Preliminary and incomplete. Please do not quote.

USING CO-MOVEMENTS TO FORECAST COMMODITY PRICES

Kenneth D. West
University of Wisconsin

Ka-Fu Wong
University of Hong Kong

October 2012

ABSTRACT

We fit a factor model to a panel of 10 real commodity prices. Prices consistently display a tendency to revert towards the factor, though the speed of mean reversion is slow. We attempt to use the model to forecast. By a mean squared error criterion, the factor model forecasts better than a random walk about half the time. Those improvements are sometimes but not always significant at traditional levels. The factor model does a little better at short (one quarter) than long (eight quarter) horizons.

West thanks the National Science Foundation for financial support. We thank Yi Zhang for research assistance.

1. INTRODUCTION

In this paper we aim to exploit co-movements across commodity prices to better predict movements in individual commodity prices. We build on two literatures. The first literature studies prediction of commodity prices, the second studies co-movements of commodity prices.

The literature on prediction of commodity prices is vast. Two recent papers that evaluate forecasts for a range of commodity prices are Groen and Pesenti (2010) and Chinn and Coibion (2012); two recent papers that evaluate forecasts for a specific commodity price are Alquist et al. (2011), who consider oil, and Bernard et al. (2008), who consider aluminum. Our reading of this literature is that simple time series models generally forecast as well or better than models relying on futures data or on measures of the level of economic activity.

A substantial literature views co-movement as a central and distinctive characteristic of commodity prices. This is evident in much popular discussion of commodity prices, e.g. July 2012 articles in the *Economist* and the *Financial Times*.[1] Scholarly studies referencing "supercycles" that may have characterized commodity price movements as long ago as the 19th century include Cuddington and Jerrett (2008) and Erten and Ocampo (2012). And scholarly literature focusing, as we do, on recent decades has looked at or used co-movements in a variety of ways: to consider whether such co-movements are excessive in a certain precise sense (Pindyck and Rotemberg (1990)); to forecast inflation (Gospodinov and Ng (2011)); to characterize in an atheoretical way the pattern of co-movements (Gomez et al. (2011)); to evaluate the extent to which oil price movements translate into movements in a wide range of commodity prices (Baffes (2007)).

Beginning with the pioneering work of Sargent and Sims (1976), factor models have been used by economists to capture co-movement of series.[2] Perhaps the commodity paper that is closest to ours in modeling strategy is Byrne et al. (2011), who fit a factor model to a panel of commodity prices and attempt to correlate the fitted factor to observable macro variables. The forecasting paper that is closest to ours is

Engel et al. (2012), who evaluate forecasts of exchange rates with techniques similar to the ones we apply.

We use quarterly data on a panel of 10 real commodity prices, consisting of oil, coal and metals. We deflate our nominal series by U.S. CPI to control for what we view as a likely but uninteresting tendency for general increases in the overall price level to be reflected in general upward movement of commodity prices. We fit a simple factor model to the panel. We compare 1, 4 and 8 quarter forecasts from our factor model to forecasts from a random walk model. By a mean squared error criterion, our factor model does better than the random walk model about half the time. The factor model does relatively well at the one quarter horizon (in the baseline model, 6 of 10 mean squared prediction errors (MSPEs) are smaller) than at the 4 quarter (5 MSPEs smaller) or 8 quarter (3 MSPEs smaller) horizons. The reductions in root MSPE were small, generally less than 5 percent. When the factor model produced a smaller MSPE, the reduction was usually statistically significant at the 10 percent level. A simulation test of whether there was at least one significant reduction in MSPE over all 30 comparisons (30 = 3 horizons × 10 commodities) also yielded a p-value less than 10 percent.

We close this introduction with some cautions. First, we take as given our focus on out of sample analysis, which we and others have found informative. But we recognize that some economists might disagree. See, e.g., Inoue and Kilian (2004) and Diebold (2012). Second, we looked at the data, and of course were familiar with research that others have done with similar data, before formulating our forecasting model. Hence our forecasting results are not true out of sample in the strict sense.

Section 2 describes our empirical models, section 3 our data and forecast evaluation techniques. Section 4 presents empirical results, section 5 robustness checks. Section 6 concludes. An appendix contains detailed empirical results that are concisely summarized in one of the tables in the paper.

2. THE FACTOR MODEL

Our baseline model uses one factor. Models with two or more factors are discussed below as part of our checks for robustness. Our first step is to estimate the factor and factor loadings from the levels of the real commodity prices. Let P_{it} be the US dollar price of commodity i. Let $P_{CPI,t}$ be the US CPI. Let p_{it} = $100 \times \ln(P_{it}/P_{CPI,t})$ be the real price of commodity i. For commodity i, i=1, ...,10, the factor model is

$$(2.1) \qquad p_{it} = \text{constant} + \delta_i f_t + v_{it}$$

$$\equiv \text{constant} + F_{it} + v_{it}.$$

The factor f is unobserved. The constant is chosen so that v_{it} has mean zero.

Some previous research (e.g., Dahl and Eglesias (2009)) has found real commodity prices to be I(1). Our procedures are applicable with or without unit roots. If prices are I(1), then f is I(1); if prices are stationary then, f is stationary. Here and throughout, we do not attempt to test for unit roots in the factor or any other variable for that matter. See Bai (2004) and Stock and Watson (2006) on estimation of factor models with unit root or with stationary data.

In contrast to research such as Groen and Pesenti (2010), the factor is not constructed from a set of fundamental variables reflecting the state of the economy or of supply and demand in a particular industry. Instead, like Byrne et al. (2011), we construct the factor from the commodity prices themselves. Using an estimate of $F_{it} \equiv \delta_i f_t$, we shall forecast p_{it}. Our supposition is that $\delta_i f_t$ is a central tendency toward which commodity prices tend to revert: after controlling for means via the "constant" in (2.1), we suppose that when p_{it} is above F_{it}, p_{it} subsequently tends to fall; when p_{it} is below F_{it}, p_{it} subsequently tends to rise. Of course the factor f_t is a linear combination of the commodity prices. As we shall see, in our data the estimated δ_i's are all positive. Thus our presumption can be restated as: when p_{it} is above a weighted averaged of commodity prices, p_{it} tends to rise; when p_{it} is below a weighted averaged of commodity prices, p_{it} tends to fall.

An example worked out in Engel et al. (2012) illustrates why a model such as (2.1) might be useful in forecast. In that example, F_{it} follows a random walk with an innovation uncorrelated with v_{it}, which is assumed to be i.i.d.. In plausibly calibrated examples, autocorrelations for Δp_{it} are very close to zero, with a random walk for p_{it} a good time series approximation. Nonetheless, for one step ahead predictions, population mean square prediction errors (MSPEs) from the factor model are, by the standards of financial forecasting, distinctly smaller than those from a random walk model.

We do not maintain that model or for that matter any particular time series process for F_{it}. What we do assume is that if the data are I(1), cointegration is such that there is a single unit root. Specifically, for a one factor model, we assume F_{it}-$p_{it} \equiv$ constant + v_{it} is stationary for all i and may be useful in predicting (stationary) future changes in p_{it}. Indeed, as noted in Diebold et al. (1994), cointegration across a set of prices implies that at least one price is not a random walk and the multivariate representation of the first difference of such a price will involve an error correction term. In our model, if the data are I(1) that error correction term is F_{it}-p_{it}. Our decision to use a single factor (i.e., a single unit root) in our baseline model was based on presumed limitations of a panel of cross-section dimension 10 (see, e.g., Ho and Sorenson (1996)). As shown below, our results are not sensitive to allowing 2 or 3 rather than 1 factor. We normalize f to have mean zero and unit variance.

