in the dependent variables. I report the optimal set of factors for each subset size $n = 1, \ldots, 5$, (relative) and $n = 1, \ldots, 6$ (absolute). The optimal selection is based on the probability "Prob." derived from the Bayes factor BF, for each of the priors chosen for the maximum Sharpe ratio $1.25; 1.5; 1.75; 2$. The table also reports the average absolute α (Avg. $|\alpha|$), the Wald (W), and the Gibbons et al. (1989) (GRS) statistic and its p-value. The period of interest is 1994–2017 and the set of factors includes the basis B, momentum M, basis-momentum BM, open-interest growth ΔOI, β, and crowding CR.

Prior ($Sharpe_{max}$)	Nb. Factors	Selected factors	Relative						Absolute					
			Avg. $\lvert\alpha\rvert$ %	W	GRS	P-value GRS	BF	Prob.	Avg. $\lvert\alpha\rvert$ %	W	GRS	P-value GRS	BF	Prob.
1.25	1	B	0.12	23.24	4.63	0.00	0.13	0.11	0.10	63.96	2.01	0.00	0.05	0.04
	2	B-CR	0.09	9.76	2.43	0.05	0.42	0.29	0.10	50.05	1.63	0.02	0.13	0.12
	3	B-BM-CR	0.06	3.70	1.23	0.30	1.05	0.51	0.09	43.79	1.47	0.05	0.36	0.26
	4	**B-M-BM-CR**	**0.03**	**0.68**	**0.34**	**0.71**	**1.54**	**0.61**	**0.09**	**40.67**	**1.41**	**0.07**	**0.56**	**0.36**
	5	B-M-BM-ΔOI-CR	0.01	0.02	0.02	0.90	1.34	0.57	0.09	39.98	1.44	0.07	0.50	0.33
	6	β-B-M-BM-ΔOI-CR	-	-	-	-	-	-	0.09	39.97	1.50	0.05	0.38	0.27
1.5	1	B					0.03	0.03					0.01	0.01
	2	B-CR					0.34	0.25					0.31	0.24
	3	B-BM-CR					1.35	0.58					2.15	0.68
	4	**B-M-BM-CR**					**2.22**	**0.69**					**4.73**	**0.83**
	5	B-M-BM-ΔOI-CR					1.66	0.62					3.80	0.79
	6	β-B-M-BM-ΔOI-CR											2.32	0.70
1.75	1	B					0.01	0.01					0.01	0.01
	2	B-CR					0.36	0.27					2.01	0.67
	3	B-BM-CR					1.84	0.65					26.79	0.96
	4	**B-M-BM-CR**					**3.03**	**0.75**					**71.52**	**0.99**
	5	B-M-BM-ΔOI-CR					1.97	0.66					52.15	0.98
	6	β-B-M-BM-ΔOI-CR											26.79	0.96
2	1	B					0.01	0.01					0.02	0.02
	2	B-CR					0.44	0.30					18.79	0.95
	3	B-BM-CR					2.49	0.71					386.69	1.00
	4	**B-M-BM-CR**					**3.97**	**0.80**					**1140.45**	**1.00**
	5	B-M-BM-ΔOI-CR					2.28	0.70					755.92	1.00
	6	β-B-M-BM-ΔOI-CR											336.32	1.00

Table III.5: **Roll-day effect**

This table reports the results of a panel regression with contract fixed effects, and variance clustering at the contract level. Panel A (B) presents the results for commodity returns (turn-over) for the nearby and first deferred contract. The dependent variables are regressed onto five dummies that takes the value "1" ("0") when the current date t is a rolling day for contract c. The variable $di_{c,t}$ takes a value "1" when the week includes the i^{th} day of the five-day roll of the SP-GSCI, $i = 1, \ldots, 5$. Panel C reports the results for the time series regressions on the eight factors, for which I assume a monthly roll such as the one of the WTI crude oil contract. "Pre" and "Post" indicate the pre- (1994–2003) and post-financialization (2004–2017) sub-periods. All coefficients are expressed in basis points with the corresponding t-statistics in parenthesis. The period of interest is 1994–2017.

Panel A: Returns

	Nearby contract						First deferred contract					
	All contracts			SP-GSCI contracts			All contracts			SP-GSCI contracts		
	Pre	Post	All	Pre	Post	All	Pre	Post	All	Pre	Post	All
$d1_{c,t}$	-2.05	25.53	16.28	4.92	28.28	20.89	-14.99	23.31	9.97	-9.42	20.94	10.99
	(-0.14)	(1.69)	(1.49)	(0.30)	(1.54)	(1.57)	(-1.03)	(1.65)	(1.03)	(-0.59)	(1.19)	(0.91)
$d2_{c,t}$	19.68	-27.95	-10.31	28.06	-28.34	-7.02	29.87	-32.67	-9.09	31.68	-30.44	-7.05
	(1.27)	(-1.86)	(-0.94)	(1.43)	(-1.62)	(-0.52)	(1.88)	(-2.26)	(-0.87)	(1.59)	(-1.82)	(-0.55)
$d3_{c,t}$	25.16	-13.37	3.35	6.32	-22.78	-10.23	10.03	-8.28	-0.35	7.12	-19.37	-7.98
	(1.53)	(-1.18)	(0.36)	(0.37)	(-2.19)	(-1.12)	(0.81)	(-0.75)	(-0.05)	(0.44)	(-1.87)	(-1.02)
$d4_{c,t}$	20.12	-40.83	-18.58	20.84	-38.59	-16.32	30.69	-46.29	-17.01	24.14	-46.60	-19.42
	(0.76)	(-2.38)	(-1.07)	(0.62)	(-2.05)	(-0.72)	(1.38)	(-2.81)	(-1.13)	(0.86)	(-2.70)	(-0.98)
$d5_{c,t}$	-18.00	39.28	18.23	-0.90	34.27	22.09	-15.47	38.42	18.81	2.45	39.28	26.36
	(-0.95)	(2.70)	(1.49)	(-0.04)	(1.84)	(1.36)	(-0.82)	(2.84)	(1.62)	(0.12)	(2.27)	(1.79)
Adj R^2 (%)	-0.07	-0.05	-0.08	-0.06	-0.03	-0.08	-0.06	-0.02	-0.08	-0.03	0.02	-0.08
Nb. obs.	13468	18824	32292	9324	13032	22356	13468	18824	32292	9324	13032	22356

Panel B: Turnover

	Nearby contract						First deferred contract					
	All contracts			SP-GSCI contracts			All contracts			SP-GSCI contracts		
	Pre	Post	All	Pre	Post	All	Pre	Post	All	Pre	Post	All
$d1_{c,t}$	-42.07	6.71	-12.42	-10.52	8.62	0.49	174.16	3.44	74.76	229.87	4.79	100.10
	(-1.22)	(3.94)	(-0.92)	(-0.73)	(4.11)	(0.08)	(1.10)	(1.73)	(1.12)	(1.02)	(1.67)	(1.04)
$d2_{c,t}$	26.54	1.53	10.79	-4.79	2.74	-0.30	-12.01	-0.01	-5.16	-23.58	0.28	-9.85
	(0.91)	(2.13)	(0.98)	(-0.96)	(3.04)	(-0.13)	(-0.76)	(-0.02)	(-0.76)	(-1.07)	(0.38)	(-1.04)
$d3_{c,t}$	1.87	1.33	1.51	4.80	2.13	3.20	-12.16	1.90	-3.99	22.08	1.79	10.67
	(0.51)	(3.30)	(0.99)	(0.94)	(5.14)	(1.51)	(-0.41)	(4.40)	(-0.31)	(1.01)	(3.44)	(1.11)
$d4_{c,t}$	-2.42	0.33	-0.51	0.66	1.23	1.11	4.91	0.24	3.18	4.86	0.58	2.11
	(-0.63)	(0.73)	(-0.36)	(0.97)	(2.85)	(2.78)	(3.17)	(0.51)	(2.78)	(2.50)	(0.90)	(3.66)
$d5_{c,t}$	-5.36	7.66	2.03	-9.39	11.09	2.46	191.18	-2.99	77.16	218.62	-2.87	91.05
	(-0.53)	(3.38)	(0.44)	(-0.65)	(4.40)	(0.38)	(1.19)	(-1.84)	(1.15)	(0.97)	(-1.30)	(0.95)
Adj R^2 (%)	-0.23	2.67	-0.09	-0.24	5.08	-0.10	-0.22	-0.02	-0.09	-0.23	0.02	-0.09
Nb. obs.	13468	18824	32292	9324	13032	22356	13468	18824	32292	9324	13032	22356

