Exchange Rates do not Predict Commodity Prices*

Lasse Bork, Aalborg University, Dept Business and Management; bork@business.aau.dk

Pablo Rovira Kaltwasser, KU Leuven, Dept Economics; pablo.rovirakaltwasser@kuleuven.be

Piet Sercu, KU Leuven, Dept Accounting, Finance and Insurance; piet.sercu@kuleuven.be

Tom Vinaimont, Nazarbayev University, Graduate School of Business; tom.vinaimont@nu.edu.kz

Tuesday 19$^{\text{th}}$ September, 2023

Abstract

Chen, Rogoff and Rossi (2010) report that, for 'commodity currencies', the exchange rate predicts the country's commodity index but not vice versa, consistent with the Engel–West model where the country's key export prices act as the fundamentals. Predictability is assessed 'against a variety of benchmarks' (the random walk, the random walk with drift, and an AR(1) process). One snag is that, commodity prices being AR(1), only that third model is valid. Deleting inappropriate benchmarks and correcting a programming error, only one out-of-sample case remains significant, not thirteen, and even that one is not robust to the test statistic. When we use a larger sample the relation becomes non-robust, at best. Commodity prices appear to be no worse than exchange rates at digesting information.

Key words: commodity currencies, Engel–West currency model, fundamentals, predictability, market efficiency

JEL classification: C22, C52, C53, F31, F47.

*We thank the Editor and two referees for very helpful suggestions. All remaining shortcomings are ours. Tom Vinaimont acknowledges support by a Social Policy grant from Nazarbayev University.

Introduction

Chen, Rogoff and Rossi (2010) (henceforth CRR) investigate, for Australia, Canada, Chile, New Zealand and South Africa, the relation between the exchange rate and a price index of the country's commodity bundles. For each country, these two variables' first log differences are highly correlated, which makes the country's export-commodity index a candidate for an exchange-rate fundamental in the Engel and West (2005) sense. Based on that model, CRR test the proposition that these exchange rates predict their underlying commodity prices but not vice versa—the Commodity Currency Hypothesis (CCH)—and conclude that exchange rates do forecast commodity prices. We disagree even if we use the same data and run the same regressions. Our doubts remain after extending the test period and working with global indices rather than country-specific ones, as in Chen, Rogoff and Rossi (2014)'s internet update: the evidence is ambiguous and non-robust.

As a hypothesis, the proposition of one-way predictability from exchange rate movements to changes in their fundamentals certainly does make sense and should hold even outside Engel and West (2005)'s rather specific model: efficient financial markets anticipate the future to the full extent the future is predictable and relevant. An implicit assumption in the CRR tests of course is that the fundamental *is* to some extent predictable. If the fundamental is an exogenous state variable, there may very well be some predictability. Many spot commodity markets, however, do rely on price discovery in futures markets, which have financial asset-like characteristics themselves and should predict the future value of their own fundamentals. Even pure spot markets react to expected future prices because output can be reduced or increased, and inventories can be released or built up, in light of anticipated future spot prices. If both currency values and commodity prices immediately incorporate news about their respective fundamental variables (*i.e.* they are both forward-looking), then one should not be able to use either to predict the other. The only viable commodity-currency connection is contemporaneous.[1] In that light, CRR's empirical results are unexpected. This motivates

[1]The simple efficient-markets argument as stated above takes the familiar assumption of constant expected returns as its Null (Fama, 1991). What should be unpredictable is changes in updated expectations for a given future date T, which is not the same as changes in spot values being unpredictable. To link expectations to spot values one needs the risk-adjusted expectation or certainty equivalent (CEQ) and the spot-forward premium: $V_t = \mathrm{CEQ}(\tilde{V}_T)(1+r^*_{t,T})/(1+r_{t,T})$, where r denotes the home-currency riskfree rate and r^* the foreign risk-free return (for a currency) or the percentage convenience yield (on a commodity). Even if expectations change unpredictably, risk premiums, risk-free rates, and net conveniences should all be mean-reverting and, therefore, partly predictable. But those components seem to have an absolutely minor impact on spot values, as the vast

our interest in the CCH tests.

The empirical literature has not produced a consensus. Some studies do concur that exchange rates predict commodities, like Hua (1998), Clements and Fry (2008), Frank and Garcia (2010), Kato (2010) and Burgess and Rohde (2011). Other studies come to the opposite conclusion, including Amano and van Norden (1995), Hatzinikolaou and Polasek (2005), Simpson (2005), Schaling, Ndlovu and Alagidede (2014) and Zhang, Dufour and Galbraith (2016). Bidirectional relations are found by *e.g.* Akram (2009) and Bashar and Kab (2013), while inconclusive results are reported by Groen and Pesenti (2011) or Chan, Tse and Williams (2011). Even the CRR authors themselves have produced mixed results, with Chen and Rogoff (2002) concluding that commodities do lead and Ferraro, Rogoff and Rossi (2015) reporting that there is no lead–lag relation either way between the oil price and oil exporters' currency values.[2]

1 Tests on the original CRR data

1.1 Data

CRR construct quarterly commodity price baskets for Australia (AU), Canada (CA), Chile (CL), New Zealand (NZ) and South Africa (ZA), using data downloaded from various sources, including the Reserve Bank of Australia, ANZ Bank and the Bank of Canada. The resulting time series, which are available on Rossi's website, all end in 2008Q1; their starting dates range from 1974Q2 (CA) to 1994Q2 (ZA); see Table 1 for details. All tests are done on quarterly changes in p and s, the logs of the price index or the exchange rate.

1.2 In-sample tests

Denote changes in the log exchange rate between times $t-1$ and t by Δs_t, and changes in the log of the country's commodity price index by Δp_t. We indicate Granger Causality (GC) from x to y as "$x \to y$." For simplicity we omit country subscripts. In-sample, CRR first test for

literature on (non-)predictability of currency, commodity, and stock markets testifies.

[2]The latter study does report predictive ability at the daily frequency. At this frequency, price asynchronicity could be a major issue, and it is not clear what the timing of their data is.

GC in the standard way, by estimating

$$(s \to p): \quad \Delta p_t \quad = \quad \beta_0 + \beta_1 \, \Delta p_{t-1} + \beta_2 \, \Delta s_{t-1} + e_{p,t}, \tag{1}$$

$$(p \to s): \quad \Delta s_t \quad = \quad \beta_3 + \beta_4 \, \Delta s_{t-1} + \beta_5 \, \Delta p_{t-1} + e_{s,t}. \tag{2}$$

Their first in-sample GC test works with a constant-coefficient regression. For robustness, they then add the Rossi (2005) test, which allows for structural breaks.