So a first stage produces a time series for $\hat{f}_t$ for factor loadings, $\hat{\delta}_i$ and $\hat{F}_{it}$, i=1,...,10. To illustrate our second stage, which produced forecasts, consider a horizon of 4 quarters. We estimate

(2.2) $$p_{it+4}-p_{it} = \alpha_i + \beta_1(\hat{F}_{it}-p_{it}) + \beta_2(\hat{F}_{it-1}-p_{it-1}) + u_{it+4},$$

where α_i is a fixed effect for commodity i. We then use $\hat{\alpha}_i$, $\hat{\beta}_1$ and $\hat{\beta}_2$ to predict. Details are given in the next section. Our presumption that commodity prices revert towards rather than away from F-p implies that we expect the estimates of β_1+β_2 to be positive.

3. DATA AND FORECASTING EVALUATION

Our data are quarterly data, 1980:1-2012:2. We arbitrarily began the out of sample period in 1990. The data source is the International Monetary Fund. We selected one series for each of the minerals in the data base. The qualifier "one" means that, for example, we only used one of the two series for coal and one of the many series for oil and gas. The qualifier "minerals" means that we dropped the majority of the series in the IMF database, because these were agricultural or food. An initial examination of the data indicated that, for most of the sample, the iron ore nominal price was adjusted once per year. This indicated a need for special treatment by an obvious model, and we dropped iron ore. We ended up with the 10 commodities listed in Table 1.

The IMF data are monthly average of daily prices rather than point in time. Now, if the real commodity price series followed random walks, averaging would produce series with first order autocorrelations of 0.25. Indeed, first order autocorrelations of the differenced real monthly series yielded estimates ranging from .20 to .36 for 9 of the 10 commodities (the exception happened to be uranium, whose first order autocorrelation was 0.06). We feared that time aggregation would drive our results. So to attenuate possible effects from time aggregation, we constructed nominal quarterly series by sampling the last month in the quarter. We converted to real using monthly CPI, all consumers.

Basic statistics for the resulting levels and first differences are in Table 2. As explained above, our modeling strategy is robust to the presence or absence of unit roots. We do note that the first order autocorrelations of the levels presented in the "ρ_1" line of panel A of Table 2 are all above 0.9. This is suggestive of near unit root or unit root behavior. In panel B, we see from the lines for mean and s.d. that the differenced series seem to have little drift: for all series, the absolute value of the point estimate of the mean is far below (s.d./$\sqrt{T}$). Hence the growth rate of the nominal commodity prices has tracked that of overall CPI. An approximate $1/\sqrt{T} \approx .09$ standard error for the first order autocorrelation ρ_1 suggests some significant first order serial correlation, in particular in coal and tin and to a lesser extent in aluminum, nickel and uranium. This justifies the lagged term $\hat{F}_{it-1}$-p_{it-1} in (2.2) above.

Let us illustrate the mechanics of our forecasting algorithm using the four quarter horizon (h=4), for the first forecast. As shown in (3.1) below, we use data from 1980:1 to 1989:4 to estimate the factor and factor loadings, and construct $\hat{F}_{it}$ for i=1,...,10. The estimation technique is principal components.

```
(3.1)   -------------data used to estimate factor--------------
        ----data used to estimate panel regression----
        |______________________________________|_____|_____|
                                            88:4  89:4  90:4
```

We then use right hand side data from 1980:2 to 1988:4 to estimate a fixed effects panel data regression

(3.2) $p_{it+4}\text{-}p_{it} = \alpha_i + \beta_1(\hat{F}_{it}\text{-}p_{it}) + \beta_2(\hat{F}_{it\text{-}1}\text{-}p_{it\text{-}1}) + u_{it+4}$, t=1980:2,...,1988:4.

We use 1989:4 data to predict the 4 quarter change in p:

(3.3) Prediction of $(p_{i,1990:4}\text{-}p_{i,1989:4}) = \hat{\alpha}_i + \hat{\beta}_1(\hat{F}_{i,1989:4}\text{-}p_{i,1989:4}) + \hat{\beta}_2(\hat{F}_{i,1989:3}\text{-}p_{i,1989:3})$.

We then add an observation to the end of the sample, and repeat. For a 4 quarter horizon, this resulted in 87 predictions (first prediction is for t+4 =1990:4, last is for 2012:2). The comparable figures for h=1 quarter and h=8 quarter horizons are 90 (first prediction is for t+1 = 1990:1, last = 2012:2) and 83 (first prediction is for t+8 = 1991:4, last=2012:2).

As is indicated by this discussion, the recursive method is used to generate predictions: observations are added to the end of the estimation sample, so that the sample size used to estimate factors and panel data regressions grows. The direct (as opposed to iterated) method is used to make multiperiod predictions.

For a given date, the factor, factor loadings and right hand side variables are identical across horizons: for given t, the same values of $\hat{F}_{it}\text{-}p_{it}$ and $\hat{F}_{it\text{-}1}\text{-}p_{it\text{-}1}$ are used for h=1, 4 and 8. However, the left hand side variable is different (h period difference in p_{it}), and regression samples are smaller for larger h. This means that for a given estimation sample, the regression coefficients ($\hat{\alpha}_i$, $\hat{\beta}_1$, $\hat{\beta}_2$) and predictions vary with h.

We compare our factor model to a random walk. In the random walk model, the forecasted change in the commodity price is zero. Our measure of forecast performance is root mean squared prediction error (RMSPE). For a random walk model, for example, MSPE for h=4 is computed as

(3.4) $\frac{1}{87}\sum_t(p_{it+4}-p_{it})^2$,

where the sum runs over the 87 predictions from t+4 = 1990:4 to t+4 = 2012:2.

We present relative RMSPE's: Theil's U-statistic, the ratio of the RMSPE from our factor model to the alternative model. In terms of point estimates, our measure of success for the factor model is to produce a U-statistic less than one. This means that the RMSPE is lower for the factor model than for the alternative model, i.e., the random walk. A U-statistic of exactly 1 indicates that the sample RMSPEs from the factor model and the alternative model are the same. As argued by Clark and West (2006), if the alternative model is a random walk, a U-statistic of 1 is evidence *against* the random walk model. If, indeed, a random walk generates the data, then the factor model introduces spurious variables into the forecasting process. In finite samples, attempts to use such variables will, in expectation, introduce noise that inflates the variability of the forecasting error of the factor model. Hence under a random walk null, we expect sample U-statistics greater than 1, even though that null implies that population ratios of RMSPEs are 1.

We report t-statistics that we use to make one sided tests on H_0: RMSPE(our model) = RMSPE(random walk) against H_A: RMSPE(our model) < RMSPE(random walk). These t-statistics are constructed in accordance with Clark and West (2006, 2007), who develop a test procedure that accounts for the potential inflation of the factor model's RMSPE noted in the previous paragraph. This procedure adjusts the differences in MSPEs and then conducts a standard t-test, a.k.a. a Diebold-Mariano-West (DMW) test. When the horizon h>1, we constructed the standard error as in West (1997), which under a random walk null is also the procedure proposed in Hodrick (1992).

