Using Common Features to Understand the Behavior of Metal-Commodity Prices and Forecast them at Different Horizons

João Victor Issler[*]   Claudia Rodrigues[†]   Rafael Burjack[‡]

December 31, 2012

Abstract

The objective of this article is to study (understand and forecast) spot metal price levels and changes at monthly, quarterly, and annual horizons. The data to be used consists of metal-commodity prices in a monthly frequency from 1957 to 2012 from the International Financial Statistics of the IMF on individual metal series. We will also employ the (relatively large) list of co-variates used in Welch and Goyal (2008) and in Hong and Yogo (2009) , which are available for download. Regarding short- and long-run comovement, we will apply the techniques and the tests proposed in the *common-feature* literature to build parsimonious VARs, which possibly entail *quasi-structural* relationships between different commodity prices and/or between a given commodity price and its potential demand determinants. These parsimonious VARs will be later used as forecasting models to be combined to yield metal-commodity prices optimal forecasts. Regarding out-of-sample forecasts, we will use a variety of

---

[*]Graduate School of Economics - EPGE, Getulio Vargas Foundation. email: `jissler@fgv.br`

[†]Investor Relations Department, Vale. email: `claudia.rodrigues@vale.com`

[‡]Graduate School of Economics - EPGE, Getulio Vargas Foundation email: `rburjack@fgvmail.br`

models (linear and non-linear, single equation and multivariate) and a variety of co-variates to forecast the returns and prices of metal commodities. With the forecasts of a large number of models (N large) and a large number of time periods (T large), we will apply the techniques put forth by the *common-feature* literature on forecast combinations.

1 Introduction

The purpose of this paper is twofold. The first is to improve our understanding of metal-commodity price variation either in the long run or in the short run by using standard time-series techniques. We rely on the *common-trend and common-cycle* approach put forward by Engle and Kozicki (1993), Vahid and Engle (1993, 1997), Engle and Issler (1995), Issler and Vahid (2001, 2006), Vahid and Issler (2002), Hecq et al. (2006), and Athanasopoulos et al. (2011). Here, non-stationary economic series are decomposed into an integrated trend component and a stationary and ergodic cyclical component, where their properties can be jointly investigated in a unified multivariate setting based on vector autoregressive (VAR) models. Trends and cycles can be common to a group of series being modelled, and these *common features* can be removed by independent linear combination[1]. Our second objective is to improve on current forecasts of metal-commodity prices taking into account the recent financialization of commodity markets and the role of information in commodity markets; see Hong and Yogo (2009, 2012) and Gargano and Timmermann (2012). Instead of relying on a specific model to forecast metal-commodity prices, we diversify out the risk of large forecast errors (and increase the information set used in forecasting) by combining forecasts of different models. This approach, first put forward by Bates and Granger (1969), has been shown to reduce forecast uncertainty of individual models in a variety of studies; see Hendry and Clements (2004) and Stock and Watson (2006). Recently, Issler and Lima (2009) have developed an optimal forecast-combination panel data approach, where forecasts of different models or survey results comprise the cross-sectional dimension. In their context, the op-

---

[1]Perhaps cointegration is the best-known example of *common features*.

timal forecast using a mean-squared error (MSE) risk function can be consistently estimated employing the bias-corrected average forecast (BCAF), which is a *common feature* of all forecast models and survey results.

From a theoretical point-of-view, commodity-price dynamics have been studied at least since Deaton and Laroque (1992, 1996) and Chambers and Bailey (1996). In the former, a strong first-order autocorrelation, consistent with a unit-root in prices, is generated by the intertemporal no-arbitrage condition for risk-neutral agents holding a speculative inventory of a given commodity. Moreover, actions of speculators generate (and modify) the serial dependence in prices even when there is no dependence in the original demand shocks, leading to demand-driven cyclical variations about trend prices. As Deaton and Laroque put it, "it is likely that demand shocks are a more plausible source of price fluctuations than has usually been supposed in the literature."

Early modern empirical work on commodity prices focused on the behavior of trend prices – Cuddington and Urzúa (1989) and Cuddington (1992) – where trends are modelled as martingale processes. Moreover, as (Deaton, 1999, p. 27) puts it, referring to the drift term in commodity prices: "what commodity prices lack in trend, they make up for in variance." Cashin et al. (2002) summarize the "stylized facts about real commodity prices: they are often dominated by long periods of doldrums punctuated by sharp upward spikes (Deaton and Laroque (1992)); they have a tendency to trend down in the long run (Grilli and Yang (1988)); shocks to commodity prices tend to persist for several years at a time (Cashin et al. (2000)); and unrelated commodity prices move together (Pindyck and Rotemberg (1990))." Regarding the cyclical behavior of commodity prices, Jerrett and Cuddington (2008) note that "a number of authors have analyzed the movement of metal prices over the business cycle as well as comovements among commodity prices (see Labys et al. (1999); Cashin et al. (1999); and Pindyck and Rotemberg (1990))."

Given these stylized facts, we investigate long-run comovement of metal-commodity prices by using cointegration analysis in two different ways: whether a group of metal-commodity prices have a similar long-run behavior (cointegrate) or if metal-commodity prices cointegrate with global industrial activity or a regional version of

3

that. The latter is motivated by the fact that economic theory views metal commodities as inputs in industrial production processes. Hence, there is a *derived demand* for metal commodities stemming from industrial production. Since there is a recent migration of industrial activity from developed countries to emerging economies, especially China and India, successful modeling of metal-commodity prices should include data from emerging markets.

Recently, there has been a renewed interest on commodity-price cycles, see Jerrett and Cuddington (2008) and IMF (2012), and our paper complements this recent effort. In our view, the study of short-run behavior of metal-commodity prices entails understanding their growth cycles using deviations from existing long-run relationships as predictors for future cyclical variation, i.e., a vector error-correction model (VECM).

Exploiting the error-correction structure has been the norm in several papers in the financial econometrics literature; see Lettau and Ludvigson (2001, 2004) and Welch and Goyal (2008). Indeed, the recent "financialization" of commodity markets – commodities being considered by financial investors as a distinct asset class[2] – supports the incorporation of financial econometric techniques into commodity-price analysis. Despite that, so far, the financial econometric literature has neglected the possibility of short-run restrictions of the form of synchronization of cycles, which come about naturally since the short-run determinants of (some) commodity prices are identical. Here, we advance with respect to previous literature: not only these two types of comovement can be jointly modelled, but the long- and short-run restrictions could also be jointly (or separately) imposed. There are also tests to determine if these restrictions are supported by the data.