Panel C: Factors

	AHP			Q			β			B		
	Pre	Post	All	Pre	Post	All	Pre	Post	All	Pre	Post	All
$d1_{c,t}$	0.24	0.31	0.28	0.37	0.22	0.28	-0.19	0.00	-0.08	0.29	0.32	0.31
	(1.77)	(2.25)	(2.85)	(2.52)	(1.81)	(3.00)	(-1.09)	(-0.03)	(-0.72)	(1.86)	(2.40)	(3.03)
$d2_{c,t}$	-1.12	0.22	-0.29	0.97	0.13	0.45	0.55	0.00	0.23	1.40	-0.11	0.47
	(-2.47)	(0.50)	(-0.91)	(1.97)	(0.34)	(1.47)	(0.91)	(0.01)	(0.60)	(2.61)	(-0.26)	(1.40)
$d3_{c,t}$	1.06	-0.57	0.06	-0.70	-0.18	-0.36	-0.41	-0.31	-0.36	-1.34	-0.14	-0.59
	(2.19)	(-1.20)	(0.19)	(-1.33)	(-0.43)	(-1.10)	(-0.64)	(-0.58)	(-0.86)	(-2.34)	(-0.30)	(-1.65)
$d4_{c,t}$	0.38	-0.24	0.04	-0.83	0.85	0.12	0.33	-0.30	-0.03	-0.42	-0.21	-0.30
	(0.94)	(-0.55)	(0.12)	(-1.90)	(2.25)	(0.43)	(0.61)	(-0.61)	(-0.08)	(-0.88)	(-0.51)	(-0.96)
$d5_{c,t}$	0.14	-0.08	-0.02	0.46	-0.84	-0.27	0.68	-0.17	0.16	0.28	0.81	0.60
	(0.27)	(-0.15)	(-0.06)	(0.83)	(-1.79)	(-0.76)	(1.00)	(-0.28)	(0.35)	(0.46)	(1.58)	(1.53)
Adj R^2 (%)	1.44	0.03	-0.23	0.62	0.23	-0.15	0.32	-0.03	-0.09	0.75	-0.26	0.08
Nb. obs.	518	724	1242	518	724	1242	518	724	1242	518	724	1242

	M			BM			OI			CR		
	Pre	Post	All	Pre	Post	All	Pre	Post	All	Pre	Post	All
$d1_{c,t}$	0.38	0.33	0.35	0.17	0.31	0.25	0.44	-0.08	0.14	0.00	0.23	0.13
	(2.02)	(2.11)	(2.92)	(1.08)	(2.38)	(2.50)	(2.94)	(-0.56)	(1.35)	(-0.07)	(1.77)	(1.65)
$d2_{c,t}$	0.29	0.25	0.28	0.68	-0.44	0.00	-0.72	0.16	-0.18	0.27	0.39	0.34
	(0.46)	(0.50)	(0.72)	(1.29)	(-1.07)	(0.00)	(-1.40)	(0.36)	(-0.54)	(1.28)	(0.92)	(1.28)
$d3_{c,t}$	-0.38	-0.48	-0.45	-0.82	0.34	-0.12	0.10	-0.65	-0.36	-0.30	0.20	-0.01
	(-0.55)	(-0.91)	(-1.08)	(-1.46)	(0.77)	(-0.33)	(0.19)	(-1.36)	(-1.00)	(-1.32)	(0.44)	(-0.04)
$d4_{c,t}$	-0.09	-0.89	-0.55	-0.31	-0.24	-0.27	0.03	-0.05	-0.01	0.13	0.06	0.08
	(-0.16)	(-1.86)	(-1.49)	(-0.65)	(-0.60)	(-0.89)	(0.07)	(-0.12)	(-0.05)	(0.66)	(0.15)	(0.33)
$d5_{c,t}$	0.58	1.16	0.91	0.12	0.96	0.63	0.16	0.39	0.29	-0.29	-0.32	-0.30
	(0.80)	(1.97)	(2.00)	(0.20)	(1.95)	(1.66)	(0.27)	(0.73)	(0.74)	(-1.19)	(-0.65)	(-0.96)
Adj R^2 (%)	-0.73	0.63	0.17	0.50	0.04	-0.04	0.12	-0.10	0.23	0.69	0.05	0.04
Nb. obs.	518	724	1242	518	724	1242	518	724	1242	518	724	1242

Table III.6: **Performance of the liquidity and insurance premiums**

This table reports the time series statistics of second-stage, cross-sectional, Fama-MacBeth regressions of $R_{c,t}$ onto lagged characteristics, (i) AHP and (ii) Q are the KRT characteristics for insurance and liquidity, (iii) CIT is the share of SP-GSCI interest in total OI, (iv) B is the basis, S is the 52-week trailing regression residuals of commodity returns on the S&P 500 returns, multiplied by "-1" when the speculators are net short, (v) $R_{c,t-1}$ is the lagged weekly return, and (vi) M, (vii) BM, and (viii) CR, are the momentum, basis-momentum, and crowding characteristics. Columns "KRT" are the specifications of Kang et al. (2020, Table VI, 2[nd] and 3[rd] columns). The column "18-CIT" restrict the estimations to the 18 contracts of the sample that belongs to the SP-GSCI. The column "3+" restrict the sample to the weeks that have at least three days under which the SP-GSCI rolls. The columns "Pre" and "Post" restrict the sample before (1994–2017) and after (2005–2017) the financialization. Finally, 1994–2020 extends the sample until December 2020. The t-statistics computed with Newey-West standard errors with four lags are reported in parenthesis. The period of interest is 1994–2017.

	KRT		Opt. risk	3+	18-CIT	26-CIT	Pre	Post	1994–2020
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$AHP_{c,t-1}$	0.51	0.43	0.34	0.55	0.31	0.38	0.25	0.43	0.41
	(3.35)	(2.67)	(1.93)	(1.98)	(1.18)	(2.04)	(0.92)	(1.84)	(2.33)
$Q_{c,t-1}$		4.66	3.80	2.32	3.78	3.70	2.63	4.82	3.20
		(5.97)	(4.91)	(1.88)	(3.38)	(4.63)	(3.10)	(3.91)	(2.49)
$CIT_{c,t-1}$					-0.14	-0.25			
					(-0.11)	(-0.29)			
$B_{c,t-1} \times 10^2$	-1.38	-1.39	-0.53	-2.70	-0.21	-0.49	0.42	-1.38	1.35
	(-2.52)	(-2.52)	(-0.85)	(-2.09)	(-0.25)	(-0.76)	(0.50)	(-1.55)	(0.65)
$S_{c,t-1} \times 10^{-2}$	-0.21	-0.16							
	(-1.72)	(-1.30)							
R_{t-1}	0.01	0.03							
	(1.19)	(3.14)							
$M_{c,t-1}$			0.11	0.12	0.09	0.09	0.21	0.01	0.03
			(0.67)	(0.42)	(0.39)	(0.54)	(0.88)	(0.06)	(0.16)
$BM_{c,t-1}$			1.52	0.86	1.38	1.51	0.62	2.27	2.32
			(3.06)	(0.88)	(1.95)	(2.98)	(0.97)	(3.09)	(2.32)
$CR_{c,t-1}$			-0.96	-1.05	-0.74	-0.96	-0.15	-1.65	-0.63
			(-3.45)	(-2.03)	(-2.09)	(-3.33)	(-0.59)	(-3.59)	(-1.78)
Intercept	0.05	0.06	0.05	-0.01	-0.01	0.03	0.05	0.05	0.01
	(0.89)	(1.11)	(0.90)	(-0.10)	(-0.10)	(0.42)	(0.79)	(0.60)	(0.15)
Avg. R^2 (%)	25.66	30.02	35.78	35.17	52.92	39.18	36.65	35.02	36.52
Avg. Adj. R^2 (%)	11.06	11.95	13.73	12.81	18.52	13.88	13.65	13.78	14.44