Of course, with ten rather than just a single commodity, it is possible that one or more test statistics will be significant even if in population the factor model's RMSPE is not smaller than that of the random walk for any of the 10 commodity prices. Since alternative in the test is one-sided, a chi-square test of the sort described in West (1996, 2006) is not appealing in terms of power. Instead, we guarded against the possibility that our multiple pairwise tests were yielding spuriously significant t-statistics by testing

H_0: RMSPE(our model) = RMSPE(random walk) for all commodities

against

H_A: RMSPE(our model) < RMSPE(random walk) for at least one commodity.

We did so by looking at the maximum of a given set of t-statistics. The "given set" was either (1)the 10 t-statistics for a given horizon of 1, 4 or 8 quarters, which yielded 3 separate p-values for the maximum t-statistic at each of the three horizons, or (2)for all 30 t-statistics, which yielded a single p-value. We instantiated this test with the logic of Hubrich and West (2010).

Specifically, we generated 1000 samples by bootstrapping under a random walk null–that is, a null that commodity price changes are i.i.d. over time though possibly correlated across commodities. We did so as follows. We demeaned changes in commodity prices. The initial vector of prices for t=1 was set to actual price data. Then, for t=2, ..., 130 we sampled with replacement from demeaned changes in commodity prices, for each t drawing a 10×1 vector of demeaned price changes. For each t, the vector that was drawn was added to the previous period's vector of prices to obtain prices at period t. Once the 130 observation sample was constructed, we replicated the steps in our empirical work. We saved the largest of the 10 t-statistics for each horizon. We sorted the t-statistics from low to high and report the fraction of samples in which the maximum t-statistic exceeded that in our sample.

4. EMPIRICAL RESULTS

Our focus is on out of sample prediction. But first a bit of summary on our sequence of in-sample regressions. Table 3 presents the factor loadings δ_i for the largest sample for which we estimated the factor model. These estimates are a little unrepresentative of the sequence of estimates in that the estimated δ_i's in Table 3 are clustered in a narrow range between 0.66 and 0.90. Other samples included an estimate below 0.2 along with another above 0.9. A central point, however, is that the Table 3 estimates are representative in that the loadings were positive. This was the case for all commodities for all samples.

As noted above, our presumption that commodity prices revert towards F-p implies that in (2.2) we expect the estimates of $\beta_1+\beta_2$ to be positive. Indeed all 260 of our estimates of $\hat{\beta}_1+\hat{\beta}_2$ were indeed positive (260=90+87+83 estimates for h=1, 4 and 8 respectively). Median values were as follows:

(4.1)	h=1	h=4	h=8
	0.070	0.301	0.546

Thus mean reversion, while present, is modest. At the one quarter horizon, for example, we see from (4.1) that if current and lagged F is (say) one percent above p, then, in sample, next quarter's Δp falls, on average, by 0.070 percent. Over 8 quarters mean reversion is barely over half the initial gap. Thus these in-sample estimates suggest mean reversion ($\hat{\beta}_1+\hat{\beta}_2$ always positive), with that mean reversion being slow.

The rest of our discussion is on out of sample results. For i=zinc, Figure 1 plots a pasted together sequence of estimates of $\hat{F}_{it} \equiv \hat{\delta}_i\hat{f}_t$ for i=zinc along with zinc price p_{it} that has been standardized to have mean 0 (for reasons explained below) and variance one (to improve readability). The sequence $\hat{F}_{it}$ has been constructed as follows: the first 40 observations are the 40 fitted values of $\hat{\delta}_i\hat{f}_t$ produced by the first (1980:1-1989:4) of the set of samples used to estimate the factor model. Each successive observation is the last fitted value from the sample that ends on that date; for example, the observation at 1990:1 is the last fitted value from the estimate from the 1980:1-1990:1 sample. Thus each observation is the value of $\hat{F}_{it}$ that was used to make predictions from that forecast base. For our model, a key determinant of whether we predict a commodity price to rise or to fall is whether the last fitted value of $\hat{F}_{it}$ is above demeaned $\hat{p}_{it}$; standardization of p_{it} preserves whether or not $\hat{F}_{it}$ is above demeaned $\hat{p}_{it}$.[3]

Let us use Figure 1 to illustrate our forecasting presumption, which was described above as "when p_{it} is above F_{it}, p_{it} subsequently tends to fall; when p_{it} is below F_{it}, p_{it} subsequently tends to rise." We will point to one forecast base where our presumption was supported, and another where it was contradicted. Zinc, and these two forecast bases, are chosen because the outcome was stark. One forecast base where our presumption is supported is 1990:2. In that quarter, the standardized level of zinc prices was about 1.5, well above the slightly positive level of $\hat{F}_{it}$. The mean reversion that we presume thus suggests that p_{it} will fall over coming quarters. One can see that zinc prices did indeed fall over the next 1, 4 and 8 quarters. For an example where our presumption is vilified rather than confirmed, consider 2001:2. One can see that p_{it} is below $\hat{F}_{it}$. Our presumption is thus that a rise toward $\hat{F}_{it}$ will occur in subsequent quarters. Nonetheless–alas!–zinc prices in fact are lower 1, 4 and 8 quarters subsequently.

Figure 2 depicts the same information, but for *i*=oil. This plot is rather more muddled, and probably is more representative of the remaining commodity prices than is Figure 1 for zinc.

One indicator of the strength of mean reversion is persistence of deviations of p_{it} from F_{it}. One can see in Figure 1 and Figure 2 that prices stay as long as several years on one side of $\hat{F}_{it}$. So mean reversion, if it is present, is slow. But indirect evidence of mean reversion comes from comparing crossings by p_{it} of $\hat{F}_{it}$ to crossings by p_{it} of its mean (the zero line). If one counts such crossings in Figure 1, one finds that zinc prices cross $\hat{F}_{it}$ 20 times and their own mean 14 times (and of course by construction prices must cross their own mean at least once). In Figure 2, the comparable counts are oil prices crossing $\hat{F}_{it}$ 19 times, their own mean 6 times. Since these prices arguably are I(1), the fact that prices cross $\hat{F}_{it}$ more often than their own mean is at best weak and qualitative evidence of mean reversion of p_{it} towards F_{it}. For quantitative evidence we turn to our forecasting results.

Table 4 presents our forecasting results. As is typical with financial data, forecast improvements, if any, are modest at the one quarter horizon. The lowest U-statistic in the *h*=1 column is for aluminum, for which the factor model lowered the RMSPE by 3.7%. For *h*=1, the factor model beat the random walk for 6 of the 10 commodities; 4 of these improvements were significant at the 10 percent level. The comparable figures for *h*=4 and *h*=8 were improvements in 5 and 3 commodities, with 3 and 2 of these significant at the 10 percent level. In these longer horizons, some of the U-statistics were not only below 1 but well below 1. The global minimum U-statistic is for zinc, *h*=8. For this commodity and horizon, the factor model lowered the RMSPE by 15.8%.

Summing over horizons, we find that the factor model had lower RMSPE than the random walk in

14 of the 30 comparisons, or about half the time. In these 14 instances, the t-test was significant at the 10 percent level in 9 comparisons.

According to the "p-value max-*t*" entries in Table 4, once we control for the fact that we did multiple comparisons, a statistically significant improvement continues to be found at the 1 and 4 quarter horizons but not at the 8 quarter horizon. Finally, when we control for multiple comparisons across all 30 entries across all three horizons, we see that the p-value, at 0.045, is less than the conventional 0.10 value. So we conclude that the improvement of the factor model in forecasts of at least one commodity and horizon does not seem to be due to chance.