Throughout their paper, the Nulls that CRR considered are the Meese and Rogoff (1983) pure random walk (RW), the random walk with drift (RWwD), and the AR(1) process. The RW Null probably refers to the empirical success of Meese-Rogoff's RW Null for exchange rates. That came as a shock to the profession, at the time, and the RW benchmark has often proven hard to beat.[3] However, the Engel–West framework that is being tested here is, specifically, about asymmetric cross-predictability and nothing more: if p is a fundamental and is predictable at all, it should be predicted by s (i.e. $\beta_2 \neq 0$), and the reverse should not hold (i.e. $\beta_5 = 0$). The Engel–West model is agnostic about autocorrelation and drift in Δp.[4] For this reason we replace the CRR Null (which we label the 'joint' Null),

$$\text{'joint' H}_0\text{:} \;\; \beta_0 = 0 = \beta_2. \tag{3}$$

by the 'single Null',

$$\text{'single' H}_0\text{:} \;\; \beta_2 = 0. \tag{4}$$

Panel A in Table 1 reports the results of the GC test obtained in the same data set as CRR (their Table I). For each country we provide the individual parameter estimates of the regression model $\Delta y_t = \beta_0 + \beta_1 \Delta y_{t-1} + \beta_2 \Delta x_{t-1} + e_t$, and, in the next row, their corresponding p-values.[5] The column '[joint]' reports the results of the GC test when the (inappropriate) joint hypothesis: $\beta_0 = \beta_1 = 0$ is tested, while the column 'single' reports the results of the GC test

---

[3]Even for individual commodities the RW works fairly well; see e.g. Wang, Liu and Wu (2020) for recent evidence. Indices exhibit more momentum.

[4]If CRR wanted to reject the RW for commodities, they could have pointed out the presence of autocorrelation in Δp. They do not; somehow, only drift is deemed to be evidence.

[5]Notice that we obtain slightly different p-values than CRR when we replicate their joint hypothesis GC test. The reason for this small difference is that CRR compute the values of the χ^2 cumulative distribution function themselves following an interpolation procedure, while our values are obtained directly from the Matlab function *chi2cdf*. CRR report the following p-values: Australia, 0.17 and 0.41; Canada, 0.06 and 0.92; Chile, 0.10 and 0.70; New Zealand, 0.11 and 0.45; and South Africa, 0.01 and 0.40.

Table 1: Granger causality test, in-sample, CRR data

	$\Delta p_t = \beta_0 + \beta_1\,\Delta p_{t-1} + \beta_2\,\Delta s_{t-1}$					$\Delta s_t = \beta_3 + \beta_4\,\Delta s_{t-1} + \beta_5\,\Delta p_{t-1}$						
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	[joint]	single	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	[joint]	single	Begin	Nobs
					A. Standard GC test, Newey–West							
AU coeff	0.007	0.424	0.003	3.344	0.005	0.001	–0.010	–0.147	1.757	1.636	1984Q2	96
p-value	**0.068**	**0.000**	0.945	0.177	0.945	0.924	0.863	0.215	0.434	0.212		
CA coeff	0.011	0.104	–0.045	4.907	0.111	0.000	0.089	–0.016	0.154	0.138	1973Q2	136
p-value	**0.030**	0.238	0.739	**0.086**	0.739	0.867	0.320	0.711	0.926	0.711		
CL coeff	0.019	0.130	–0.622	4.366	4.177	0.004	0.129	–0.003	0.688	0.005	1989Q4	74
p-value	0.162	0.249	**0.046**	0.115	**0.042**	0.458	0.268	0.943	0.710	0.943		
NZ coeff	0.005	0.335	–0.102	4.608	2.395	–0.002	0.109	–0.131	1.591	1.067	1987Q2	84
p-value	0.188	**0.001**	0.127	**0.098**	0.123	0.737	0.368	0.305	0.453	0.302		
ZA coeff	0.021	0.108	–0.118	8.536	1.121	0.014	0.122	–0.082	1.816	0.102	1994Q2	56
p-value	**0.006**	0.519	0.292	**0.014**	0.287	0.184	0.317	0.744	0.404	0.743		
(p-values)					B. Rossi's GC test (structural breaks)							
AU				**0.021**	0.836				**0.000**	**0.072**	1984Q2	96
CA				**0.057**	1.000				0.361	1.000	1973Q2	136
CL				0.224	0.295				**0.000**	1.000	1989Q4	74
NZ				**0.072**	0.146				**0.099**	**0.060**	1987Q2	84
ZA				**0.000**	**0.000**				**0.000**	0.324	1994Q2	56

Description: Using CRR's data, we run Granger Causality (GC) regressions between five countries' quarterly changes in log commodity-price indices p and in exchange rates s. Rossi's GC variant allows for breaks in the relation. Starting dates for data are in the column 'Begin'; the end date is 2008Q2. In Part A, the columns $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ report the coefficient estimates of the full-sample OLS regressions and their corresponding p-values. The columns '[joint]' and 'single' report the results of the GC test for, respectively, the joint hypothesis: $\beta_0 = \beta_1 = 0$ and single hypothesis: $\beta_1 = 0$. The joint test is inappropriate, which is signalled by the square brackets around the column's label. Panel B reports similar tests for the mirror model, p predicting s. All the reported p-values are based on the Newey and West (1987) procedure with bandwidth $T^{1/3}$. Part B reports the p-values of the Rossi (2005) GC test, which adopts re-estimated coeffficients when instability is indicated.

Interpretation: In each set of five 'single' tests, only one suggests a detectable link, rather than three or four in the joint tests. The false positives from the joint test indicated drift, that is, not cross-predictive ability. The GC and Rossi tests disagree on where the model does well.

when the single hypothesis: $\beta_1 = 0$ is tested. The single-Null T-test provides evidence in favor of the CCH just once (Chile), and the Wald test on β_2, added in the column 'single', tells us almost the same, as one would expect. The support for the asymmetric cross-correlation hypothesis is weak, in short; self-prediction by commodity prices actually does a better job, with two significant β_1 estimates. CRR's 'joint' Wald test, in contrast, comes up with three significant instances, not one, but their two extra cases are instances with drift, not cross-predictability.

A similar conclusion is obtained when we apply the Rossi (2005) test. In this test, the parameters are allowed to change over time in exactly the same way as in CRR, in order to account for potential structural breaks: $\Delta y_t = \beta_{0,t} + \beta_{1,t}\Delta y_{t-1} + \beta_{2,t}\Delta x_{t-1} + e_t$. The p-values are shown in Panel B of Table 1, corresponding to Table III in CRR. We again report the 'single' Wald test for $\beta_{2,t} = 0$ next to the CRR 'joint' one for $\beta_{2,t} = 0 = \beta_0$. The joint test misleadingly suggests four significant outcomes, while the correct single test finds just one. Further bad news, from a CCH perspective, is that when we apply the Rossi (2005) test we find two significant instances where the cross-predictability runs the wrong way, namely from commodities to exchange rates. In sum, our results indicate that the in-sample support in favor of the CCH is weak at best, regardless of whether we use the standard GC test or the Rossi (2005) version.

1.3 Out-of-sample tests

To complement their in-sample tests, CRR produce a series of one-period-ahead out-of-sample forecasts, evaluated via the Clark and McCracken (2001) test of equal forecasting power for nested models, *i.e.* models with/without the predictor of interest. In the out-of-sample tests one issue again relates to CRR's testing the CCH "against a variety of benchmarks" (namely the random walk, the random walk with drift and an AR(1) process) as if all of them are equally valid. They are not. First, as before, one should not count evidence of drift as evidence in favor of the CCH (the RW comparison, in Table 2). Second, CRR's commodity-return data exhibit strong AR(1) traits, but in the RW or RWwD benchmark model Δp_{t-1} is omitted, thus allowing the predictor Δs_{t-1} to proxy for the missing item. That is, even the RW-with-drift benchmark is unreliable. That leaves the AR(1) model as the sole correct benchmark in the GC class. In addition, there turns out to be a programming slip-up in the Clark-McCracken matlab routine of CRR, posted on the journal's website: a squaring operation is inadvertently done twice and results in an overstatement in the significance levels of their forecasting results. We correct these flaws. Lastly, we add the Clark and West (2007) test as a robustness test to Clark-McCracken's version.[6]

Table 2 replicates and complements the main forecasting results of CRR (Table IV: Tests for out-of-sample forecasting ability). The column DSFE reports the Difference in Mean Squared Forecasting Errors when the predictor of interest is added to the benchmark model, while the column RSFE reports the ratio of the competing Mean Squared Forecasting Errors. A negative

[6] This test takes into account that the fact that, under the Null, the larger model is inherently noisier because it estimates a parameter that plays no genuine role. For that reason the mean of the Clark-McCracken T-test is negative under the Null. Clark-West avoids this.

