DEPARTMENT OF ECONOMICS AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

Economic significance of commodity return forecasts from the fractionally cointegrated VAR model

Sepideh Dolatabadim, Paresh Kumar Narayan, Morten Ørregaard Nielsen and Ke Xu

CREATES Research Paper 2018-35

Department of Economics and Business Economics
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

Economic significance of commodity return forecasts from the fractionally cointegrated VAR model*

Sepideh Dolatabadi
Queen's University

Paresh Kumar Narayan
Deakin University

Morten Ørregaard Nielsen†
Queen's University and CREATES

Ke Xu
Queen's University

January 5, 2017

Abstract

Based on recent evidence of fractional cointegration in commodity spot and futures markets, we investigate whether a fractionally cointegrated model can provide statistically and/or economically significant forecasts of commodity returns. Specifically, we propose to model and forecast commodity spot and futures prices using a fractionally cointegrated vector autoregressive (FCVAR) model that generalizes the more well-known (non-fractional) CVAR model to allow fractional integration. We derive the best linear predictor for the FCVAR model and perform an out-of-sample forecast comparison with the non-fractional model. In our empirical analysis to daily data on 17 commodity markets, the fractional model is found to be superior in terms of in-sample fit and also out-of-sample forecasting based on statistical metrics of forecast comparison. We analyze the economic significance of the forecasts through a dynamic trading strategy based on a portfolio with weights derived from a mean-variance utility function. Although there is much heterogeneity across commodity markets, this analysis leads to statistically significant and economically meaningful profits in most markets, and shows that profits from both the fractional and non-fractional models are higher on average and statistically more significant than profits derived from a simple moving-average strategy. The analysis also shows that, in spite of the statistical advantage of the fractional model, the fractional and non-fractional models generate very similar profits with only a slight advantage to the fractional model on average.

Keywords: commodity markets, economic significance, forecasting, fractional cointegration, futures markets, price discovery, trading rule, vector error correction model.

JEL Classification: C32, G11.

1 Introduction

The forecastability of commodity market returns is a very active area of research in financial economics. In particular, recent research has shown that commodity spot and futures prices are

*We are grateful to the participants at the Conference on Recent Developments in Financial Econometrics and Applications at Deakin University, Australia (December, 2014) as well as Peter Extercate and Michał Popiel for many useful comments and suggestions. Financial support from the Canada Research Chairs program, the Social Sciences and Humanities Research Council of Canada (SSHRC), and the Center for Research in Econometric Analysis of Time Series (CREATES, funded by the Danish National Research Foundation, DNRF78) is gratefully acknowledged.

†Corresponding author. Email: mon@econ.queensu.ca

fractionally cointegrated; see, *inter alia*, Baillie and Bollerslev (1994), Lien and Tse (1999), Maynard and Phillips (2001), Coakley, Dollery, and Kellard (2011), and Dolatabadi, Nielsen, and Xu (2016). The implication is that a fractionally cointegrated model may provide a better statistical fit when modeling and forecasting commodity prices and returns. Relatedly, the understanding of how commodity market return forecasts can be used to devise trading strategies appears as of yet to be rather limited.

In this paper, we make two contributions to this literature. Our first contribution is methodological. We propose to model and forecast commodity spot and futures prices using the recently developed fractionally cointegrated vector autoregressive (FCVAR) model of Johansen (2008) and Johansen and Nielsen (2012). Specifically, we derive the best linear predictor for the FCVAR model and show that it takes a relatively simple form due to the autoregressive structure of the model. We thus demonstrate how to forecast commodity spot and futures prices and returns based on the FCVAR model, and we evaluate these using statistical measures of forecast performance. Our second contribution is to investigate the economic significance of the FCVAR model forecasts through a dynamic trading strategy based on a portfolio of two assets with portfolio weights derived from a mean-variance utility function and from return forecasts. Throughout, we compare with forecasts from the more standard (non-fractional) cointegrated vector autoregressive (CVAR) model of Johansen (1995).

We apply the FCVAR model to spot and futures prices of 17 commodities and demonstrate that it provides superior statistical in-sample fit compared with the more standard CVAR model. We also estimate price discovery from both models, see Hasbrouck (1995), Gonzalo and Granger (1995), Figuerola-Feretti and Gonzalo (2010), and Dolatabadi, Nielsen, and Xu (2015). This tells us whether price discovery is dominated by the commodity spot or futures market, which may be important from a forecasting point of view since historical information from the dominant market could be useful in forecasting prices and returns in the non-dominant market. In any case, both the FCVAR and CVAR models are joint models of the two prices series, and as such they automatically take into account the price discovery information in modeling and forecasting. With both the CVAR and FCVAR models we find that there is significant price discovery in both the spot and futures markets for many commodities, although the general tendency is that the futures market has a larger share of the price discovery process, as much theory predicts (e.g., Hasbrouck, 1995).

In our empirical analysis we consider both short horizon ($h = 1$) and longer horizon ($h = 5$ and $h = 21$) forecasting. Using a variety of out-of-sample statistical forecasting evaluation metrics, we find that the FCVAR model tends to outperform the CVAR model. Specifically, in terms of statistical tests of forecast superiority at the short horizon, these favor the FCVAR model in almost all cases and are statistically significant at standard levels in most but not all cases. At longer horizons, most statistical tests continue to favor the FCVAR model, although fewer are now statistically significant. Among those that are statistically significant for longer horizon forecasting ($h = 5$ or $h = 21$), 22 out of 23 favor the FCVAR model. Thus, the FCVAR model has superior statistical in-sample fit as well as out-of-sample forecasting performance, when considering purely statistical measures of forecast comparison.

As an additional metric of forecast performance and comparison, we also examine the economic—as opposed to purely statistical—significance of return forecasts. We do this by investigating whether the return forecasts can generate significant excess returns when implemented in a dynamic portfolio trading strategy. For our main empirical analysis we find that using return forecasts from both FCVAR and CVAR models in simple mean-variance trading strategies leads to statistically significant and economically meaningful profits in most commodity markets, , although there is much heterogeneity in profits across different markets. Furthermore, in spite of the advantage of the FCVAR model in terms of statistical measures, we find that profits are very similar on average whether based on forecasts from the FCVAR model or the CVAR model, although with a slight advantage on average to the FCVAR model profits.

Our finding that profits from commodity markets are statistically significant and economically meaningful is consistent with a broad range of studies which show, using different approaches, that commodity markets are profitable. For example, Miffre and Rallis (2007), Szakmary, Shen and Sharma (2010), and Narayan, Ahmed, and Narayan (2014) show profitability using technical trading and momentum trading strategies. However, given the profitability of these approaches, limited focus has been on using a model-based forecasting approach to estimate profits. An exception is Narayan, Narayan, and Sharma (2013), and, as in their study, we also include a brief comparison of the returns from our forecasting based approach with those from a simple technical trading rule given by a moving-average crossover strategy. Our results show that the forecasting based approach delivers higher and more statistically significant excess returns on average, as well as higher Sharpe ratios.

In spite of the limited attention to model-based forecasting approaches, there is a clear acceptance of the fact that a forecasting based trading model that draws its profitability analysis from a utility function, such as a mean-variance utility function, has theoretical appeal, see e.g. Marquering and Verbeek (2004) and Campbell and Thompson (2008). On the basis of this evidence, commodity markets are treated as an investment class. As the focus on theoretically motivated profitability analysis gains momentum, following, for example, the works mentioned above, the emphasis on and hence demand for appropriate forecasting models will increase.

We note from the outset that, although trading strategies based on commodity spot prices are not really feasible, because it would be too expensive to take possession of the commodity, we nonetheless consider simultaneous modeling of commodity spot and futures prices. In terms of applying these as forecasting models for futures returns, it has no relevance whether spot prices can be traded on or not, and hence this point is irrelevant for all our results regarding futures markets, futures price and return forecasting, and trading strategies involving commodity futures. For trading strategies involving commodity spot markets, these can still be considered a useful metric for comparison of forecast performance in terms of economic significance, even if the trading strategies are infeasible; a related point was also made in, e.g., Graham-Higgs, Rambaldi, and Davidson (1999), Wang (2000), and Narayan, Narayan, and Sharma (2013). Thus, even if portfolios involving commodity spot positions are infeasible, we consider such "artificial portfolios" as a means of forecast evaluation and comparison.

Finally, to demonstrate the robustness of our empirical results, our analysis is conducted with several different variations. First, in the forecasting models, we forecast returns over both short and long horizons. Second, we use more than one out-of-sample statistical forecast evaluation technique. Third, when estimating profits using the mean-variance investor utility function, where the choice of the investor's risk-aversion coefficient influences portfolio weights, we consider low, medium, and high risk-aversion investors. Fourth, we also calculate Sharpe ratios and compare with a simple moving-average crossover trading rule. As a final robustness analysis, when estimating profits we consider several alternative restrictions on short-selling and leverage/borrowing. In general, all these results confirm (i) that portfolio returns are statistically different from zero and economically meaningful in many commodity markets, and (ii) that portfolio returns derived from CVAR and FCVAR model forecasts are similar, although the latter are slightly higher on average.

The remainder of the paper is organized as follows. The econometric model is explained in the next section, where, in particular, the best linear predictor is derived and forecasting is discussed in Section 2.3. The following section discusses the economic equilibrium model and shows how fractional cointegration can arise from an economic model which thus provides a link between economic theory and econometric modeling. In Section 4 we discuss the commodity data and conduct some preliminary data analysis. Section 5 contains the empirical results, and is divided into subsections on estimation, statistical forecast comparison, economic significance of forecasts, and robustness of economic significance. Finally, in Section 6 we provide some concluding remarks.

2 Econometric methodology: fractionally cointegrated VAR model

Our empirical analysis applies the FCVAR model, see Johansen (2008) and Johansen and Nielsen (2012), as well as its non-fractional counterpart. The FCVAR model is a generalization of Johansen's (1995) CVAR model to allow for fractionally integrated (or just fractional) time series.

2.1 Fractional integration and cointegration

Fractional time series models are based on the fractional difference operator

$$\Delta^d X_t = \sum_{n=0}^{\infty} \pi_n(-d) X_{t-n}, \tag{1}$$

where the fractional coefficients $\pi_n(u)$ are defined in terms of the binomial expansion $(1-z)^{-u} = \sum_{n=0}^{\infty} \pi_n(u) z^n$, i.e.,

$$\pi_n(u) = \frac{u(u+1)\cdots(u+n-1)}{n!}.$$

For details and for many intermediate results regarding this expansion and the fractional coefficients, see, for example, Johansen and Nielsen (2015, Appendix A). Efficient calculation of fractional differences, which we apply in our estimation, is discussed in Jensen and Nielsen (2014).

With the definition of the fractional difference operator in (1), a time series X_t is said to be fractional of order d, denoted $X_t \in I(d)$, if $\Delta^d X_t$ is fractional of order zero, i.e., if $\Delta^d X_t \in I(0)$. The latter property can be defined in the frequency domain as having spectral density that is finite and non-zero near the origin or in terms of the linear representation coefficients if the sum of these is non-zero and finite, see, for example, Johansen and Nielsen (2012, p. 2672). An example of a process that is fractional of order zero is the stationary and invertible ARMA model. Finally, then, a p-dimensional time series $X_t \in I(d)$ for which one or more linear combinations are fractional of a lower order, i.e., for which there exists a $p \times r$ matrix β such that $\beta' X_t \in I(d-b)$ with $b > 0$, is said to be (fractionally) cointegrated.

2.2 The FCVAR model and interpretation

For a time series Y_t of dimension p, the well-known CVAR model is given in error correction form as

$$\Delta Y_t = \alpha\beta' Y_{t-1} + \sum_{i=1}^{k} \Gamma_i \Delta Y_{t-i} + \varepsilon_t = \alpha\beta' L Y_t + \sum_{i=1}^{k} \Gamma_i L^i \Delta Y_t + \varepsilon_t,$$

where, as usual, ε_t is p-dimensional independent and identically distributed with mean zero and covariance matrix Ω. The simplest way to derive the FCVAR model from the CVAR is to replace the difference and lag operators, Δ and $L = 1 - \Delta$, in (2) by their fractional counterparts, Δ^b and $L_b = 1 - \Delta^b$, respectively, and apply the resulting model to $Y_t = \Delta^{d-b} X_t$. We then obtain

$$\Delta^d X_t = \alpha\beta' \Delta^{d-b} L_b X_t + \sum_{i=1}^{k} \Gamma_i \Delta^d L_b^i X_t + \varepsilon_t,$$

where Δ^d is the fractional difference operator and $L_b = 1 - \Delta^b$ is the fractional lag operator.[1] When the so-called "restricted constant term", denoted ρ, is included, the CVAR and FCVAR models are given by

$$\Delta Y_t = \alpha(\beta' Y_{t-1} + \rho') + \sum_{i=1}^{k} \Gamma_i \Delta Y_{t-i} + \varepsilon_t \tag{2}$$

and

$$\Delta^d X_t = \alpha \Delta^{d-b} L_b(\beta' X_t + \rho') + \sum_{i=1}^{k} \Gamma_i \Delta^d L_b^i X_t + \varepsilon_t, \tag{3}$$

respectively, see Johansen and Nielsen (2012) or Dolatabadi, Nielsen, and Xu (2016).

Model (3) nests Johansen's (1995) CVAR model in (2) as the special case $d = b = 1$. Some of the parameters are well-known from the CVAR model and these have the usual interpretations also in the FCVAR model. The most important of these are the long-run parameters α and β, which are $p \times r$ matrices with $0 \leq r \leq p$. The rank r is termed the cointegration, or cofractional, rank. The columns of β constitute the r cointegration (cofractional) vectors such that $\beta' X_t$ are the cointegrating combinations of the variables in the system, i.e. the long-run equilibrium relations. The parameters in α are the adjustment or loading coefficients which represent the speed of adjustment towards equilibrium for each of the variables. The restricted constant ρ is interpreted as the mean level of the long-run equilibria when these are stationary, i.e. $E\beta' X_t + \rho' = 0$. Finally, the short-run dynamics of the variables is governed by the parameters $(\Gamma_1, \ldots, \Gamma_k)$ in the autoregressive augmentation.

The FCVAR model has two additional parameters compared with the CVAR model, namely the fractional parameters d and b. Here, d denotes the fractional integration order of the observable time series. As would presumably be the case for most—if not all—financial asset prices, we assume in our study that these are integrated of order $d = 1$. That is, we consider $d = 1$ to be fixed and known, and therefore not estimated. On the other hand, the parameter b is estimated jointly with the remaining parameters, and determines the degree of fractional cointegration, i.e. the reduction in fractional integration order of $\beta' X_t$ compared to X_t itself.

The FCVAR model (3) thus has the same main structure as the standard CVAR model (2), in that it allows for modeling of both cointegration and adjustment towards equilibrium, but is more general since it accommodates fractional integration and fractional cointegration.

We note that the fractional difference as defined in (1) is an infinite series, but any observed sample will include only a finite number of observations. This makes calculation of the fractional differences as defined in (1) impossible. In practice, therefore, the summation in (1) would need to be truncated at $n = t - 1$. This truncation is analyzed by Johansen and Nielsen (2012, 2015), who argue that the effects of the truncation can be alleviated by conditioning the (maximum likelihood) statistical analysis on a number of initial values, denoted N. Conditional inference is quite standard in autoregressive models; for example, conditional maximum likelihood estimation of an AR(k) model leads to least squares estimation, which is commonly applied. Furthermore, making the estimation conditional on a number of initial values also alleviates the effect of a non-zero starting point for the first observation on the process, i.e., for X_1. In our empirical work we follow this suggestion (setting $N = 10$) and apply the version of the FCVAR model given in (3) and the CVAR model in (2).

[1] Both the fractional difference and fractional lag operators are defined in terms of their binomial expansion in the lag operator, L, as in (1). Note that the expansion of L_b has no term in L^0 and thus only lagged disequilibrium errors appear on the right-hand side of the error correction model.

The asymptotic analysis of the FCVAR model is provided in Johansen and Nielsen (2010, 2012), where it is shown that the maximum likelihood estimator of $(b, \alpha, \Gamma_1, \ldots, \Gamma_k)$ is asymptotically normal, while the maximum likelihood estimator of (β, ρ) is asymptotically mixed normal when $b > 1/2$ and asymptotically normal when $b < 1/2$. The important implication is that standard asymptotic inference can be applied to all these parameters.

Likelihood ratio (trace-type) tests for cointegration rank can be calculated as well, and hypotheses on the cointegration rank can be tested in the same way as in the CVAR model. In the FCVAR model, the asymptotic distribution of the tests for cointegration rank depends on the unknown (true value of the) scalar parameter b, which complicates empirical analysis compared to the CVAR model. However, the distribution can be simulated on a case-by-case basis. The calculation of maximum likelihood estimators and test statistics is discussed in detail in Johansen and Nielsen (2012) and Nielsen and Popiel (2014), with the latter providing Matlab computer programs that we apply in our empirical analysis.