To depict what underlies a U-statistic of various values, Figure 3 plots realizations vs. predictions for zinc, while Figure 4 does the same for oil. Zinc was chosen because predictions were unusually good for zinc. Oil was chosen because it is probably the commodity of greatest interest, and because our predictions were unusually poor for oil at the h=4 and h=8 horizons. To make the figures readable, the scale of the horizontal axis, which has predictions, and the scale of the vertical axis, which has realizations, are not comparable: as one would expect, predictions are far less variable than are realizations.

In Figure 3, an increasingly positive correlation between predictions and realizations is apparent, at least to our eyes. This is consistent with the fall in U-statistics for zinc in Table 4, from 0.981 to 0.903 to 0.842 as the horizon goes from h=1 to h=4 to h=8. In Figure 4, no such increase is evident for oil forecasts. We observe that the figures suggest a modest positive correlation between prediction and realization. Such a correlation, which is the foundational motivation for our modeling strategy, is indeed present: the sample correlation between prediction and realization is 0.12 for h=1, 0.10 for h=4 and 0.14 for h=8. For h=4 and h=8, the RMSPE is nonetheless larger for the factor model than the random walk because of a nonzero correlation between the factor model prediction and factor model prediction error.

Indeed, in 29 of our 30 sets of predictions (30 = 10 commodities × 3 horizons), the correlation between factor model forecast and realization was positive. (The exception happened to be tin, h=8.) The median correlations between factor model forecast and realizations were about 0.13 (h=1), 0.16 (h=4) and 0.22 (h=8). These values are modest, but the fact that the correlation is consistently positive is confirmation of our basic supposition that commodity prices can be expected to mean revert towards the common factor.

5. ROBUSTNESS

We report two robustness tests. First, to check the sensitivity of our results to particular sample periods, Figure 5 graphs recursively computed U-statistics for the h=1 and h=8 horizons for zinc and oil. Note that the vertical scale is larger for the h=8 graph than for the h=1 graph. The initial value in the graphs–1990:1 (h=1) or 1991:4 (h=8)–is computed from a single observation. The number of observations used in computing the U-statistics increases through the sample. As noted above, the number of observations used to compute the final value in 2012:2 is 90 for h=1 and 83 for h=8. The final values in the graphs, in 2012:2, are the ones reported in the table and text above. For example, the 2012:2 value for zinc is 0.847.

Of course, the initial values in the graphs fluctuate quite a bit. But once a couple of years worth of observations have been accumulated, the values settle down. It does not appear that our results would change qualitatively if we knocked a few years off the sample.

For our second robustness check, we repeated the forecast comparison varying the basic specification as follows: we allowed 1 or 3 rather than 2 terms in $\hat{F}_{it}$-p_{it} in (2.2); next, keeping 2 terms in $\hat{F}_{it}$-p_{it} in (2.2), we allowed 2 or 3 rather than 1 factor. Table 5 concisely summarizes results, with details analogous to Table 4 presented in the Appendix. Line (1) in Table 5 repeats results in Table 4, for convenience of comparison. We see that varying the number of terms in $\hat{F}_{it}$-p_{it} or the number of factors little affects the results.

6. CONCLUSIONS

We fit a factor model to a panel of commodity prices. We find that commodity prices tend to mean revert towards the factor. Such mean reversion is sufficiently strong and precise that in about half of our pseudo out of sample comparisons, the factor model predicted better than a random walk by a mean squared error criterion.   This improvement in prediction generally was significant at traditional levels.

Tasks for future research include expanding the set of commodities, using higher frequency data, and combining factor information with industry and macroeconomic data.

FOOTNOTES

1.*Economist* July 28, 2012, “Downhill cycling: A peak may be in sight for commodity prices,” www.economist.com/node/21559647; *Financial Times* July 17, 2012, “Supercycle runs out of steam – for now”, www.ft.com/intl/cms/s/0/ba9b6d96-cb3d-11e1-916f-00144feabdc0.html#axzz27mmmjhV0.

2. We use “factor” model as shorthand for the class of models that include what, in a more nuanced discussion, would instead be called index models or principal components models.

3. This statement is subject to the qualification that in Figure 1 we demeaned using the mean of the entire sample, while the demeaning consistent with fixed uses the mean of the estimation sample.

APPENDIX

This appendix contains details on the results summarized in Table 5. The format is identical to Table 4. See the notes to Table 4 for details.

Table A1
Forecast Comparisons, Number of Factors =1, Number of Terms in $\hat{F}_{it}$-p_{it}=1

Commodity	----		Horizon		----	
	h=1		*h*=4		*h*=8	
	U	*t*	U	*t*	U	*t*
Aluminum	0.980	2.061	0.918	2.503	0.881	1.690
Coal	1.000	0.581	0.986	0.781	1.021	0.464
Copper	1.008	0.405	1.024	0.521	1.036	0.399
Lead	0.999	0.650	0.984	0.792	1.029	0.404
Nickel	0.995	0.895	0.949	1.455	0.885	1.195
Rubber	1.005	0.449	1.015	0.597	1.020	0.533
Tin	1.031	-0.038	1.080	0.287	1.171	0.094
Uranium	1.044	-0.167	1.109	0.144	1.133	0.379
Zinc	0.987	1.349	0.916	2.197	0.838	2.027
Oil	0.995	0.856	1.028	0.391	1.115	0.183
#U<1 / t>1.282	5	2	5	3	3	2
p-value max-t		[0.175]		[0.074]		[0.156]
p-value max-t, *h*=1,4,8		[0.066]				

Table A2
Forecast Comparisons, Number of Factors =1, Number of Terms in $\hat{F}_{it}$-p_{it}=3

Commodity	----		Horizon		----	
	h=1		*h*=4		*h*=8	
	U	*t*	U	*t*	U	*t*
Aluminum	0.974	1.796	0.962	1.616	0.921	1.442
Coal	0.973	1.454	1.006	0.691	1.028	0.416
Copper	1.016	0.226	1.046	0.292	1.056	0.313
Lead	0.992	0.794	0.998	0.585	1.046	0.346
Nickel	0.975	1.662	0.942	1.390	0.898	1.108
Rubber	1.013	0.326	1.040	0.306	1.037	0.430
Tin	1.016	0.635	1.091	0.278	1.180	0.008
Uranium	1.030	0.516	1.103	0.331	1.156	0.300
Zinc	0.967	2.399	0.898	2.483	0.851	1.857
Oil	1.023	-0.210	1.072	0.056	1.136	0.130
#U<1 / t>1.282	5	4	4	3	3	2
p-value max-t		[0.087]		[0.061]		[0.222]
p-value max-t, *h*=1,4,8		[0.064]				

Table A3
Forecast Comparisons, Number of Factors =2, Number of Terms in $\hat{F}_{it}$-p_{it}=2

Commodity	----		Horizon		----	
	h=1		*h*=4		*h*=8	
	U	*t*	U	*t*	U	*t*
Aluminum	0.963	2.568	0.949	1.873	0.902	1.550
Coal	0.964	1.520	0.992	0.842	1.027	0.434
Copper	1.002	0.543	1.036	0.376	1.047	0.347
Lead	1.006	0.426	0.991	0.678	1.039	0.372
Nickel	0.979	1.766	0.940	1.475	0.893	1.136
Rubber	0.997	0.811	1.027	0.476	1.029	0.489
Tin	1.004	0.827	1.083	0.324	1.179	0.053
Uranium	1.035	0.054	1.101	0.299	1.147	0.347
Zinc	0.981	1.964	0.904	2.491	0.844	1.938
Oil	0.996	0.642	1.052	0.210	1.128	0.159
#U<1 / t>1.282	6	4	5	3	3	2
p-value max-t		[0.049]		[0.054]		[0.186]
p-value max-t, *h*=1,4,8		[0.047]				