Table 2: Replication CRR Out-of-sample Forecast Tests

		Panel A: s predicts p (‘$s \to p$’)					Panel B: p predicts s (‘$p \to s$)				
	Benchmark	DSFE	RSFE	[p-CMC CRR]	p-CMC correct	p-CW	DSFE	RSFE	[p-CMC CRR]	p-CMC correct	p-CW
AU	AR1	1.813	1.027	**0.010**	1.000	0.956	0.241	1.011	1.000	1.000	0.395
	[RWwD]	–0.142	0.996	**0.100**	1.000	0.239	0.064	1.003	1.000	1.000	0.328
	[RW]	–2.115	0.878	**0.010**	**0.050**	**0.002**	0.537	1.028	**0.100**	1.000	0.479
CA	AR1	1.051	1.030	**0.050**	1.000	0.756	1.634	1.023	1.000	1.000	0.930
	[RWwD]	1.047	1.026	1.000	1.000	0.737	1.794	1.024	**0.050**	1.000	0.945
	[RW]	–0.010	1.000	1.000	1.000	0.275	0.594	1.012	1.000	1.000	0.566
CL	AR1	–0.163	0.992	**0.050**	1.000	0.289	1.188	1.042	**0.050**	1.000	0.754
	[RWwD]	–0.431	0.972	**0.050**	1.000	0.168	0.906	1.044	1.000	1.000	0.623
	[RW]	–0.448	0.958	**0.010**	**0.100**	0.122	0.998	1.083	1.000	1.000	0.616
NZ	AR1	0.320	1.020	**0.010**	**0.100**	0.153	0.233	1.013	1.000	1.000	0.277
	[RWwD]	–1.613	0.857	**0.010**	**0.010**	**0.005**	0.232	1.017	**0.050**	1.000	0.149
	[RW]	–0.752	0.950	**0.010**	**0.050**	**0.022**	0.156	1.010	**0.050**	1.000	0.152
ZA	AR1	1.346	1.123	**0.010**	1.000	0.913	1.571	1.142	1.000	1.000	0.881
	[RWwD]	1.686	1.083	**0.010**	1.000	0.952	1.372	1.150	1.000	1.000	0.811
	[RW]	–1.393	0.822	**0.010**	**0.050**	**0.018**	2.092	1.270	1.000	1.000	0.933

Description: The table summarises the out-of-sample tests for one-period-ahead predictive power on the CRR data studied in the preceding table, replicating the main forecasting results of CRR. The Granger model $y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 x_{t-1} + e_t$ is competing against the random walk model (RW: $\beta_0 = \beta_1 = \beta_2 = 0$), the random walk with drift (RWwD: $\beta_1 = \beta_2 = 0$), and the AR(1) model ($\beta_2 = 0$) which is, in fact, the sole valid benchmark. The column DSFE reports the Difference in Mean Squared Forecasting Errors between the CRR and the corresponding benchmark model, while the column RSFE reports the Relative Mean Squared Forecasting Errors. The column p-CW reports the p-values of the Clark and West (2007) test while the columns p-CMC CRR and p-CMC correct report the p-values of the Clark and McCracken (2001) test, first as computed by CRR, and then corrected. As in CRR, a p-value reported as 0.1 actually indicates a p-value in [0.10, 0.05[, and a reported 0.05 actually means [0.05, 0.01[. A benchmark or a test that is invalid is indicated by square brackets around its label: [RW], [RWwD], and [p-CMC-CRR].

Interpretation: In the $s \to p$ panel, Panel A, the column with p-values for the CMC-CRR test reports thirteen out of fifteen outcomes as significant at 10 percent or better. The number shrinks to six when the coding mistake is corrected, to one when invalid benchmarks (RW, RWwD) are ignored, and to zero when the CW test is adopted instead of the CMC variant. Out of sample, the evidence in this data is absent, that is. Panel B reveals no reverse predictive power either.

DSFE (or, equivalently, a RSFE below unity) indicates a successful prediction. The columns p-CMC_CRR and p-CMC_correct report the p-values of the Clark and McCracken (2001) test, calculated with and without the programming mishap. Following CRR, a p-value reported as 0.1 actually indicates a p-value below 0.1 but larger than 0.05, and a unit p-value means no improvement (or possibly even a worsening) in the forecasts when the extra regressor is added. Lastly, in the column p-CW we add p-values for the Clark and West (2007) alternative to the Clark and McCracken (2001) test. We provide test results against all three benchmarks used by CRR, even though the RW and RWwD benchmarks are invalid. We also show results using two versions of the Clark and McCracken (2001) test, once with the coding error and once without. To help identify valid versus invalid tests, we again add square brackets around the corresponding row or column labels: [RW], [RWwD], and [p-CMC-CRR].

Reading, in Table 2, the column p-CMC_CRR across all tests, we see that in 13 out of 15 cases the CCH model seems to outperform "a variety of benchmarks" at the 10-percent level. Correcting for the programming mistake, the score shrinks to six. Retaining only the AR1 tests as valid, only one success case survives the Clark and McCracken (2001) test, and that sole success is not robust: the Clark and West (2007) test statistic does not identify any case of significant forecasting power. In short, we find no basis to claim that exchange rates demonstrate "a surprisingly robust forecasting power over global commodity prices, both in-sample and out-of-sample", as CRR conclude.

2 Tests on an updated data set

In an online update of the original CRR study, Chen, Rogoff and Rossi (2014) add five years of data and seem to have corrected the coding error. CRR report that "the original messages in CRR (2010) continue to hold, though they are perhaps more nuanced than we initially recognized." In this section we verify whether an even longer dataset adds even more grounds for nuance or not.

2.1 Data

The tests in the preceding Section use the same data as CRR. However, the data set we work with for our update in the present section does not have the same coverage of commodities. CRR used indices and subindices published by the Reserve Bank of Australia, ANZ bank, the Bank of Canada, *etc*. The indices and subindices published by these sources today are not fully comparable to the indices used by CRR, as both the list of constituents and their weights change over time.

We collect commodity price data for the same countries as CRR from the following sources: Reserve Bank of Australia, ANZ Bank, Refinitiv and IMF, and construct indices on the basis of CRR's weights. These data end in 2021Q4, thus adding 54 quarters to the original time series. In Appendix Table A.1 we compare the indices (CRR *versus* our update) in terms of constituents and weights, as per April 2022.