One of the advantages of the *common-trend and common cycle* method is parsimony. As is well known, VARs have been increasingly used in multivariate analysis

---

[2]Commodities have gained recent notoriety among financial investors. According to the Investment Company Institute, total net assets of commodity exchange trade funds grew $10,000\%$ between 2004 and 2010 achieving more than $100 billion. From a financial perspective, precious metals have been used for hedging against exposure to crises and to inflation risk. Furthermore, commodity prices are connected to economic cycles. For example, crude oil price increases have been related to economic recessions, see Hamilton (2011).

of economic data. However, one of the shortcomings of these models is the excessive number of parameters. For example, a $VAR\left(p\right)$ for n series has $n^2 \cdot p$ parameters in the conditional mean. One can easily see the burden on degrees of freedom if the number of series being modelled is large. Cointegration certainly reduces the number of parameters, but these reductions are mild. On the other hand, short-run restrictions – or common cycles – have a much greater potential to reduce the number of parameters in the dynamic representation. For example, when dealing with post-war quarterly data, and a VAR with three variables and eight lags, there are seventy five mean parameters to be estimated from about two hundred data points on each variable. If the three-variable system has one known cointegrating vector, the number of free parameters falls from seventy five to sixty nine when estimating a VECM. *Common-cyclical features* show more potential in reducing the number of conditional-mean parameters. If the three variables in the VECM share one common cycle, then the number of mean parameters falls from sixty nine to twenty seven.

There are two advantages of having synchronized cycles is a multivariate data set when the *common-trend common cycle* method is employed. The first is the existence of contemporaneous relationships between the stationary series in the system (level for stationary series and first differences for integrated series). They appear because the (observed) cyclical determinants of the series involved are the same, implying the existence of linear combinations of the data that are orthogonal to these cyclical determinants, thus being white noise. These linear combinations can be thought as *quasi-structural* contemporaneous relationships, since, from a regression point of view, they involve nothing else but contemporaneous series[3]. As discussed above, the second advantage is parsimony, which implies more efficient parameter estimates and more precise out-of-sample forecasting models. The latter has been shown to be relevant from an empirical point of view; see Issler and Vahid (2001, 2006), Vahid and Issler (2002), and Athanasopoulos et al. (2011).

This brings us to the second objective of this paper, which is to improve on

---

[3]This is the case labelled strong-form serial-correlation common features. A less stringent restrictions allows a few lags to be in the linear combination as well, thus being labelled weak-form serial-correlation common features. Both are discussed in Section 2 and in the Appendix.

current forecasts of metal-commodity prices, taking into account the financialization of commodity markets and the role of information in commodity markets. Given the importance of the commodities market in world economy, and its importance for decision making among investors, it is not surprising that there are many studies relating the commodity prices with different state variables. Recently, Hong and Yogo (2009) finds evidence of limited in-sample predictability of commodity spot price and futures returns and Gargano and Timmermann (2012) finds no or little prediction power on a monthly horizon using the default return spread, growth in money supply, and the T-bill rate. They also find that the explanatory power of different state variables used in predicting commodity prices varies over time, with money supply showing some predictability after the great crisis, probably due to *Quantitative Easing* in its different versions. Despite these recent disappointing results, we believe that we can improve on current forecasts by using the optimal forecast-combination method of Issler and Lima (2009).

As is well known, Bates and Granger (1969) made the econometrics profession aware of the benefits of forecast combination when a limited number of forecasts is considered. The widespread use of different combination techniques has lead to an interesting puzzle from the econometrics point of view – the *forecast combination puzzle*: if we consider a fixed number of forecasts ($N < \infty$), combining them using equal weights ($1/N$) fare better than using "optimal weights" constructed to outperform any other forecast combination in the mean-squared error (MSE) sense. Regardless of how one combine forecasts, if the series being forecast is stationary and ergodic, and there is enough diversification among forecasts, we should expect that a weak law-of-large-numbers (WLLN) applies to well-behaved forecast combinations. This argument was considered in Palm and Zellner (1992) who asked the question "*to pool or not to pool*" forecasts? Timmermann (2006) used risk diversification – a principle so keen to finance – to defend pooling of forecasts. Of course, to obtain this WLLN result, at least the number of forecasts has to diverge ($N \to \infty$), which entails the use of asymptotic panel-data techniques. This is exactly the approach in Issler and Lima, with the added twist that now $N, T \to \infty$, with $T \to \infty$ prior than N: the sequential asymptotic approach developed by Phillips and Moon (1999),

6

denoted by $(T, N \to \infty)_{seq}$.

The reason why forecast combination works well is because it takes advantage of the principle of risk diversification: idiosyncratic forecast errors vanish because of the WLLN works as the number of forecasts being combined increases without bounds. However, the forecast combination puzzle also works against forecast combinations because of the *curse of dimensionality*: as N increases, if one has to estimate "optimal weights" to combine forecasts with a fixed number of observations, then, the estimates of these weights become increasingly volatile and are inconsistent.

Issler and Lima solve the curse of dimensionality by imposing equal weights that need not be estimated $(1/N)$, and perform bias correction to take MSE down to its minimum, identifying, in the limit, the conditional expectation of the series being forecast: if y_t is the series being forecast, and h is the horizon, then, what is being identified is the latent variable $\mathbb{E}_{t-h}(y_t)$, where $\mathbb{E}_{t-h}(\cdot)$ is the conditional expectation operator using all information available (observable or not) up to period $t - h$. Here, we are able to expand the information content of every individual model.