Table III.7: **Panel regressions**

This table reports panel data regression predictability results for the AHP, Q, and CIT characteristics. In Panel A, I report the results for the full sample period (columns 1 and 2), and the pre- and post-financialization periods ("Pre" and "Post"). For each sample, I report the results with and without the factor risk adjustment (B-M-BM-CR), common to all contracts, and add characteristics/characteristics interactions time series average FE, when the Haussman (1978) test rejects the null of no FE; see Hoechle et al. (2020). Panel B replicates the study above using the DCOT data available from 2007, for AHP^{PROD}, Q^{PROD}, and adds the variable CIT^{SWAP}, which is the index pressure computed from the swap dealers positions of the DCOT. I do not report the factor coefficients. Panel B also extends the study to 2020. The t-statistics computed with Driscoll and Kraay (1998) standard errors (controlling for heteroscedasticity and autocorrelation at the contract level) are in parenthesis. For the Haussman test, ** and *** indicate the significance levels of 5% and 1%, respectively.

Panel A: COT

	1994–2017		Pre		Post	
	(1)	(2)	(3)	(4)	(5)	(6)
$AHP_{c,t-1}$	0.33	0.33	0.48	0.50	0.25	-0.37
	(2.28)	(2.24)	(2.26)	(2.36)	(1.34)	(-0.76)
$Q_{c,t-1}$	2.78	2.80	2.52	2.69	3.19	4.79
	(5.73)	(5.75)	(4.83)	(4.98)	(3.26)	(3.31)
$CIT_{c,t-1}$	-0.26	-0.18	2.93	1.55	-0.69	-0.86
	(-0.19)	(-0.14)	(0.71)	(0.41)	(-0.48)	(-0.50)
B_t		0.56		2.69		-2.31
		(0.30)		(1.35)		(-0.69)
M_t		2.30		0.39		-1.52
		(1.28)		(0.18)		(-0.51)
BM_t		-1.95		-0.51		-3.97
		(-0.96)		(-0.28)		(-1.23)
CR_t		-1.37		-0.79		0.65
		(-0.80)		(-0.29)		(0.21)
Constant	0.01	0.01	-0.02	-0.03	0.02	-0.55
	(0.18)	(0.19)	(-0.37)	(-0.38)	(0.31)	(-2.01)
FE						✓
Haussman	1.28	15.70	2.87	13.93	6.00	25.16**
Adj. R^2 (%)	0.15	0.17	0.21	0.36	0.11	0.33

Panel B: DCOT

	2007–2017	2007–2020
	(1)	(2)
$AHP^{PROD}_{c,t-1}$	-0.39	-0.34
	(-0.84)	(-0.84)
$Q^{PROD}_{c,t-1}$	4.65	4.78
	(3.20)	(3.81)
$CIT^{SWAP}_{c,t-1}$	0.10	-0.13
	(0.12)	(-0.18)
Constant	-0.14	-0.11
	(-1.07)	(-1.02)
FE	✓	✓
Haussman	9.97***	14.02***
Adj. R^2 (%)	0.21	0.22

Appendix

Table III.A1: **Contract description**

This table reports the specifications of the 26 contracts. The reported characteristics are the trading venues, tickers, underlying commodities, units, maturity months, and the inception date of the corresponding electronically traded contract.

Commodity	Trading venue	Ticker	SP-GSCI	Underlying	Trading months	Electronification date
WTI crude oil	NYMEX/ICE	CL	✓	bbl (1,000)	FGHJKMNQUVXZ	5/9/2006
Heating oil	NYMEX	HO	✓	gal (42,000)	FGHJKMNQUVXZ	5/9/2006
Natural gas	NYMEX/ICE	NG	✓	MMBtu (10,000)	FGHJKMNQUVXZ	5/9/2006
Platinum	NYMEX	PL	✓	ozt (50)	FGHJKMNQUVXZ	4/12/2006
Palladium	NYMEX	PA		ozt (100)	HMUZ	4/12/2006
Silver	CMX	SI	✓	ozt (5,000)	FHKNUZ	4/12/2006
Copper	CMX	HG		lbs (25,000)	FGHJKMNQUVXZ	4/12/2006
Gold	CMX	GC	✓	ozt (100)	GJMQVZ	4/12/2006
Wheat	CBT	W	✓	bu (5,000)	HKNUZ	1/8/2006
Kansas wheat	KBT	KW	✓	bu (5,000)	HKNUZ	13/01/2008
Minn. Wheat	CME	MWE		bu (50,000)	HKNUZ	15/12/2004
Corn	CBT	C	✓	bu (5,000)	HKNUZ	1/8/2006
Oats	CBOT	O		bu (5,000)	HKNUZ	1/8/2006
Soybeans	CBT	S	✓	bu (5,000)	FHKNQUX	1/8/2006
Soybean oil	CBOT	BO		lbs (60,000)	FHKNQUVZ	1/8/2006
Soybean meal	CBOT	SM		sh tn (100)	FHKNQUVZ	1/8/2006
Rough rice	CBOT	RR		CWT (2,000)	FHKNUX	1/8/2006
Cotton	ICE-US	CT	✓	lbs (50,000)	HKNVZ	2/2/2007
Orange juice	ICE	OJ	✓	lbs (15,000)	FHKNUX	2/2/2007
Lumber	CME	LB		m3 (110,000)	FHKNUX	20/10/2008
Cocoa	ICE-US	CC	✓	MT (10)	HKNUZ	2/2/2007
Raw sugar	ICE-US	SB	✓	lbs (112,000)	HKNVZ	2/2/2007
Coffee	ICE-US	KC	✓	lbs (37,500)	HKNUZ	2/2/2007
Lean hogs	CME	LH	✓	lbs (40,000)	GJMNQVZ	4/12/2006
Live cattle	CME	LC	✓	lbs (40,000)	GJMQVZ	4/12/2006
Feeder cattle	CME	FC	✓	lbs (50,000)	FHJKQUVX	24/11/2008

Maturity month code: F = January, G = February, H = Mars, J = April, K = May, M = June, N = July, Q = August, U = September, V = October, X = November, Z = December.

Table III.A2: **Futures returns on the first deferred contract**

This table reports the annualized means and standard deviations, and the skewness and kurtosis of the weekly returns on the 26 first deferred commodity futures contracts. The period of interest is 1994–2017, and two sub-periods before (1994–2003) and after (2004–2010) the financialization. The table reports the statistics for the equally-weighted portfolios for each group (Energy, Metal, Agriculture, Soft and Livestock) and for the 26 contracts equally-weighted portfolio. The tickers are defined in the Appendix Table A.1.