Table 3: Summary of results: significant test results per test, dataset and period
# of significant (10%) cases for $s \to p$ / # of significant (10%) cases for $p \to s$

	test	hypothesis	CRR (-2008)	New(-2008)	Total Period
In-sample	standard	['joint']	3 / 0	2 / 0	3 / 1
		'single'	1 / 0	1 / 0	3 / 1
	Rossi 2005	['joint']	4 / 4	4 / 3	3 / 2
		'single'	1 / 2	1 / 1	3 / 0
	benchmark	test	CRR (-2008)	New(-2008)	Total Period
Out out sample	[any]	[CMC CRR]	13 / 5	10 / 5	9 / 4
		[CMC correct]	6 / 0	2 / 0	9 / 3
	AR(1)	CMC	1 / 0	0 / 0	3 / 1
		CW	0 / 0	0 / 0	3 / 0

Description: In- and out-of-sample tests are now applied to a longer data set, where 56 more quarterly observations are added (2008Q3 to 2021Q4). The Granger regression is again competing against the the random walk, the random walk with drift, or the AR(1) alternatives—RW, RWwD and AR(1)—of which only the latter is valid. The table lists the numbers of significant test outcomes, at a significance level of 10 percent, per data set, test, and period, using the Newey-West T-test in-sample and a Clark-MacCracken test out-of-sample. The three data bases are the CRR set (ending 2008Q2), our best replication over CRR's period, and our best replication in the total period (ending 2021Q4). An inappropriate hypothesis, benchmark or test is indicated by square brackets around the row's label. In the out-of-sample tests, the benchmark labelled 'any' refers to outperformance at 10 percent of either the RW, RWwD or AR(1) model. "CMC" refers to the Clark and McCracken (2001) test, as coded by CRR (with a programming slip-up that inflates significance) and then in the correct form. "CW' refers to Clark and West (2007)'s alternative to Clark and McCracken (2001)'s test.

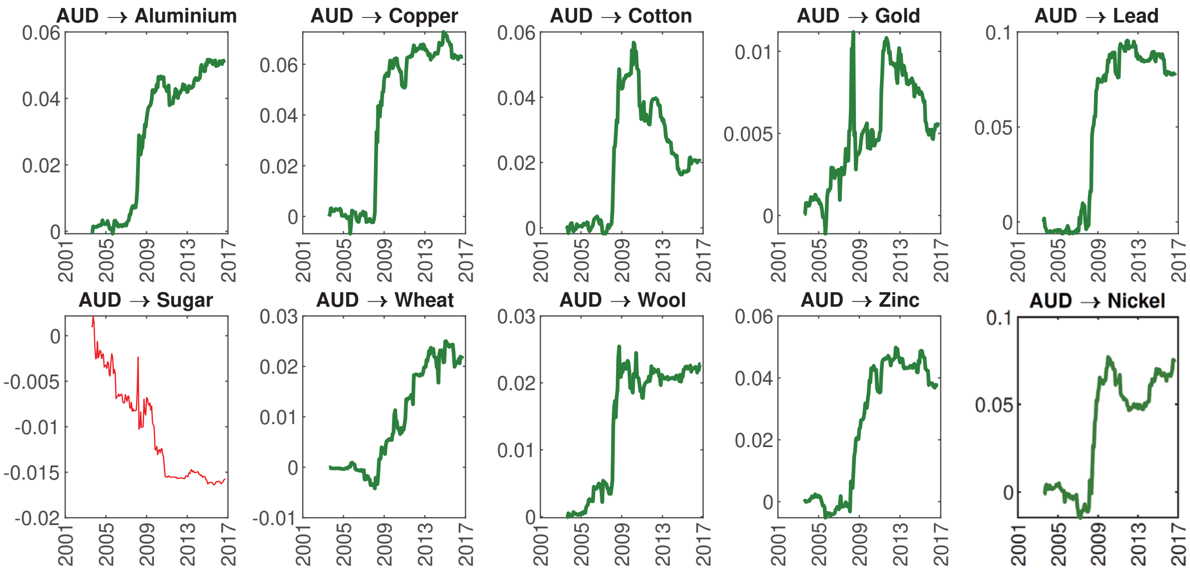
Interpretation: The results from the New(-2008) data set are very similar to those from the CRR data: one reliable in-sample occurrence of s predicting p, and two or zero out-of-sample depending on the test. In the total period, however, the evidence of predictability from s to p is stronger, with consistently three occurrences not matched by instances of reverse causality. Predictive power may have improved, that is.

2.2 Results

To verify whether results based on later data are due to the different commodity coverage rather than to the newly added years, we report results for two periods. The first period refers to the one used by CRR(2010) that ends in 2008, which we label New(-2008). The second period starts at the same time as CRR and ends on December 31, 2021, which we label Total Period. The key results for the new data set can be found in Tables A.2 (in-sample) and A.3 (out-of-sample).[7] Much of our discussion focuses on the overview of all those results, for both data sets and periods, in Table 3. The figures shown in each data×test cell are the numbers of significant outcomes of $s \to p$ and $p \to s$, respectively, separated by a slash, using the T-test

[7]Like in Tables 1 and 2, square brackets around a column's or row's heading indicate that the test or benchmark or hypothesis is inappropriate. As of now, significant p-values are shown in bold only for appropriate tests.

Figure 1: Cumulative gains, out-of-sample, from adding the CCH regressor to the AR(1) model



Description: The graphs relate to commodity-by-commodity out-of-sample tests of s predicting p. For each of the ten commodities in our Australia index we show the cumulative gains from adding the CCH regressor (Δs_{t-1}) to the AR(1) model for Δp, for the total period. A rising pattern means the CCH model helped.

Interpretation: In nine cases, the gain is significant, but the results are clearly concentrated in the 2008IV period, the financial crisis outlier. Before and after, there is no predictive ability.

in-sample and a Clark-MacCracken test out-of-sample.

We first compare the two datasets for the same pre-2009 period, labeled CRR(-2008) and New(-2008), respectively. Judging by the numbers of significant rejections of the (single) Null reported in Table 3, the in-sample evidence in the alternative New(-2008) data is fairly comparable to that from the CRR series, except that when we apply the Rossi (2005) test we find just one case for $p \rightarrow s$ rather than two. Out out sample, the number of false alarms because of the error in the test or the wrong benchmark is somewhat lower, but the bottom line is unaltered: the Clark-West test still finds zero cases of causality either way. In short, we find no material differences between the two data sets until 2008.

In the full data set the number of degrees of freedom rises by 56, so if the weak evidence from the pre-2009 data set would have been driven by a small sample size, we should now see an improvement. That does seem to be the case: we now have three instances of $s \rightarrow p$ instead of one, whether we look at the standard GC test, the Rossi variant, or the CMC and CW tests. The three significant $s \rightarrow p$ outcomes (AU, CL, NZ) are counterbalanced by one or zero $p \rightarrow s$ case. Summing up, while there is no evidence in the new -2008 data whatsoever, in the -2021 sample the results do point into the CCH direction.

The contrast is puzzling. A closer look reveals that the predictability is closely associated with the outlier episode around Lehman Brothers' implosion, when both the USD and, a bit later, most commodities plunged. Figure 1 shows the cumulative difference in the squared prediction errors of the competing models, for each of the ten goods that constitute our AUD export index. A rising pattern means the CCH model helped. In all cases but sugar, a Clark and West (2007) test (available on request) identifies the gains as significant at 10 percent. Yet the gains are concentrated around the Lehman outlier; we see unremarkable results in the periods before and after. That is, the three cases of $s \rightarrow p$ in Table 3 reflect one momentous event rather than three independent empirical regularities. Also, the apparent $s \rightarrow p$ predictability does not fit in with our results for the world-index tests added by CRR (2014), as the next section shows: there, the general impression is general ambiguity and non-robustness.