We now outline the contents of this paper. As stated above, its objective is to study (understand and forecast) spot metal price (log) levels and its changes at monthly, quarterly, and annual horizons. The data to be used consists of metal-commodity prices in a monthly frequency from 1957 to 2012 from the London Metal Exchange (LME) on individual metal series and inventories. We will also employ the (relatively large) list of co-variates used in Welch and Goyal (2008) and in Hong and Yogo (2012), which are available for download. Regarding short- and long-run comovement, we will apply the techniques and the tests proposed in the *common feature* literature to build parsimonious VARs, which possibly entail *quasi-structural* relationships between different commodity prices and/or between a given commodity price and its potential demand determinants; see the theoretical details in Section 2. These parsimonious VARs will be later used as forecasting models to be combined to yield metal-commodity prices optimal forecasts. Regarding out-of-sample forecasts, we will use a variety of models (linear and non-linear, single equation and multivariate) and a variety of co-variates to forecast the returns and prices of metal commodities. With the forecasts of a large number of models (N large) and a large

number of time periods (T large), we will apply the techniques discussed in Section 3, with results reported in detail in Section 4. Section 5 concludes. The Appendix contains additional material regarding short- and long-run comovement and the optimal forecast-combination technique.

2 Cointegration and Common Cycles for Metal Prices

The Appendix contains a brief summary of the techniques used here. For an in-depth theoretical discussion of these issues see Engle and Granger (1987), Vahid and Engle (1993), Vahid and Engle (1997), Hecq et al. (2006), and Athanasopoulos et al. (2011).

Assume that y_t is a n-vector of $I(1)$ metal prices (or log metal prices), which can be represented by a vector autoregression (VAR) model in levels:

$$y_t = \Gamma_1 y_{t-1} + \ldots + \Gamma_p y_{t-p} + \epsilon_t. \tag{1}$$

If elements of y_t cointegrate, Engle and Granger (1987) showed that the system (1) can be written as a Vector Error-Correction model (VECM):

$$\Delta y_t = \Gamma_1^* \Delta y_{t-1} + \ldots + \Gamma_{p-1}^* \Delta y_{t-p+1} + \gamma \alpha' y_{t-1} + \epsilon_t \tag{2}$$

where γ and α are full rank matrices of order $n \times r$, r is the rank of the cointegrating space, $-\left(I - \sum_{i=1}^{p} \Gamma_i\right) = \gamma\alpha'$, and $\Gamma_j^* = -\sum_{i=j+1}^{p} \Gamma_i$, $j = 1, \ldots, p-1$.

For our purposes, testing for cointegration will be used to verify whether metal-price data share common trends (or have long-run comovement). As is well known, metals are an important input in industrial processes, and thus it is expected that most metals would have their long-run prices linked to global industrial factors.

Testing for common trends among y_t will use the maximum-likelihood approach in Johansen (1991). A key issue to assure that inference is done properly in this case is

to estimate the lag length of the VAR (1) consistently, i.e., to estimate p consistently. Athanasopoulos et al. discuss how this can be achieved by using a combination of information criteria. An alternative to way to infer p is to perform diagnostic testing to rule out the risk of underestimation of p, which leads to inconsistent estimates for the parameters in (2).

Vahid and Engle (1993) show that the dynamic representation of y_t (2) may be restricted if there exist white noise independent linear combinations of the series Δy_t, i.e., if the Δy_t share *common cycles*. These white noise linear combinations of the series Δy_t can be expressed using cofeature vectors $\tilde{\alpha}_i'$, stacked in an $s \times n$ matrix $\tilde{\alpha}'$, which eliminate all serial correlation in Δy_t. Thus, $\tilde{\alpha}'\Delta y_t = \tilde{\alpha}'\epsilon_t$. This is what Hecq, Palm and Urbain (2006) have labelled *strong-form serial-correlation common features*:

$$\tilde{\alpha}'\Gamma_1^* = \tilde{\alpha}'\Gamma_2^* = \ldots = \tilde{\alpha}'\Gamma_{p-1}^* = 0, \text{ and} \qquad (3)$$

$$\tilde{\alpha}'\gamma = 0. \qquad (4)$$

If we only impose restrictions (3), but not (4), we obtain what they have labelled *weak-form serial-correlation common features*: $\tilde{\alpha}'[\Delta y_t - \gamma\alpha' y_{t-1}] = \tilde{\alpha}'\epsilon_t$, i.e., we only inherit an unpredictable linear combination of Δy_t once we *control* for the long-run deviations $\alpha' y_{t-1}$ stemming from cointegration.

We continue the discussion of common cycles in the case of strong-form serial-correlation common features $((3)$ and $(4))$, given that the weak-form case can be immediately inferred from it[4]. Since cofeature vectors are identified only up to an invertible transformation, without loss of generality, we can consider $\tilde{\alpha}$ to be of the form:

$$\tilde{\alpha} = \begin{bmatrix} I_s \\ \tilde{\alpha}_{(n-s)\times s}^* \end{bmatrix}$$

Completing the system by adding the unconstrained VECM equations for the re-

---

[4]The Appendix contains a more complete discussion.

maining $n - s$ elements of $\Delta\, y_t$, we obtain a *quasi-structural* model,

$$\left[\begin{array}{cc} I_s & \tilde{\alpha}^{*\prime} \\ \underset{(n-s)\times s}{\mathbf{0}} & I_{n-s} \end{array} \right] \Delta\, y_t = \left[\begin{array}{c} \underset{s\times(np+r)}{\mathbf{0}} \\ \Gamma_1^{**} \; \cdots \; \Gamma_{p-1}^{**} \; \gamma^* \end{array} \right] \left[\begin{array}{c} \Delta\, y_{t-1} \\ \vdots \\ \Delta\, y_{t-p+1} \\ \alpha' y_{t-1} \end{array} \right] + v_t. \qquad (5)$$

Since $\left[\begin{array}{cc} I_s & \tilde{\alpha}^{*\prime} \\ \underset{(n-s)\times s}{\mathbf{0}} & I_{n-s} \end{array} \right]$ is invertible, we can recover (2) from (5). However, that the latter has $s \cdot (np + r) - s \cdot (n - s)$ fewer parameters, thus, being over-identified.

One way to test for the existence of s serial-correlation common features (SCCF) is to use of canonical-correlation analysis; see Engle and Kozicki (1993) and the detailed discussion in Issler and Vahid (2001). The null is that the first smallest s canonical correlations are jointly zero and the test statistic is $-T \sum_{i=1}^{s} \log (1 - \lambda_i)$, where λ_i, $i = 1, \cdots, n$, are the sample squared canonical correlations between $\{\Delta\, y_t\}$ and $\{\alpha' y_{t-1}, \Delta\, y_{t-1}, \Delta\, y_{t-2}, \cdots, \Delta\, y_{t-p+1}\}$. The limiting distribution of this test statistic is χ^2 with $s\,(np + r) - s\,(n - s)$ degrees of freedom.