	1994–2017				1994–2003				2004–2017			
Ticker	Mean (%)	S.d. (%)	Skewness	Kurtosis	Mean (%)	S.d. (%)	Skewness	Kurtosis	Mean (%)	S.d. (%)	Skewness	Kurtosis
CL	10.19	32.04	-0.10	4.72	21.37	29.76	-0.33	4.58	2.18	33.56	0.04	4.74
HO	8.78	30.05	0.10	4.30	13.57	28.41	-0.15	3.66	5.34	31.17	0.25	4.59
NG	-7.02	41.01	0.21	3.94	18.73	43.44	0.15	3.96	-25.45	39.01	0.23	3.85
PL	7.31	20.87	-0.03	5.70	11.09	19.43	0.30	5.63	4.19	21.99	-0.20	5.58
PA	16.06	32.51	0.40	5.82	18.93	34.57	0.87	7.07	13.77	30.79	-0.14	4.02
SI	6.00	28.76	-0.01	6.28	-0.73	20.31	0.53	4.99	10.81	33.52	-0.13	5.38
HG	9.91	24.71	0.01	5.64	5.60	20.71	0.00	3.80	12.97	27.21	-0.01	5.67
GC	3.71	16.54	0.50	9.34	-2.21	13.40	2.08	21.45	7.93	18.45	0.00	6.20
W	-5.73	27.85	0.50	4.74	-4.81	22.41	0.46	4.73	-6.39	31.18	0.49	4.29
KW	-1.78	26.56	0.45	4.59	-1.43	22.26	0.57	6.04	-2.03	29.26	0.40	3.90
MWE	2.83	24.19	0.39	6.28	0.29	19.97	0.88	7.42	4.64	26.81	0.22	5.52
C	-3.75	26.11	0.34	5.62	-7.82	20.91	0.37	4.88	-0.83	29.28	0.30	5.19
O	1.19	30.23	0.50	7.54	-2.53	26.21	0.23	4.54	3.77	32.75	0.57	7.98
S	6.21	22.84	0.06	4.13	3.64	19.78	0.16	4.04	8.04	24.81	0.01	3.93
BO	-1.44	23.15	0.21	4.06	-3.52	20.17	0.21	3.84	0.04	25.07	0.20	3.91
SM	11.27	25.80	0.19	3.97	9.14	22.49	0.49	4.50	12.79	27.94	0.07	3.62
RR	-6.22	23.49	0.08	4.24	-10.82	23.90	0.15	5.14	-2.92	23.20	0.04	3.52
CT	0.19	25.48	0.10	4.89	-0.38	22.76	0.17	4.34	0.60	27.27	0.07	4.89
OJ	-0.57	30.28	0.23	5.11	-12.09	25.16	0.01	4.82	7.68	33.43	0.24	4.75
LB	-6.39	26.87	0.27	3.46	-4.06	25.55	0.08	2.86	-8.05	27.79	0.37	3.75
CC	1.70	27.86	0.68	6.44	0.63	29.08	1.03	7.94	2.46	26.98	0.36	4.94
SB	4.99	27.66	-0.06	4.77	2.98	25.12	0.22	4.16	6.43	29.36	-0.18	4.87
KC	1.45	34.98	0.50	6.21	4.07	40.07	0.62	6.78	-0.42	30.86	0.27	3.80
LH	-0.43	27.92	0.15	10.06	0.52	27.92	0.21	16.16	-1.12	27.94	0.11	5.71
LC	1.52	13.77	-0.47	7.90	2.59	12.84	-1.40	15.50	0.73	14.42	0.01	4.31
FC	1.98	15.07	-0.33	5.75	-1.23	12.92	-0.79	9.76	4.28	16.43	-0.19	4.30
Energy	3.98	27.80	0.00	3.84	17.89	27.02	-0.09	3.63	-5.97	28.28	0.07	3.99
Metal	8.71	18.95	-0.14	5.39	5.95	13.63	0.23	3.43	10.69	21.99	-0.21	4.76
Agriculture	0.24	18.59	0.29	4.66	-1.89	15.37	0.44	4.49	1.77	20.60	0.22	4.32
Soft	0.23	15.04	-0.01	3.90	-1.48	13.42	0.32	3.56	1.45	16.10	-0.16	3.88
Livestock	1.17	14.76	-0.23	6.23	0.63	14.27	-0.36	9.53	1.55	15.11	-0.15	4.32
Average	2.27	12.37	-0.10	5.33	2.31	9.17	0.17	3.19	2.24	14.24	-0.14	4.83

Table III.A3: **Relative importance of the financialization in the term structure**

This Table reports the relative importance of the open interest of index investment in the term structure of the 26 commodities. Panel A reports the importance of the SP-GSCI open interest in the total open interest and total commercial long positions from the "legacy" report of the CFTC, in which commercial longs include the swap dealers positions, proxy for index investment. Panel B reports the relative weights of nearby and first deferred open interest and trading volume, in the total open interest and trading volume of the term structure.

	Nearby OI ratio			First deferred OI ratio			Nearby volume ratio			First deferred volume ratio			SP-GSCI in total OI			SP-GSCI in "commercials" long		
	1994–2017	1994–2003	2004–2017	1994–2017	1994–2003	2004–2017	1994–2017	1994–2003	2004–2017	1994–2017	1994–2003	2004–2017	1994–2017	1994–2003	2004–2017	1994–2017	1994–2003	2004–2017
CL	17.05	19.88	16.40	17.02	18.57	16.67	45.44	43.19	45.79	24.61	30.59	23.69	13.21	4.89	15.11	27.29	7.19	34.39
HO	18.95	21.59	17.99	23.48	22.36	23.89	36.30	44.23	34.37	30.27	30.56	30.20	9.39	4.19	11.29	16.83	7.29	20.47
NG	11.82	13.75	11.40	16.02	12.22	16.84	40.60	49.82	39.03	22.50	19.75	22.97	4.78	2.33	5.31	10.63	3.31	13.46
PL	8.53	39.46	2.14	26.72	48.03	22.32	9.14	45.03	2.34	36.05	46.93	33.99	0.23	1.34	0.00	1.01	2.95	0.00
PA	8.59	38.74	4.48	26.91	47.70	24.08	7.72	44.77	4.27	35.11	45.50	34.14						
SI	0.38	0.68	0.29	18.37	19.29	18.11	1.45	2.58	1.17	37.43	35.12	38.01	1.27	0.37	1.53	4.12	0.95	5.38
HG	1.57	2.74	1.25	17.37	20.53	16.51	2.19	6.31	1.50	33.82	30.27	34.41						
GC	0.25	0.30	0.23	19.78	17.92	20.21	1.12	2.03	0.97	38.68	36.33	39.05	2.76	1.14	3.14	7.75	1.78	10.82
W	31.25	33.79	30.77	36.14	38.74	35.64	41.49	40.33	41.71	38.45	41.43	37.89	14.05	6.61	15.48	25.98	8.58	31.18
KW	30.81	31.63	30.59	37.27	41.36	36.20	39.70	38.32	40.06	39.15	42.64	38.23	8.59	3.52	9.90	14.82	3.12	22.67
MWE	27.69	34.12	27.01	36.72	42.35	36.12	38.28	39.88	38.07	39.22	45.06	38.44						
C	26.68	25.45	26.96	32.58	33.32	32.41	38.63	36.63	39.00	34.14	34.43	34.08	6.73	2.40	7.72	11.25	2.09	16.40
O	35.33	33.91	36.47	40.23	38.87	41.33	44.65	43.40	45.90	39.54	39.61	39.48						
S	22.73	22.97	22.67	29.67	30.89	29.39	35.03	34.30	35.20	33.14	35.62	32.56	3.94	1.17	4.58	7.22	1.33	9.76
BO	18.63	17.89	18.85	30.15	30.07	30.17	30.17	29.71	30.27	34.64	36.18	34.33						
SM	18.33	18.40	18.30	28.01	28.17	27.95	30.54	31.22	30.35	33.32	34.03	33.13						
RR	35.08	27.43	37.72	42.45	38.78	43.72	41.31	36.22	43.05	44.47	43.83	44.69						
CT	25.88	24.13	26.37	38.55	34.41	39.70	36.15	34.60	36.69	38.65	37.27	39.13	7.65	2.58	9.07	14.88	4.53	18.18
OJ	40.09	37.62	42.10	39.03	36.33	41.23	48.21	46.58	49.56	40.18	39.28	40.93	1.52	3.33	0.05	3.43	7.09	0.11
LB	43.29	52.12	40.31	42.80	34.69	45.53	50.65	55.18	47.92	39.42	36.51	41.18						
CC	19.68	17.51	20.50	35.79	30.97	37.63	34.78	36.23	34.35	44.59	44.11	44.73	2.82	0.64	3.66	5.11	0.97	7.13
SB	38.63	41.00	38.22	26.19	27.43	25.97	47.54	54.37	46.42	30.06	28.19	30.37	6.47	3.00	7.06	12.26	5.61	13.41
KC	25.17	28.54	24.45	46.08	42.08	46.94	37.19	39.15	36.62	45.04	46.79	44.52	4.31	1.65	4.87	8.84	3.48	9.95
LH	21.28	30.19	20.02	30.93	34.34	30.45	31.32	38.55	30.08	33.54	38.53	32.68	13.58	7.45	14.45	33.74	21.55	35.20
LC	16.29	22.44	14.66	40.19	35.34	41.47	26.97	34.79	24.97	39.86	36.79	40.64	11.34	4.64	13.11	27.38	11.85	31.20
FC	25.16	27.26	24.45	31.08	28.42	31.99	27.84	31.21	26.92	35.10	33.30	35.59	10.57	1.56	13.64	35.69	5.56	45.27