3 World-market tests

In the variant test that CRR (2014) add, the country-specific commodity indices are replaced by world indices, the same for all five commodity exporters. They adopt six world indices. The PNFUEL series (Non-Fuel Price Index) from the IMF, back-calculated by CRR for the pre-1991 periods, is advanced as the prime proxy; for robustness checks they then work with data from Commodity Research Bureau-BLS, Reuters/Jeffries, Moody's, Dow Jones-AIG, and Goldman Sachs, all obtained from Global Financial Data. Their data end in 2013Q3.

The motivation is not entirely clear. There is no discussion why replacing a country-specific index by a global index is expected to improve rather than harm the data's relevance and power, why the omission of oil is desirable, *a priori*,[8] nor why PNFUEL is also otherwise the superior choice. The choice for a common p series across the five countries does increase cross-dependencies in the test results, given that there are strong correlations already among the Δs series. If the *a priori* justification is not very clear, the results are: the alternative data do deliver a pro-CCH picture (their Table V(a).D). In the standard GC tests on PNFUEL data, with either two, three, or four exchange rate regressors (AU-NZ, AU-NZ-CA, AU-NZ-CA-CL), CRR reject the Null in their in-sample tests. It is not stated whether the Null is single or double, and there is an obvious dependence issue across the tests, given the overlaps in the

[8] The tests on oil prices by Ferraro *et al.* (2015) reject predictability either way, whether changes are computed quarterly or monthly.

Table 4: p **values in in-sample GC tests with CRR (2014)'s global PNFUEL index data paired with country exchange-rates**

series	Standard GC		Rossi GC	
	$s \to p$	$p \to s$	$s \to p$	$p \to s$
AU	0.001	0.049	0.000	0.122
NZ	0.000	0.072	0.000	0.218
CA	0.042	0.057	0.047	0.044
CL	0.000	0.107	0.000	0.187
ZA	0.054	0.307	0.409	1.000

Description: The tests are similar to the in-sample tests summarised in Table 1 except that the country-specific indices are replaced by CRR (2014)'s global PNFUEL index over their sample period, 1980Q1 to 2013Q3. We report the p-values from Newey-West T-tests of the 'single' hypothesis in each direction.

Interpretation: In the standard-GC and Rossi tests, respectively, five or four exchange rates show evidence of $s \to p$ predictabive power, in-sample, for the PNFUEL index. There are instances suggesting reverse predicability too, but they are less frequent (three and one cases, respectively).

regressor sets. So we re-run the regressions on their data, using, each time, each regressor separately rather than two, three or four of them. Our findings, in Table 4, are fully in line with theirs. All five slope coefficients in the 'single' $s \to p$ tests are significant, and in four cases the significance survives in Rossi's variant tests.

The marked shift in the results may de due to addition of a turbulent period, which includes Lehman's 2008 implosion. In line with the idea of unruly data, CRR's results contain a few cases of clear reverse causality too: three in the standard test, and one in the Rossi variant. Still, this counter-evidence remains weaker compared to the $s \to p$ results. Did the world change in 2008, or do we see outlier effects, or are the new price data different? We re-run CRR's tests on our longer post-2008 series and on our longer total data set that ends in 2021, for the three indices to which we had access, namely PNFUEL, Moody's and S&P. Lastly, we also add a regular Newey-West T-test obtained from the R package *sandwich*, which should largely agree with CRR's standard GC test.[9]

The results are reported in Table 5. The two Newey-West tests are in reassuringly broad agreement. Across indices, however, we do observe marked disagreements. In the post-2008 period, most blatantly, the PNFUEL index comes up with 8/15 cases of $s \to p$ across the three tests and S&P with 9, but Moody's with none at all. Equally confusing, for PNFUEL and S&P

[9]This is just a robustness check, without any priors as to which is better. Our test adopts the default settings of the package: pre-whitening, Bartlett kernel, automatic determination of lag through the function *bwNeweyWest* which follows Newey and West (1987), while CRR use the Bayesian Information Criterion to determine the lag.

Table 5: **In-sample GC Tests with world indices rather than country-specific indices: p-values**

	2008Q1 to 2021Q4						various starting dates -2021Q4					
	PNFUEL		Moody's		S&P		PNFUEL		Moody's		S&P	
	$s \to p$	$p \to s$	$s \to p$	$p \to s$	$s \to p$	$p \to s$	$s \to p$	$p \to s$	$s \to p$	$p \to s$	$s \to p$	$p \to s$
Panel A. CRR's GC test, their code												
AU	0.003	0.131	0.202	0.009	0.078	0.525	0.000	0.175	0.018	0.076	0.005	0.235
NZ	0.024	0.175	0.531	0.084	0.204	0.891	0.001	0.160	0.019	0.074	0.065	0.838
CA	0.724	0.028	0.210	0.000	0.719	0.191	0.029	0.064	0.830	0.008	0.149	0.273
CL	0.002	0.123	0.371	0.959	0.005	0.040	0.000	0.094	0.016	0.607	0.001	0.079
ZA	0.552	0.039	0.460	0.000	0.595	0.009	0.084	0.170	0.389	0.002	0.895	0.021
sgnf	3	2	0	4	2	2	5	2	3	4	3	2
Panel B. Standard Newey-West T, R package *sandwich*												
AU	0.000	0.060	0.172	0.013	0.094	0.535	0.000	0.196	0.025	0.083	0.007	0.243
NZ	0.018	0.311	0.585	0.141	0.246	0.861	0.001	0.204	0.027	0.065	0.071	0.836
CA	0.451	0.023	0.177	0.000	0.374	0.155	0.028	0.072	0.835	0.009	0.164	0.272
CL	0.006	0.013	0.300	0.825	0.002	0.018	0.000	0.095	0.013	0.628	0.001	0.056
ZA	0.256	0.016	0.353	0.000	0.339	0.000	0.079	0.181	0.397	0.002	0.897	0.020
sgnf	3	4	0	3	2	2	5	2	3	4	3	2
Panel C. Rossi's GC test with stuctural breaks, CRR code												
AU	0.025	0.163	0.822	0.028	0.049	0.000	0.000	0.503	0.020	0.440	0.000	0.534
NZ	0.144	0.358	0.348	0.000	0.000	0.000	0.000	0.085	0.080	0.572	0.269	0.670
CA	1.000	0.000	0.844	0.000	0.000	0.000	0.044	0.347	0.000	0.035	0.027	0.162
CL	0.000	0.571	0.256	1.000	0.000	0.079	0.000	0.234	0.033	1.000	0.000	0.000
ZA	0.353	0.259	0.747	0.000	0.000	0.000	0.424	0.714	1.000	0.000	1.000	0.217
sgnf	2	1	0	4	5	5	4	1	4	2	3	1
sum	8	7	0	11	9	9	14	5	10	10	9	5

Description In these in-sample tests we study three alternative global commodity index data instead of the country's specific index: either the IMF's no-fuel index PNFUEL, Standard&Poor's commodity index, or Moody's version. The three GC tests are (i) the GC test as coded by CRR(2014), (ii) a standard Newey-West T-test with the default settings of the R package 'sandwich', and (iii) Rossi's GC test that allows for structural breaks. The tests are run first for the post-CRR period 2008Q1 to 2021Q4 and then for the entire period (with starting dates depending on the country, as in Table 1). The numbers reported are p-values. The rows 'sgnf' report the number of significant outcomes at the 10 percent level.