One possible drawback of the canonical-correlation approach is that it assumes homoskedastic data, and that may not hold for metal-price (and other macroeconomic and financial data) collected at high frequency. In this case, a GMM approach is more robust, since inference can be conducted with Heteroskedastic and Auto-Correlation (HAC) robust estimates of the variance-covariance matrices of parameter estimates. The vector of instruments comprise the series in $\alpha' y_{t-1}, \Delta\, y_{t-1}, \Delta\, y_{t-2}, \cdots, \Delta\, y_{t-p+1}$, collected in a vector Z_{t-1}. GMM estimation exploits the following moment restric-

tion:

$$
\begin{aligned}
\mathbf{0} \;=\;& \mathbb{E}\left[v_t \otimes Z_{t-1}\right] = \tag{6} \\
=\;& \mathbb{E}\left[\left(\left[\begin{array}{cc} I_s & \tilde{\alpha}^{*\prime} \\ \underset{(n-s)\times s}{\mathbf{0}} & I_{n-s} \end{array}\right]\Delta\,y_t - \left[\begin{array}{cc} \underset{s\times(np+r)}{\mathbf{0}} \\ \Gamma_1^{**} \;\ldots\; \Gamma_{p-1}^{**} \;\gamma^* \end{array}\right]\left[\begin{array}{c} \Delta\,y_{t-1} \\ \vdots \\ \Delta\,y_{t-p+1} \\ \alpha' y_{t-1} \end{array}\right]\right) \otimes Z_{t-1}\right],
\end{aligned}
$$

i.e., the orthogonality between all the elements in v_t and all the elements in Z_{t-1}. The test for common cycles is an over-identifying restriction test – the J test proposed in Hansen (1982) – which has an asymptotic χ^2 distribution with degrees of freedom equal to the number of over-identifying restrictions. The over-identifying restrictions test checks whether the errors of the system are orthogonal to all the instruments in Z_{t-1}.

3 A Theoretical Approach for Forecast Combination

The techniques discussed in this section are appropriate for forecasting a weakly stationary and ergodic univariate process $\{y_t\}$ using a large number of forecasts that will be combined to yield an optimal forecast in the mean-squared error (MSE) sense. These forecasts are the result of several econometric models that need to be estimated prior to forecasting. We label forecasts of y_t, computed using conditioning sets lagged h periods, by $f_{i,t}^h$, $i = 1, 2, \ldots, N$. Therefore, $f_{i,t}^h$ are h-step-ahead forecasts and N is the number of models estimated to forecast $f_{i,t}^h$.

Issler and Lima consider 3 consecutive distinct time sub-periods. The first sub-period E is labeled the "estimation sample", where models are usually fitted to forecast y_t subsequently. The number of observations in it is $E = T_1 = \kappa_1 \cdot T$, comprising $(t = 1, 2, \ldots, T_1)$. The sub-period R (for regression) is labeled the post-model-estimation or "training sample", where realizations of y_t are usually confronted

with forecasts produced in the estimation sample, and weights and bias-correction terms are estimated. It has $R = T_2 - T_1 = \kappa_2 \cdot T$ observations in it, comprising $(t = T_1 + 1, \ldots, T_2)$. The final sub-period is P (for prediction), where genuine out-of-sample forecast is entertained. It has $P = T - T_2 = \kappa_3 \cdot T$ observations in it, comprising $(t = T_2 + 1, \ldots, T)$.

Forecasts $f_{i,t}^h$'s are approximations to the optimal forecast $(\mathbb{E}_{t-h}(y_t))$ as follows:

$$f_{i,t}^h = \mathbb{E}_{t-h}(y_t) + k_i^h + \varepsilon_{i,t}^h, \tag{7}$$

where k_i^h is the individual model time-invariant bias for h-step-ahead prediction and $\varepsilon_{i,t}^h$ is the individual model error term in approximating $\mathbb{E}_{t-h}(y_t)$, where $\mathbb{E}(\varepsilon_{i,t}^h) = 0$ for all i, t, and h. Here, the optimal forecast is a *common feature* of all individual forecasts and k_i^h and $\varepsilon_{i,t}^h$ arise because of forecast misspecification.

We can always decompose the series y_t into $\mathbb{E}_{t-h}(y_t)$ and an unforecastable component ζ_t^h, such that $\mathbb{E}_{t-h}(\zeta_t^h) = 0$ in:

$$y_t = \mathbb{E}_{t-h}(y_t) + \zeta_t^h. \tag{8}$$

Combining (7) and (8) yields, the well known two-way decomposition, or error-component decomposition, of the forecast error $f_{i,t}^h - y_t$:

$$\begin{aligned} f_{i,t}^h &= y_t + \mu_{i,t}^h, & i = 1, 2, \ldots, N, \text{ and } T > T_1, \\ \mu_{i,t}^h &= k_i^h + \eta_t^h + \varepsilon_{i,t}^h, \text{ where } -\zeta_t^h = \eta_t^h \end{aligned} \tag{9}$$

From the perspective of combining forecasts, the components k_i^h, $\varepsilon_{i,t}^h$ and η_t^h play very different roles. If we regard the problem of forecast combination as one aimed at diversifying risk, i.e., a finance approach, then, on the one hand, the risk associated with $\varepsilon_{i,t}^h$ can be diversified, while that associated with η_t^h cannot. On the other hand, in principle, diversifying the risk associated with k_i^h can only be achieved if a bias-correction term is introduced in the forecast combination, which reinforces its usefulness.