Table III.A4: **Optimal subset of eight factors**

This table reports the optimal factor selection based on the Bayesian procedure of Barillas and Shanken (2018). The left columns ("Relative") reports the statistics for the relative test of Barillas and Shanken (2017), in which only factors are included in the set of dependent variables of the time series regressions. The right columns ("Absolute"), reports the statistic for which both remaining factors and the 26 commodity returns are in the dependent variables. I report the optimal set of factors for each subset size $n = 1, \dots, 7$, (relative) and $n = 1, \dots, 8$ (absolute). The optimal selection is based on the probability "Prob." derived from the Bayes factor BF, for each of the $k = 1, 1.25, 1.75, 2$ priors chosen for the maximum Sharpe ratio. The table also reports the average absolute α (Avg. $|\alpha|$), the Wald (W) and the Gibbons et al. (1989) (GRS) statistic and its p-value. The period of interest is 1994–2017 and the set of factors includes the basis B, momentum M, basis-momentum BM, open-interest growth ΔOI, β, crowding CR, average hedging pressure AHP, and net trading Q.

Prior ($Sharpe_{max}$)	Nb. Factors	Selected factors	Relative						Absolute					
			Avg. $\lvert\alpha\rvert$ %	W	GRS	P-value GRS	BF	Prob.	Avg. $\lvert\alpha\rvert$ %	W	GRS	P-value GRS	BF	Prob.
1.25	1	Q	0.18	34.61	4.91	0.00	0.02	0.02	0.13	76.48	2.25	0.00	0.00	0.00
	2	B-Q	0.11	19.55	3.24	0.00	0.08	0.07	0.11	60.92	1.85	0.00	0.03	0.03
	3	B-Q-CR	0.10	10.83	2.15	0.06	0.39	0.28	0.10	51.91	1.63	0.02	0.16	0.14
	4	B-BM-Q-CR	0.07	5.13	1.27	0.28	1.16	0.54	0.10	46.02	1.49	0.04	0.57	0.36
	5	B-M-BM-Q-CR	0.05	2.13	0.71	0.55	1.80	0.64	0.09	42.93	1.44	0.06	0.97	0.49
	6	**B-M-BM-AHP-Q-CR**	**0.03**	**0.66**	**0.33**	**0.72**	**1.83**	**0.65**	**0.09**	**41.41**	**1.44**	**0.07**	**1.05**	**0.51**
	7	B-M-BM-AHP-Q-ΔOI-CR	0.02	0.04	0.04	0.84	1.47	0.59	0.09	40.77	1.47	0.06	0.86	0.46
	8	β-B-M-BM-AHP-Q-ΔOI-CR							0.09	40.73	1.52	0.04	0.59	0.37
1.5	1	Q					0.00	0.00					0.00	0.00
	2	B-Q					0.04	0.04					0.08	0.08
	3	B-Q-CR					0.45	0.31					1.93	0.66
	4	B-BM-Q-CR					1.94	0.66					14.13	0.93
	5	B-M-BM-Q-CR					3.15	0.76					30.53	0.97
	6	**B-M-BM-AHP-Q-CR**					**2.88**	**0.74**					**32.23**	**0.97**
	7	B-M-BM-AHP-Q-ΔOI-CR					1.89	0.65					22.67	0.96
	8	β-B-M-BM-AHP-Q-ΔOI-CR											12.18	0.92
1.75	1	Q					0.00	0.00					0.00	0.00
	2	B-Q					0.05	0.04					0.92	0.48
	3	B-Q-CR					0.68	0.40					48.80	0.98
	4	B-BM-Q-CR					3.28	0.77					519.70	1.00
	5	B-M-BM-Q-CR					5.10	0.84					1212.05	1.00
	6	**B-M-BM-AHP-Q-CR**					**4.11**	**0.80**					**1196.31**	**1.00**
	7	B-M-BM-AHP-Q-ΔOI-CR					2.29	0.70					730.62	1.00
	8	β-B-M-BM-AHP-Q-ΔOI-CR											324.84	1.00
2	1	Q					0.00	0.00					0.00	0.00
	2	B-Q					0.06	0.05					14.26	0.93
	3	B-Q-CR					1.06	0.52					1259.97	1.00
	4	B-BM-Q-CR					5.34	0.84					16352.66	1.00
	5	B-M-BM-Q-CR					7.74	0.89					38433.34	1.00
	6	**B-M-BM-AHP-Q-CR**					**5.54**	**0.85**					**34981.22**	**1.00**
	7	B-M-BM-AHP-Q-ΔOI-CR					2.68	0.73					18831.22	1.00
	8	β-B-M-BM-AHP-Q-ΔOI-CR											7171.01	1.00

Bibliography

Abdi, F., Ranaldo, A., 2017. A simple estimation of bid-ask spreads from daily close, high, and low prices. Review of Financial Studies 30, 4437–4480.

Acharya, V. V., Lochstoer, L. A., Ramadorai, T., 2013. Limits to arbitrage and hedging: Evidence from commodity markets. Journal of Financial Economics 109, 441–465.

Adams, Z., Gluck, T., 2015. Financialization in commodity markets: A passing trend or the new normal? Journal of Banking and Finance 60, 93–111.

Admati, A. R., Pfleiderer, P., 1991. Sunshine trading and financial market equilibrium. Review of Financial Studies 4, 443–481.

Albuquerque, R., 2012. Skewness in stock returns: Reconciling the evidence on firm versus aggregate returns. Review of Financial Studies 25, 1630–1673.

Andersen, T. G., Bollerslev, T., 1998a. Deutsche Mark-Dollar volatility: Intraday activity patterns, macroeconomic announcements and long run dependencies. Journal of Finance 53, 219–265.

Andersen, T. G., Bollerslev, T., 1998b. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. International Economic Review 39, 885–905.

Andersen, T. G., Bollerslev, T., Diebold, F. X., 2007. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. Review of Economics and Statistics 89, 701–720.

Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71, 579–625.

Andersen, T. G., Bollerslev, T., Huang, X., 2011a. A reduced form framework for modeling volatility of speculative prices based on realized variation measures. Journal of Econometrics 160, 176–189.

Andersen, T. G., Bollerslev, T., Meddahi, N., 2011b. Realized volatility forecasting and market microstructure noise. Journal of Econometrics 160, 220–234.

Anderson, R. W., 1985. Some determinants of the volatility of futures prices. Journal of Futures Markets 5, 331–348.

Anderson, R. W., Danthine, J.-P., 1981. Cross hedging. Journal of Political Economy 89, 1182–1196.

Anderson, R. W., Danthine, J. P., 1983. The time pattern of hedging and the volatility of futures prices. Review of Economic Studies 50, 249–266.

Asness, C. S., Moskowitz, T. J., Pedersen, L. H., 2013. Value and momentum everywhere. Journal of Finance 68, 929–985.

Aulerich, N. M., Irwin, S. H., Garcia, P., 2013. Bubbles, food prices, and speculation: Evidence from the cftc's daily large trader data files. Available at: `http://dx.doi.org/10.3386/w19065`

Aït-Sahalia, Y., 2002. Telling from discrete data whether the underlying continuous-time model is a diffusion. Journal of Finance 57, 2075–2112.

Aït-Sahalia, Y., 2004. Disentangling diffusion from jumps. Journal of Financial Economics 74, 487–528.

Aït-Sahalia, Y., Mykland, P. A., Zhang, L., 2005. How often to sample a continuous-time process in the presence of market microstructure noise. Review of Financial Studies 18, 351–416.

Bai, J., Bali, T. G., Wen, Q., 2019. Common risk factors in the cross-section of corporate bond returns. Journal of Financial Economics 131, 619–642.