2.3 Forecasting from the FCVAR model

We now discuss how to forecast log-prices, that is X_t, as well as returns, $r_t = \Delta X_t$, from the FCVAR model. Because the model is autoregressive, the best linear predictor takes a simple form and is relatively straightforward to calculate. We note that $\Delta X_{t+1} = X_{t+1} - X_t$ for $t \geq 1$ and rearrange (3) with $d = 1$ as

$$X_{t+1} = X_t + \alpha \Delta^{1-b} L_b(\beta' X_{t+1} + \rho') + \sum_{i=1}^{k} \Gamma_i \Delta L_b^i X_{t+1} + \varepsilon_{t+1}. \tag{4}$$

Since $L_b = 1 - \Delta^b$ is a lag operator, so that $L_b^i X_{t+1}, i \geq 1$, is known at time t, this equation can be used to calculate forecasts from the model.

We let conditional expectation given the information set at time t be denoted $E_t(\cdot)$, and the best linear predictor of any variable Z_{t+1} given information available at time t be denoted $\hat{Z}_{t+1|t} = E_t(Z_{t+1})$. Clearly, we then have that the forecast of the innovation for period $t+1$ at time t is $\hat{\varepsilon}_{t+1|t} = E_t(\varepsilon_{t+1}) = 0$, and $\hat{X}_{t+1|t}$ is then easily found from (4). Inserting also coefficient estimates based on data available up to time t, denoted[2] $(\hat{b}, \hat{\alpha}, \hat{\beta}, \hat{\rho}, \hat{\Gamma}_1, \ldots, \hat{\Gamma}_k)$, we have that

$$\hat{X}_{t+1|t} = X_t + \hat{\alpha} \Delta^{1-\hat{b}} L_{\hat{b}}(\hat{\beta}' X_{t+1} + \hat{\rho}') + \sum_{i=1}^{k} \hat{\Gamma}_i \Delta L_{\hat{b}}^i X_{t+1}. \tag{5}$$

This defines the forecast of log-prices for period $t+1$ given information at time t. From (5) we can derive the forecast of returns as

$$\hat{r}_{t+1|t} = \hat{X}_{t+1|t} - X_t. \tag{6}$$

We note that, after constructing a series of one-step ahead log-price forecasts, $\hat{X}_{t+1|t}$ for a range of t, the return forecast (6) is different from $\Delta \hat{X}_{t+1|t}$, which may seem the obvious forecast of returns based on forecasts of log-prices, given the definition of returns as the first difference of log-prices. However, since X_t is known at time t, clearly (6) is the appropriate forecast of returns.

Multi-period ahead forecasts can be generated recursively. That is, to calculate the h-step ahead forecast, we first generalize (5) as

$$\hat{X}_{t+j|t} = \hat{X}_{t+j-1|t} + \hat{\alpha} \Delta^{1-\hat{b}} L_{\hat{b}}(\hat{\beta}' \hat{X}_{t+j|t} + \hat{\rho}') + \sum_{i=1}^{k} \hat{\Gamma}_i \Delta L_{\hat{b}}^i \hat{X}_{t+j|t}, \tag{7}$$

[2] To emphasize that these estimates are based on data available at time t, they could be denoted by a subscript t. However, to avoid cluttering the notation we omit this subscript and let it be understood in the sequel.

where $\hat{X}_{s|t} = X_s$ for $s \leq t$. Then forecasts are calculated recursively from (7) for $j = 1, 2, \ldots, h$ to generate h-step ahead forecasts of log-prices, $\hat{X}_{t+h|t}$. Given these, h-step ahead forecasts of returns are calculated as in (6) using the recursively generated log-price forecasts on the right-hand side. We will apply the return forecast (6) in our empirical analysis below for several forecast horizons, h, and we will compare with the similarly obtained return forecast based on the CVAR model (2).

3 Economic equilibrium model

The economic model for the dynamics of spot and futures prices that will provide the theoretical foundation for our empirical analysis is a variation of the equilibrium model for spot and futures prices developed by Figuerola-Feretti and Gonzalo (2010, henceforth FG), which in turn builds on Garbade and Silber (1983). The particular variation that we consider, where fractional cointegration between spot and futures log-prices can be derived from an economic model, was developed by Dolatabadi, Nielsen, and Xu (2015, 2016). We first briefly review the FG model by presenting the two cases of their model separately: (i) infinite elasticity of supply of arbitrage services and (ii) finite elasticity of supply of arbitrage services. In the third subsection we then discuss the Dolatabadi, Nielsen, and Xu (2015, 2016) variation that will establish a natural connection to the FCVAR model described in Section 2 above.

3.1 FG equilibrium model with infinite elasticity of supply of arbitrage services

We begin with the following set of standard market conditions, which are collectively referred to as Assumption A.

- A.1 No taxes or transaction costs.
- A.2 No limitations on borrowing.
- A.3 No costs other than financing a futures position (short or long) and storage costs.
- A.4 No limitations on short sale in the spot market.

We denote the log-spot price of a commodity in period t by s_t and the contemporaneous log-futures price for a one-period-ahead futures contract by f_t, while r_{ft} and c_t denote the continuously compounded interest rate and storage cost, respectively, for period t. The time series behavior of these variables is described in the following conditions, which are collectively referred to as Assumption B.

- B.1 $r_{ft} = \bar{r}_f + u_{rt}$, where $\bar{r}_f$ denotes the mean of r_{ft} and u_{rt} denotes an I(0) process with mean zero and finite positive variance.
- B.2 $c_t = \bar{c} + u_{ct}$, where $\bar{c}$ denotes the mean of c_t and u_{ct} denotes an I(0) process with mean zero and finite positive variance.
- B.3 Δs_t is an I(0) process with mean zero and finite positive variance.

Under Assumption A, no-arbitrage equilibrium conditions imply

$$f_t = s_t + r_{ft} + c_t, \tag{8}$$

so that, imposing also Assumption B,

$$f_t - s_t = \bar{r}_f + \bar{c} + u_{rt} + u_{ct},$$

which implies that s_t and f_t are both I(1) and cointegrate to I(0) with cointegration vector $(1, -1)$.

3.2 FG equilibrium model with finite elasticity of supply of arbitrage services

Finite elasticity of the supply of arbitrage services reflects the existence of factors such as basis risk, convenience yields, constraints on storage space and other factors that make arbitrage transactions risky. FG focus on convenience yield, in particular, which is the benefit associated with storing the commodity instead of holding the futures contract (Kaldor, 1939). A general defintion of convenience yield due to Brennan and Schwartz (1985) is "the flow of services that accrues to an owner of the physical commodity but not to an owner of a contract for future delivery of the commodity". Accordingly, FG then give backwardation an economic interpretation as "the present value of the marginal convenience yield of the commodity inventory". When this is negative, the market is said to be in contango.

With the convenience yield denoted by y_t, the no-arbitrage condition (8) is then modified to

$$f_t + y_t = s_t + r_{ft} + c_t. \tag{9}$$

It is commonplace to characterize convenience yield as a (linear or nonlinear) function of s_t and f_t. In particular, FG approximate y_t by a linear combination of s_t and f_t, i.e. $y_t = \gamma_1 s_t - \gamma_2 f_t$ with $\gamma_i \in (0,1)$ for $i = 1,2$. Imposing Assumption B then implies the equilibrium condition

$$s_t + \beta_2 f_t + \rho = u_{rt} + u_{ct}, \tag{10}$$

where β_2 and ρ are simple functions of the model parameters. In particular, β_2 can take three different values (with the interpretations assuming a small enough value of ρ):

(i) $-\beta_2 > 1$: long-run backwardation ($s_t > f_t$).
(ii) $-\beta_2 < 1$: long-run contango ($s_t < f_t$).
(iii) $-\beta_2 = 1$: neither backwardation nor contango in the long run.

Note that the equilibrium condition (10) is often stated with only s_t on the left-hand side, which is the reason why we interpret the cointegration coefficient in terms of $-\beta_2$. However, the econometric model specifies the equilibrium in terms of $\beta' X_t + \rho$ and to unify our notation, we have specified the equilibrium (10) in the same way.

Thus, the equilibrium model of FG admits the (empirically warranted) theoretical possibility of having a cointegration coefficient $-\beta_2$ different from unity. We next describe Dolatabadi, Nielsen, and Xu's (2015, 2016) variation of the economic model that will link it to the econometric FCVAR model.

3.3 Fractionally cointegrated equilibrium model

The above analysis makes it clear that the I(0) term, $u_{rt} + u_{ct}$, in the equilibrium (cointegrating) relationship (10) stems from Assumptions B.1 and B.2, where interest rates and storage costs are assumed to be I(0). While storage costs are basically unobserved, interest rates are observed and are typically not found to be I(0).

To obtain a model with fractional cointegration, we therefore replace Assumption B by the following conditions, which are collectively referred to as Assumption C.

C.1 $r_{ft} = \bar{r}_f + v_{rt}$, where $\bar{r}_f$ denotes the mean of r_{ft} and v_{rt} denotes an I($1-b$) process with $b > 1/2$, mean zero, and finite positive variance.

C.2 $c_t = \bar{c} + v_{ct}$, where $\bar{c}$ denotes the mean of c_t and v_{ct} denotes an I($1-b$) process with $b > 1/2$, mean zero, and finite positive variance.

C.3 Δs_t is an I(0) process with mean zero and finite positive variance.

Here, Assumptions C.1 and C.2 generalize B.1 and B.2 to fractionally integrated interest rates and storage costs.

To simplify notation we assume that interest rates and storage costs have the same order of fractional integration, i.e. that both v_{rt} and v_{ct} are I$(1-b)$. Also, the assumption that $b > 1/2$ ensures that the processes v_{rt} and v_{ct} are stationary, since then $1 - b < 1/2$. Neither of these assumptions are critical, nor even necessary for the economic equilibrium model, but they facilitate interpretation of the parameters in the FCVAR model. For example, with two different fractional integration orders of v_{rt} and v_{ct}, the sum $v_{rt} + v_{ct}$ would be fractional of the highest of the two orders. Furthermore, if $b < 1/2$, then v_{rt} would not be stationary and in that case we would define $\bar{r}_f$ simply as a constant, rather than interpret it as the mean of r_{ft}, and v_{rt} would denote an I$(1-b)$ process initialized at zero. Similarly for v_{ct}.

We now impose Assumption C on (9) instead of Assumption B, which results in the equilibrium condition

$$s_t + \beta_2 f_t + \rho = v_{rt} + v_{ct}. \tag{11}$$

Hence, replacing Assumption B in the FG model with Assumption C implies the same cointegration vector, but the equilibrium condition differs from that in the FG model in that the long-run equilibrium errors are fractionally integrated of order $1 - b$ rather than I(0). More generally, it follows that s_t and f_t are fractionally cointegrated such that the FCVAR model of Section 2 is directly applicable to this economic model.

3.4 Price discovery

We now briefly review how to analyze price discovery within the FCVAR model based on the discussion in Dolatabadi, Nielsen, and Xu (2015). The analysis applies the permanent-transitory (PT) decomposition of Gonzalo and Granger (1985) to the FCVAR model. As described in detail in FG, there is "a perfect link between an extended Garbade and Silber (1983) theoretical model and the PT decomposition".

In the notation of the previous subsections, we let $X_t = (s_t, f_t)'$, where s_t and f_t denote the log-spot and log-futures prices at time t, respectively. According to the PT decomposition, X_t may be decomposed into a transitory (stationary) part, $\beta' X_t$, and a permanent part, $W_t = \alpha_\perp' X_t$, using the identity

$$\begin{aligned} X_t &= (\beta_\perp(\alpha_\perp'\beta_\perp)^{-1}\alpha_\perp' + \alpha(\beta'\alpha)^{-1}\beta')X_t \\ &= A_1 W_t + A_2 \beta' X_t, \end{aligned}$$

where $\alpha_\perp$ is such that $\alpha_\perp'\alpha = \alpha'\alpha_\perp = 0$.

Here, W_t is the common permanent component of X_t. In the case of spot and futures log-prices, W_t is interpreted as the long-run dominant (fundamental or efficient) market price, in the sense that information that does not affect W_t will not have a permanent effect on X_t. Thus, the proportions of price discovery attributable to each market may be inferred from the elements of the parameter $\alpha_\perp$, after being normalized so that the elements sum to unity. For further details, we refer the reader to Gonzalo and Granger (1995), FG, and Dolatabadi, Nielsen, and Xu (2015).

An alternative, yet strongly related, interpretation of the coefficient α is that of an adjustment coefficient that measures how the disequilibrium errors in previous periods feed into today's changes in X_t. Under this interpretation, the natural question to ask about the adjustment coefficients is whether some coefficients in α are zeros, in which case the variable in question is weakly (or long-run) exogenous for the parameters α and β. For example, if $\alpha_2 = 0$, futures prices do not react to the disequilibrium error, i.e. the transitory component, implying that futures prices are the main contributors to price discovery.

Table 1: Data description for commodity spot and futures prices

Commodity	CRB	Spot	Futures	Start	Volume	Open int.	Vol. 2012	O.I. 2012
Canola	WC	Vancouver	WCE	3/30/83	6255	53,266	18,649	209,368
Cocoa	CC	New York	ICE	3/30/83	5441	53,443	23,763	177,720
Coffee	KC	New York	ICE	3/30/83	8337	47,609	24,387	139,664
Copper	HG	New York	NYMEX	3/30/83	12,627	68,929	68,716	147,796
Corn	C-	Chicago	CBOT	3/30/83	61,239	320,828	322,417	1,227,418
Crude oil	CL	DJES	NYMEX	3/30/83	202,147	573,095	582,596	1,477,764
Gasoline	RB	DOE	NYMEX	12/3/84	43,308	116,930	148,492	313,488
Gold	GC	Composite	NYMEX	3/30/83	52,332	198,512	179,622	421,913
Heating oil	HO	DJES	NYMEX	3/30/83	38,857	129,577	141,131	303,555
Palladium	PA	New York	NYMEX	3/30/83	899	7531	4353	21,528
Platinum	PL	Engelhard	NYMEX	7/29/85	2697	14,869	9040	46,519
Silver	SI	Composite	NYMEX	3/30/83	20,017	101,307	53,797	115,476
Soybean	S-	Central IL	CBOT	3/30/83	50,882	165,613	211,884	692,521
Soy meal	SM	Decatur, IL	CBOT	3/30/83	18,142	75,344	75,682	236,697
Soy oil	BO	Decatur, IL	CBOT	3/30/83	20,198	91,680	108,576	344,871
Sugar	SB	New York	ICE	3/30/83	23,694	166,924	109,728	699,786
Wheat	W-	St. Louis	CBOT	3/30/83	21,605	102,263	117,191	445,275

Notes: This table provides information on the data sources and commodity markets for the spot and futures prices. The first four columns show the CRB identifier symbol, spot data source, futures data source, and sample start date for each commodity. The end date is 10/12/12 for all data series. The next two columns show the daily volume (number of contracts) and open interest averaged over the entire sample period, while the last two columns show the daily volume and open interest averaged over the 2012 part of the sample.

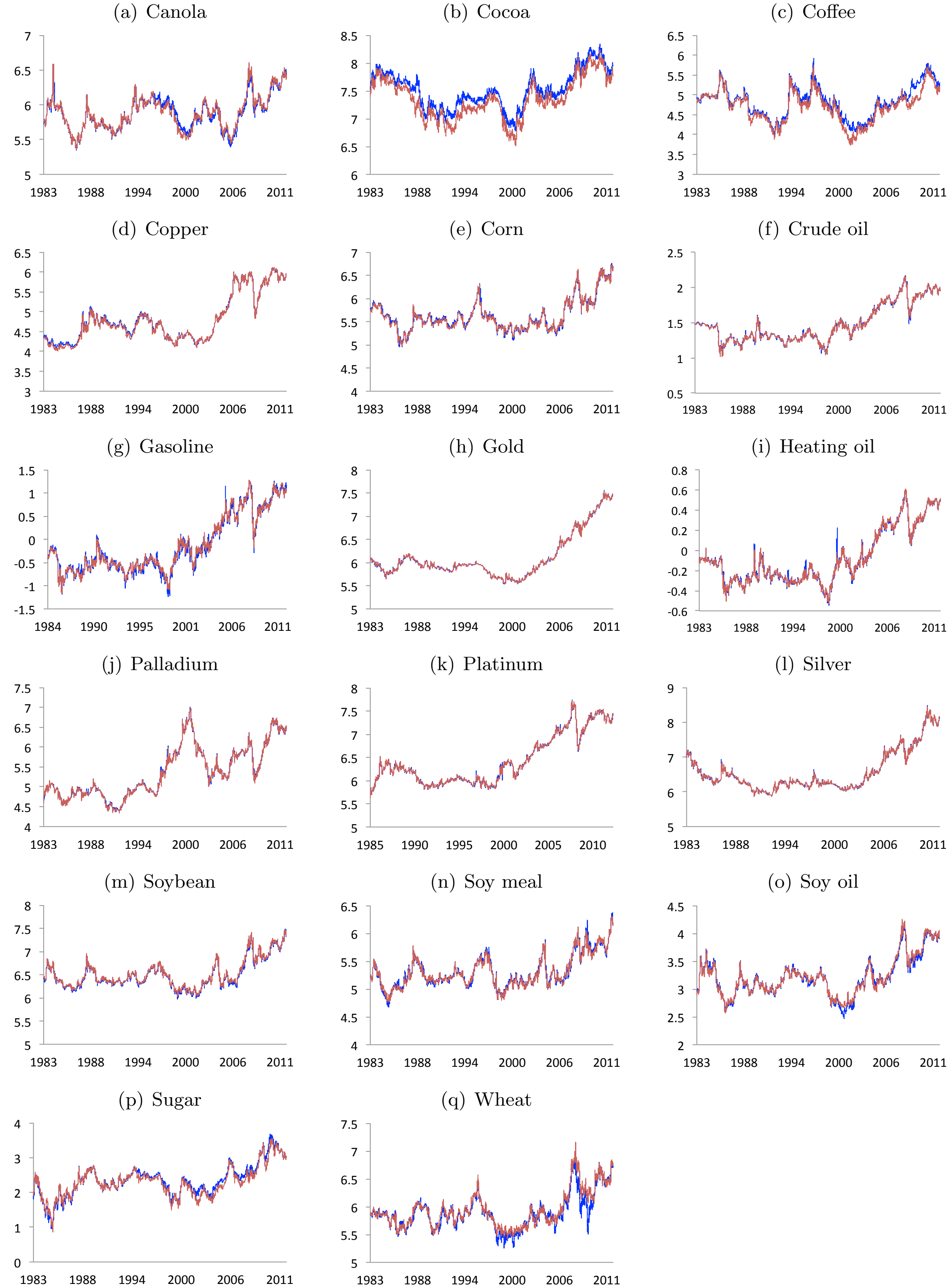
4 Data description and preliminary analysis

In our empirical analysis we have data on 17 commodity spot and futures markets. These are canola, cocoa, coffee, copper, corn, crude oil, gasoline, gold, heating oil, palladium, platinum, silver, soybean, soybean meal, soybean oil, sugar, and wheat. The spot and futures price series are both closing prices and are obtained from the Commodity Research Bureau (CRB) database. The same data set (except gasoline and heating oil) was used by Narayan, Ahmed and Narayan (2014) to analyze momentum-based trading strategies in commodity futures markets.