Table A4
Forecast Comparisons, Number of Factors =3, Number of Terms in $\hat{F}_{it}$-p_{it}=2

Commodity	----		Horizon		----	
	h=1		*h*=4		*h*=8	
	U	*t*	U	*t*	U	*t*
Aluminum	0.963	2.559	0.947	1.883	0.898	1.549
Coal	0.964	1.513	0.993	0.832	1.028	0.427
Copper	1.002	0.544	1.036	0.377	1.047	0.347
Lead	1.006	0.422	0.991	0.676	1.039	0.371
Nickel	0.978	1.782	0.939	1.486	0.892	1.145
Rubber	0.997	0.804	1.027	0.469	1.029	0.479
Tin	1.004	0.815	1.084	0.312	1.182	0.043
Uranium	1.035	0.053	1.101	0.299	1.147	0.347
Zinc	0.981	1.927	0.906	2.449	0.846	1.908
Oil	0.996	0.649	1.051	0.211	1.126	0.161
#U<1 / t>1.282	6	4	5	3	3	2
p-value max-t		[0.052]		[0.064]		[0.202]
p-value max-t, *h*=1,4,8		[0.048]				

REFERENCES

Alquist, Ron, Lutz Kilian and Robert J. Vigfusson, 2011, "Forecasting the Price of Oil", manuscript, University of Michigan.

Baffes, John, 2007, "Oil Spills on Other Commodities," *Resources Policy* 32. 126–134.

Bai, Jushan, 2004, "Estimating Cross-Section Common Stochastic Trends in Nonstationary Panel Data," *Journal of Econometrics*, 137-183.

Bernard, Jean-Thomas, Lynda Khalaf, Maral Kichian and Sebastien Mcmahon, "Forecasting Commodity Prices: GARCH, Jumps, and Mean Reversion," 2008, *Journal of Forecasting* 27, 279–291.

Byrne, Joseph P. , Giorgio Fazio, and Norbert Fiess, 2011, "Primary Commodity Prices: Co-movements, Common Factors and Fundamentals," World Bank Policy Research Working Paper 5578.

Chinn, Menzie D. and Olivier Coibion, 2012, "The Predictive Content of Commodity Futures," manuscript, University of Wisconsin.

Clark, Todd E. and Kenneth D. West, 2006, "Using Out-of-Sample Mean Squared Prediction Errors to Test the Martingale Difference Hypothesis," *Journal of Econometrics* 135 (1-2), 155-186.

Clark, Todd E. and Kenneth D. West, 2007, "Approximately Normal Tests for Equal Predictive Accuracy in Nested Models," *Journal of Econometrics*, 138(1), 291-311.

Cuddington John T. and Daniel Jerrett, 2008, "Super Cycles in Real Metals Prices?", *IMF Staff Papers* 55(4), 541-565.

Dahl, Christian M. and Emma M. Iglesias, 2009, "Volatility spill-overs in commodity spot prices: New empirical results," *Economic Modelling* 26 (3), 601-607.

Diebold, Francis X., 2012, "Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests," manuscript, University of Pennsylvania.

Diebold, Francis X.,, Javier Gardeazabal, and Kamil Yilmaz, 1994, " On Cointegration and Exchange Rate Dynamics," *Journal of Finance* XLIX, 727-735.

Engel, Charles, Nelson M. Mark and Kenneth D. West, 2012, "Factor Model Forecasts of Exchange Rates," NBER Working Paper No. 18382.

Erten, Bilge and José Antonio Ocampo, 2012, "Super-cycles of commodity prices since the mid-nineteenth century" DESA Working Paper No. 110.

Gomez, David Matesanz, Guillermo J. Ortega, Benno Torgler, and German Dabat, 2011, "Co-movements in commodity prices: A note based on network analysis," manuscript, CREMA.

Groen, Jan J. J., and Paolo A. Pesenti, 2010, "Commodity Prices, Commodity Currencies, and Global Economic Developments," NBER Working Paper 15743.

Gospodinov, Nikolay and Serena Ng, 2011, "Commodity Prices, Convenience Yields and Inflation," manuscript, Concordia University.

Hodrick, Robert J., 1992, "Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement," *Review of Financial Studies* 5, 357-386.

Ho, Mun S. and Bent E. Sørensen, 1996, "Finding Cointegration Rank in High Dimensional Systems Using the Johansen Test: An Illustration Using Data Based Monte Carlo Simulations," *Review of Economics and Statistics*, 726-732.

Hubrich, Kirstin and Kenneth D. West, 2010, "Forecast Comparisons for Small Nested Model Sets," *Journal of Applied Econometrics*. 25, 574-594.

Inoue, Atsushi and Lutz Kilian, 2004, In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use?, *Econometric Reviews* 23(4), 371–402,

Pindyck, Robert S. and Julio J. Rotemberg, 1990, "The Excess Co-movement of Commodity Prices," *Economic Journal* 100, 1173-1189.

Sargent, Thomas J. and Christopher A. Sims, 1977, "Business Cycle Theory Without Pretending to Have Too Much Economic Theory," 45-109 in *New Methods in Business Cycle Research: Proceedings From a Conference*, Federal Reserve Bank of Minneapolis.

Stock, James H. and Mark W. Watson, 2006, "Forecasting with Many Predictors," 515-550 in *Handbook of Economic Forecasting*, Vol 1, G. Elliott, C.W.J. Granger and A. Timmermann (eds), Amsterdam: Elsevier.

West, Kenneth D., 1996, "Asymptotic Inference About Predictive Ability," *Econometrica* 64 , 1067-1084.

West, Kenneth D., 1997, "Another Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Journal of Econometrics* 76, 171-191.

West, Kenneth D., 2006, "Forecast Evaluation," 100-134 in *Handbook of Economic Forecasting*, Vol. 1, G. Elliott, C.W.J. Granger and A. Timmerman (eds), Amsterdam: Elsevier.

.