Inpretation: While the results from CRR's code and the R package are in broad agreement, the conclusions depend disconcertingly on the index and period. In the post-CRR period, PNFUEL detects marginally more in-sample $s \to p$ cases than $p \to s$ ones, but Moody's finds only $p \to s$ instances, and S&P is agnostic. When we add in the pre-2008 data, where earlier tests found nothing, $s \to p$ does better for PNFUEL and S&P, but Moody's is agnostic.

the 8 or 9 $s \to p$ cases come with 7 or 9 reverse-causality cases, while for Moody's the tally is 0 times $s \to p$ against 11 times $p \to s$. It is hard to say this supports the CCH. Across three tests, three indices, and five countries, there is no agreement either about a dominant direction of causality when we compare the total period to the post-2008 one. In the total data, all in all, the picture does look fairly clear, with 33/45 tests suggesting $s \to p$ causality, against 20/45 the other way around. In the post-2008 data, however, we see a reversed picture (17/45 significant $s \to p$ cases against 27/45 for $p \to s$). Note also that the shifting patterns over time contradict

Table 6: **Out-of-sample with world indices**

	PNFUEL				Moody's				S&P			
	$s \to p$		$p \to s$		$s \to p$		$p \to s$		$s \to p$		$p \to s$	
	RSFE	p-CW	RSFE	p-CW	RSFE	p-CW	RSFE	p-CW	RSFE	p-CW	RSFE	p-CW
	Panel A. (Various starting dates) till 2008											
AU	0.930	**0.023**	1.018	0.838	0.981	0.198	1.019	0.735	1.030	0.241	1.021	0.317
CA	0.992	0.141	1.009	0.845	1.059	0.856	1.000	0.342	0.991	0.111	1.012	0.674
CL	0.835	**0.010**	1.067	0.942	0.861	**0.015**	0.999	0.288	1.018	0.385	1.005	0.380
NZ	0.987	**0.083**	1.048	0.721	0.959	**0.049**	1.038	0.854	1.095	0.878	1.052	0.722
ZA	1.070	0.671	1.083	0.535	1.295	0.918	1.047	0.466	1.134	0.938	1.037	0.290
sgnf		3		0		2		0		0		0
	Panel B. (Various starting dates) till 2021Q4											
AU	0.926	**0.013**	1.008	0.267	0.991	0.195	0.988	0.101	0.950	**0.070**	1.033	0.576
CA	1.017	0.420	0.993	**0.083**	1.013	0.767	0.934	**0.002**	1.022	0.660	1.023	0.435
CL	0.953	**0.042**	1.011	0.337	0.978	0.155	1.030	0.713	0.905	**0.018**	1.049	0.804
NZ	0.978	0.130	1.007	0.369	1.003	0.298	1.009	0.626	1.000	0.295	1.044	0.963
ZA	1.003	0.305	1.006	0.101	1.008	0.796	0.934	**0.008**	1.004	0.555	1.011	0.126
sgnf		2		1		0		2		2		0

Description: In these out-of-sample tests we study, over the longer data period, three alternative global commodity index data instead of the country's specific index: either the IMF's no-fuel index PNFUEL, Standard&Poor's commodity index, or Moody's version. The tests are run first for the post-CRR period 2008Q1 to 2021Q4 and then for the entire period (with starting dates depending on the country, as in Table 1). The table reports the Difference of Mean Square Forecast Errors for Δp_t or Δs_t made by AR(1) models with versus without the cross-predictor added, *i.e.* Δs_{t-1} or Δp_{t-1}, alongside the p-values of out-of-sample Clark-West tests of one-period-ahead forecasting ability.

Interpretation: Again no robust conclusion seems possible. PNFUEL tends to favor $s \to p$, S&P agrees for the full period but finds nothing in recent data, and Moody's changes their veiw over time.

our findings with the country indices, where the older data provided no support whatsoever either way while the newer ones did suggest a $s \to p$ effect. The country-index data do have a different commodity coverage and different weights, but why this induces these particular patterns is not at all obvious. The finding does add to the picture of non-robustness, though.

We close with the out-of-sample tests. The total sample part, Panel B in Table 6, is again ambiguous: across the three indices, four cases of $s \to p$ are almost perfectly counterbalanced by three cases of $p \to s$, with Moody's tallies even being zero against two. Unlike what we saw for the country-specific indices, the early period now provides somewhat better results, from the CCH perspective: PNFUEL and Moody's come up with, in all, 5 successes and zero failures. But S&P disagrees: nothing is significant.

The general picture from these alternative data, then, is that the CRR (2014) PNFUEL-index sample was hardly representative. In the larger data set we observe general non-robustness among world indices, between global indices and local ones, and across periods.

4 Concluding remarks

The CRR conclusion that commodity prices are predicted by exchange rates and not *vice versa* reflects a programming wobble in an out-of-sample test, adopts an irrelevant Null (mixing drift with cross-predictability), and is not robust to the test statistic (Clark-McCracken versus Clark-West). Correcting for all that, we find that the evidence is very weak. Out out sample, for instance, at the 10% level their code finds 13 statistics suggesting outperformance against three simple benchmarks while, depending on the test, we find one or zero. In the CRR(2010) data, nothing demonstrates "a surprisingly robust forecasting power over global commodity prices, both in-sample and out-of-sample." In the extended series as a whole, -2021Q4, there is some evidence, but it can be traced to an outlier episode around the Lehman collapse; and the picture from global indices, a new test advanced in CRR (2014), is non-robust rather than in line with the country-specific results.

To financial economists this conclusion probably is not unexpected. The CCH makes perfect sense, provided the fundamental *is* predictable. However, commodity pricing is strongly forward-looking: spot prices are often closely following short-tenor futures prices, and in most spot markets the producers and/or users can shift supply and demand over time. That induces the same financial logic as in currency markets: the spot price reflects all what is predictable about the fundamentals, and never reacts to news with a lag. Reversing the Engel–West argument or invoking the standard efficient-markets tenet, one therefore expects that p should predict s, provided the latter is predictable at all. With both s and p following the same financial logic, then, we expect very little cross-predictive power either way.

References

Akram, Q. Farooq, "Commodity prices, interest rates and the dollar," *Energy Economics*, 2009, *31* (6), 838–851.

Amano, Robert A. and Simon van Norden, "Terms of trade and real exchange rates: the Canadian evidence," *Journal of International Money and Finance*, 1995, *14* (1), 83 – 104.

Bashar, Omar K. M. R. and Sarkar H. Kab, "Relationship between Commodity Prices and Exchange Rate in Light of Global Financial Crisis: Evidence from Australia," *International Journal of Trade, Economics and Finance*, 2013, *4* (5), 265–269.

Burgess, Kieran and Nicholas Rohde, "Can Exchange Rates Forecast Commodity Prices? Evidence using Australian Data," *Economics Bulletin*, 2011, *33* (1), 511 – 518.

Chan, Kalok, Yiuman Tse, and Michael Williams, "The Relationship between Commodity Prices and Currency Exchange Rates: Evidence from the Futures Markets," NBER Working Papers 11859, National Bureau of Economic Research, Inc 2011.

Chen, Yu-Chin and Kenneth Rogoff, "Commodity currencies," *Journal of International Economics*, 2002, *60* (1), 133 – 160.

__ , Kenneth S. Rogoff, and Barbara Rossi, "Can Exchange Rates Forecast Commodity Prices?," *The Quarterly Journal of Economics*, August 2010, *125* (3), 1145–1194.