Bai, J., Lumsdaine, R. L., Stock, J. H., 1998. Testing for and dating common breaks in multivariate time series. Review of Economic Studies 65, 395–432.

Bailey, W., Chan, K. C., 1993. Macroeconomic influences and the variability of the commodity futures basis. Journal of Finance 48, 555–573.

Bakshi, G., Gao, X., Rossi, A. G., 2019. Understanding the sources of risk underlying the cross-section of commodity returns. Management Science 65, 619–641.

Baltas, N., 2019. The impact of crowding in alternative risk premia investing. Financial Analysts Journal 75 (3), 89–104.

Barillas, F., Shanken, J., 2017. Which alpha? Review of Financial Studies 30, 1316–1338.

Barillas, F., Shanken, J., 2018. Comparing asset pricing models. Journal of Finance 73, 715–754.

Barndorff-Nielsen, O. E., Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society: Series B 64, 253–280.

Basak, S., Pavlova, A., 2016. A model of financialization of commodities. Journal of Finance 71, 1511–1555.

Basu, D., Miffre, J., 2013. Capturing the risk premium of commodity futures: The role of hedging pressure. Journal of Banking and Finance 37, 2652–2664.

Bayer, S., Dimitriadis, T., 2020. Regression-based expected shortfall backtesting. Journal of Financial Econometrics, forthcoming.

Beckmann, J., Belke, A., Czudaj, R., 2014. Does global liquidity drive commodity prices? Journal of Banking and Finance 48, 224–234.

Bekaert, G., Harvey, C. R., Lumsdaine, R. L., 2002. Dating the integration of world equity markets. Journal of Financial Economics 65, 203–247.

Bekierman, J., Manner, H., 2018. Forecasting realized variance measures using time-varying coefficient models. International Journal of Forecasting 34, 276–287.

Bessembinder, H., 1992. Systematic risk, hedging pressure and risk premiums in futures markets. Review of Financial Studies 5, 637–667.

Bessembinder, H., 1993. An empirical analysis of risk premia in futures markets. Journal of Futures Markets 13, 611–630.

Bessembinder, H., Carrion, A., Tuttle, L., Venkataram, K., 2016. Liquidity, resiliency and market quality around predictable trades: Theory and evidence. Journal of Financial Economics 121, 142–166.

Bessembinder, H., Coughenour, J. F., Seguin, P. J., Monroe-Smoller, M., 1996. Is there a term structure of futures volatilities? Reevaluating the Samuelson hypothesis. Journal of Derivatives 4 (2), 45–58.

Black, F., 1976. The pricing of commodity contracts. Journal of Financial Economics 3, 167–179.

Bodie, Z., Rosansky, V. I., 1980. Risk and return in commodity futures. Financial Analysts Journal 36 (3), 33–39.

Boehmer, E., Musumecci, J., Poulsen, A. B., 1991. Event-study methodology under conditions of event-induced variance. Journal of Financial Economics 30, 253–272.

Bollerslev, T., Hood, B., Huss, J., Pedersen, L. H., 2018. Risk everywhere: Modeling and managing volatility. Review of Financial Studies 31, 2729–2773.

Bollerslev, T., Patton, A. J., Quaedvlieg, R., 2016. Exploiting the errors: A simple approach for improved volatility forecasting. Journal of Econometrics 192, 1–18.

Boons, M., Porras Prado, M., 2019. Basis-momentum. Journal of Finance 74, 239–279.

Boons, M., Roon *de*, F., Szymanowska, M., 2014. The price of commodity risk in stock and futures markets. Available at: `http://dx.doi.org/10.2139/ssrn.1785728`

Breeden, D. T., 1980. Consumption risk in futures markets. Journal of Finance 35, 503–520.

Brennan, M. J., 1958. The supply of storage. American Economic Review 48, 50–72.

Brunetti, C., Reiffen, D., 2014. Commodity index trading and hedging costs. Journal of Financial Markets 21, 153–180.

Brunnermeier, M. K., Pedersen, L. H., 2005. Predatory trading. Journal of Finance 60, 1825–1863.

Buccheri, G., Corsi, F., 2019. HARK the SHARK: Realized volatility modeling with measurement errors and nonlinear dependencies. Journal of Financial Econometrics, forthcoming.

Buyuksahin, B., Haigh, M. S., Harris, J. H., Overdahl, J. A., Robe, M. A., 2009. Fundamentals, trader activity and derivative pricing. Available at: `http://dx.doi.org/10.2139/ssrn.966692`

Buyuksahin, B., Harris, J. H., 2011. Do speculators drive crude oil futures prices? Energy Journal 32, 167–202.

Buyuksahin, B., Robe, M. A., 2014. Speculators, commodities and cross-market linkages. Journal of International Money and Finance 42, 38–70.

Carpantier, J.-F., 2010. Commodities inventory effect. Available at: `https://econpapers.repec.org/RePEc:cor:louvco:2010040`

Carpantier, J.-F., Dufays, A., 2012. Commodities volatility and the theory of storage. Available at: https://EconPapers.repec.org/RePEc:cor:louvco:2012037

Carpantier, J.-F., Samkharadze, B., 2012. The asymmetric commodity inventory effect on the optimal hedge ratio. Journal of Futures Markets 33, 868–888.

Carter, C. A., Rausser, G. C., Schmitz, A., 1983. Efficient asset portfolios and the theory of normal backwardation. Journal of Political Economy 91, 319–331.

Casas, I., Ferreira, E., Orbe, S., 2019. Time-varying coefficient estimation in SURE models. Application to portfolio management. Journal of Financial Econometrics, forthcoming.

Casas, I., Mao, X., Veiga, H., 2018. Reexamining financial and economic predictability with new estimators of realized variance and variance risk premium. CREATES Research Paper 2018-10. Available at: https://ideas.repec.org/p/aah/create/2018-10.html

Casassus, J., Collin-Dufresne, P., 2005. Stochastic convenience yield implied from commodity futures and interest rates. Journal of Finance 55, 2283–2331.

Casassus, J., Liu, P., Tang, K., 2013. Economic linkages, relative scarcity, and commodity futures returns. Review of Financial Studies 26, 1324–1362.

Chang, E. C., 1985. Returns to speculators and the theory of normal backwardation. Journal of Finance 40, 193–208.

Chen, R., Yang, L., Hafner, C., 2004. Nonparametric multistep-ahead prediction in time series analysis. Journal of the Royal Statistical Society: Series B 66, 669–686.

Chen, X. B., Gao, J., Li, D., Silvapulle, P., 2018. Nonparametric estimation and forecasting for time-varying coefficient realized volatility models. Journal of Business and Economic Statistics 36, 88–100.

Chen, Y.-J., Duan, J.-C., Hung, M.-W., 2000. Volatility and maturity effects in the NIKKEI index futures. Journal of Futures Markets 19, 895–909.

Cheng, I.-H., Kirilenko, A., Xiong, W., 2015. Convective risk flows in commodity futures markets. Review of Finance 19, 1733–1781.

Chng, M. T., 2009. Economic linkages across commodity futures: Hedging and trading implications. Journal of Banking and Finance 33, 958–970.

Cipollini, F., Gallo, G. M., Otranto, E., 2017. On heteroskedasticity and regimes in volatility forecasting. Available at: http://dx.doi.org/10.2139/ssrn.3037550

Cochrane, J. H., 2011. Presidential address: Discount rates. Journal of Finance 66, 1047–1108.

Cootner, P. H., 1960. Returns to speculators: Telser versus Keynes. Journal of Political Economy 68, 396–404.

Corsi, F., 2009. A simple approximate long-memory model of realized volatility. Journal of Financial Econometrics 7, 174–196.

Corsi, F., Reno, R., 2012. Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. Journal of Business and Economic Statistics 30, 368–380.

Couperier, O., Leymarie, J., 2020. Backtesting expected shortfall via multi-quantile regression. Available at: https://halshs.archives-ouvertes.fr/halshs-01909375v5/file/Backtesting%20ES%20via%20Multi-Quantile%20Regression.pdf

Daskalaki, C., Kostakis, A., Skiadopoulos, G., 2014. Are there common factors in individual commodity futures returns? Journal of Banking and Finance 40, 346–363.