Figure 1

$\hat{F}_{it}$ and standardized p_{it}, i=zinc

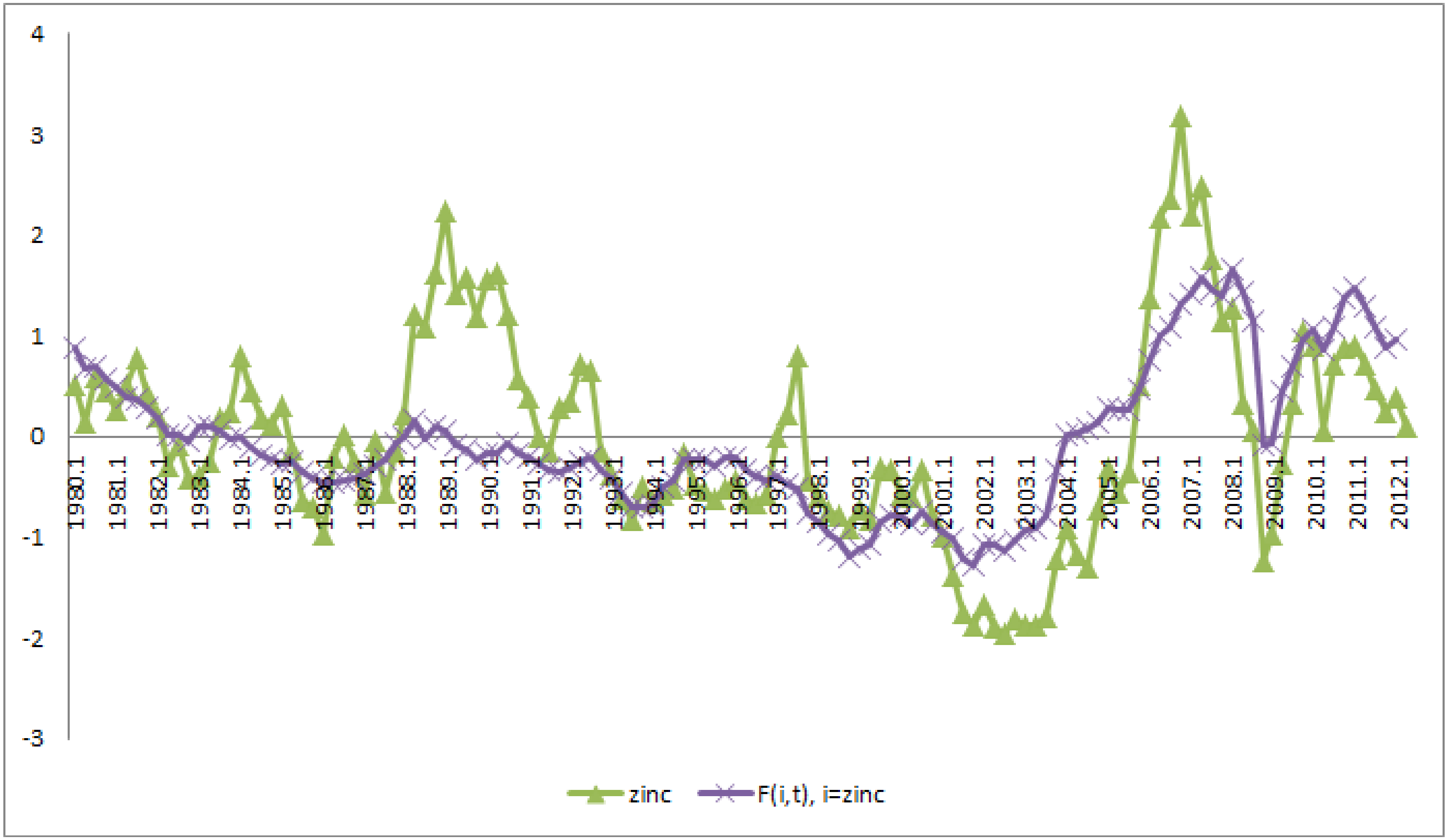


Figure 2

$\hat{F}_{it}$ and standardized p_{it}, i=oil

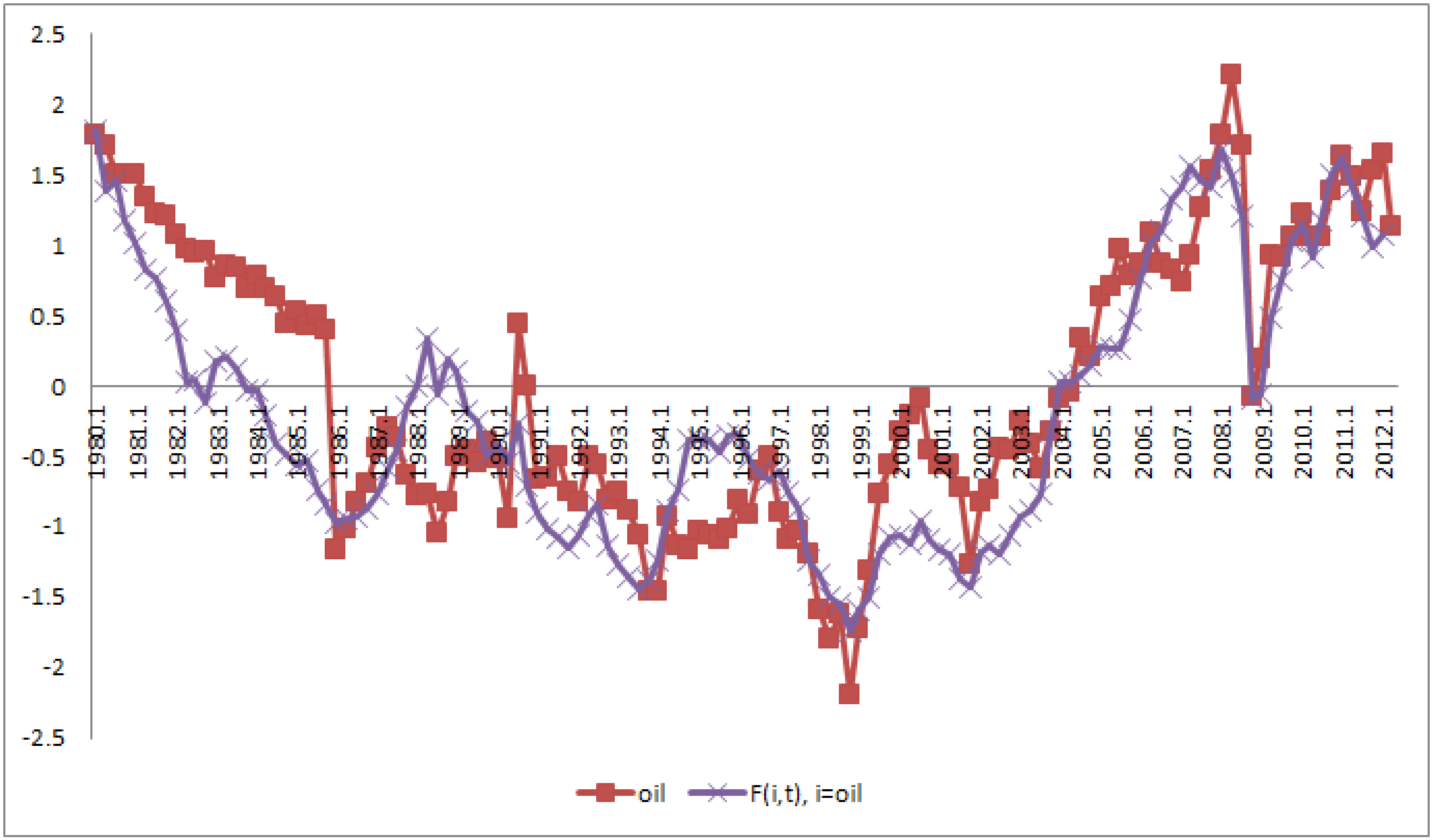


Figure 3

Actual vs. Predicted Change in Real Zinc Prices

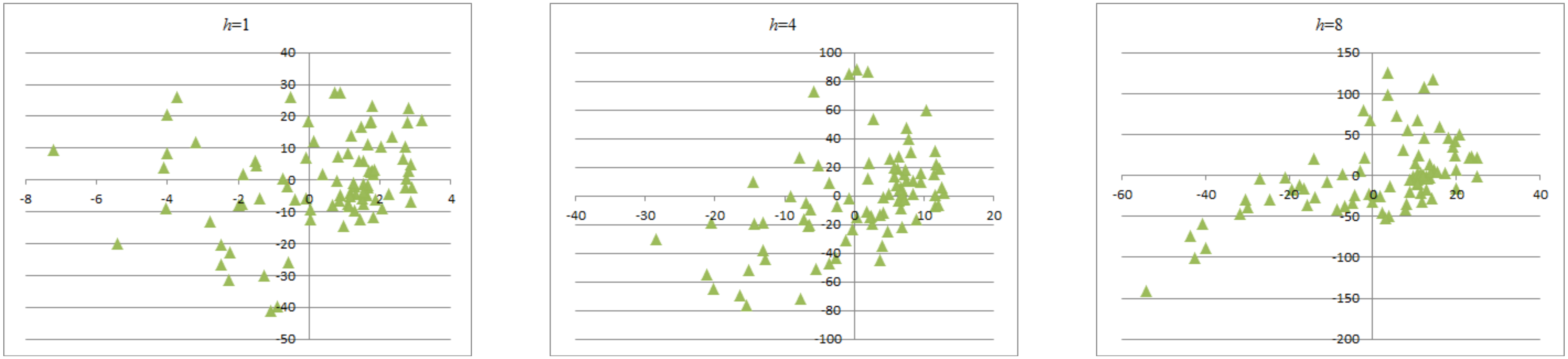


Figure 4

Actual vs. Predicted Change in Real Oil Prices

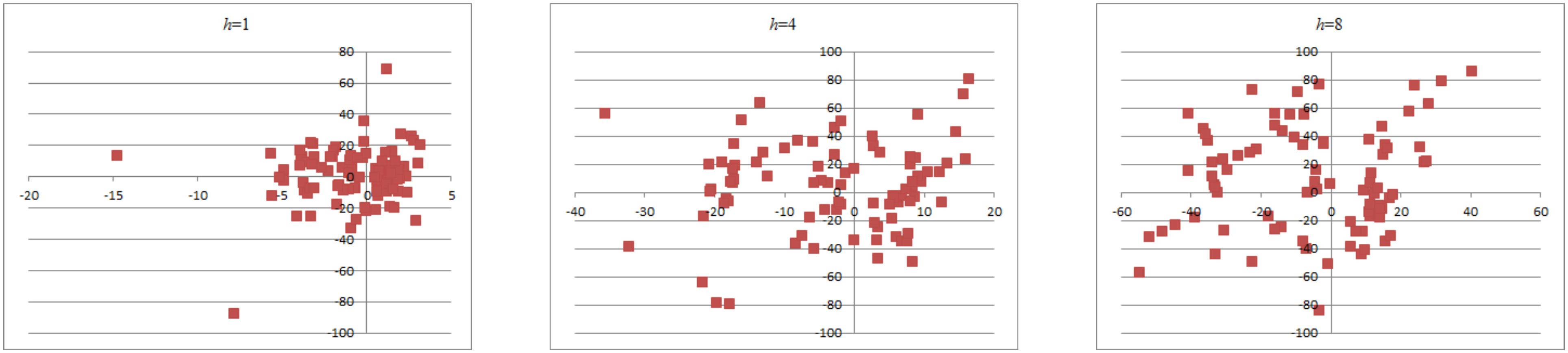
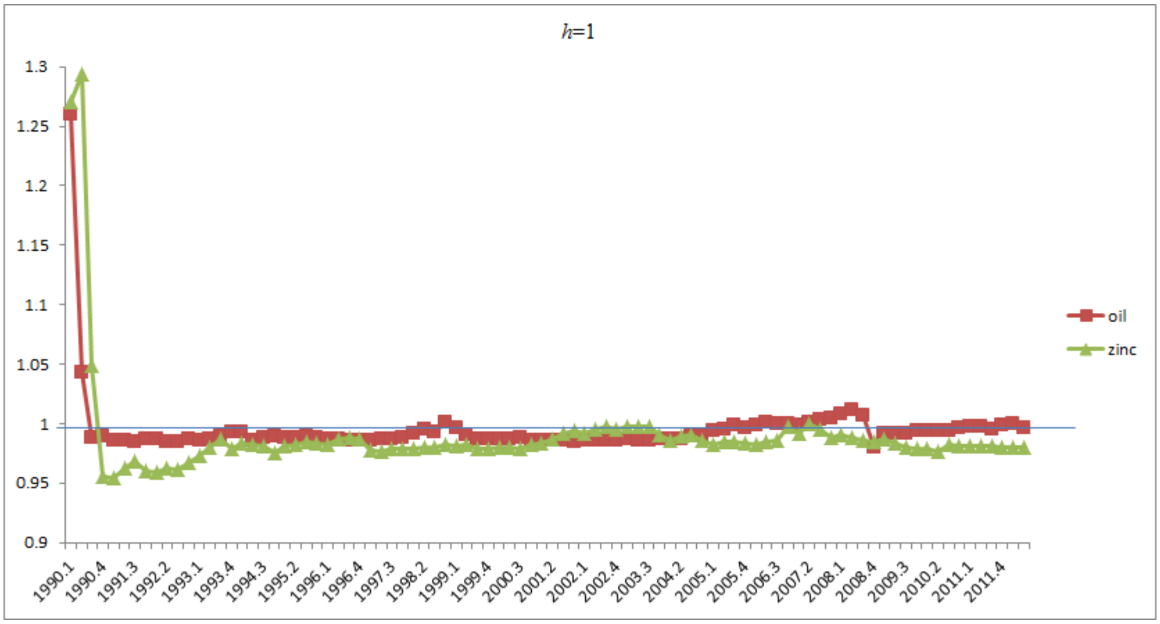


Figure 5

Recursive U-statistics, Zinc and Oil



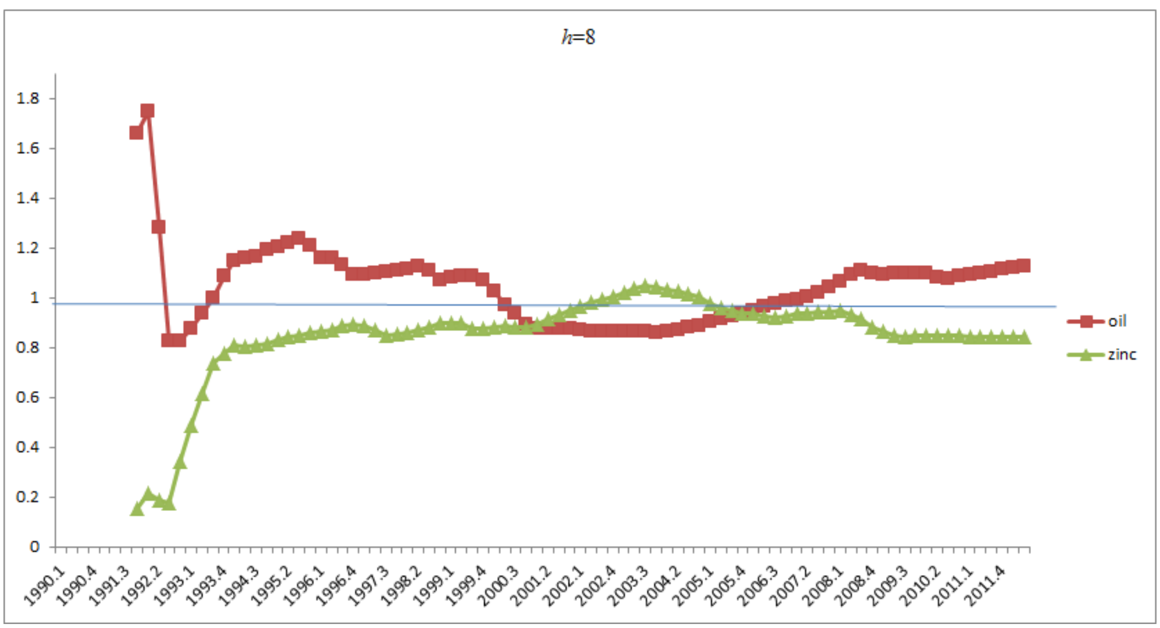


Table 1

List of commodities

Aluminum	99.5% minimum purity, LME spot price, CIF UK ports, US$ per metric ton
Coal	Thermal coal, 12,000- btu/pound, less than 1% sulfur, 14% ash, FOB Newcastle/Port Kembla, US$ per metric ton
Copper	Grade A cathode, LME spot price, CIF European ports, US$ per metric ton
Lead	99.97% pure, LME spot price, CIF European Ports, US$ per metric ton
Nickel	Melting grade, LME spot price, CIF European ports, US$ per metric ton
Rubber	Singapore Commodity Exchange, No. 3 Rubber Smoked Sheets, 1st contract, US cents per pound
Tin	Standard grade, LME spot price, US$ per metric ton
Uranium	NUEXCO, Restricted Price, Nuexco exchange spot, US$ per pound
Zinc	High grade 98% pure, US$ per metric ton
Oil	West Texas Intermediate 40 API, Midland Texas, US$ per barrel