Issler and Lima propose the following non-parameteric consistent estimates for the

components k_i^h, B^h, η_t^h, and $\varepsilon_{i,t}^h$: $\widehat{k}_i^h = \frac{1}{R} \sum_{t=T_1+2}^{T_2} f_{i,t}^h - \frac{1}{R} \sum_{t=T_1+2}^{T_2} y_t$, $\widehat{B^h} = \frac{1}{N} \sum_{i=1}^{N} \widehat{k}_i^h$, $\widehat{\eta}_t^h = \frac{1}{N} \sum_{i=1}^{N} f_{i,t}^h - \widehat{B^h} - y_t$, $\widehat{\varepsilon}_{i,t}^h = f_{i,t}^h - y_t - \widehat{k}_i^h - \widehat{\eta}_t^h$. They show that, under the set of conditions listed in this Appendix, the the *feasible* bias-corrected average forecast (BCAF) $\frac{1}{N} \sum_{i=1}^{N} f_{i,t}^h - \widehat{B^h}$ obeys:

$$
\underset{(T,N\to\infty)_{seq}}{\mathrm{plim}} \left(\frac{1}{N} \sum_{i=1}^{N} f_{i,t}^h - \widehat{B^h} \right) = y_t + \eta_t^h = \mathbb{E}_{t-h}(y_t),
$$

where $\mathrm{plim}_{(T,N\to\infty)_{seq}}$ is the probability limit using the sequential asymptotic framework of Phillips and Moon (1999). Thus, the feasible BCAF is an optimal forecasting device.

They also show that there is an infinite number of optimal forecast combinations using deterministic weights $\{\omega_i\}_{i=1}^{N}$, such that $\omega_i \neq 0$, $\omega_i = O(N^{-1})$ uniformly, with $\sum_{i=1}^{N} \omega_i = 1$ and $\lim_{N\to\infty} \sum_{i=1}^{N} \omega_i = 1$. This allows the discussion of the well-known *forecast combination puzzle*: if we consider a fixed number of forecasts ($N < \infty$), combining them using equal weights ($1/N$) fare better than using "optimal weights" constructed to outperform any other forecast combination in the mean-squared error (MSE) sense. Optimal population weights, constructed from the variance-covariance structure of models with stationary data, are optimal. Thus, the forecast-combination puzzle must be a consequence of the lack of consistency in estimating them, and can arise when N, the number of models being combined, is high relative to the number of observations used in estimating them by OLS – R.

Finally, there is one interesting case in which we can dispense with estimation in combining forecasts: when the mean bias is zero, i.e., $B^h = 0$, there is no need to estimate B^h and the BCAF is simply equal to $\frac{1}{N} \sum_{i=1}^{N} f_{i,t}^h$, the sample average of all forecasts. This is the ultimate level of parsimony. To be able to test the null that $B^h = 0$, Issler and Lima developed a robust t-ratio test that takes into account the cross-sectional dependence in k_i^h.

3.1 Forecast Combination for Nested Models

The potential problem of nested models is that the innovations from nested models can exhibit high cross-sectional dependence, preventing a weak law-of-large numbers (WLLN) to hold. We introduce nested models by considering a continuous set of models splitting the total number of models N into M classes (or blocks), each of them containing m nested models, so that $N = mM$. In the index of forecasts, $i = 1, \ldots, N$, we group nested models contiguously. Hence, models within each class (block) are nested but models across classes (blocks) are non-nested.

The number of classes and the number of models within each class to be functions of N, respectively as follows: $M = N^{1-d}$ and $m = N^d$, where $0 \leq d \leq 1$. Notice that this setup considers all the relevant cases: (i) $d = 0$ corresponds to the case in which all models are non-nested; $d = 1$ corresponds to the case in which all models are nested and; (iii) the intermediate case $0 < d < 1$ gives rise to N^{1-d} blocks of nested models, all with size N^d.

Regarding the interaction across blocks of nested models, it is natural to impose that the correlation structure of innovations across classes is such that it does not prevent a weak law-of-large numbers (WLLN) to hold, although we expect it not to hold within every block of nested models. Keeping some nested models poses no problem, since the mixture of models will still deliver the optimal forecast. From a practical point of view, the choice of $0 \leq d < 1$ seems to be superior and is sufficient to guarantee optimality of forecasts combinations as before.

4 Empirical analysis

4.1 Data

In our empirical analysis, we employed data of different frequencies and different sources building a very comprehensive dataset of metal prices and potential covariates that can be used in quasi-structural models and also for forecasting. We have now a library of data on three different frequencies: monthly, quarterly and annual.

On a monthly basis, the metal-price data consists of commodity prices for a variety of metals (or derived products) – Aluminium, Copper, Lead, Nickel, Tin and Zinc – provided by the International Financial Statistics (IFS) of the IMF. Metal-price data in this frequency is available from 1957 to 2012. Nominal price data were deflated using the consumer price index (CPI) for the U.S., which was extracted from the FRED database of the St. Louis FED. We have also a measure of industrial production in a monthly basis, constructed by J.P. Morgan, which is used in building quasi-structural models. It includes Chinese and India's industrial production. Because of that, this series is available only from 1992:01 through 2012:09. In forecasting, we used co-variates which are potentially correlated to the prices of these metals or derived products. These are mostly composed by financial indices downloaded from the library kept by Welch and Goyal (2008) and by Hong and Yogo (2012), available from 1965 to 2008. This list includes: global, U.S., and Chinese industrial production, the primary metals coincident and leading indices, provided by the United States Geological Service (USGS), and a few financial-sector co-variates, such as: VIX – a volatility index, the U.S. real effective exchange rate and the S&P500 index.

On a quarterly basis, the price data are for the same metals (or derived products) listed for monthly frequency, available from the IFS. Nominal price data were deflated using the CPI for the U.S. We also employed the (relatively large) list of co-variates used in Welch and Goyal (2008) and in Hong and Yogo (2012).

On an annual basis, metal-price data were provided by the United States Geological Service (USGS), from 1900 through 2010. Annual prices were deflated by the U.S. CPI – a series put together by the St. Louis and Minneapolis Federal Reserve Economic Database. Actual annual CPI data covers the period 1913-2010, whereas the period 1900-13 uses FED estimates. We also employed the list of annual co-variates used in Welch and Goyal (2008) and in Hong and Yogo (2012) and a list of financial indices and real economic variables, such as Angus Maddison's historical GDP, and Shiller's U.S. per capita real consumption.

Due to time constraints, our analysis of common-cyclical features will focus on the GMM tests proposed in Section 2, which is an appropriate testing strategy under

unknown heterogeneity and dependence of the moment restrictions in question, which ultimately depend on these same features regarding the data being used. The same is not true for the canonical-correlation analysis, therefore, we will not present it here in this version of the paper. Also due to time constraints, we only investigated here common-cyclical features between metal prices and industrial production on a monthly frequency, leaving for future research looking into this issue on a quarterly and annual basis.