Some facts about the data sources and commodity markets are collected in Table 1. First of all, the second column of Table 1 lists the CRB commodity identifier symbol. Secondly, we note that the spot and futures price series come from different sources/exchanges. These are listed in the third and fourth columns of the table for the spot prices and futures prices, respectively. The next column shows the start date for our sample period, which varies by commodity although for most commodities it is March 30, 1983. For all commodities, the end date is October 12, 2012. This gives a total of 7708 observations, except for gasoline (7270 observations) and platinum (7100 observations). Our choice of commodities is determined by the availability of time-series data for both spot and futures prices in the CRB database. There are several commodities for which either spot or futures prices are unavailable or only available for a short time period (natural gas, for example, only has futures price data starting in 1990 and spot prices in 1993). For other commodities (cotton and orange juice, for example) there are many months or even years of missing data.

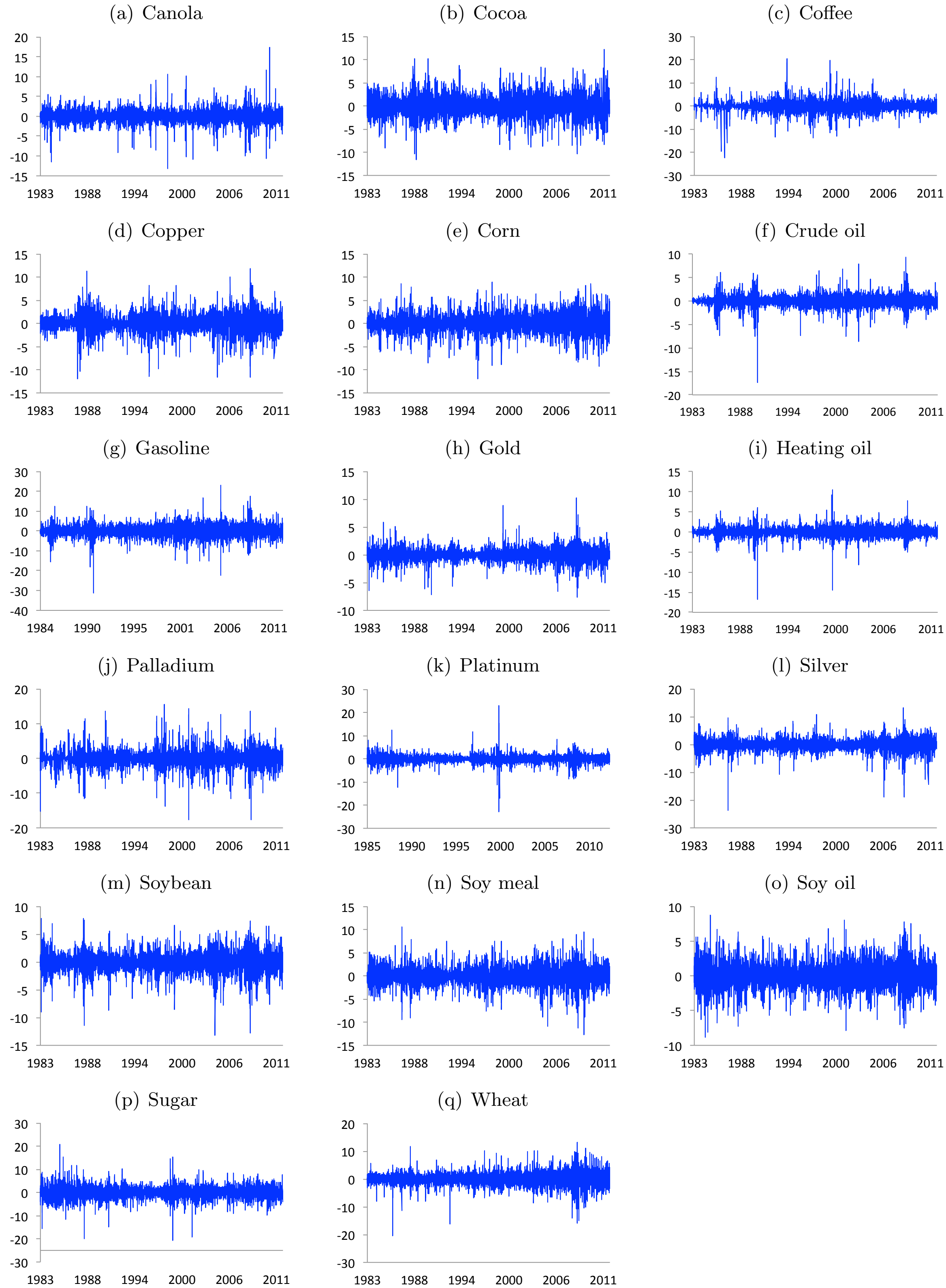
Some further characteristics of the commodity futures markets by way of volume and open

Figure 1: Daily commodity log-prices



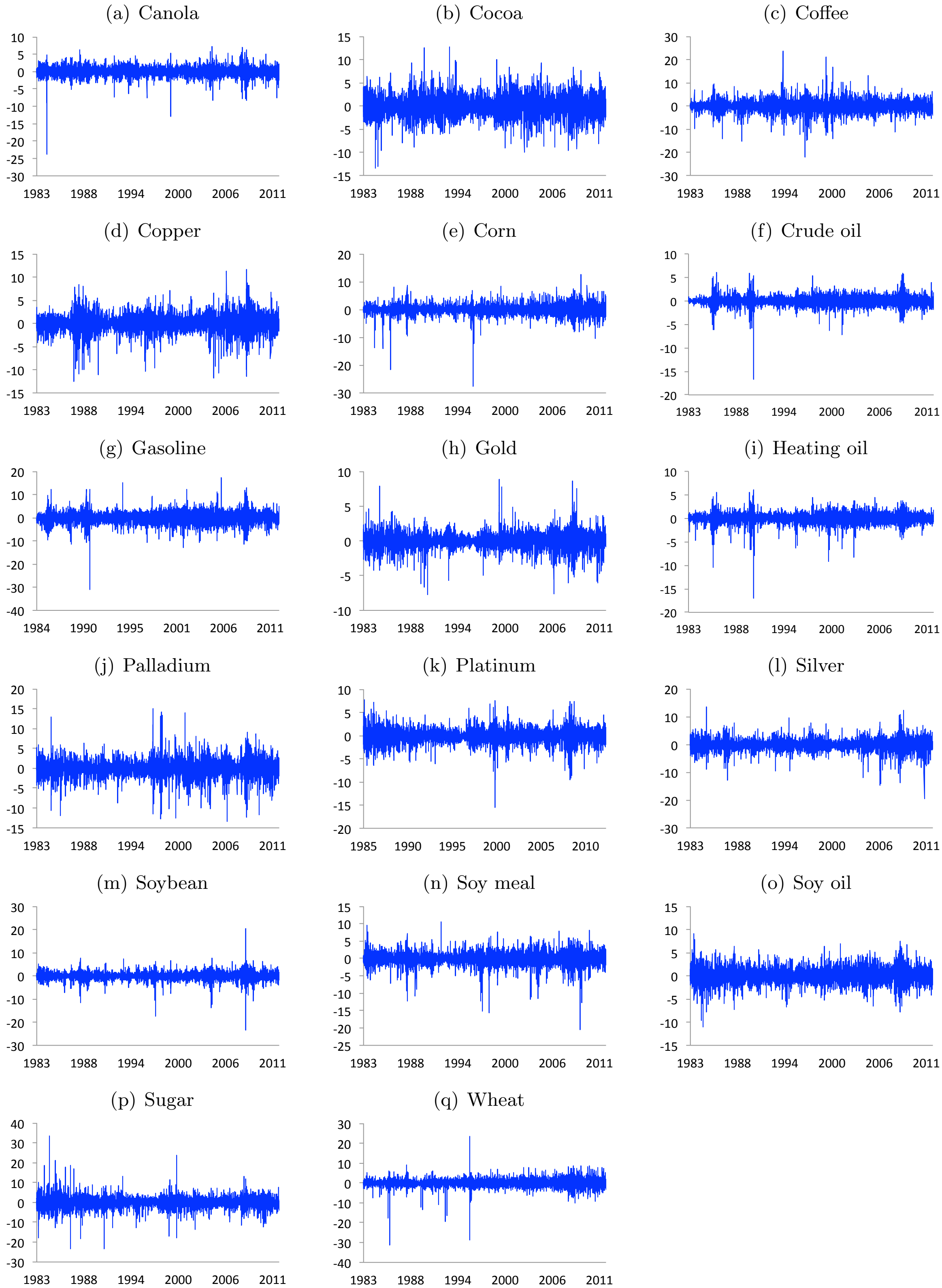
Notes: Each plot shows daily commodity log-prices. The blue lines are spot prices and the red lines are futures prices. The sample start dates vary by commodity, see Table 1, and the end date is 10/12/12 for all commodities.

Figure 2: Daily commodity spot returns



Notes: Each plot shows daily commodity spot returns in percentage. The sample start dates vary by commodity, see Table 1, and the end date is 10/12/12 for all commodities.

Figure 3: Daily commodity futures returns



Notes: Each plot shows daily commodity futures returns in percentage. The sample start dates vary by commodity, see Table 1, and the end date is 10/12/12 for all commodities.

interest are presented in the next two columns of Table 1. These statistics are daily time-series averages for each commodity, where average volume reflects the number of futures contracts traded each day and average open interest reflects the total number of outstanding futures contracts held by market participants each day. In other words, open interest can be used to gauge the liquidity situation of the futures market. The higher the number of open interest contracts the larger the market activity and hence liquidity. The volume of contracts is largest for crude oil at an average of over 200,000 contracts per day over the full sample period and over 580,000 contracts per day in 2012. Based on trading volume, crude oil makes up over one-third of all commodity contracts traded in the market. Corn, gold, and soybean contracts each make up 8–10% of the market, while heating oil and gasoline make up another approximately 7% of the market each. The rest of the commodities each contribute less than 5% to total volume of trade in the commodity futures markets. Similarly, open interest data suggest that roughly a quarter of all outstanding contracts belong to the crude oil market, followed by corn (14%), soybean (9%), and gold (8%).

The commodity spot and futures log-price series and corresponding spot and futures returns are plotted in Figures 1–3. Returns are computed as the first difference of the log-price series, and the displayed returns series in Figures 2 and 3 are multiplied by 100 to yield a (continuously compounded) percentage return. Three tendencies emerge from the figures. First, the log-price series do not appear to have obvious time trends in Figure 1. This finding was supported by statistical tests in the estimation of our models below, where any trend included was statistically insignificant and therefore removed (not reported). Second, the spot and futures log-price series appear to move together in the long-run, supporting the notion that they are cointegrated. Third, there is clearly heterogeneity among the commodities. For example, the variance of returns in Figures 2 and 3 varies substantially across commodities.

In Table 2 we present some descriptive statistics for each of the commodity spot and futures returns series. Again, returns are multiplied by 100 to yield a continuously compounded percentage return. The statistics in Table 2 confirm the tendencies observed in the figures. The sample mean returns for the spot market vary from 0.004% to 0.031% per day and in the futures market from 0.003% to 0.030% per day. A similar disparity in sample standard deviation, skewness, and kurtosis is found. The implication here is that these specific commodities can potentially offer investors quite different risk-return trade-offs when considered from an investment portfolio point of view. The last column in the table reports the first-order autocorrelation coefficient for each series. These are all quite small, ranging from -0.085 to 0.095, but several are in fact statistically significant due to the large sample size. Nonetheless, the small autocorrelation coefficients suggest that all returns are clearly stationary $I(0)$ processes, thus confirming our modeling choice of fixing $d = 1$ in the FCVAR analysis, but also suggest that past returns alone will likely not be very good predictors of returns in the future.

5 Empirical results and economic significance

This section has three parts. In the first part, we present estimation results for the FCVAR and CVAR models based on the first 75% of the sample. We use a relatively large fraction of the total sample for estimation because the fractional models tend to require large sample sizes for reliable estimation. In the second part, we present and discuss results for out-of-sample forecasting for the remaining part of the sample, based on statistical measures of forecast accuracy and comparison. The third part of the results is about the economic significance of return forecasts. In other words, this is where we evaluate the forecasting models by asking: how beneficial are these forecasting models to investors?

Table 2: Selected descriptive statistics of daily commodity spot and futures returns

	Spot market returns					Futures market returns				
Commodity	Mean	S.d.	Skew.	Kurt.	AC	Mean	S.d.	Skew.	Kurt.	AC
Canola	0.009	1.437	−0.282	13.968	0.002	0.009	1.295	−1.169	21.957	0.095
Cocoa	0.005	1.784	−0.018	6.625	−0.085	0.004	1.929	−0.007	5.962	0.004
Coffee	0.004	1.931	−0.042	17.243	0.003	0.003	2.279	0.012	11.817	0.000
Copper	0.020	1.656	−0.294	8.725	−0.039	0.021	1.722	−0.426	8.035	−0.032
Corn	0.011	1.664	−0.139	6.546	0.034	0.012	1.659	−1.003	21.212	0.053
Crude oil	0.006	1.052	−0.792	19.979	−0.044	0.006	0.939	−0.877	20.787	−0.010
Gasoline	0.020	2.647	−0.350	10.526	0.014	0.018	2.255	−0.457	11.274	0.015
Gold	0.019	0.995	−0.131	10.097	−0.029	0.019	1.003	−0.103	10.153	−0.006
Heating oil	0.008	1.047	−0.812	22.937	0.043	0.008	0.997	−1.364	21.283	−0.006
Palladium	0.022	1.958	−0.106	12.200	0.019	0.024	1.961	−0.234	8.802	0.082
Platinum	0.025	1.502	−0.331	26.744	−0.035	0.025	1.405	−0.584	9.091	0.040
Silver	0.015	1.794	−1.063	15.324	−0.016	0.015	1.802	−0.697	10.896	−0.009
Soybean	0.012	1.503	−0.568	7.950	−0.009	0.011	1.539	−0.934	20.509	0.018
Soy meal	0.012	1.699	−0.200	6.556	0.030	0.012	1.678	−0.904	12.272	0.056
Soy oil	0.013	1.570	0.020	5.137	0.020	0.013	1.509	−0.047	5.661	0.058
Sugar	0.015	2.231	−0.133	9.567	−0.039	0.014	2.514	0.218	15.249	−0.044
Wheat	0.011	2.049	−0.273	9.052	−0.019	0.011	1.857	−1.217	29.765	−0.021

Notes: This table reports selected descriptive statistics for the 17 commodity spot and futures return series. Specifically, the table reports the sample mean, standard deviation, skewness, kurtosis, and the first-order sample autocorrelation of returns (all given in terms of percentage returns).