1. The data and the data description are from the IMF commodity price data base (www.imf.org/external/np/res/commod/index.aspx). "LME" is the London Metal Exchange.

Table 2

Summary Statistics

A. Levels, $\ln(P_{it}/P_{CPI,t})$

	Aluminum	Coal	Copper	Lead	Nickel	Rubber	Tin	Uranium	Zinc	Oil
mean	2.36	-1.24	2.84	1.58	4.07	-1.03	4.02	-2.13	2.06	-1.59
s.d.	0.28	0.41	0.43	0.46	0.45	0.45	0.57	0.65	0.32	0.49
median	2.31	-1.26	2.79	1.53	3.95	-1.05	3.87	-2.23	2.02	-1.75
ρ_1	0.90	0.96	0.95	0.94	0.93	0.95	0.98	0.98	0.91	0.94
ρ_2	0.77	0.89	0.88	0.89	0.82	0.88	0.94	0.95	0.80	0.88
ρ_3	0.62	0.83	0.82	0.81	0.71	0.82	0.91	0.92	0.67	0.84

B. Growth rates, $\Delta p_{it} \equiv 100\times\Delta\ln(P_{it}/P_{CPI,t})$

	Aluminum	Coal	Copper	Lead	Nickel	Rubber	Tin	Uranium	Zinc	Oil
mean	-0.85	-0.19	0.09	-0.42	-0.12	-0.21	-0.74	-0.53	-0.10	-0.25
s.d.	12.19	11.41	14.16	15.64	17.64	14.37	12.04	13.15	13.83	17.09
median	-1.18	-0.74	0.84	-1.64	-1.63	-1.42	-1.24	-1.12	-1.56	-0.06
ρ_1	0.16	0.36	0.07	-0.09	0.18	0.13	0.26	0.16	0.07	0.00
ρ_2	0.05	-0.01	-0.04	0.27	0.06	-0.05	-0.02	0.10	0.12	-0.24
ρ_3	-0.11	-0.05	-0.17	-0.18	0.03	0.06	-0.14	0.16	-0.05	0.05

Notes:

1. The quarterly data run from 1980:1-2012:2 in panel A (130 observations), 1980:2-2012:2 (129 observations) in panel B. The data are constructed by sampling last month of quarter for the monthly data source given in the notes to Table 1.

2. The symbols ρ_1, ρ_2 and ρ_3 denote the first, second and third autocorrelations.

Table 3

Factor Loadings, 1980:1-2012:1 Sample

	Alum.	Coal	Copper	Lead	Nickel	Rubber	Tin	Uranium	Zinc	Oil
$\hat{\delta}_i$	0.66	0.88	0.90	0.93	0.79	0.90	0.81	0.88	0.73	0.82

Notes:

1. Let p_{it} be the 100×log of the real price of commodity i. The fitted model is p_{it} = const. + $\hat{\delta}_i \hat{f}_t + \hat{v}_{it} \equiv \hat{F}_{it} + \hat{v}_{it}$; $\hat{f}_t$ is the estimated factor.

Table 4

Forecast Comparison

Commodity	Horizon *h*=1 U	*h*=1 *t*	*h*=4 U	*h*=4 *t*	*h*=8 U	*h*=8 *t*
Aluminum	0.963	2.578	0.947	1.920	0.896	1.588
Coal	0.964	1.526	0.991	0.851	1.026	0.436
Copper	1.002	0.549	1.035	0.388	1.045	0.357
Lead	1.006	0.427	0.990	0.679	1.038	0.374
Nickel	0.979	1.771	0.939	1.489	0.891	1.149
Rubber	0.997	0.811	1.027	0.473	1.028	0.484
Tin	1.003	0.837	1.080	0.326	1.176	0.049
Uranium	1.035	0.060	1.100	0.299	1.146	0.342
Zinc	0.981	1.972	0.903	2.503	0.842	1.950
Oil	0.996	0.638	1.052	0.205	1.127	0.153
#U<1 / *t*>1.282	6	4	5	3	3	2
p-value max-*t*		[0.031]		[0.047]		[0.180]

p-value max-*t*, *h*=1,4,8: [0.034]

Notes:

1. The "U" columns present the U-statistic: RMSPE Model/RMSPE random walk. When U<1 the factor model had a smaller MSPE than did a random walk model. The number of commodities for which this was less than 1 is given in the #U<1 entries.

2. The "*t*" columns present a *t*-test of H_0: U=1 (equality of RMSPEs) against one-sided H_A: U<1 (RMSPE factor model is smaller), using the Clark and West (2006) procedure. The number of commodities for which this test rejected equality at the 10 percent level is given in the *t*>1.282 entry.

3. "p-value max-*t*" gives the fraction of 1000 bootstrap samples in which the maximum t-statistic exceeded the maximum of the 10 t-statistics in the sample for a given horizon. For example, the entry [0.047] for *h*=4 indicates that in 47 of the 1000 bootstrap samples, the maximum of the 10 *h*=4 Clark-West (2006) t-statistics for H_0: U=1exceeded 2.503; 2.503 is the critical value because it is the maximum of the 10 sample *t*-statistics for *h*=4. The entry for "p-value max-*t*, *h*=1,4,8" presents the fraction of bootstrap samples in which the maximum of 30 t-statistics for *h*=1, 4 and 8 exceeded 2.578, which is the maximum of the 30 t-statistics presented in the table.

Table 5

Forecast Comparison, Alternative Specifications

	(1) No. factors	(2) No. terms in $\hat{F}_{it}$-p_{it} in (2.2)	(3)	(4)	(5)	(6) p-value max-t, *h*=1, 4, 8
			----- #U<1/#t>1.282 /[p-value max t] -----			
			h=1	*h*=4	*h*=8	
(1)	1	2	6 / 4 / [0.031]	5 / 3 / [0.047]	3 / 2 / [0.180]	[0.034]
(2)	1	1	5 / 2 / [0.153]	5 / 3 / [0.042]	3 / 2 / [0.144]	[0.043]
(3)	1	3	5 / 4 / [0.062]	5 / 3 / [0.044]	3 / 2 / [0.207]	[0.051]
(4)	2	2	6 / 4 / [0.032]	5 / 3 / [0.041]	3 / 2 / [0.182]	[0.035]
(5)	3	2	6 / 4 / [0.033]	5 / 3 / [0.049]	3 / 2 / [0.189]	[0.036]

Notes:

1. For convenience of comparison, the results in line (1) repeat ones already reported in Table 4.

2. See notes to Table 4.