__ , __ , and __ , "Can Exchange Rates Forecast Commodity Prices? An Update," mimeo 2014.

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

__ and Michael W. McCracken, "Tests of Equal Forecast Accuracy and Encompassing for Nested Models," *Journal of Econometrics*, November 2001, *105* (1), 85–110.

Clements, Kenneth W. and Rene Fry, "Commodity currencies and currency commodities," *Resources Policy*, 2008, *33* (2), 55 – 73.

Engel, Charles and Kenneth D. West, "Exchange Rates and Fundamentals," *Journal of Political Economy*, June 2005, *113* (3), 485–517.

Fama, Eugene F., "Efficient Capital Markets: II," *The Journal of Finance*, 1991, *46* (5), 1575–1617.

Ferraro, Domenico, Kenneth Rogoff, and Barbara Rossi, "Can oil prices forecast exchange rates? An empirical analysis of the relationship between commodity prices and exchange rates," *Journal of International Money and Finance*, 2015, *54*, 116–141.

Frank, Julieta and Philip Garcia, "How Strong are the Linkages among Agricultural, Oil, and Exchange Rate Markets?," 2010. NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting and Market Risk.

Groen, Jan J. J. and Paolo A. Pesenti, "Commodity Prices, Commodity Currencies, and Global Economic Developments," NBER Working Papers, National Bureau of Economic Research, Inc 2011.

Hatzinikolaou, Dimitris and Metodey Polasek, "The Commodity Currency View of the AUD: a Multivariate Cointegration Approach," *Journal of Applied Economics*, 2005, *8* (1), 81–99.

Hua, Ping, "On Primary Commodity Prices: The Impact of Macroeconomic/Monetary Shocks," *Journal of Policy Modeling*, 1998, *20* (6), 767 – 790.

Kato, Haruko, "Changes in the Relationship between Currencies and Commodities," 2010. Bank of Japan Review Series 12E2.

Meese, Richard A and Kenneth Rogoff, "Empirical exchange rate models of the seventies: Do they fit out of sample?," *Journal of international economics*, 1983, *14* (1-2), 3–24.

Newey, Whitney K. and Kenneth D. West, "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 1987, *55* (3), 703–08.

Rossi, Barbara, "Optimal Tests For Nested Model Selection With Underlying Parameter Instability," *Econometric Theory*, October 2005, *21* (05), 962–990.

Schaling, Eric, Xolani Ndlovu, and Paul Alagidede, "Modelling the rand and commodity prices: A Granger causality and cointegration analysis," *South African Journal of Economic and Management Sciences*, 2014, *17* (5), 673–690.

Simpson, John L., "The Relationship Between Commodity Prices and the Australian Dollar," 2005. EFMA 2002 London Meetings.

Wang, Yudong, Li Liu, and Chongfeng Wu, "Forecasting commodity prices out-of-sample: Can technical indicators help?," *International Journal of Forecasting*, 2020, *36* (2), 666–683.

Zhang, Hui Jun, Jean-Marie Dufour, and John W. Galbraith, "Exchange rates and commodity prices: Measuring causality at multiple horizons," *Journal of Empirical Finance*, 2016, *36* (Supplement C), 100 – 120.

Appendix Table

For a discussion and interpretation of Tables A.2 and A.3, see Table 3 and the comments thereon in the main text.

Table A.1: **Data (CRR versus current)**

CRR data		2022Q1 data		CRR data		2022Q1 data		CRR data		2022Q1 data	
Australia				**Canada**				**New Zealand**			
'Rural'	*29.0*		*9.3*	*Agriculture*	*14.7*		*16.0*	*Agricultural*	*31.2*		*25.5*
Wheat	8.3	Wheat	1.3			Barley		Beef	9.4	Beef	11.3
Beef	7.9	Beef/veal	3.6	Beef	7.8			Lamb	12.5	Lamb	11.9
Wool	4.1	Wool	1.1	Canola	1.2	Canola		Skins	1.6	Skins	0.6
Cotton	2.8	Cotton	0.6			Cattle				Venison	0.5
Sugar	2.5	Sugar	0.5	Corn	0.5	Corn		Wool	7.7	Wool	1.2
Barley	1.9	Barley	0.4	Hogs	1.8	Hogs		Dairy	*35.8*		*41.4*
Canola	1	Canola	0.4			Potatoes		Wholemeal MP	10.6	Wholemilk powder	
		Lamb/mutton	1.4	Wheat	3.4	Wheat		Skim MP	3.7	Skimmilk powder	
Rice	0.5			*Metals/minerals*	*15.9*		*23.2*	Butter	6.5	Butter	
Base metals	*15.7*		*4.0*	Gold	2.3	Gold		Cheese	8.3	Cheese	
Aluminum	8.1	Aluminium	1.4	Silver	0.3	Silver		Casein	6.7	Casein	
Copper	2.8	Copper	1.3	Nickel	2.4	Nickel		*Horticultural*	*6.8*		*10.9*
Nickel	2.6	Nickel	0.5	Copper	2	Copper		Apples	3.1	Apples	2.8
Zinc	1.5	Zinc	0.5	Aluminum	5	Aluminium		Kiwi	3.7	Kiwifruit	8.1
Lead	0.7	Lead	0.3	Zinc	2.3	Zinc		*Forestry*	*11.2*		*14.3*
Bulk commodities	*33.7*		*53.3*	Potash	1.6	Potash		Logs	3.5	Logs	9.4
Coking coal	14.7	Metallurgical coal	13.7			Lead		Sawn timber	4.6	Sawn timber	3
Steaming coal	9.7	Thermal coal	8.1			Iron		Pulp	3.1	Pulp	1.9
Iron ore	9.3	Iron ore	31.5	Coal	*1.8*	Coal	*0.9*	*Seafood*	*6.7*		*4.9*
Other resources	*21.6*		*33.4*	*Crude oil*	*21.4*		*41.1*	Fish	6.7		
Gold	9.4	Gold	7.6			Brent	1.5			Hoki	
		Copper ore	2.2			Western Cnd Slct	19.0			Orange roughy	
		Crude oil	3			West Texas Interm	20.7			Mussels	
Alumina	7.4	Alumina	3.5	*Fisheries*	*1.3*		*1.4*	*Metals*	*8.3*		*3.1*
LNG	4.8	LNG	17	Fish	1.3	Fish		Aluminum	8.3	Aluminum	3.1
				Forestry	*34.1*		*11.1*				
Chile				Pulp	12.8	Pulp		**South Africa**			
Copper	100	Copper	100	Lumber	13.6	Lumber		Coal	22	Coal	22
				Newsprint	7.7	Newsprint		Gold	48	Gold	48
				Nat. gas	*10.7*	Nat. gas	*6.1*	Platinum	30	Platinum	30

Note CRR data are taken from Chen's website. More recent data are from the Reserve Bank of Australia, Bank of Canada, ANZ Bank, Refinitiv and IMF. We show the CRR weights and the 2022Q1 weights (which are not always available at the individual good level). Constituents and weights change repeatedly.