5.1 Estimation results

Before we can estimate the FCVAR model and apply it in forecasting, we have to make some model selection choices. First, as discussed in Section 2 above, we apply estimation conditional on $N = 10$ initial values for all our FCVAR results, corresponding to conditioning on the first two weeks of observations. Experimentation with different values of N showed little effect. For the CVAR model we applied estimation conditional on $k+1$ initial values, such that maximum likelihood estimation is reduced rank regression (Johansen, 1995). Second, we have to specify the lag length, k, in the vector error correction models (2) and (3). For the CVAR model we select the lag length to minimize the Bayesian Information Criterion (BIC) based on the model that has full rank $r = p$, where p is the dimension of the system.[3] For the FCVAR model, we apply several different statistics to select the lag length, namely the BIC, the LR test statistic for significance of Γ_k, and univariate Ljung-Box Q tests (with $m = 10$ lags) for each of the two residual series, in each case based on the model that has full rank $r = p$. In addition, we examined the unrestricted estimates of b which, when the lag length is misspecified, will sometimes be very far from what should be expected. Specifically, due to a non-identification issue in the FCVAR model with misspecified lag length, it is sometimes found that, e.g., $\hat{b} = 0.05$ or similar, see Johansen and Nielsen (2010, Section 2.3) for a theoretical discussion of this phenomenon. For each commodity, we use the BIC as an initial rough guide to choose the lag length, and starting from there we find the nearest lag length which satisfies the criteria (i) Γ_k is significant based on the LR test, (ii) the unrestricted estimate of b is reasonable

[3]When calculating the BIC for different values of k, we use $N = 10$ initial values for all k to have the same effective sample size—and hence facilitate comparison—across different values of k.

(very widely interpreted), and (iii) the Ljung-Box Q tests for serial correlation in the two residual series do not show signs of misspecification. Third, after choosing the number of lags, we select the cointegrating rank, r, by sequentially testing (using the LR trace statistic) the hypotheses $r = 0, 1, 2$ until rejection, choosing the last non-rejected hypothesis to be the cointegration rank. The critical values for the rank tests are simulated case-by-case for the FCVAR model, and for the CVAR model we used Johansen (1995, Table 15.2).

Table 3 reports results[4] from estimation of CVAR and FCVAR models for commodity spot and futures log-prices, i.e. with $X_t = (s_t, f_t)'$ in the notation of Section 2. For the estimation we use only the first 75% of the sample, and reserve the remainder for out-of-sample forecasting. This leaves $T = 5781$ observations in the estimation sample, except for gasoline ($T = 5452$) and platinum ($T = 5325$).

First of all, the second column of Table 3 shows the chosen lag-order (k) for each commodity. It is clear that fewer lags are usually needed in the lag-augmentation for the FCVAR model compared with the CVAR model (only crude oil has more lags in the FCVAR model specification). This is expected since the FCVAR model includes the additional parameter b to accommodate serial dependence. Although the FCVAR model in this way includes one additional parameter (b), each additional lag included in the CVAR model requires four additional parameters (Γ_i) to be estimated, and hence the CVAR model in most cases includes a larger number of parameters than the FCVAR model.

The third column of Table 3 shows the estimated fractional parameter, $\hat{b}$. The point estimates $\hat{b}$ range from 0.194 to 0.955, showing a wide variety of fractional cointegration properties across the different commodities. Relating the estimates of b to the theoretical model in Section 3.3, we note that the heterogeneity in these estimates derives from heterogeneity in the storage cost equation in Assumption C.2, since the interest rate in Assumption C.1 presumably will be the same for different commodities. As discussed in Section 2 above, the CVAR model is nested within the FCVAR model by imposing the hypothesis $b = 1$. Thus, we may test the CVAR model against the more general FCVAR model by testing the restriction $b = 1$. In Table 3, we use one, two, and three asterisks on the estimates $\hat{b}$ to denote when the fractional parameter is significantly different from unity at the 10%, 5%, and 1% level, respectively. From these tests we note that the CVAR model is rejected against the FCVAR model for 15 of 17 commodities at the 1% level, and for one additional commodity at the 10% level. Only palladium appears to be well-described by a CVAR model in terms of in-sample fit when judged by this statistical test. Thus, the FCVAR model provides a better statistical in-sample fit in most cases.[5]

The next six columns of Table 3 report estimates of the cointegration coefficient, $-\hat{\beta}_2$, the restricted constant term, $-\hat{\rho}$, the adjustment coefficients, $\hat{\alpha}_1$ and $\hat{\alpha}_2$, and the price discovery coefficients, $\hat{\alpha}_{\perp,1}$ and $\hat{\alpha}_{\perp,2}$. The estimates $-\hat{\beta}_2$ of the cointegration coefficients are close to unity, as expected from an efficient markets hypothesis point of view, although they may deviate from unity still. The latter reflects a market in long-run backwardation (when $-\beta_2 > 1$) or contango (when $-\beta_2 < 1$), see Section 3 above. Generally, the estimates $-\hat{\beta}_2$ suggest that backwardation is more common across markets. This is especially the case for estimates from the FCVAR model, which indicate backwardation in 14 of 17 commodity markets, with the remaining three having estimates $-\hat{\beta}_2$ very close to unity. As discussed in the economic equilbrium model in Section 3, the extent of backwardation or contango is related to the convenience yield, which is expected to differ across different commodities.

[4]Full estimation results are available from the authors upon request.

[5]We note that the log-likelihood—and hence the BIC—for the CVAR and FCVAR models are based on different effective sample sizes, even for the same commodity, because the number of initial values are different. Therefore, these measures cannot be used directly as a means of statistical comparison of the CVAR and FCVAR in-sample fit.

Table 3: Estimation results for CVAR and FCVAR models of commodity prices

Commodity	k	$\hat{b}$	$-\hat{\beta}_2$	$-\hat{\rho}$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_{\perp,1}$	$\hat{\alpha}_{\perp,2}$
Panel A: CVAR model								
Canola	3	1	1.036	−0.199	−0.019	0.000	−0.002	1.002***
Cocoa	3	1	0.916	0.777	0.000	0.012	0.982**	0.018
Coffee	2	1	0.956	0.303	−0.018	0.004	0.171	0.829***
Copper	5	1	0.993	0.060	−0.025	0.004	0.143	0.857***
Corn	1	1	1.057	−0.298	−0.034	0.015	0.305***	0.695***
Crude oil	1	1	1.006	−0.007	−0.247	0.035	0.124**	0.876***
Gasoline	1	1	0.993	−0.015	−0.046	−0.011	−0.329**	1.329***
Gold	2	1	1.000	−0.003	−0.506	0.148	0.226***	0.774***
Heating oil	6	1	1.005	0.004	−0.075	−0.022	−0.414**	1.414***
Palladium	2	1	1.002	−0.002	−0.057	0.037	0.392***	0.608***
Platinum	7	1	0.999	0.006	−0.076	0.107	0.586***	0.414***
Silver	3	1	0.993	0.042	−0.352	0.106	0.232***	0.768***
Soybean	1	1	1.029	−0.201	−0.036	0.061	0.628***	0.372***
Soy meal	3	1	1.042	−0.211	−0.019	0.007	0.255	0.745***
Soy oil	4	1	1.179	−0.552	0.001	0.012	1.062**	−0.062
Sugar	3	1	1.119	−0.219	−0.002	0.011	0.862***	0.138
Wheat	1	1	1.110	−0.652	−0.012	0.010	0.471***	0.529***
Panel B: FCVAR model								
Canola	2	0.776***	1.031	−0.177	−0.063	−0.004	−0.066	1.066***
Cocoa	2	0.752***	0.966	0.407	0.014	0.043	1.502**	−0.502
Coffee	1	0.528***	1.034	−0.131	−0.116	0.091	0.441**	0.559***
Copper	0	0.341***	1.053	−0.194	−0.920	0.056	0.057	0.943***
Corn	1	0.923*	1.057	−0.297	−0.044	0.021	0.325***	0.675***
Crude oil	3	0.194***	1.007	−0.008	−44.874	2.691	0.057	0.943***
Gasoline	1	0.735***	1.027	0.005	−0.125	−0.027	−0.270*	1.270***
Gold	1	0.795***	1.002	−0.015	−0.878	0.181	0.171***	0.829***
Heating oil	4	0.694***	1.014	0.006	−0.206	−0.055	−0.364	1.364***
Palladium	2	0.955	1.002	0.000	−0.066	0.043	0.392***	0.608***
Platinum	1	0.664***	0.996	0.022	−0.507	0.213	0.296***	0.704***
Silver	1	0.672***	0.998	0.011	−1.174	−0.027	−0.023	1.023***
Soybean	1	0.872***	1.028	−0.198	−0.050	0.098	0.661***	0.339***
Soy meal	1	0.518***	1.138	−0.700	−0.115	0.103	0.472**	0.528***
Soy oil	2	0.488***	1.294	−0.891	−0.041	0.046	0.529	0.471
Sugar	3	0.606***	1.090	−0.188	−0.003	0.086	0.969**	0.031
Wheat	0	0.717***	1.140	−0.820	−0.027	0.048	0.640***	0.360***

Notes: This table reports estimation results for CVAR (Panel A) and FCVAR (Panel B) models applied to the first 75% of the sample of commodity spot and futures log-prices. The columns include lag-order, k, estimates of the fractional parameter, $\hat{b}$, the cointegration coefficient, $-\hat{\beta}_2$, the restricted constant term, $-\hat{\rho}$, the adjustment coefficients, $\hat{\alpha}_1$ and $\hat{\alpha}_2$, and the price discovery coefficients, $\hat{\alpha}_{\perp,1}$ and $\hat{\alpha}_{\perp,2}$. The latter are normalized to add to unity. For the fractional parameter we let one, two, and three asterisks denote significant difference from unity at the 10%, 5%, and 1% level, respectively, and for the price discovery coefficients, asterisks denote significant difference from zero.

As discussed briefly in Sections 2.2 and 3.4, the adjustment coefficients α_1 and α_2 determine the speed of adjustment towards equilibrium for the two variables when the system is in disequilibrium. In our model, $\alpha_1 < 0$ and $\alpha_2 > 0$ imply adjustment towards equilibrium, with the other sign indicating adjustment away from equilibrium. For example, we note that the estimates for gasoline suggest that the futures prices do not adjust towards equilibrium (and in fact adjust away from equilibrium, although this is only mildly significant, see below), whereas spot prices do adjust towards equilibrium. This behavior is found only for a few commodities. Note also that the system can still exhibit overall adjustment towards the equilibrium, even if one price series does not adjust towards equilibrium. This happens, of course, if the price series that does adjust towards equilibrium does so at a faster rate than the other price series moves away from equilibrium. For most commodities, however, both prices adjust towards equilibrium but at different speeds.

The final two columns for each model show the price discovery coefficients, $\hat{\alpha}_{\perp,1}$ and $\hat{\alpha}_{\perp,2}$, normalized to add to unity. The first is the proportion of price discovery in the spot market, and the second is the proportion of price discovery in the futures market. For the price discovery coefficients we test the hypotheses that they are zero and let one, two, and three asterisks denote significance at the 10%, 5%, and 1% level, respectively.[6] Thus, to test whether, for example, the futures market is dominant in the sense that price discovery takes place exclusively in the futures market, one would test the equivalent hypothesis that there is no price discovery in the spot market, which is formulated as $\alpha_{\perp,1} = 0$. We note that, corresponding to those α coefficients that do not indicate adjustment towards equilibrium, there are a few negative $\hat{\alpha}_{\perp,i}$ coefficients. However, only two are significantly negative in the CVAR results and only one is significantly negative (and only at the 10% level) in the FCVAR results.

It is seen that, according to point estimates of the price discovery coefficients from both the CVAR and FCVAR models, the futures market dominates price discovery for most commodities, as expected from theory (e.g., Hasbrouck, 1995), with average futures market price discovery coefficients of 0.665 and 0.659 from the CVAR and FCVAR models, respectively. However, in general there is significant price discovery taking place in both the spot and futures markets for many commodities.

In particular, according to the CVAR model, the spot market is dominant in the price discovery process for three commodities (cocoa, soy oil, and sugar), while the futures market is dominant for six commodities (canola, coffee, copper, gasoline, heating oil, and soy meal). We include gasoline and heating oil in this list because their price discovery coefficients for the spot market are negative, even though they are significant at the 5% level.

From the FCVAR model, the conclusions are similar. Specifically, the spot market is dominant in the price discovery process for two commodities (cocoa and sugar). One possible reason for the empirically observed spot market dominance could be the fall in futures trading during and after the recent crisis, where futures trading volume shrank substantially. On the other hand, the futures market is dominant for six commodities (canola, copper, crude oil, gasoline, heating oil, and silver), where again gasoline has a significantly negative price discovery coefficient for the spot market suggesting that the futures market is dominant for gasoline.

Overall, the CVAR and FCVAR models thus agree that there is strong statistical evidence that the spot market is dominant in the price discovery process for cocoa and sugar, while the futures market is dominant for canola, copper, gasoline, and heating oil. For the remaining commodities, there is evidence of price discovery taking place in both the spot and futures markets, although

[6] The hypotheses on $\alpha_\perp$ are in fact tested by testing the mirror hypotheses on α. For example, as is obvious from the definition of $\alpha_\perp$, the hypothesis $\alpha_{\perp,1} = 0$ is equivalent to the mirror hypothesis $\alpha_2 = 0$, and the latter hypothesis is straightforward to test within the CVAR and FCVAR models, see also the discussion in Dolatabadi, Nielsen, and Xu (2015).

with some differences across the CVAR and FCVAR models.

Our findings on price discovery connect with the literature on price discovery in commodity markets; see, e.g., Figuerola-Ferretti and Gonzalo (2010), Dolatabadi, Nielsen, and Xu (2015), and the papers cited therein. In this literature there are several studies which show that price discovery is dominated by the futures market. Our study confirms this broad view, but at the same time points to a few commodities where price discovery is not dominated by the futures market. Thus, while our results are consistent in spirit with the literature, suggesting that for most commodities futures market dictates price discovery, this evidence is not completely general—a finding consistent with Dolatabadi, Nielsen, and Xu (2015).

These price discovery results are not trivial outcomes because the dominance of one market over another indicates the market which has the highest information content. This has implications for investors because the market which has most information can then be used to forecast the market which has less information. In univariate regression-style forecasting models, one would then consider using past information from the dominant market to forecast prices or returns in the non-dominant market. However, our FCVAR (and CVAR) models are joint models for spot and futures price series, and will therefore forecast both series simultaneously and hence automatically take the price discovery information in both markets into account.

5.2 Statistical out-of-sample forecast comparison

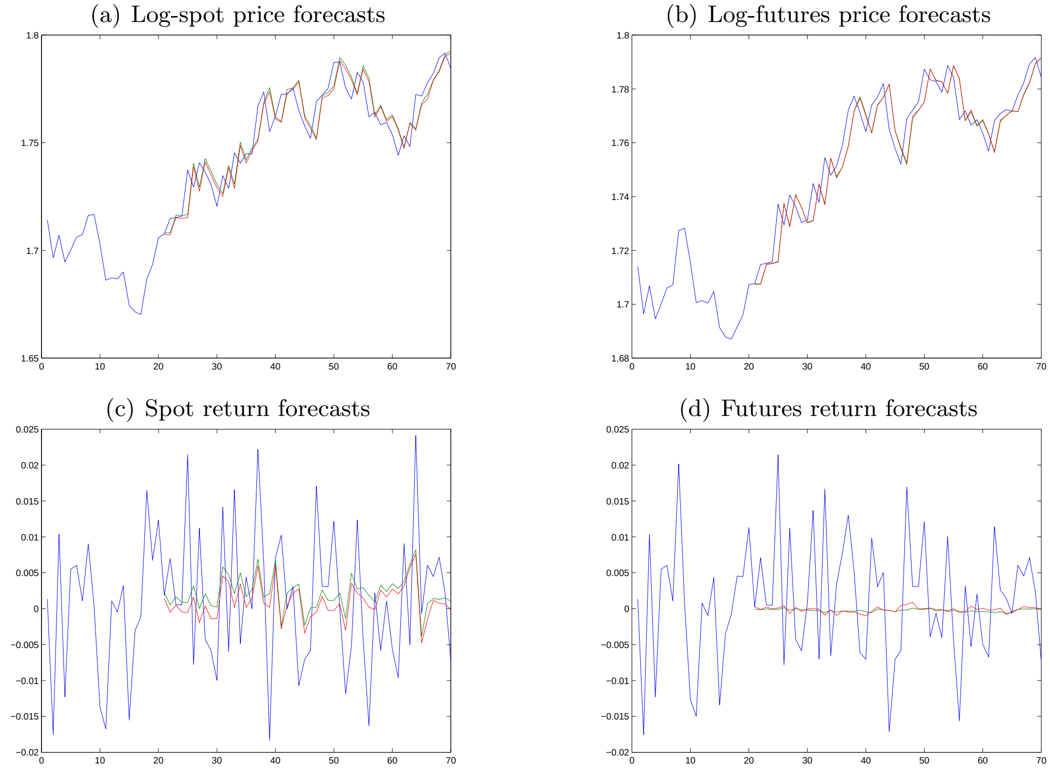
In this subsection we move on to out-of-sample forecasting. Specifically, starting from the estimation results in the previous subsection, we recursively generate one-step ahead (daily) return forecasts, re-estimating the model each period. We generate a total of 1927 out-of-sample return forecasts in this manner (only 1818 for gasoline and 1775 for platinum), to match the remaining 25% of our observations. This allows us to compare our forecasts with the actually observed out-of-sample returns series.