Cointegration analysis investigates the existence of long-run relationships among economic data. As is well known, this requires the use of long-span data. Higher frequency at the expense of span is not a substitute for it. Thus, we should put more emphasis on cointegration tests using annual data, given it has the longest span – 110 years.

Regarding the forecasting exercise, the focus will be on monthly and annual frequencies alone, the former being appropriate to short-term forecasting and the latter to long-term forecasting. Time constraints also put a burden on our exercise, since we have experienced convergence problems for some of our algorithms computing average forecasts. For that reason, the number of forecasts on these averages had to reduced.

4.2 Cointegration and Common Cycles for Metal Prices

4.2.1 Monthly Frequency

Data for (log) prices of metals (or derived products) – Aluminium, Copper, Lead, Nickel, Tin and Zinc – are available from 1957:01 through 2012:03, whereas data for (log) Global Industrial Production (seasonally adjusted) is available from 1992:01 through 2012:09. All these series show signs of containing a unit root, which is confirmed for all of them using Phillips and Perron (1989) test[5]. We first investigate whether prices for metal commodities cointegrate and share common cycles with

---

[5]A slight caveat involves (log) aluminium prices, which rejects the null of a unit root at 5% significance when a constant is included, but rejects when a constant and trend are included. It also rejects Kwiatkowski et al. (1992) stationarity test. Thus, we chose to model it as a $I(1)$ process.

global industrial production. The analysis is pairwise, one commodity price at a time. Results are presented in Table 1.

Regarding cointegration, with the exception of aluminium, we find no evidence of a long-run relationship between metal prices and industrial production for the last 20 years. On the other hand, results for common cycles are very different using the GMM approach of Section 2 (robust to heteroskedasticity and serial correlation of unknown form). At 5% significance, we found evidence of strong-form *common-cyclical features* between industrial production and the following metals: copper, nickel, tin, and zinc. In addition to that, we also found evidence of strong-form *common-cyclical features* between industrial production and aluminium.

To motivate the findings of common-cyclical features in Table 1, we focus on the well-known association between copper prices and economic activity, which has been a conventional wisdom for financial analysts for a long time. Denote by P_t^{copper} the current price of copper, and by Y_t current industrial production. Using the Ljung-Box test at 10% significance, our empirical results found that the following linear combination is unpredictable:

$$\Delta \ln \left(P_t^{copper} \right) - \underset{(1.50)}{7.52} \times \Delta \ln \left(Y_t \right) + \underset{(0.006)}{0.015}, \tag{10}$$

with robust standard errors of parameter estimates in parenthesis. Notice that both $\Delta \ln \left(P_t^{copper} \right)$ and $\Delta \ln \left(Y_t \right)$ show strong signs of serial correlation using this same tool. Thus,

$$\mathbb{E}_t \left(\Delta \ln \left(Y_t \right) \right) = \mathbb{E}_t \left(\frac{\Delta \ln \left(P_t^{copper} \right)}{7.52} \right) + 0.002.$$

Therefore, it is possible to forecast global industrial production growth using the growth rate of copper prices, which is a very useful result since copper prices are traded daily but global industrial production is only available with some lag vis-a-vis current time[6].

Next, we analyze the pairwise behavior of metal commodity prices alone. Results are also presented in Table 1. They showed no signs of pairwise common cycles for

---

[6]The linear combination in (10) has an estimated standard deviation of 0.071.

Table 1: Common-Feature Tests (Monthly)

		Strong-form SCCF			Weak-form SCCF	
$\Delta y_{1,t}$	$\Delta y_{2,t}$	$\tilde{\alpha}^*$	J-statistic	Cointegration	$\tilde{\alpha}^*$	J-statistic
Aluminum	Lead	0.193*	0.0118	-	-	-
		(0.106)	[0.05]			
Copper	Aluminum	1.15***	0.0353	-	-	-
		(0.029)	[0.000]			
Aluminum	Industrial Production	5.316***	0.0427	(1,-0.201)	3.784***	0.0275
		(0.969)	[0.036]		(0.949)	[0.085]
Aluminum	Tin	0.168	0.0130	-	-	-
		(0.146)	[0.035]			
Nickel	Aluminum	0.735***	0.1407	-	-	-
		(0.186)	[0.000]			
Zinc	Aluminum	0.283	0.0344	(1,-0.463)	0.295	0.0344
		(0.188)	[0.012]		(0.19)	[0.007]
Lead	Copper	0.241**	0.0275	(1,-0.99)	0.274***	0.0262
		(0.097)	[0.001]		(0.096)	[0.001]
Lead	Industrial Production	4.052*	0.0331	-	-	
		(2.101)	[0.047]			
Tin	Lead	0.38***	0.0320	(1,-1.625)	0.411***	0.0303
		(0.094)	[0.000]		(0.097)	[0.000]
Nickel	Lead	0.23	0.0265	(1,-0.673)	0.32	0.0257
		(0.189)	[0.000]		(0.194)	[0.000]
Zinc	Lead	0.383***	0.0255	(1,-0.488)	0.467***	0.0256
		(0.146)***	[0.000]		(0.156)	[0.000]
Copper	Industrial Production	7.523***	0.0310	-	-	-
		(1.504)	[0.189]			
Copper	Tin	0.841***	0.0403	-	-	-
		(0.238)	[0.000]			
Copper	Nickel	0.319**	0.0358	(1,-1.67)	0.365***	0.0355
		(0.128)	[0.000]		(0.133)	[0.000]
Zinc	Copper	0.442***	0.0479	(1,-0.281)	0.418***	0.0431
		(0.094)	[0.000]		(0.093)	[0.000]
Tin	Industrial Production	5.23***	0.0096	-	-	-
		(1.603)	[0.512]			
Tin	Nickel	0.122*	0.0361	(1,-5.198)	0.165**	0.0379
		(0.07)	[0.001]		(0.077)	[0.000]
Tin	Zinc	0.284***	0.0377	(1,-4.035)		0.0362
		(0.095)	[0.006]			[0.005]
Nickel	Industrial Production	6.034***	0.0292	-	-	-
		(1.728)	[0.219]			
Zinc	Nickel	0.287***	0.0233	(1,-0.922)	0.278***	0.0238
		(0.092)	[0.000]		(0.086)	[0.000]
Zinc	Industrial Production	5.827***	0.0337	-	-	-
		(1.601)	[0.329]			

Notes: GMM estimation using equation (6) for Strong-form SCCF and the analogue equation for
Weak-form SCCF at the appendix.