Table A.2: Granger causality test, in-sample; new dataset, two test periods

	$\Delta p_t = \beta_0 + \beta_1\,\Delta p_{t-1} + \beta_2\,\Delta s_{t-1}$					$\Delta s_t = \beta_3 + \beta_4\,\Delta s_{t-1} + \beta_5\,\Delta p_{t-1}$				
	$\hat\beta_0$	$\hat\beta_1$	$\hat\beta_2$	[joint]	single	$\hat\beta_3$	$\hat\beta_4$	$\hat\beta_5$	[joint]	single
Panel A. Standard GC test, Newey-West, -2008 period										
AU	0.007	0.435	−0.063	7.145	1.487	0.001	0.000	−0.165	1.797	1.666
p-val	0.041	0.003	0.226	0.028	0.223	0.892	0.996	0.200	0.407	0.197
AU	0.011	0.096	−0.062	4.322	0.143	0.000	0.107	−0.012	0.085	0.072
p-val	0.040	0.223	0.705	0.115	0.705	0.879	0.249	0.789	0.958	0.788
CL	0.021	0.110	−0.615	3.735	2.999	0.004	0.122	0.000	0.833	0.000
p-val	0.121	0.363	0.088	0.155	**0.083**	0.437	0.239	0.995	0.659	0.995
NZ	0.004	0.291	−0.046	1.407	0.403	−0.002	0.137	−0.089	1.095	0.855
p-val	0.310	0.012	0.527	0.495	0.526	0.690	0.242	0.358	0.578	0.355
ZA	0.012	0.589	−0.019	4.961	0.038	0.011	0.169	0.083	1.525	0.070
p-val	0.039	0.002	0.845	0.084	0.845	0.409	0.231	0.793	0.467	0.792
Panel B. Rossi's GC test (structural breaks), -2008 period										
AU				0.000	0.583				0.000	0.151
AU				0.053	1.000				0.437	1.000
CL				0.161	0.488				0.000	1.000
NZ				0.056	0.141				0.866	0.854
ZA				0.000	**0.000**				0.000	**0.000**
Panel C. Standard GC test, Newey-West, total period										
AU	0.007	0.324	−0.304	13.704	9.781	0.001	0.041	0.060	0.323	0.276
p-val	0.124	0.000	0.002	0.001	**0.002**	0.897	0.687	0.600	0.851	0.600
AU	0.007	0.116	0.056	1.849	0.084	0.002	0.010	−0.064	4.185	4.176
p-val	0.262	0.173	0.773	0.397	0.772	0.440	0.900	0.042	0.123	**0.041**
CL	0.016	0.012	−0.687	7.984	7.164	0.007	0.060	0.037	4.292	1.037
p-val	0.124	0.898	0.008	0.018	**0.007**	0.098	0.344	0.310	0.117	0.308
NZ	0.005	0.215	−0.241	5.829	4.048	−0.001	0.089	0.016	0.107	0.070
p-val	0.186	0.020	0.046	0.054	**0.044**	0.827	0.447	0.792	0.948	0.791
ZA	0.010	0.262	−0.082	2.948	1.126	0.015	0.010	−0.140	5.553	2.104
p-val	0.106	0.041	0.291	0.229	0.289	0.052	0.909	0.150	0.062	0.147
Panel D. Rossi's GC test (structural breaks), total period										
AU				0.000	**0.000**				1.000	1.000
AU				0.143	1.000				0.398	0.239
CL				0.099	**0.049**				0.000	0.750
NZ				0.071	**0.073**				0.887	0.777
ZA				0.612	0.884				0.059	0.448

Description Compared to Table 1, the data is an alternative set that, unlike CRR's original, is available till end 2021. Tests are run first on the -2008 period (CRR's) and then for the full period.

Interpretation: The table provides the inputs for Table 3 where the results are discussed.

Table A.3: Replication CRR Out-of-sample Forecast Tests; new dataset, two periods

		$s \to p$					$p \to s$				
	Benchmark	DSFE	RSFE	p-CW	[p-CMC CRR]	p-CMC correct	DSFE	RSFE	p-CW	[p-CMC CRR]	p-CMC correct
						Panel A: -2008 period					
AU	AR1	0.660	1.048	0.462	0.010	1.000	0.264	1.016	0.369	1.000	1.000
	[RW]	–2.135	0.838	0.001	0.010	0.010	0.498	1.034	0.417	0.100	1.000
	[RWwD]	–0.370	0.976	0.114	0.010	1.000	0.133	1.008	0.313	1.000	1.000
CA	AR1	1.181	1.039	0.826	1.000	1.000	1.583	1.025	0.913	0.100	1.000
	[RW]	0.118	1.004	0.362	1.000	1.000	0.619	1.014	0.562	1.000	1.000
	[RWwD]	1.057	1.032	0.770	1.000	1.000	1.547	1.025	0.901	0.010	1.000
CL	AR1	0.314	1.018	0.434	1.000	1.000	1.168	1.034	0.754	0.050	1.000
	[RW]	–0.231	0.979	0.174	0.050	1.000	0.820	1.074	0.595	1.000	1.000
	[RWwD]	0.021	1.001	0.313	1.000	1.000	0.925	1.036	0.678	0.100	1.000
NZ	AR1	1.352	1.069	0.761	0.100	1.000	0.686	1.023	0.615	1.000	1.000
	[RW]	0.280	1.020	0.334	0.010	1.000	0.667	1.027	0.548	1.000	1.000
	[RWwD]	1.261	1.056	0.761	0.100	1.000	0.708	1.019	0.599	1.000	1.000
ZA	AR1	2.727	1.079	0.996	0.050	1.000	1.407	1.129	0.903	1.000	1.000
	[RW]	–1.940	0.768	0.006	0.010	0.010	1.926	1.275	0.950	1.000	1.000
	[RWwD]	0.501	1.015	0.591	0.010	1.000	1.318	1.154	0.862	1.000	1.000
						Panel B: Total period					
AU	AR1	–1.738	0.913	**0.004**	0.010	**0.010**	0.867	1.069	0.627	1.000	1.000
	[RW]	–2.456	0.855	0.000	0.010	0.010	1.230	1.084	0.774	1.000	1.000
	[RWwD]	–2.146	0.867	0.001	0.010	0.010	1.017	1.066	0.715	1.000	1.000
CA	AR1	1.858	1.014	0.957	1.000	1.000	–0.037	0.999	0.217	1.000	1.000
	[RW]	1.567	1.024	0.864	1.000	1.000	0.170	1.008	0.193	0.050	0.100
	[RWwD]	1.730	1.018	0.918	1.000	1.000	–0.092	0.996	0.146	0.050	0.100
CL	AR1	–0.420	0.982	**0.058**	0.100	**0.050**	1.540	1.032	0.885	1.000	1.000
	[RW]	–0.171	0.988	0.065	0.050	0.050	1.722	1.057	0.918	1.000	1.000
	[RWwD]	–0.400	0.970	0.062	0.010	0.050	2.141	1.050	0.967	1.000	1.000
NZ	AR1	–1.295	0.915	**0.011**	0.010	**0.010**	1.348	1.036	0.851	1.000	1.000
	[RW]	–1.419	0.879	0.006	0.010	0.010	1.259	1.045	0.614	1.000	1.000
	[RWwD]	–1.445	0.886	0.008	0.010	0.010	0.988	1.031	0.544	1.000	1.000
ZA	AR1	0.630	1.014	0.558	1.000	1.000	–1.251	0.978	**0.028**	0.010	1.000
	[RW]	1.123	1.087	0.491	1.000	1.000	–0.857	0.960	0.041	0.050	0.100
	[RWwD]	–0.091	0.996	0.180	1.000	1.000	–1.399	0.971	0.014	1.000	1.000

Description: Compared to Table 2, the data is an alternative set that, unlike CRR's original, is available till end 2021. Tests are run first on the -2008 period (CRR's) and then for the full period.

Interpretation: The table provides the inputs for Table 3 where the results are discussed.