In Figure 4 we show forecasts as well as subsequently realized values for daily forecasting ($h = 1$) of crude oil (a) log-spot prices, (b) log-futures prices, (c) spot returns, and (d) futures returns. We choose crude oil for the illustration because it is the most heavily traded commodity in our dataset, see Table 1. Each subplot shows the last 20 observations in the estimation sample (i.e. X_t or $r_t = \Delta X_t$ for $t = 5762, \ldots, 5781$) together with the first 50 out-of-sample one-step ahead forecasts ($\hat{X}_{t+1|t}$ or $\hat{r}_{t+1|t}$ for $t = 5781, \ldots, 5830$) and the corresponding realized values (X_{t+1} or r_{t+1} for $t = 5781, \ldots, 5830$). In each subplot there are three lines: The blue line denotes the data observations, the red line shows the recursive FCVAR forecasts, and the green line the recursive CVAR forecasts. It is noted that the log-price forecasts track the subsequently realized observations quite well, whereas the returns are clearly predicted much less accurately. This is, of course, expected from no-arbitrage theory of efficient markets. However, we do note a slightly better forecasting performance of spot returns compared with futures returns, at least for this particular part of the sample, which is also in accordance with our finding of price discovery in the futures market for crude oil, see Table 3.

The difference between FCVAR model forecasts and CVAR model forecasts shows most clearly in the forecasts of the equilibrium error series, $\beta' X_t$, which is depicted in Figure 5. Here we show the last (daily) observation on the model equilibrium error (i.e. $\hat{\beta}' X_t$ for $t = 5781$), together with the recursive h-period ahead forecasts of these, $\hat{\beta}' \hat{X}_{t+h|t}$ for $t = 5781$ and $h = 1, \ldots, 100$, generated from (a) the CVAR model and (b) the FCVAR model. In each panel, the horizontal line indicates the mean of the equilibrium relation given by $-\hat{\rho}$. Again, the forecasts are depicted for crude oil as an illustration.

It is clear from Figure 5 that the CVAR model equilibrium error forecasts return to their mean value much more quickly than the FCVAR model equilibrium error forecasts. This reflects

Figure 4: Daily crude oil forecasts



Notes: The four plots show forecasts as well as subsequently realized values for (a) log-spot prices, (b) log-futures prices, (c) spot returns, and (d) futures returns. Each plot shows the last 20 observations in the estimation sample together with the first 50 out-of-sample one-step ahead forecasts and the subsequently realized values. In each subplot there are three lines: The blue denotes the data observations, the red line are the recursive FCVAR forecasts, and the green line the recursive CVAR forecasts.

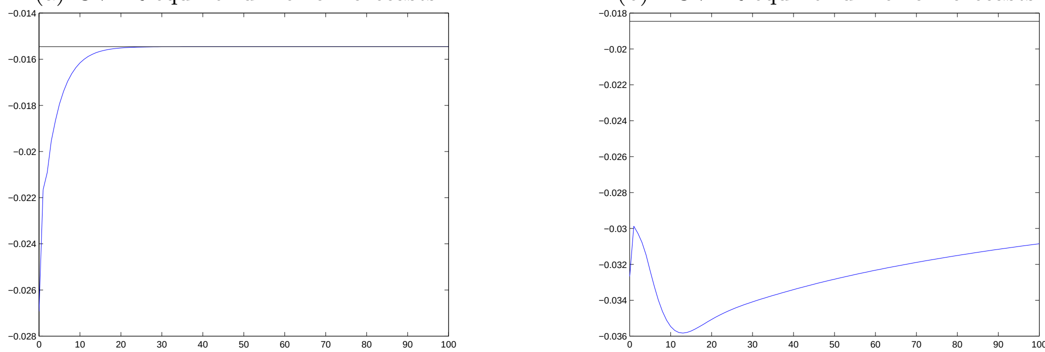
the $I(0)$ nature of $\beta' X_t$ in the CVAR model and the fractional integration nature of $\beta' X_t$ in the FCVAR model, where $\beta' X_t$ is estimated to be $I(0.806)$, see Table 3, which is nonstationary but mean-reverting. It would be natural to expect that this feature of the multi-step ahead forecasts generated by the FCVAR model may impact the forecasting performance of the FCVAR model relative to the CVAR model at longer horizons, which we will investigate below.

In Table 4 we report some out-of-sample forecast comparison statistics for the one-step ahead (daily, $h = 1$) forecasts calculated from either the FCVAR model or the CVAR model. In particular, we first report the Clark and West (2007, Section 2) test statistic for equal predictive ability, which is a modification of the Diebold and Mariano (1995) test statistic to account for the fact that the CVAR model is nested within the FCVAR model class[7] (see also Giacomini and White, 2006, for the point about nested model classes). The CW statistic is asymptotically standard normally distributed and favors the FCVAR model forecasts when it is positive. The null hypothesis of the nested CW test is that the CVAR model forecasts are at least as good as the FCVAR model forecasts, and the alternative hypothesis is that the FCVAR forecasts are superior. Note, therefore, that this is a one-sided test, and we report one-sided significance using asterisks in the table. The next statistic is the relative root mean squared error (RMSE) of the two forecasts (from the FCVAR and CVAR models, respectively), and this is calculated such that negative values favor the FCVAR model forecasts. Finally, we report the out-of-sample R^2 for both sets of forecasts.

[7] We are grateful to Peter Extercate for bringing this point to our attention.

Figure 5: Daily crude oil equilibrium error forecasts

(a) CVAR equilibrium error forecasts (b) FCVAR equilibrium error forecasts



Notes: The two plots show the last observation of the model equilibrium error, i.e. $\hat{\beta}' X_t$, together with the recursive h-period ahead forecasts, $\hat{\beta}' \hat{X}_{t+h|t}$ for $h = 1, \ldots, 100$, generated from (a) the CVAR model and (b) the FCVAR model. In each panel, the mean is indicated by a horizontal line given by $-\hat{\rho}'$.

The results in Table 4 clearly favor the FCVAR model forecasts. Specifically, the CW statistic favors the FCVAR model forecasts in 15/17 spot markets and 16/17 futures markets, and is significant at the 10% level or better in 13 and 11 of those cases, respectively. None of the three negative CW statistics are significant even at the 10% level. The relative RMSE prefers the FCVAR model forecasts for 14/17 commodities in both the spot futures markets.

Where the CW statistic and relative RMSE are both statistical measures of forecast comparison, the final columns in Table 4 report the out-of-sample R^2 for both sets of return forecasts. It is seen from these columns that the forecastability of returns vary greatly across commodities, and also between spot and futures markets for the same commodity. Comparing the FCVAR and CVAR forecasts, the out-of-sample R^2 values support the conclusions from the previous columns with R^2 being higher on average for the FCVAR model. Also, in most cases the out-of-sample R^2 is higher for the spot market than for the futures market, confirming earlier results on the relative forecastability of returns from the two markets.

Before moving on to analyzing economic significance, we investigate the robustness of the Table 4 results by considering forecasting at longer horizons. Specifically, we consider forecasting at the weekly and monthly horizons based on daily data, i.e., horizons of $h = 5$ and $h = 21$ periods ahead. The motivation is that these horizons could correspond to an investor that rebalances the portfolio weekly or monthly and hence needs only forecasts at those horizons. With the same motivation, therefore, we consider only non-overlapping forecasts. That is, the forecast is calculated every h periods (days) for h steps ahead. This yields a total of 385 one-week ahead ($h = 5$) forecasts and 91 one-month ahead ($h = 21$) forecasts of h-day returns, except for gasoline (363 and 86, respectively) and platinum (355 and 84, respectively).

The out-of-sample forecasting results for these longer horizons are presented in Tables 5 (weekly, $h = 5$) and 6 (monthly, $h = 21$), which are both laid out exactly like Table 4. The results for out-of-sample forecast comparisons using statistical measures are generally similar for these horizons as for the daily horizon presented in Table 4, although not quite as favorable towards the FCVAR as in Table 4 and with fewer statistically significant CW statistics. For the weekly horizon, the FCVAR model is preferred to the CVAR model by the CW statistic for 12/17 (spot markets) and 13/17 (futures markets) commodities, although only six of these are significant at the 10% level for the spot markets and seven for the futures markets. For the monthly horizon, the FCVAR model is preferred by the CW statistic for 11/17 (spot markets) and 10/17 (futures markets) commodities,

Table 4: Statistical out-of-sample forecast comparison at daily ($h = 1$) horizon

	CW statistic		Relative RMSE		R^2_{OOS} CVAR		R^2_{OOS} FCVAR	
Commodity	Spot	Futures	Spot	Futures	Spot	Futures	Spot	Futures
Canola	2.3240**	2.0699**	−0.1932	−0.1516	0.4010	−0.0008	0.4033	0.0023
Cocao	4.9334***	2.4235***	−0.6563	−0.1389	0.0037	−0.0030	0.0167	−0.0002
Coffee	10.1160***	0.4626	−2.3397	0.0455	0.4386	0.0003	0.4646	−0.0006
Copper	2.4182***	1.1569	−0.2294	−0.0591	−0.0013	−0.0014	0.0033	−0.0003
Corn	0.7305	0.8105	−0.0182	−0.0181	0.0009	0.0030	0.0013	0.0033
Crude oil	2.0993**	2.1385**	−0.3290	−0.2452	−0.0461	0.0026	−0.0393	0.0075
Gasoline	2.1831**	−0.3623	−0.1532	0.0202	0.0177	0.0021	0.0207	0.0017
Gold	1.3540*	0.5399	−0.0672	−0.0079	−0.0399	0.0426	−0.0385	0.0428
Heating oil	2.2040**	1.5268*	−0.1608	−0.1013	−0.0129	−0.0059	−0.0096	−0.0039
Palladium	3.4676***	2.0645**	−0.0962	−0.0239	0.1165	0.0091	0.1182	0.0096
Platinum	4.2108***	2.4841***	−0.8827	−0.3388	0.1241	−0.0032	0.1395	0.0036
Silver	1.6530**	2.0524**	−0.1440	−0.1923	−0.0740	0.0204	−0.0709	0.0241
Soybean	−0.2842	1.7922**	0.0553	−0.3238	0.0103	−0.0080	0.0092	−0.0015
Soy meal	0.6216	0.5917	0.0220	0.0229	0.0055	0.0055	0.0051	0.0051
Soy oil	1.3304*	1.4648*	−0.0726	−0.0888	0.0048	−0.0064	0.0062	−0.0046
Sugar	−0.8063	1.6087*	0.2434	−0.0854	0.0511	−0.0015	0.0465	0.0003
Wheat	1.5979*	1.4354*	−0.1292	−0.0668	−0.0030	−0.0035	−0.0005	−0.0021

Notes: This table reports out-of-sample forecast comparison statistics for one-step ahead ($h = 1$) return forecasts. The statistics reported are the Clark and West (2007) test statistic, the relative RMSE, and the out-of-sample R^2. The CW statistic is asymptotically standard normally distributed and positive values favors the FCVAR model. Statistical significance (one-sided) of the CW statistic at the 10%, 5%, and 1% level is denoted by one, two, and three asterisks, respectively. The relative RMSE is calculated such that it favors FCVAR model when it is negative.

with four of these being significant in the spot markets and five in the futures markets. Of course, part of the explanation here is that the number of non-overlapping monthly forecasts is smaller than the number of daily forecasts, and therefore it is more difficult to distinguish between the monthly forecasts from the two different models in a statistically significant manner. Among those CW statistics that are statistically significant for longer horizon forecasting ($h = 5$ or $h = 21$) in Tables 5 and 6, 22 out of 23 favor the FCVAR model forecasts over the CVAR forecasts.

For the out-of-sample R^2, the general tendency is, not surprisingly, that it is smaller for the longer horizon forecasts. However, both the out-of-sample R^2 and the relative RMSE in Tables 5 and 6 generally still favor the FCVAR model forecasts over the CVAR forecasts.

Generally, comparing the results across Tables 4–6, our findings clearly show that it is more difficult to predict returns at longer horizons in both the spot and futures commodity markets, at least for the horizons considered here. We would expect this to have implications for the profitability and economic significance of longer horizon forecasts, and we return to this point below.

5.3 Mean-variance utility function based profits

Another possible metric of comparison and evaluation of forecasting performance is economic—rather than purely statistical—significance. That is, whether the forecasts can generate significant returns when incorporated into a dynamic trading strategy. In calculating these metrics, we also investigate whether the improved statistical in-sample fit and forecast performance of the FCVAR model relative to the CVAR model translate into economically significant profits, and whether

Table 5: Statistical out-of-sample forecast comparison at weekly ($h = 5$) horizon

	CW statistic		Relative RMSE		R^2_{OOS} CVAR		R^2_{OOS} FCVAR	
Commodity	Spot	Futures	Spot	Futures	Spot	Futures	Spot	Futures
Canola	0.7481	1.9031**	−0.0849	−0.2551	0.1040	−0.0026	0.1055	0.0025
Cocao	3.1027***	0.9849	−0.9886	−0.1030	0.0653	0.0030	0.0836	0.0050
Coffee	4.0184***	0.1238	−2.4138	0.2639	0.0167	0.0081	0.0636	0.0029
Copper	1.7589**	0.5624	−0.3404	−0.0578	0.0083	−0.0027	0.0150	−0.0015
Corn	0.9508	1.4693*	−0.1422	−0.1047	−0.0079	−0.0025	−0.0050	−0.0004
Crude oil	1.8899**	2.1041**	−0.8968	−1.0455	−0.0138	−0.0026	0.0042	0.0181
Gasoline	0.1395	0.2749	0.3515	0.0041	0.0424	0.0144	0.0356	0.0144
Gold	−0.6288	−2.7025	0.0859	0.2793	−0.0121	0.0119	−0.0139	0.0063
Heating oil	0.6003	−0.9668	−0.0343	0.2178	−0.0069	0.0033	−0.0062	−0.0010
Palladium	0.1160	2.4649***	−0.0003	−0.0816	0.0360	0.0014	0.0360	0.0030
Platinum	7.3564***	7.3241***	−5.3669	−3.4374	−0.0871	−0.0586	0.0265	0.0129
Silver	−0.4919	−0.6082	0.3041	0.4550	−0.0208	0.0151	−0.0271	0.0061
Soybean	−0.4930	1.7129**	0.2480	−0.5934	0.0112	−0.0205	0.0062	−0.0084
Soy meal	−0.4758	−0.0447	0.3978	0.1358	0.0144	−0.0035	0.0065	−0.0063
Soy oil	0.8123	0.4126	−0.1027	−0.0009	0.0051	−0.0061	0.0071	−0.0061
Sugar	−0.7432	1.6662**	0.5411	−0.3672	−0.0003	0.0303	−0.0112	0.0374
Wheat	1.4391*	0.2732	−0.5161	0.1065	−0.0113	−0.0054	−0.0008	−0.0075

Notes: This table reports out-of-sample forecast comparison statistics for one-week ahead ($h = 5$) non-overlapping return forecasts. The statistics reported are the Clark and West (2007) test statistic, the relative RMSE, and the out-of-sample R^2. The CW statistic is asymptotically standard normally distributed and positive values favors the FCVAR model. Statistical significance (one-sided) of the CW statistic at the 10%, 5%, and 1% level is denoted by one, two, and three asterisks, respectively. The relative RMSE is calculated such that it favors FCVAR model when it is negative.

the relatively strong forecastability in some markets compared with others translate into different economic significance across markets.

The economic significance question is important because the statistical superiority of a model over its competitors is just a first step in informing investors. An equally important question is: how can investors benefit from a statistically superior model? This question is directly based on the ability to forecast returns, that is, whether an investor can use forecasts from the model to devise successful trading strategies and make superior profits compared with forecasts from alternative models. In other words, these trading strategies should deliver statistically significant and meaningful profits.