Standard errors are in parentheses and p-values are in brackets.

the growth rates of metal-commodity prices.

4.2.2 Quarterly Frequency

On a quarterly frequency, data for (log) prices of metals (or derived products) – Aluminium, Copper, Lead, Nickel, Tin and Zinc – are available from 1957:01 through 2012:01. Table 2 presents results of pairwise cointegration between metal prices, reporting overwhelming evidence of cointegration between prices of different metal commodities. In this case, reduced-form models are vector error-correction models (VECMs) and *quasi-structural* models entail, in the first equation, a linear combination of two metal price growth rates. The second equation of the structural system is just a reduced-form counterpart.

We found evidence of serial-correlation common features among commodity-price growth when testing employed the GMM setup described in Section 2. For all possible 15 pairwise cases, we found strong-form SCCF for 6 of them – Aluminium-Copper, Aluminium-Lead, Aluminium-Nickel, Copper-Nickel, Lead-Zinc, and Nickel-Tin. It seems that Aluminium and Nickel are well synchronized with other metal commodities. Similar results are also obtained for weak-form SCCF.

4.2.3 Annual Frequency

On an annual basis, metal-price data were provided by the United States Geological Service (USGS), from 1900 through 2010, deflated by the U.S. CPI. Table 4 presents results of pairwise cointegration between metal prices, reporting overwhelming evidence of cointegration between prices of different metal commodities. Given the longer span of this annual database vis-a-vis the monthly and quarterly databases – more than twice as long – cointegrating evidence here should receive more weight vis-a-vis previous evidence.

Table 3 presents the results of common-feature tests for annual data. From all possible 15 cases, we found cointegration among 10 pairs of metal-commodity prices. One interesting issue is the long-run behavior of real metal-commodity prices: while three of them displayed an obvious increase in prices (Copper, Nickel, and Zinc)

Table 2: Common-Feature Tests (Quarterly)

$\Delta y_{1,t}$	$\Delta y_{2,t}$	Strong-form SCCF $\tilde{\alpha}^*$	Strong-form SCCF J-statistic	Cointegration	Weak-form SCCF $\tilde{\alpha}^*$	Weak-form SCCF J-statistic
Aluminum	Lead	0.241**	0.0604	(1,-0.278)	0.144	0.0543
		(0.114)	[0.225]		(0.123)	[0.232]
Copper	Aluminum	0.433***	0.0647	(1,-0.134)	0.355***	0.0658
		(0.128)	[0.177]		(0.136)	[0.117]
Aluminum	Tin	0.377	0.0473	(1,-0.203)	0.429*	0.0430
		(0.241)	[0.004]		(0.244)	[0.002]
Aluminum	Nickel	0.391***	0.0326	(1,9.784)	0.366***	0.0319
		(0.091)	[0.314]		(0.108)	[0.227]
Zinc	Aluminum	0.67**	0.0356	(1,-0.836)	0.467	0.0313
		(0.275)	[0.02]		(0.317)	[0.009]
Lead	Copper	0.562***	0.0447	(1,-1.013)	0.7***	0.0364
		(0.169)	[0.007]		(0.214)	[0.005]
Lead	Tin	0.701***	0.0583	(1,-0.548)	0.904***	0.0456
		(0.137)	[0.049]		(0.16)	[0.079]
Nickel	Lead	-0.126	0.0594	(1,-0.677)	0.078	0.0542
		(0.176)	[0.045]		(0.186)	[0.038]
Zinc	Lead	0.477***	0.0528	(1,-2.593)	0.497***	0.0573
		(0.136)	[0.076]		(0.141)	[0.029]
Copper	Tin	0.782***	0.0520	-	-	-
		(0.199)	[0.01]			
Copper	Nickel	0.681***	0.0525	(1,-0.412)	0.618***	0.0408
		(0.247)	[0.077]		(0.237)	[0.115]
Zinc	Copper	0.904***	0.0517	(1,-0.269)	0.732***	0.0435
		(0.142)	[0.024]		(0.161)	[0.023]
Tin	Nickel	0.604***	0.0573	(1,-0.192)	0.962***	0.0451
		(0.189)	[0.053]		(0.246)	[0.081]
Zinc	Tin	0.479**	0.0505	(1,-0.249)	0.714**	0.0459
		(0.232)	[0.004]		(0.292)	[0.002]
Zinc	Nickel	0.604***	0.0303	(1,-0.517)	0.552***	0.0234
		(0.18)	[0.036]		(0.186)	[0.024]

Notes: GMM estimation using equation (6) for Strong-form SCCF and the analogue equation for
Weak-form SCCF at the appendix.

Standard errors are in parentheses and p-values are in brackets.

in 110 years – more than double from 1900 to 2010 – the other three displayed an obvious decrease over time (Aluminium, Lead, and Tin) of about 70%-90%.

We also found overwhelming evidence of SCCF among commodity-price growth when testing employed the GMM setup described in Section 2. For all possible 15 pairwise cases, we found strong-form SCCF for 14 of them, the only exception being the pair Tin-Zinc. Similar results are also obtained for weak-form SCCF. One point to note is that annual data for metal-prices showed much more synchronization than did quarterly and monthly data. This is a sign that some high-frequency fluctuations that are not synchronized tend to disappear with time aggregation.