Deaton, A., Laroque, G., 1992. On the behaviour of commodity prices. Review of Economic Studies 59, 1–23.

Deaton, A., Laroque, G., 1996. Competitive storage and commodity price dynamics. Journal of Political Economy 104, 896–923.

Dewally, M., Ederington, L. H., Fernando, C. S., 2013. Determinants of trader profits in commodity futures markets. Review of Financial Studies 26, 2648–2683.

Diebold, F. X., Mariano, R. S., 1995. Comparing predictive accuracy. Journal of Business and Economic Statistics 13, 253–263.

Dincerler, C., Khokher, Z., Simin, T., 2020. An empirical analysis of commodity convenience yields. Quarterly Journal of Finance 10, 2050009. Available at: https://doi.org/10.1142/S2010139220500093

Driscoll, J. C., Kraay, A. C., 1998. Consistent covariance matrix estimation with spatially dependent panel data. Review of Economics and Statistics 80, 549–560.

Dubois, M., Maréchal, L., 2021. The valuation effects of index investment in commodity futures. This manuscript.

Dusak, K., 1973. Futures trading and investor returns: An investigation of commodity market risk premiums. Journal of Political Economy 81, 1387–1406.

Ederington, L. H., Lee, J. H., 2002. Who trades futures and how: Evidence from the heating oil futures market. Journal of Business 75, 353–373.

Erb, C. B., Harvey, C. R., 2006. The strategic and tactical value of commodity futures. Financial Analysts Journal 62 (2), 69–97.

Erb, C. B., Harvey, C. R., 2016. Conquering misperceptions about commodity futures investing. Financial Analysts Journal 72 (4), 26–35.

Etula, E., 2013. Broker-dealer risk appetite and commodity returns. Journal of Financial Econometrics 11, 486–521.

Fama, E. F., French, K. R., 1987. Commodity futures prices: Some evidence on forecast power, premiums, and the theory of storage. Journal of Business 60, 55–73.

Fama, E. F., French, K. R., 1988. Business cycles and the behavior of metals prices. Journal of Finance 43, 1075–1093.

Fama, E. F., MacBeth, J. D., 1973. Risk, return, and equilibrium: Empirical tests. Journal of Political Economy 81, 607–636.

Fama, F. E., French, K. R., 1993. Common risk factors in the returns of stocks and bonds. Journal of Financial Economics 33 (1), 3–56.

Fama, F. E., French, K. R., 2015. A five-factor asset pricing model. Journal of Financial Economics 116, 1–22.

Fung, W., Hsieh, D. A., 1997. Investment style in the returns of CTAs. Journal of Portfolio Management 24 (1), 30–41.

Galloway, T. M., Kolb, R. W., 1996. Futures prices and the maturity effect. Journal of Futures Markets 16, 809–828.

Geman, H., Nguyen, V.-N., 2005. Soybean inventory and forward curve dynamics. Management Science 51, 1076–1091.

Geman, H., Ohana, S., 2009. Forward curves, scarcity and price volatility in oil and natural gas markets. Energy Economics 31, 576–585.

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

Gibbons, M. R., Ross, S. A., Shanken, J., 1989. A test of the efficiency of a given portfolio. Econometrica 57, 1121–1152.

Gibson, R., Schwartz, E. S., 1990. Stochastic convenience yield and the pricing of oil contingent claims. Journal of Finance 45, 959–976.

Gorton, G., Hayashi, F., Rouwenhorst, K. G., 2012. The fundamentals of commodity futures returns. Review of Finance 17, 35–105.

Gorton, G., Rouwenhorst, K. G., 2006. Facts and fantasies about commodity futures. Financial Analysts Journal 62 (2), 47–68.

Gow, I. D., Ormazabal, G., Taylor, D. J., 2010. Correcting for cross-sectional and time-series dependence in accounting research. Accounting Review 85, 483–512.

Grammatikos, T., Saunders, A., 1986. Futures price variability: A test of maturity and volume effects. Journal of Business 59, 319–330.

Grauer, F. L. A., Litzenberger, R. H., 1979. The pricing of commodity futures contracts, nominal bonds and other risky assets under commodity price uncertainty. Journal of Finance 34, 69–83.

Gray, R. W., 1961. The search for a risk premium. Journal of Political Economy 69, 250–260.

Greer, R. J., 2000. The nature of commodity index returns. Journal of Alternative Investments 3 (1), 45–52.

Groot *de*, W., Karstanje, D., Zhou, W., 2014. Exploiting commodity momentum along the futures curves. Journal of Banking and Finance 48, 79–93.

Grossman, S. J., Miller, M. H., 1988. Liquidity and market structure. Journal of Finance 43, 617–633.

Hamilton, J. D., Wu, J. C., 2014. Risk premia in crude oil futures prices. Journal of International Money and Finance 42, 9–37.

Hamilton, J. D., Wu, J. C., 2015. Effects of index-funds investing on commodity futures prices. International Economic Review 56, 187–205.

Hansen, P. R., Lunde, A., Nason, J. M., 2011. The model confidence set. Econometrica 79, 453–497.

Harvey, S., Leybourne, S., Newbold, P., 1997. Testing the equality of prediction mean squared errors. International Journal of Forecasting 13, 281–291.

Haugom, E., Langeland, H., Molnar, P., Westgaard, S., 2014. Forecasting volatility of the U.S. oil market. Journal of Banking and Finance 47, 1–14.

Haussman, J. A., 1978. Specification tests in econometrics. Econometrica 46, 1251–1271.

Hazuka, T. B., 1984. Consumption betas and backwardation in commodity markets. Journal of Finance 39, 647–655.

Henderson, B. J., Pearson, N. D., Wang, L., 2015. New evidence on the financialization of commodity markets. Review of Financial Studies 28, 1285–1311.

Hicks, J. R., 1946. Value and capital: An inquiry into some fundamental principles of economic theory. In: *Value and Capital: An Inquiry into Some Fundamental Principles of Economic Theory*, Clarendon Press, Oxford.

Hirshleifer, D., 1988. Residual risk, trading costs, and commodity futures risk premia. Review of Financial Studies 1, 173–193.

Hirshleifer, D., 1989. Determinants of hedging and risk premia in commodity futures markets. Journal of Financial and Quantitative Analysis 24, 313–331.

Hirshleifer, D., 1990. Hedging pressure and futures price movements in a general equilibrium model. Econometrica 58, 411–428.

Hoechle, D., Schmid, M., Zimmermann, H., 2020. Does unobservable heterogeneity matter for portfolio-based asset pricing tests? Available at: http://dx.doi.org/10.2139/ssrn.3569485

Hong, G., Sarkar, S., 2008. Commodity betas with mean reverting output prices. Journal of Banking and Finance 32, 1286–1296.

Hong, H., 2000. A model of returns and trading in futures markets. Journal of Finance 55, 959–988.

Hong, H., Yogo, M., 2012. What does futures market interest tell us about the macroeconomy and asset prices? Journal of Financial Economics 105, 473–490.

Hou, A. J., Norden, L. L., 2018. Vix futures calendar spreads. Journal of Futures Markets 38, 822–838.

Jagannathan, R., 1985. An investigation of commodity futures prices using the consumption-based intertemporal capital asset pricing model. Journal of Finance 40, 175–191.

Jegadeesh, N., Titman, S., 1993. Returns to buying winnners and selling losers: Implications for stock market efficiency. Journal of Finance 48, 65–91.

Kaldor, N., 1939. Speculation and economic stability. Review of Economic Studies 7, 1–27.

Kang, W., Rouwenhorst, K. G., Tang, K., 2020. A tale of two premiums: The role of hedgers and speculators in commodity futures markets. Journal of Finance 75, 377–417.

Kang, W., Rouwenhorst, K. G., Tang, K., 2021. Crowding and factor returns. Available at: `http://dx.doi.org/10.2139/ssrn.3803954`

Karafiath, I., 1994. On the efficiency of least squares regression with security abnormal returns as the dependent variable. Journal of Financial and Quantitative Analysis 29, 279–300.