In the stock return forecasting literature, a mean-variance utility function is typically utilized to derive a dynamic trading strategy for investors. We assume that the investor rebalances the portfolio every h days, where we analyze in particular $h = 1$, $h = 5$, and $h = 21$, corresponding to daily, weekly, and monthly rebalancing, respectively. The investor can invest in two assets: the risk-free asset with (continuously compounded) return from period t to period $t+h$ given by $r_{f,t+h}$ and the risky asset with (continuously compounded) return given by $r_{t+h} = s_{t+h} - s_t$ in the case where the risky asset is a spot position or $r_{t+h} = f_{t+h} - f_t$ in the case where the risky asset is a futures position. The investor then forms a portfolio with weight w_{t+h} on the risky asset and this

Table 6: Statistical out-of-sample forecast comparison at monthly ($h = 21$) horizon

	CW statistic		Relative RMSE		R^2_{OOS} CVAR		R^2_{OOS} FCVAR	
Commodity	Spot	Futures	Spot	Futures	Spot	Futures	Spot	Futures
Canola	0.1086	1.3962*	0.3256	−0.4392	−0.0218	−0.0073	−0.0285	0.0015
Cocao	0.5633	0.1983	−0.2073	0.1141	−0.0136	−0.0037	−0.0094	−0.0060
Coffee	2.3997***	−0.7523	−4.3002	1.2826	−0.1338	0.0258	−0.0384	0.0007
Copper	−0.5860	0.1951	0.3697	−0.0186	0.0148	−0.0021	0.0075	−0.0017
Corn	0.6826	1.5109*	−0.3148	−0.2279	0.0017	−0.0132	0.0080	−0.0086
Crude oil	−1.8314**	−0.5001	2.3319	0.5496	−0.0227	0.0013	−0.0709	−0.0096
Gasoline	0.8510	0.3442	−0.3066	−0.0589	0.0759	0.0419	0.0816	0.0430
Gold	−0.0048	−1.5281	0.0059	0.1915	0.0001	−0.0178	0.0000	−0.0217
Heating oil	0.9596	−0.9572	−0.5101	0.6360	−0.0629	−0.0249	−0.0521	−0.0380
Palladium	−0.5635	1.3000*	0.0672	−0.1058	−0.0156	−0.0014	−0.0170	0.0007
Platinum	1.6461**	1.9065**	−0.6423	−0.9022	0.0029	−0.0206	0.0156	−0.0023
Silver	−0.4319	−0.3645	0.1744	0.2706	0.0078	0.0090	0.0044	0.0036
Soybean	−0.1114	2.5335***	0.3023	−1.2456	0.0176	−0.0557	0.0117	−0.0296
Soy meal	0.3061	−0.2431	0.2976	0.3945	0.0043	0.0136	−0.0017	0.0058
Soy oil	0.8103	1.4102*	−0.4503	−0.7284	0.0066	−0.0380	0.0155	−0.0229
Sugar	0.3643	−0.8353	−0.0360	1.5687	0.0128	0.0594	0.0136	0.0297
Wheat	1.5656*	0.4133	−1.8860	−0.0312	−0.0377	−0.0155	0.0011	−0.0148

Notes: This table reports out-of-sample forecast comparison statistics for one-month ahead ($h = 21$) non-overlapping return forecasts. The statistics reported are the Clark and West (2007) test statistic, the relative RMSE, and the out-of-sample R^2. The CW statistic is asymptotically standard normally distributed and positive values favors the FCVAR model. Statistical significance (one-sided) of the CW statistic at the 10%, 5%, and 1% level is denoted by one, two, and three asterisks, respectively. The relative RMSE is calculated such that it favors FCVAR model when it is negative.

portfolio yields a return of

$$\begin{aligned} r_{p,t+h} &= w_{t+h} r_{t+h} + (1 - w_{t+h}) r_{f,t+h} - \theta |w_{t+h} - w_t| \\ &= w_{t+h}(r_{t+h} - r_{f,t+h}) + r_{f,t+h} - \theta |w_{t+h} - w_t|, \end{aligned} \tag{12}$$

where θ denotes a transactions cost, which is applied to the value of the fraction of the portfolio that is being traded, hence the multiplication by $|w_{t+h} - w_t|$. The size of transactions costs in commodity markets is subject to some discussion in the literature. We follow Locke and Venkatesh (1997), who write (p. 239) that "Overall, transaction costs appear to be relatively low for futures trading, as these numbers translate to between 0.0004% and 0.033% of notional value, much less than the 1% or so often cited for equities." Specifically, we take the average of their suggested range and use $\theta = 0.000167$ in (12), i.e., we use 0.0167% of nominal value as the transactions cost for trading in both futures and spot markets. We note from the outset that, as also discussed in the introduction, trading in the futures market is much more practical than trading in the spot market. Indeed, the latter may not even be feasible for some commodities. Thus, for futures market trading in particular, these transactions costs appear reasonable.

Following Marquering and Verbeek (2004) and Campbell and Thompson (2008), among others, the weight on the risky asset is determined by maximizing the investor's mean-variance utility function,

$$U(r_{p,t+h}) = E_t(r_{p,t+h}) - \frac{1}{2}\gamma Var_t(r_{p,t+h}), \tag{13}$$

where $E_t(\cdot)$ and $Var_t(\cdot)$ denote conditional mean and variance given information at time t and γ is the investor's coefficient of relative risk aversion. Maximizing $U(r_{p,t+h})$ with respect to w_{t+h} yields the optimal weight

$$w^*_{t+h} = \frac{E_t(r_{t+h}) - r_{f,t+h}}{\gamma Var_t(r_{t+h})}, \tag{14}$$

noting that the risk-free rate carries no risk and hence does not contribute to the variance of the portfolio. Following the literature, we further constrain the optimal weight and impose $w^*_{t+h} \geq -0.5$ (at most 50% short-selling) and $w^*_{t+h} \leq 1.5$ (at most 50% borrowing/leverage). In the next subsection we consider robustness to alternative restrictions on short-selling and leverage.

To summarize the calculation of profitability of the return forecasts, that is their economic significance, three steps are performed: (i) forecast returns, (ii) compute portfolio weights, and (iii) calculate portfolio returns. The first step involves calculating (6) to forecast commodity (spot or futures) returns at each time period, as explained in Section 2.3.

In the second step we calculate portfolio weights from (14) given the return forecasts $E_t(r_{t+h}) = \hat{r}_{t+h|t}$. For the risk-free return $r_{f,t+h}$ we use the return on the US three-month Treasury bill, which is assumed known at time t (since it is risk-free). The risk-aversion coefficient is set at $\gamma = 6$, corresponding to an investor that takes a medium level of risk, and for robustness we also consider in the next subsection a higher risk-aversion investor ($\gamma = 12$) and a lower risk-aversion investor ($\gamma = 3$). Finally, following standard practice, we estimate the time-varying variance of the risky asset by a GARCH(1,1) model using all observations available at time t.

Third, given the portfolio weights, portfolio returns are computed from (12) for each period. These are then aggregated across time and reported as an annualized average portfolio excess return to facilitate comparison across different values of the rebalancing horizon, h. Because futures market trading is more practical than spot market trading, we begin with the commodity futures markets.

In Table 7 we report the annualized average excess portfolio return (multiplied by 100 to yield a continuously compounded percentage return) for commodity futures markets. The returns are reported as excess returns above and beyond the risk-free rate (the average return on which was 1.753% per annum over the out-of-sample forecasting period). The results are presented for a medium risk-aversion investor, that is, with risk-aversion coefficient $\gamma = 6$ in (13) and (14), and with weights restricted to the interval $[-0.5, 1.5]$, corresponding to at most 50% short-selling and borrowing/leverage. We report results for daily ($h = 1$), weekly ($h = 5$), and monthly ($h = 21$) rebalancing and for forecasts based on both the CVAR and FCVAR models. Standard errors are reported in parentheses. For each commodity and each rebalancing horizon, we conduct a statistical test of the null hypothesis that excess portfolio return is zero against the two-sided alternative that excess portfolio return is different from zero, and we interpret this as a (statistical) test of economic significance of the return forecasts.

The results in Table 7 show several clear tendencies. First of all, excess portfolio returns average about 9.0% and 10.8% per annum with daily rebalancing, based on CVAR and FCVAR forecasts, respectively, 0.2% and 0.9% with weekly rebalancing, and 1.9% and 2.5% with monthly rebalancing. Compared with the average annual return of just under 1.8% on the risk-free asset and the very small average returns on the commodities themselves, which in many cases is only very slightly higher than that of the risk-free asset, as reported in Table 2, we find that the excess returns in Table 7 with daily rebalancing are impressively large for most commodities. Secondly, with daily rebalancing, the excess returns are significantly positive in 14 of 34 cases, whereas only one of the negative returns is significant (and only at the 10% level). For longer horizons, there are fewer significant returns because the returns are smaller and their standard errors higher. Thirdly, as shown in the last row of the table, by far the highest returns on average are found with daily

Table 7: Annualized excess portfolio returns for commodity futures markets

	Daily		Weekly		Monthly	
Commodity	CVAR	FCVAR	CVAR	FCVAR	CVAR	FCVAR
Canola	14.042***	16.732***	2.778	0.426	−3.707	−4.566
	(5.430)	(5.202)	(4.799)	(3.760)	(6.370)	(3.713)
Cocoa	−6.157*	−0.547	0.666	−3.621	6.824	0.568
	(3.659)	(3.747)	(4.320)	(6.469)	(8.222)	(9.337)
Coffee	2.565	3.910	−6.106	−0.328	−10.748	2.572
	(4.240)	(4.655)	(5.908)	(7.818)	(9.563)	(9.584)
Copper	−3.810	−0.574	−4.338	−1.852	7.636	2.386
	(3.858)	(0.952)	(3.222)	(2.077)	(5.827)	(3.848)
Corn	14.163**	15.545**	−12.351*	−9.570	−3.461	−3.397
	(6.066)	(6.330)	(7.053)	(7.548)	(14.955)	(15.914)
Crude oil	−1.040	4.362	0.612***	−0.101	0.684***	0.496***
	(1.778)	(3.592)	(0.054)	(0.068)	(0.101)	(0.093)
Gasoline	6.705	7.133	8.915	9.493	29.981*	25.199
	(4.833)	(4.688)	(6.844)	(6.679)	(17.561)	(16.170)
Gold	36.877***	36.726***	0.816	0.600	8.067	7.425
	(4.765)	(4.680)	(5.546)	(5.131)	(6.811)	(7.173)
Heating oil	−1.289	0.696	−0.653***	−0.354***	0.108	0.323***
	(4.047)	(3.566)	(0.097)	(0.088)	(0.117)	(0.105)
Palladium	26.677***	27.852***	−1.351	−1.879	7.067	6.935
	(6.483)	(6.594)	(5.878)	(5.572)	(7.730)	(6.992)
Platinum	1.322	6.199**	−3.099	−1.673	9.072	15.287**
	(3.629)	(2.784)	(4.062)	(3.651)	(6.819)	(6.269)
Silver	38.723***	32.990***	1.455	−3.333	4.737	8.816**
	(6.306)	(4.788)	(10.523)	(3.987)	(11.122)	(4.234)
Soybean	−3.796	−1.052	−2.159	2.166	−11.097	−8.732
	(4.362)	(5.055)	(6.880)	(7.496)	(10.621)	(10.996)
Soy meal	13.839**	13.379**	−2.907	6.155	−11.337**	−10.342
	(6.513)	(6.078)	(5.796)	(6.615)	(5.325)	(8.705)
Soy oil	2.272	4.173	8.841	0.582	−12.082	−4.102
	(6.447)	(6.052)	(5.375)	(5.185)	(8.978)	(5.258)
Sugar	8.066	14.399*	22.104***	25.621**	−10.072	−6.563
	(7.912)	(8.467)	(8.359)	(10.245)	(15.294)	(16.114)
Wheat	3.496	1.162	−10.276	−7.763	21.121	9.477
	(5.229)	(5.398)	(7.501)	(9.294)	(12.938)	(13.676)
Average	8.980	10.770	0.174	0.857	1.929	2.458

Notes: This table reports annualized average excess portfolio percentage returns for commodity futures markets for a medium risk-aversion investor ($\gamma = 6$) with weights restricted to the interval $[-0.5, 1.5]$. The results are reported for daily ($h = 1$), weekly ($h = 5$), and monthly ($h = 21$) rebalancing and for forecasts based on both the CVAR and FCVAR models. Standard errors are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is denoted by one, two, and three asterisks, respectively. The final row is the average return across all commodities.

rebalancing. Thus, it would appear that more opportunities to rebalance clearly outweighs the additional transactions costs imposed.

Comparing portfolio returns using weights calculated from CVAR and FCVAR based forecasts show that each model produces similar returns on average, although there are sometimes substantial differences for individual commodities. In fact, the FCVAR forecasts generate higher returns for 13 out of 17 commodities with daily rebalancing. With weekly and monthly rebalancing, the FCVAR forecasts generate higher returns in 10/17 commodities and 9/17 commodities, respectively. Quite surprisingly, given our expectations of the FCVAR model as a superior long-horizon forecasting model, the CVAR and FCVAR models basically perform equally well with longer horizon rebalancing. Of course, as noted in Tables 4–6, the accuracy of the forecasts is worse at longer horizons, and this is clearly part of the reason why profits are lower with longer horizon rebalancing. Thus, overall, the two models perform quite similarly in terms of futures market profits for a medium risk-aversion investor.

The results for the spot markets are presented in Table 8, which is laid out as in Table 7. These results show even larger excess returns with daily rebalancing than for the futures markets, which reflects the earlier finding from Figure 4 and Tables 4–6 that futures returns are more difficult to forecast than spot returns. It is also not too surprising given the finding from Table 3 that price discovery is primarily in the futures market for most commodities, suggesting that portfolio returns from the trading strategy may be higher in the spot markets than in the futures markets. In particular, we notice from Table 2 that there is no substantial difference between average returns in the spot and futures markets, so the different profits in the two markets cannot be attributed simply to differences in the unconditional average return in the two markets. There are four obvious exceptions, namely crude oil, gold, heating oil, and silver, where excess portfolio returns in the spot markets are lower than in the futures markets. This is not too surprising since the statistical forecast evaluation also showed that for these commodities, futures returns are more predictable than spot returns, as seen from the out-of-sample R^2 statistics in Table 4.

Since the spot market positions are not as easily tradable as the futures market positions, and since the transactions costs more closely match those found in futures markets, we will not focus too much on the spot market profits results, although these are still useful as a metric of forecast comparison between the CVAR and FCVAR models. Generally, the results in Table 8 are similar to those found in Table 7. Excess returns are highest with daily rebalancing, and this is also where the most significant returns are found. For a few commodities, these are in the hundreds of percent per annum. Also, the comparison between CVAR and FCVAR forecasts based on average returns shown the last row of the table again shows that the two models produce very similar returns on average, with a slight advantage to the FCVAR model in the daily and monthly rebalancing cases and a slight advantage to the CVAR model with weekly rebalancing.

As an additional performance metric, and an additional metric of comparison between the CVAR and FCVAR models, we report in Table 9 the estimated Sharpe ratios of excess portfolio returns for the commodity spot and futures markets with daily rebalancing, i.e., corresponding to the excess returns shown in the first two columns of Tables 7 and 8. Of course, the Sharpe ratio is defined simply as the average excess return for a commodity divided by the standard deviation of its excess return, and the reported Sharpe ratios are annualized. Thus, the Sharpe ratio allows us to investigate whether our positive excess returns arise because we are picking up some risk premium (i.e., allowing for higher risk), since the Sharpe ratio shows the performance of a portfolio adjusted for its risk. Jobson and Korkie (1981, p. 893) show using the Delta method that the asymptotic standard deviation of the Sharpe ratio, say $\hat{S}$, can be estimated by $s.e.(\hat{S}) = (1 + \hat{S}^2/2)^{1/2}/T^{1/2}$, which we report in parentheses. As usual, statistical significance is denoted by asterisks. Since the Sharpe ratios are fairly trivially calculated from the excess returns already presented, and

Table 8: Annualized excess portfolio returns for commodity spot markets

	Daily		Weekly		Monthly	
Commodity	CVAR	FCVAR	CVAR	FCVAR	CVAR	FCVAR
Canola	213.317***	209.812***	−11.277	−15.609	−31.636	−21.274
	(9.436)	(9.237)	(17.420)	(16.861)	(24.099)	(21.505)
Cocoa	38.237***	44.565***	6.299	3.763	−7.236	3.089
	(8.354)	(8.347)	(13.653)	(13.765)	(14.43)	(10.201)
Coffee	189.528***	203.214***	−20.420**	−26.835**	−5.803	−7.660
	(7.534)	(7.551)	(9.647)	(10.751)	(9.791)	(9.290)
Copper	2.171	9.339*	6.801	8.163	−7.620**	−11.765*
	(5.528)	(5.375)	(5.012)	(8.418)	(3.841)	(6.090)
Corn	20.802**	18.228**	13.662	1.606	9.211	4.107
	(8.550)	(7.640)	(16.969)	(14.547)	(32.283)	(23.750)
Crude oil	−2.398	0.152	−1.206***	−0.816***	−0.118	0.004
	(6.141)	(6.004)	(0.130)	(0.124)	(0.168)	(0.142)
Gasoline	35.462***	37.033***	26.439*	31.775*	71.912***	51.143**
	(9.450)	(10.290)	(15.947)	(18.325)	(27.360)	(25.841)
Gold	1.481	3.081	−4.146	−3.547	31.451**	29.186*
	(6.876)	(7.072)	(8.562)	(9.151)	(15.472)	(15.603)
Heating oil	−3.760	−0.739	−0.913***	−0.554***	−0.452**	−0.128
	(4.540)	(4.328)	(0.117)	(0.098)	(0.206)	(0.157)
Palladium	119.516***	119.772***	−2.883	−0.186	5.831	5.460
	(11.495)	(11.397)	(11.939)	(11.542)	(8.862)	(7.974)
Platinum	29.971***	31.787***	−7.316	−2.421	−6.068	3.809
	(4.358)	(4.209)	(6.641)	(6.694)	(11.043)	(9.081)
Silver	−1.480	−9.107	−10.603	−15.683	37.507	26.621*
	(10.523)	(11.099)	(9.604)	(10.832)	(27.947)	(15.009)
Soybean	20.292***	14.973***	11.153	6.761	2.326	−1.766
	(6.158)	(4.879)	(10.179)	(7.167)	(10.262)	(4.424)
Soy meal	16.737**	14.153*	16.537*	4.633	−8.751	−3.818
	(8.015)	(8.222)	(10.005)	(8.946)	(23.589)	(15.889)
Soy oil	17.370**	14.659**	11.073	3.613	−23.826	−13.026
	(7.245)	(6.356)	(7.919)	(5.923)	(15.467)	(10.303)
Sugar	94.437***	90.336***	−13.574	−17.619	0.312	−7.855
	(9.944)	(10.335)	(8.579)	(11.959)	(6.516)	(12.984)
Wheat	2.486	2.259	28.017*	6.269	−49.919	−10.141
	(6.197)	(3.203)	(16.775)	(8.844)	(38.537)	(12.648)
Average	46.716	47.266	2.802	−0.982	1.007	2.705

Notes: This table reports annualized average excess portfolio percentage returns for commodity spot markets for a medium risk-aversion investor ($\gamma = 6$) with weights restricted to the interval $[-0.5, 1.5]$. The results are reported for daily ($h = 1$), weekly ($h = 5$), and monthly ($h = 21$) rebalancing and for forecasts based on both the CVAR and FCVAR models. Standard errors are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is denoted by one, two, and three asterisks, respectively. The final row is the average return across all commodities.

Table 9: Sharpe ratios of excess portfolio returns for commodity markets (daily rebalancing)

	Futures markets		Spot markets	
Commodity	CVAR	FCVAR	CVAR	FCVAR
Canola	0.950***	1.182***	8.304***	8.344***
	(0.368)	(0.368)	(0.391)	(0.391)
Cocoa	−0.618*	−0.054	1.681***	1.961***
	(0.367)	(0.367)	(0.368)	(0.369)
Coffee	0.222	0.309	9.241***	9.886***
	(0.367)	(0.367)	(0.396)	(0.400)
Copper	−0.363	−0.221	0.144	0.638*
	(0.367)	(0.367)	(0.367)	(0.367)
Corn	0.858**	0.902**	0.894**	0.876**
	(0.368)	(0.368)	(0.368)	(0.368)
Crude oil	−0.215	0.446	−0.143	0.009
	(0.367)	(0.367)	(0.367)	(0.367)
Gasoline	0.525	0.575	1.419***	1.361***
	(0.378)	(0.378)	(0.379)	(0.379)
Gold	2.843***	2.882***	0.079	0.160
	(0.370)	(0.370)	(0.367)	(0.367)
Heating oil	−0.117	0.072	−0.304	−0.063
	(0.367)	(0.367)	(0.367)	(0.367)
Palladium	1.512***	1.551***	3.819***	3.860***
	(0.368)	(0.368)	(0.372)	(0.373)
Platinum	0.139	0.852**	2.632***	2.891***
	(0.383)	(0.383)	(0.385)	(0.386)
Silver	2.256***	2.531***	−0.052	−0.301
	(0.369)	(0.370)	(0.367)	(0.367)
Soybean	−0.320	−0.076	1.210***	1.127***
	(0.367)	(0.367)	(0.368)	(0.368)
Soy meal	0.780**	0.809**	0.767**	0.632*
	(0.368)	(0.368)	(0.368)	(0.367)
Soy oil	0.129	0.253	0.881**	0.847**
	(0.367)	(0.367)	(0.368)	(0.368)
Sugar	0.374	0.625*	3.488***	3.211***
	(0.367)	(0.367)	(0.372)	(0.371)
Wheat	0.246	0.079	0.147	0.259
	(0.367)	(0.367)	(0.367)	(0.367)
Average	0.541	0.748	2.012	2.100

Notes: This table reports estimated annualized Sharpe ratios of excess portfolio returns for commodity markets with daily ($h = 1$) rebalancing corresponding to the portfolio returns in Tables 7 and 8. Standard errors are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is denoted by one, two, and three asterisks, respectively. The final row is the average Sharpe ratio across all commodities.

to conserve space, we only report Sharpe ratios for our benchmark case. Of course, the usual disadvantage of the Sharpe ratio remains that it relies on the notions that risk equals variance and that all variance is bad, such that the Sharpe ratio penalizes strategies with upside potential the same as strategies with similar magnitude downside risk. Thus, any conclusions drawn from the Sharpe ratios should be considered in this light.

The Sharpe ratios reported in Table 9 reflect the heterogeneity of excess returns across commodities in Tables 7 and 8. In particular, the Sharpe ratios for the futures markets vary from -0.618 to 2.882 and for the spot markets from -0.304 to 9.886. Only one of the negative Sharpe ratios is significant (and only at the 10% level, corresponding to the significantly negative return for cocoa in Table 7), whereas many of the positive Sharpe ratios are very significant. In particular, 15 of the 34 Sharpe ratios are significant for the futures markets and 23 of the 34 Sharpe ratios are significant for the spot markets.

In terms of the Sharpe ratios, it would appear that portfolios based on FCVAR model forecasts has an advantage over those based on CVAR model forecasts, especially for the futures markets. Throughout Table 9, the Sharpe ratio for the FCVAR based portfolio is higher than that for the CVAR based portfolio in 16 of 17 futures markets and 10 of 17 spot markets. This finding is confirmed by the average Sharpe ratios reported in the final row of Table 9.

Generally, we observe that both portfolio excess profits and their Sharpe ratios are quite heterogeneous across commodities, whether based on spot or futures markets. This is not surprising. In fact, reading the literature on commodity futures markets, all studies that we are aware of (see examples in the introduction) find heterogeneous profits across different commodities. While the goal of our paper is not to explain heterogeneity in profits, our analysis does offer some insights into why profits are heterogeneous. In Table 1 we present evidence showing that trading volume and liquidity of commodity futures are very heterogeneous. Trading volume and liquidity are fundamental to the nature and magnitude of profits. Furthermore, in Table 2 we present evidence that shows how commodities differ on commonly observed statistical features, such as mean, standard deviation, and higher moments of returns. Taken together, these differences in commodity fundamentals and statistical features suggest potentially quite different risk-return relationships, and it would appear that, when analyzed using portfolios generated from a mean-variance utility function, this translates into heterogeneous portfolio profits across commodities.

5.4 Robustness of economic significance

In this section, we investigate the robustness of our findings in three directions: (i) to the choice of risk-aversion coefficient, γ, and (ii) in comparison to a simple moving-average crossover trading rule, and (iii) to the restrictions on the portfolio weights. So far we have considered an investor who takes a medium level of risk, and specifically has risk-aversion parameter $\gamma = 6$, and trading regulations allowing short-selling and borrowing/leverage up to 50%. The choice of γ has obvious implications for portfolio returns via the calculation of weights in (14), from which it is noticed that a lower risk-aversion investor ($\gamma = 3$) will place a higher weight on the risky asset and vice versa for a higher risk-aversion investor ($\gamma = 12$), although of course the investor remains risk averse for any $\gamma > 0$. A natural question is whether portfolio returns constructed as above are still statistically and economically significant when the investor is more or less risk averse. Another related question is how the portfolio returns compare with returns obtained from a simple technical trading rule such as a moving-average crossover strategy. Although such a trading rule assumes risk neutrality of the investor, it is still of interest to compare overall profits from our forecasting based strategy with a simple moving-average strategy as a type of benchmark. Similarly to the first point, close investigation of the results in the previous subsection reveals that optimal weights are quite often restricted by the 50% bound on short-selling and borrowing/leverage, and it would be interesting to examine how sensitive the above results are to alternative restrictions on the optimal weights reflecting either more restrictive assumptions (no short-selling and borrowing/leverage) or less restrictive assumptions (up to 100% short-selling and borrowing/leverage).

In our first set of robustness results, presented in Table 10, we report annualized excess portfolio returns for commodity spot and futures markets with daily rebalancing ($h = 1$) for both a lower risk-aversion investor ($\gamma = 3$) and a higher risk-aversion investor ($\gamma = 12$) with weights restricted to the interval $[-0.5, 1.5]$. Overall, the results presented here support our earlier findings.

First of all, excess portfolio returns are large and positive in almost all cases, and many returns are statistically significant. Secondly, and not surprisingly, average portfolio returns are higher for the lower risk-aversion investor than for the higher risk-aversion investor, with the results for the medium risk-aversion investor in Tables 7 and 8 falling in between. This suggests that the restrictions imposed on the optimal weights may often be binding, which will be confirmed below. Thirdly, returns from the spot markets are once more found to be higher in general than returns from the futures markets, with the exception of crude oil, gold, heating oil, and silver as above. In fact, the lower risk-aversion investor is able to obtain about a 50% excess return per annum in the gold and silver futures markets.

Comparing excess returns based on CVAR and FCVAR forecasts in Table 10 using the average across all 17 commodities shows that the FCVAR model produces slightly higher excess returns on average for both the lower risk-aversion and the higher risk-aversion investor and for both the spot and futures markets. More specifically, focusing on the futures markets, where trading is more practical, the FCVAR model outperforms the CVAR model for 14 of 17 commodities when $\gamma = 3$ and for 13 of 17 commodities when $\gamma = 12$. Thus, for the majority of commodity futures markets, the FCVAR model forecasts produces higher portfolio returns.

Next we investigate how the portfolio excess returns compare with returns obtained from a simple technical trading rule given by a moving-average (MA) crossover strategy. As above, we let the risky asset return from period t to period $t+1$ be denoted by r_{t+1}, and define the two moving averages, $y_t^S = S^{-1} \sum_{s=1}^{S} r_{t-s}$ and $y_t^L = L^{-1} \sum_{l=1}^{L} r_{t-l}$. Then the weight w_{t+1} on the risky asset between period t and period $t+1$ is maximized (within the weight restrictions) if $y_{t+1}^S \geq y_{t+1}^L$, which is interpreted as a buy signal. Similarly, the weight is minimized (again, within the weight restrictions) if $y_{t+1}^S < y_{t+1}^L$, which is interpreted as a sell signal. With these definitions, the excess return of the MA crossover trading rule is then given as usual by (12). Note that the moving averages y_{t+1}^S and y_{t+1}^L, and hence the weight w_{t+1}, are defined in terms of lagged values of returns, such that w_{t+1} is known and can feasibly be implemented at period t such as to define a portfolio return from period t to period $t+1$. We remark at this point that the MA crossover trading rule assumes risk neutrality of the investor, and therefore comparing it with our mean-variance trading strategy, which assumes risk-aversion, is not quite an apples-to-apples comparison. However, following the literature, e.g. Narayan, Narayan, and Sharma (2013), it is still of interest to compare overall profits from our forecasting based mean-variance strategy with a simple technical trading rule such as the MA crossover strategy as a type of benchmark.

In Table 11 we report average excess returns and Sharpe ratios for commodity spot and futures markets using the MA crossover trading rule with $S = 5$ and $L = 50$.[8] The results are presented for the benchmark case with daily rebalancing. It is clear from the table that the MA strategy excess returns are smaller on average than our forecasting based returns given in the previous tables, they are less statistically signficant, and their Sharpe ratios are substantially smaller. Moreover, because the MA strategy assumes a risk-neutral investor, whereas our mean-variance trading strategy assumes a risk-averse investor, the most relevant comparison should be with our trading strategy

[8] We tried also other values of S and L with qualitatively similar results.

Table 10: Annualized daily excess portfolio returns for alternative risk coefficients

	Lower risk-aversion ($\gamma = 3$)				Higher risk-aversion ($\gamma = 12$)			
	Spot markets		Futures markets		Spot markets		Futures markets	
Commodity	CVAR	FCVAR	CVAR	FCVAR	CVAR	FCVAR	CVAR	FCVAR
Canola	217.200***	213.876***	16.683**	21.239***	199.847***	196.809***	8.978**	10.069***
	(9.538)	(9.513)	(6.951)	(6.802)	(8.770)	(8.602)	(3.504)	(3.347)
Cocoa	2.459	12.209	−6.183	−1.118	2.928	7.346**	−1.990	−0.386
	(8.455)	(7.708)	(6.477)	(1.740)	(3.462)	(3.397)	(2.049)	(0.527)
Coffee	43.667***	49.239***	−10.053*	−2.935	35.504***	40.463***	−2.681	0.056
	(9.917)	(9.969)	(5.905)	(5.805)	(6.690)	(6.734)	(2.016)	(2.159)
Copper	194.244***	209.350***	1.377	3.402	175.775***	187.932***	1.346	2.229
	(7.708)	(7.759)	(6.761)	(6.409)	(7.184)	(7.219)	(2.159)	(2.703)
Corn	27.831**	23.984**	21.781**	23.310**	11.881**	9.703**	8.543**	9.240**
	(12.035)	(11.027)	(8.812)	(9.233)	(4.922)	(4.420)	(3.983)	(4.091)
Crude oil	−3.731	0.060	−1.814	7.555	−0.695	0.244	−0.236	2.290
	(6.657)	(6.424)	(2.486)	(4.962)	(5.223)	(5.267)	(1.312)	(2.149)
Gasoline	38.279***	42.372***	12.029	13.034*	26.995***	31.328***	4.252	4.248
	(12.484)	(13.542)	(7.409)	(7.465)	(6.466)	(7.075)	(2.912)	(2.759)
Gold	1.863	2.451	49.086***	48.040***	0.355	2.451	25.540***	25.890***
	(7.815)	(7.988)	(5.970)	(5.825)	(5.657)	(5.787)	(3.599)	(3.567)
Heating oil	−4.639	−1.006	−0.843	−0.008	−2.347	−0.231	−1.122	0.220
	(5.436)	(5.283)	(5.039)	(4.698)	(3.338)	(2.948)	(2.771)	(2.218)
Palladium	123.770***	124.800***	38.578***	40.022***	102.359***	102.123***	16.267***	16.942***
	(12.786)	(12.687)	(9.055)	(9.218)	(9.416)	(9.336)	(4.173)	(4.200)
Platinum	30.840***	32.429***	0.200	6.633*	28.774***	30.297***	1.638	4.954**
	(4.486)	(4.357)	(4.178)	(3.476)	(4.064)	(3.902)	(2.714)	(1.961)
Silver	−1.256	−11.248	59.202***	48.506***	−2.913	−8.000	22.530***	21.024***
	(12.331)	(13.206)	(9.245)	(6.595)	(8.215)	(8.447)	(3.589)	(3.136)
Soybean	22.161***	14.621**	−4.968	0.507	15.260***	11.833***	−3.720	−0.098
	(8.370)	(6.795)	(6.079)	(6.781)	(3.973)	(3.315)	(3.019)	(3.413)
Soy meal	22.883**	20.831**	15.138*	17.995**	13.319**	10.967**	10.698**	9.791**
	(10.720)	(10.877)	(8.785)	(8.039)	(5.234)	(5.300)	(4.478)	(4.148)
Soy oil	21.473**	18.789**	5.366	6.511	13.557***	12.271***	−0.010	0.972
	(9.937)	(8.727)	(8.412)	(7.980)	(4.740)	(4.239)	(3.940)	(3.688)
Sugar	102.521***	96.460***	10.986	18.920	77.431***	74.422***	4.783	7.054
	(11.420)	(11.851)	(11.209)	(11.723)	(8.010)	(8.285)	(4.859)	(5.083)
Wheat	5.086	3.973	8.418	3.620	1.408	1.151	1.440	0.302
	(11.317)	(5.922)	(7.477)	(7.890)	(3.140)	(1.615)	(3.338)	(3.277)
Average	49.685	50.188	12.646	15.014	41.143	41.830	5.662	6.753

Notes: This table reports annualized average excess portfolio percentage returns for commodity spot and futures markets for daily ($h = 1$) rebalancing with weights restricted to the interval $[-0.5, 1.5]$. The results are reported for a lower risk-aversion investor ($\gamma = 3$) and a higher risk-aversion investor ($\gamma = 12$) and for forecasts based on both the CVAR and FCVAR models. Standard errors are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is denoted by one, two, and three asterisks, respectively. The final row is the average return across all commodities.

Table 11: Annualized excess returns and Sharpe ratios for MA strategy (daily rebalancing)

	Excess returns		Sharpe ratios	
Commodity	Spot markets	Futures markets	Spot markets	Futures markets
Canola	−17.621*	7.942	−0.614*	0.331
	(10.539)	(8.821)	(0.367)	(0.367)
Cocoa	−4.462	17.549	−0.140	0.498
	(11.720)	(12.939)	(0.367)	(0.367)
Coffee	17.611*	−3.958	0.684*	−0.116
	(9.463)	(12.555)	(0.367)	(0.367)
Copper	−3.013	−4.100	−0.080	−0.108
	(13.833)	(13.907)	(0.367)	(0.367)
Corn	−6.812	4.991	−0.183	0.134
	(13.685)	(13.726)	(0.367)	(0.367)
Crude oil	−2.993	−9.568	−0.153	−0.545
	(7.202)	(6.447)	(0.367)	(0.367)
Gasoline	−13.144	0.771	−0.272	0.018
	(18.259)	(15.980)	(0.378)	(0.378)
Gold	24.533***	19.087**	1.035	0.824**
	(8.708)	(8.505)	(0.368)	(0.368)
Heating oil	−7.139	−3.749	−0.413	−0.234
	(6.343)	(5.875)	(0.367)	(0.367)
Palladium	17.292	17.795	0.452	0.469
	(14.066)	(13.938)	(0.367)	(0.367)
Platinum	16.622	23.731**	0.569	0.837**
	(11.174)	(10.845)	(0.383)	(0.383)
Silver	19.342	23.538	0.475	0.586
	(14.964)	(14.757)	(0.367)	(0.367)
Soybean	4.531	−1.951	0.146	−0.056
	(11.399)	(12.836)	(0.367)	(0.367)
Soy meal	20.712	9.441	0.585	0.263
	(13.003)	(13.161)	(0.367)	(0.367)
Soy oil	16.633	19.303*	0.554	0.680*
	(11.035)	(10.425)	(0.367)	(0.367)
Sugar	12.030	13.461	0.306	0.308
	(14.436)	(16.079)	(0.367)	(0.367)
Wheat	−21.176	7.669	−0.413	0.181
	(18.821)	(15.571)	(0.367)	(0.367)
Average	4.291	8.350	0.149	0.239

Notes: This table reports annualized average excess percentage returns and estimated Sharpe ratios for commodity spot and futures markets using an MA crossover trading rule with $S = 5$ and $L = 50$. The results are reported for daily ($h = 1$) rebalancing and with weights restricted to the interval $[-0.5, 1.5]$. Standard errors are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is denoted by one, two, and three asterisks, respectively. The final row is the average (excess return and Sharpe ratio) across all commodities.

with the least risk-averse investor, i.e. $\gamma = 3$, given in Table 10. Compared with those results, the MA strategy appears much less profitable.