Kelly, B. T., Palhares, D., Pruitt, S., 2020. Modeling corporate bond returns. Available at: `http://dx.doi.org/10.2139/ssrn.3720789`

Kelly, B. T., Pruitt, S., Su, Y., 2019. Characteristics are covariances: A unified model of risk and return. Journal of Financial Economics 134, 501–524.

Keynes, J. M., 1930. A treatise on money: The applied theory of money. In: MacMillian (ed.), *A Treatise on Money: The Applied Theory of Money*, MacMillian, London, 3–46.

Khoury, N., Yourougou, P., 1993. Determinants of agricultural futures price volatilities: Evidence from Winnipeg commodity exchange. Journal of Futures Markets 13, 345–356.

Kogan, L., Livdan, D., Yaron, A., 2009. Oil futures prices in a production economy with investment constraints. Journal of Finance 64, 1345–1375.

Koijen, R. S. J., Moskowitz, T. J., Pedersen, L. H., Vrugt, E. B., 2018. Carry. Journal of Financial Economics 127, 197–225.

Kolari, J. W., Pynnonen, S., 2010. Event study testing with cross-sectional correlation of abnormal returns. Review of Financial Studies 23, 3996–4025.

Liu, L. Y., Patton, A. J., Sheppard, K., 2015. Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. Journal of Econometrics 187, 293–311.

Marshall, B. R., Nguyen, N. H., Visaltanachoti, 2012. Commodity liquidity measurement and transaction costs. Review of Financial Studies 25, 599–638.

Marshall, B. R., Nguyen, N. H., Visaltanachoti, 2013. Liquidity commonality in commodities. Journal of Banking and Finance 37, 11–20.

Martinez, V., Gupta, P., Tse, Y., Kittiakarasakun, J., 2011. Electronic versus open outcry trading in agricultural commodities futures markets. Review of Financial Economics 20, 28–36.

Masters, M. W., 2008. Testimony before the Committee on Homeland Security and Governmental Affairs United States senate, May 20. Available at: `https://www.hsgac.senate.gov/download/052008masters`

Masters, M. W., White, A. K., 2008. The accidental Hunt brothers, how institutional investors are driving up food and energy prices. Available at: `https://www.loe.org/images/content/080919/Act1.pdf`

Mayers, D., 1973. Nonmarketable assets and the determination of capital asset prices in the absence of a riskless asset. Journal of Business 46, 258–267.

McLean, R. D., Pontiff, J., 2016. Does academic research destroy stock return predictability? Journal of Finance 71, 5–32.

Merton, R. C., 1987. A simple model of capital market equilibrium with incomplete information. Journal of Finance 42, 483–510.

Miffre, J., 2016. Long-short commodity investing: A review of the literature. Journal of Commodity Markets 1, 3–13.

Miffre, J., Rallis, G., 2007. Momentum strategies in commodity futures markets. Journal of Banking and Finance 31, 1863–1886.

Milonas, N. T., 1986. Price variability and the maturity effect in futures markets. Journal of Futures Markets 6, 443–460.

Moosa, I. A., Bollen, B., 2001. Is there a maturity effect in the price of the S&P 500 futures contract? Applied Economic Letters 8, 693–695.

Moskowitz, T. J., Ooi, Y. H., Pedersen, L. H., 2012. Time series momentum. Journal of Financial Economics 104, 228–250.

Mou, Y., 2011. Limits to arbitrage and commodity index investment: Front-running the goldman roll. Available at: http://dx.doi.org/10.2139/ssrn.1716841

Nadaraya, E. A., 1964. On estimating regression. Theory of Probability and its Applications 9, 141–142.

Newey, W. K., West, K. D., 1994. Automatic lag selection in covariance matrix estimation. Review of Economic Studies 61, 631–653.

Ng, V. K., Pirrong, S. C., 1994. Fundamentals and volatility: Storage, spreads and the dynamics of metals prices. Journal of Business 67, 203–230.

Patton, A. J., 2011. Volatility forecast comparison using imperfect volatility proxies. Journal of Econometrics 160, 246–256.

Patton, A. J., Sheppard, K., 2015. Good volatility, bad volatility: Signed jumps and the persistence of volatility. Review of Economics and Statistics 97, 683–697.

Petersen, M. A., 2009. Estimating standard errors in finance panel data sets: Comparing approaches. Review of Financial Studies 22, 435–480.

Pindyck, R. S., 1994. Inventories and the short-run dynamics of commodity prices. RAND Journal of Economics 25, 141–159.

Pindyck, R. S., 2004. Volatility and commodity price dynamics. Journal of Futures Markets 24, 1029–1047.

Raman, V., Robe, M. K., Yadav, P. K., 2017. The third dimension of financialization: Electronification, intraday institutional trading, and commodity market quality. Available at: http://dx.doi.org/10.2139/ssrn.3068248–.

Rampini, A. A., Sufi, A., Viswanathan, S., 2014. Dynamic risk management. Journal of Financial Economics 111, 271–296.

Robinson, P. M., 1995. Log-periodogram regression of time series with long range dependence. Annals of Statistics 23, 1048–1072.

Roll, R., 1984. A simple implicit measure of the effective bid-ask spread in an efficient market. Journal of Finance 39, 1127–1139.

Roll, R., Ross, S. A., 1980. An empirical investigation of the arbitrage pricing theory. Journal of Finance 35, 1073–1103.

Roon *de*, F. A., Nijman, T. E., Veld, C., 2000. Hedging pressure effects in futures markets. Journal of Finance 55, 1437–1456.

Routledge, B. R., Seppi, D. J., Spatt, C. S., 2000. Equilibrium forward curves for commodities. Journal of Finance 55, 1297–1338.

Rutledge, D. J. S., 1976. A note on the variability of futures prices. Review of Economics and Statistics 58, 118–120.

Samuelson, P. A., 1965. Proof that properly anticipated prices fluctuate randomly. Industrial Management Review 6, 41–49.

Samuelson, P. A., 1976. Is real-world price a tale told by the idiot of chance? Review of Economics and Statistics 58, 120–123.

Sanders, D. R., Irwin, S. H., 2011. The impact of index funds in commodity futures markets: A systems approach. Journal of Alternative Investments 14 (1), 40–49.

Sanders, D. R., Irwin, S. H., 2013. Measuring index investment in commodity futures markets. Energy Journal 34, 105–127.

Schwartz, E. S., 1997. The stochastic behavior of commodity prices: Implications for valuation and hedging. Journal of Finance 52, 923–973.

Slyusar, V. I., 1999. A family of face products of matrices and its properties. Cybernetics and Systems Analysis 35, 379–384.

Stoll, H. R., 1979. Commodity futures and spot price determination and hedging in capital market equilibrium. Journal of Financial and Quantitative Analysis 14, 873–894.

Stoll, H. R., Whaley, R. E., 2010. Commodity index investing and commodity futures prices. Journal of Applied Finance 1, 7–46.

Szymanowska, M., Roon *de*, F., Nijman, T., Goorbergh *van den*, R., 2014. An anatomy of commodity futures risk premia. Journal of Finance 69, 453–482.

Tang, K., Xiong, W., 2012. Index investment and the financialization of commodities. Financial Analysts Journal 68 (6), 54–74.

Tufano, P., 1996. Who manages risk? an empirical examination of risk management practices in the gold mining industry. Journal of Finance 51, 1097–1137.

Watson, G. S., 1964. Smooth regression analysis. Indian Journal of Statistics: Series A 26, 359–372.

Working, H., 1933. Wheat studies. Food Research Institute 9 (6), 185–240.

Working, H., 1949. The theory of price of storage. American Economic Review 39, 1254–1262.

Yan, L., Irwin, S. H., Sanders, D. R., 2019. Is the supply curve for commodity futures contracts upward sloping? Available at: `http://dx.doi.org/10.2139/ssrn.3360787`

Yang, F., 2013. Investment shocks and the commodity basis spread. Journal of Financial Economics 110, 164–184.