In our final set of robustness results, presented in Table 12, we report annualized excess portfolio returns for commodity spot and futures markets with daily rebalancing ($h = 1$) for a medium risk-aversion investor ($\gamma = 6$) with weights restricted either to the interval $[0, 1]$, where no short-selling or borrowing/leverage is allowed or to the interval $[-1, 2]$, where up to 100% short-selling and borrowing/leverage is allowed. These results again support our earlier findings, and also illustrate the importance of the restrictions placed on the optimal portfolio weights.

When the optimal portfolio weights are restricted to $[0, 1]$, returns are somewhat smaller than in the benchmark case in Tables 7 and 8 (where the interval is given by $[-0.5, 1.5]$). On the other hand, when the optimal weights are restricted to $[-1, 2]$, which is less restrictive than the benchmark case in Tables 7 and 8, portfolio returns are much higher on average than in Tables 7 and 8. This shows the importance of the restrictions on the optimal weights. As with the earlier results, a few excess returns are negative but these are not statistically significant. The empirical result from Table 12 that eliminating (allowing more) short-selling and borrowing reduces (increases) profits is due to the fact that negative price movements (i.e., negative returns) are often predicted quite accurately by the models thus generating a short position that is as big as allowed within the weight restrictions. Some of these negative returns are very large, and hence the profits made from the associated short positions can be very large as well.

Comparing the performance of the CVAR and FCVAR models in Table 12 shows a very similar picture to that in Table 10. Using the average across all 17 commodities, the FCVAR model produces slightly higher excess returns on average in both the spot and futures markets and for both the more and the less restrictive weights. Again focusing on the futures markets, where trading is more practical, the FCVAR model outperforms the CVAR model for 14 of 17 commodities whether weights are restricted to $[0, 1]$ or are restricted to $[-1, 2]$. Thus, for the majority of commodities, the FCVAR model forecasts once more produce higher portfolio returns in the futures markets.

The overall implication is that our evidence that the FCVAR is a statistically superior model mostly extends to its economic importance, even though the differences in excess returns between the CVAR and FCVAR models are mostly quite small on average. Therefore, certainly statistically and to some extent also economically, the FCVAR model offers investors a better guide to undertaking investment portfolio decisions.

6 Concluding remarks

This paper has analyzed the link between statistical models of forecasting for commodity prices and returns and their implications for investors. Identifying suitable forecasting models for asset returns is at the forefront of research in asset pricing. This is so because the accuracy of forecasts have direct implications for investors' decision making, particularly with regard to portfolio choice. In this paper we take a step in this direction by proposing an FCVAR model for forecasting commodity spot and futures returns, based on recent empirical evidence of fractional cointegration in commodity spot and futures markets. We derive the best linear predictor for the FCVAR model and perform an out-of-sample forecast comparison with forecasts from the more standard CVAR model. In our empirical analysis to 17 commodity spot and futures markets, the fractional model is found to be statistically superior in terms of both in-sample fit and out-of-sample forecasting.

In terms of economic significance of the forecasts, we analyze this through a dynamic trading strategy based on a portfolio with weights derived from a mean-variance utility function. This analysis leads to statistically significant and economically meaningful profits in most commodity markets, and shows that excess returns from both the FCVAR and CVAR models are substantially higher on average, and statistically more significant, than excess returns from a simple moving-

Table 12: Annualized daily excess portfolio returns for alternative weight restrictions

	Weights in $[0,1]$				Weights in $[-1,2]$			
	Spot markets		Futures markets		Spot markets		Futures markets	
Commodity	CVAR	FCVAR	CVAR	FCVAR	CVAR	FCVAR	CVAR	FCVAR
Canola	110.428***	108.366***	9.976***	11.073***	311.796***	306.644***	17.238**	19.856***
	(6.308)	(6.234)	(3.862)	(3.707)	(12.859)	(12.599)	(6.782)	(6.511)
Cocoa	4.166	6.325	−1.729	0.112	3.391	12.348**	−3.980	−0.771
	(4.293)	(4.205)	(2.855)	(0.195)	(6.494)	(6.300)	(4.099)	(1.053)
Coffee	22.362***	25.209***	−3.856	−1.600	54.396***	64.245***	−5.451	0.063
	(5.799)	(5.816)	(2.789)	(2.561)	(10.933)	(10.993)	(4.010)	(4.309)
Copper	96.182***	103.297***	2.101	2.930	279.279***	298.721***	2.709	4.105
	(5.051)	(5.138)	(3.482)	(3.188)	(10.469)	(10.420)	(4.317)	(5.209)
Corn	15.988**	15.208**	11.340***	12.519***	22.870**	19.265**	16.966**	18.197**
	(6.837)	(6.156)	(4.349)	(4.602)	(9.464)	(8.515)	(7.888)	(8.089)
Crude oil	−0.260	1.373	−0.672	3.303	−2.792	−0.829	−0.472	4.669
	(4.137)	(3.913)	(0.870)	(2.774)	(8.062)	(8.119)	(2.624)	(4.259)
Gasoline	18.063***	17.586**	4.452	5.293	49.838***	53.914***	8.688	8.713
	(6.457)	(7.415)	(3.246)	(3.436)	(12.301)	(12.920)	(5.755)	(5.450)
Gold	5.062	5.641	24.049***	22.846***	−1.791	1.313	46.537***	46.650***
	(4.708)	(4.844)	(3.292)	(3.160)	(9.186)	(9.365)	(6.214)	(6.119)
Heating oil	−1.599	0.459	0.221	1.235	−5.482	−1.498	−1.505	0.760
	(3.159)	(3.014)	(2.911)	(2.577)	(5.768)	(5.482)	(5.110)	(4.339)
Palladium	65.158***	65.956***	17.816***	18.741***	166.335***	165.970***	32.535***	33.902***
	(7.832)	(7.749)	(4.435)	(4.564)	(15.381)	(15.290)	(8.345)	(8.400)
Platinum	15.974***	16.750***	0.790	3.334*	43.791***	46.825***	2.345	8.647**
	(2.781)	(2.685)	(2.482)	(1.883)	(6.111)	(5.945)	(4.720)	(3.662)
Silver	4.048	−1.404	25.397***	19.547***	−5.612	−16.779	44.150***	41.638***
	(7.060)	(7.718)	(4.820)	(3.151)	(14.193)	(14.393)	(7.072)	(6.221)
Soybean	12.789***	8.566**	0.666	1.926	26.803***	20.756***	−7.042	−0.517
	(4.824)	(3.733)	(2.659)	(3.064)	(7.123)	(5.811)	(5.994)	(6.696)
Soy meal	10.317*	11.377*	5.935	9.500**	23.614**	19.179*	20.559**	18.485**
	(6.027)	(6.196)	(4.629)	(4.031)	(9.788)	(9.977)	(8.487)	(7.993)
Soy oil	10.708*	8.789*	2.571	3.434	23.080***	19.463***	1.174	2.364
	(5.731)	(4.854)	(4.773)	(4.473)	(8.447)	(7.536)	(7.700)	(7.282)
Sugar	51.447***	48.738***	5.817	10.866*	131.696***	125.969***	9.908	14.788
	(6.818)	(7.052)	(5.872)	(6.355)	(13.138)	(13.582)	(9.564)	(9.977)
Wheat	3.279	1.912	5.330	3.501	2.827	2.303	3.085	0.577
	(5.604)	(3.020)	(3.405)	(3.709)	(6.279)	(3.230)	(6.626)	(6.547)
Average	26.124	26.126	6.482	7.562	66.120	66.930	11.026	13.066

Notes: This table reports annualized average excess portfolio percentage returns for commodity spot and futures markets for a medium risk-aversion investor ($\gamma = 6$) with daily ($h = 1$) rebalancing. The results are reported for weights restricted to the intervals $[0,1]$ and $[-1,2]$ and for forecasts based on both the CVAR and FCVAR models. Standard errors are reported in parentheses. Statistical significance at the 10%, 5%, and 1% level is denoted by one, two, and three asterisks, respectively. The final row is the average return across all commodities.

average crossover strategy. The results also show that, in spite of the statistical advantage of the FCVAR model, excess returns from the FCVAR and CVAR models are very similar although with a slight advantage to the fractional model on average in terms of both portfolio excess returns and their Sharpe ratios. Our results are robust on several fronts. First, our out-of-sample forecasting evaluation exercise applies a number of statistical metrics. Second, we analyze several forecasting horizons. Third, we show that our results on profitability are robust to an investor's level of risk aversion as measured by the coefficient of relative risk aversion that enters in the utility function and hence in the calculation of portfolio weights. Fourth, our results are superior to results obtained using a simple technical moving-average crossover trading rule. Finally, our results are robust to alternative restrictions on the optimal weights, reflecting alternative restrictions on short-selling and on borrowing/leverage.

References

1. Baillie, R.T., and Bollerslev, T. (1994) The long memory of the forward premium, *Journal of International Money and Finance*, 13, 565–571.
2. Brennan, M.J., and Schwartz, E.S. (1985) Evaluating natural resource investments, *Journal of Business*, 58, 135–157.
3. Campbell, J.Y., and Thompson, S.B., (2008) Predicting the equity premium out of sample: can anything beat the historical average?, *Review of Financial Studies*, 21, 1509–1531.
4. Clark, T.E., and West, K.D., (2007) Approximately normal tests for equal predictive accuracy in nested models, *Journal of Econometrics*, 138, 291–311.
5. Coakley, J., Dollery, J., and Kellard, N., (2011) Long memory and structural breaks in commodity futures markets, *Journal of Futures Markets*, 31, 1076–1113.
6. Diebold, F.X., and Mariano, R.S., (1995) Comparing predictive accuracy, *Journal of Business and Economic Statistics*, 13, 253–263.
7. Dolatabadi, S., Nielsen, M.Ø., and Xu, K., (2015) A fractionally cointegrated VAR analysis of price discovery in commodity futures markets, *Journal of Futures Markets*, 35, 339–356.
8. Dolatabadi, S., Nielsen, M.Ø., and Xu, K., (2016) A fractionally cointegrated VAR model with deterministic trends and application to commodity futures markets, *Journal of Empirical Finance*, forthcoming.
9. Figuerola-Ferretti, I., and Gonzalo, J., (2010) Modelling and measuring price discovery in commodity markets, *Journal of Econometrics*, 158, 95–107.
10. Garbade, K.D., and Silber, W.L. (1983), Price movements and price discovery in futures and cash markets, *Review of Economics and Statistics*, 65, 289–297.
11. Giacomini, R., and White, H., (2006) Tests of conditional predictive ability, *Econometrica*, 74, 1545–1578.
12. Gonzalo, J., and Granger, C.W.J., (1995) Estimation of common long-memory components in cointegrated systems, *Journal of Business and Economic Statistics*, 13, 27–36.
13. Graham-Higgs, J., Rambaldi, A., and Davidson, B., (1999) Is the Australian wool futures market efficient as a predictor of spot prices, *Journal of Futures Markets*, 19, 565–582
14. Hasbrouck, J., (1995) One security, many markets: determining the contribution to price discovery, *Journal of Finance*, 50, 1175–1199.
15. Jensen, A.N., and Nielsen, M.Ø., (2014) A fast fractional difference algorithm, *Journal of Time Series Analysis*, 35, 428–436.
16. Jobson, J.D., and Korkie, B.M., (1981) Performance hypothesis testing with the Sharpe and Treynor measures, *Journal of Finance*, 36, 889–908.
17. Johansen, S., (1995) *Likelihood-based inference in cointegrated vector autoregressive models.* New York: Oxford University Press.

18. Johansen, S., (2008) A representation theory for a class of vector autoregressive models for fractional processes, *Econometric Theory*, 24, 651–676.
19. Johansen, S., and Nielsen, M.Ø., (2010) Likelihood inference for a nonstationary fractional autoregressive model, *Journal of Econometrics*, 158, 51–66.
20. Johansen, S., and Nielsen, M.Ø., (2012) Likelihood inference for a fractionally cointegrated vector autoregressive model, *Econometrica*, 80, 2667–2732.
21. Johansen, S., and Nielsen, M.Ø., (2015) The role of initial values in conditional sum-of-squares estimation of nonstationary fractional time series models, *Econometric Theory*, forthcoming.
22. Kaldor, N. (1939) Speculation and economic stability, *Review of Economic Studies*, 7, 1–27.
23. Lien, D., and Tse, Y.K., (1999) Fractional cointegration and futures hedging, *Journal of Futures Markets*, 19, 457–474.
24. Locke, P.R., and Venkatesh, P.C. (1997) Futures market transaction costs, *Journal of Futures Markets*, 17, 229–245.
25. Marquering, W., and Verbeek, M., (2004) The economic value of predicting stock index returns and volatility, *Journal of Financial and Quantitative Analysis*, 39, 407–429.
26. Maynard, A., and Phillips, P.C.B., (2001) Rethinking an old empirical puzzle: econometric evidence on the forward discount anomaly, *Journal of Applied Econometrics*, 16, 671–708.
27. Miffre, J., and Rallis, G., (2007) Momentum strategies in commodity futures markets, *Journal of Banking and Finance*, 31, 1863–1886.
28. Narayan, P.K., Narayan, S., and Sharma, S.S., (2013) An analysis of commodity markets: what gain for investors?, *Journal of Banking and Finance*, 37, 3878–3889.
29. Narayan, P.K., Ahmed, H.A., and Narayan, S., (2014) Do momentum-based trading strategies work in the commodity futures markets?, *Journal of Futures Markets*, 35, 868–891.
30. Nielsen, M.Ø., and Popiel, M.K., (2014) A Matlab program and user's guide for the fractionally cointegrated VAR model, QED working paper 1330, Queen's University.
31. Szakmary, A.C., Shen, Q., and Sharma, S.C., (2010) Trend-following trading strategies in commodity futures: a re-examination, *Journal of Banking and Finance*, 34, 409–426.
32. Wang, J., (2000) Trading and hedging in S&P500 spot and futures markets using genetic programming, *Journal of Futures Markets*, 20, 911–942.

Research Papers 2018

CREATES
Center for Research in Econometric Analysis of Time Series

2018-17: Søren Johansen and Morten Ørregaard Nielsen: Nonstationary cointegration in the fractionally cointegrated VAR model

2018-18: Giorgio Mirone: Cross-sectional noise reduction and more efficient estimation of Integrated Variance

2018-19: Kim Christensen, Martin Thyrsgaard and Bezirgen Veliyev: The realized empirical distribution function of stochastic variance with application to goodness-of-fit testing

2018-20: Ruijun Bu, Kaddour Hadri and Dennis Kristensen: Diffusion Copulas: Identification and Estimation

2018-21: Kim Christensen, Roel Oomen and Roberto Renò: The drift burst hypothesis

2018-22: Russell Davidson and Niels S. Grønborg: Time-varying parameters: New test tailored to applications in finance and macroeconomics

2018-23: Emilio Zanetti Chini: Forecasters' utility and forecast coherence

2018-24: Tom Engsted and Thomas Q. Pedersen: Disappearing money illusion

2018-25: Erik Christian Montes Schütte: In Search of a Job: Forecasting Employment Growth in the US using Google Trends

2018-26: Maxime Morariu-Patrichi and Mikko Pakkanen: State-dependent Hawkes processes and their application to limit order book modelling

2018-27: Tue Gørgens and Allan H. Würtz: Threshold regression with endogeneity for short panels

2018-28: Mark Podolskij, Bezirgen Veliyev and Nakahiro Yoshida: Edgeworth expansion for Euler approximation of continuous diffusion processes

2018-29: Isabel Casas, Jiti Gao and Shangyu Xie: Modelling Time-Varying Income Elasticities of Health Care Expenditure for the OECD

2018-30: Yukai Yang and Luc Bauwens: State-Space Models on the Stiefel Manifold with A New Approach to Nonlinear Filtering

2018-31: Stan Hurn, Nicholas Johnson, Annastiina Silvennoinen and Timo Teräsvirta: Transition from the Taylor rule to the zero lower bound

2018-32: Sebastian Ankargren, Måns Unosson and Yukai Yang: A mixed-frequency Bayesian vector autoregression with a steady-state prior

2018-33: Carlos Vladimir Rodríguez-Caballero and Massimiliano Caporin: A multilevel factor approach for the analysis of CDS commonality and risk contribution

2018-34: James G. MacKinnon, Morten Ørregaard Nielsen, David Roodman and Matthew D. Webb: Fast and Wild: Bootstrap Inference in Stata Using boottest

2018-35: Sepideh Dolatabadim, Paresh Kumar Narayan, Morten Ørregaard Nielsen and Ke Xu: Economic significance of commodity return forecasts from the fractionally cointegrated VAR model