Finally, we illustrate our empirical results for the case of Aluminium and Lead at annual frequency. Let $\Delta y_t = \left(\Delta \ln \left(P_t^{Al} \right), \Delta \ln \left(P_t^{Lead} \right) \right)'$ be a 2×1 vector containing the instantaneous growth rates of the prices of Aluminium and Lead, respectively. Then, our estimated *quasi-structural* model took the form[7]:

$$
\begin{bmatrix} 1 & \underset{(0.12)}{-0.62} \\ 0 & 1 \end{bmatrix} \Delta y_t = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ \underset{(0.12)}{0.30} & \underset{(0.17)}{-0.23} & \underset{(0.10)}{0.10} & \underset{(0.14)}{0.07} & \underset{(0.01)}{-0.09} \end{bmatrix} \begin{bmatrix} \Delta y_{t-1} \\ \Delta y_{t-2} \\ \alpha' y_{t-1} \end{bmatrix},
$$

where the first equation is the white-noise linear combination of $\Delta \ln \left(P_t^{Al} \right)$ and $\Delta \ln \left(P_t^{Lead} \right)$. Notice the reduced-rank structure of the matrix of estimated parameters. It yields a parsimonious representation for Δy_t: instead of estimating 12 parameters in the VECM, we only estimated 7 in the *quasi-structural form*. As theory and experience has taught us, the latter forecasts much better than the former.

4.3   Forecasting Metal Prices by Combining Model Forecasts

In order to show the empirical applicability of the forecasting techniques discussed above, we conducted an experiment to assess the predictability performance of different models of a metal-commodity prices, where forecast accuracy is measured by the root of the mean-squared forecast error. The metal-price data used here is the same one used in the cointegration and common-cycle analyses of the previous sec-

---

[7]Disregarding constant terms.

Table 3: Common-Feature Tests (Annual)

		Strong-form SCCF			Weak-form SCCF	
$\Delta y_{1,t}$	$\Delta y_{2,t}$	$\tilde{\alpha}^*$	J-statistic	Cointegration	$\tilde{\alpha}^*$	J-statistic
Aluminum	Lead	0.621***	0.0112	(1,3.63)	0.601**	0.0111
		(0.116)	[0.876]		(0.239)	[0.752]
Aluminum	Copper	-0.552***	0.0515	-	-	-
		(0.21)	[0.135]			
Aluminum	Tin	0.02**	0.0454	(1,0.122)	0.012	0.0301
		(0.008)	[0.778]		(0.008)	[0.867]
Aluminum	Nickel	-0.54***	0.0638	(1,-0.399)	0.043	0.0394
		(0.188)	[0.142]		(0.251)	[0.235]
Aluminum	Zinc	-0.264*	0.0730	(1,-0.298)	0.121	0.0429
		(0.151)	[0.096]		(0.132)	[0.201]
Copper	Lead	-0.286**	0.0116	(1,2.557)	-0.241	0.0115
		(0.117)	[0.533]		(0.363)	[0.262]
Tin	Lead	-	-	-	-	-
		-	-		-	-
Lead	Nickel	-0.618***	0.0836	(1,0.311)	0.134	0.0283
		(0.167)	[0.06]		(0.177)	[0.382]
Zinc	Lead	-0.095	0.0297	-	-	-
		(0.1)	[0.36]		-	-
Copper	Tin	-0.033	0.0579	(1,2.818)	-0.095	0.0535
		(0.075)	[0.402]		(0.132)	[0.334]
Copper	Nickel	0.223*	0.0367	-	-	-
		(0.134)	[0.265]		-	-
Copper	Zinc	0.009	0.0408	(1,-0.973)	0.193	0.0260
		(0.077)	[0.354]		(0.132)	[0.422]
Nickel	Tin	-0.564***	0.0569	(1,2.95)	-0.634**	0.0548
		(0.099)	[0.189]		(0.248)	[0.116]
Tin	Zinc	-0.758***	0.0901	(1,0.372)	-0.465***	0.0301
		(0.111)	[0.045]		(0.118)	[0.355]
Nickel	Zinc	0.498***	0.0561	-	-	-
		(0.084)	[0.116]		-	-

Notes: GMM estimation using equation (6) for Strong-form SCCF and the analogue equation for
Weak-form SCCF at the appendix.
Standard errors are in parentheses and p-values are in brackets.

tion, although we only focused on results for monthly and annual data alone. The former is appropriate to examine short-term forecast accuracy, whereas the latter is appropriate for long-term accuracy.

Our target variables in forecasting are commodity prices for a variety of metals (or derived products) – Aluminium, Copper, Lead, Nickel, Tin and Zinc – extracted from the IFS when monthly frequency is used and from the USGS when annual data is employed. For some of the estimated models, we used co-variates which are potentially correlated to the metal prices. Some are related to the economic activity, such as: the global industrial production, the U.S. industrial production, the Chinese industrial production, the primary metals coincident index (USGS), a leading index of metals price (USGS), and some other financial-sector co-variates, such as: VIX – a volatility index, the U.S. real effective exchange rate and the S&P500 index.

Our monthly data set covers the period from January 1965 to December 2008, comprising 528 observations ($T = 528$). Correlations between each co-variate and the metals price index are presented in Table 4. Our annual data set covers data from 1900 to 2010, comprising 111 observations ($T = 111$). Since there is evidence of a unit root for the metals price series and the co-variates, all series in Table 4 were transformed to instantaneous growth rates prior to computing correlations:

<div align="center">Table 4: Correlations Between the Metal Prices and Co-variates</div>

	Aluminum	Copper	Lead	Nickel	Tin	Zinc	BASISM	ISRETM	LRETB10	LTY	RV	SP500	TBL	USIP	VSRETM
Aluminum	1	0.427	0.671	0.605	0.420	0.110	-0.080	0.026	-0.041	-0.216	-0.193	-0.461	-0.095	-0.540	-0.210
Copper	-	1	0.561	0.217	0.249	0.302	0.257	0.043	-0.079	0.135	-0.062	-0.508	0.257	-0.548	-0.084
Lead	-	-	1	0.333	0.672	0.310	-0.091	0.028	-0.045	0.157	-0.128	-0.518	0.209	-0.540	-0.105
Nickel	-	-	-	1	0.165	0.378	-0.117	-0.007	-0.050	0.007	-0.024	-0.073	0.021	-0.087	-0.068
Tin	-	-	-	-	1	0.130	-0.052	-0.022	0.026	0.427	-0.157	-0.585	0.303	-0.498	-0.039
Zinc	-	-	-	-	-	1	-0.165	-0.051	0.032	0.238	0.001	0.019	0.231	0.000	0.041
BASISM	-	-	-	-	-	-	1	0.142	-0.141	0.015	0.075	-0.033	0.166	-0.009	0.116
ISRETM	-	-	-	-	-	-	-	1	-0.090	-0.086	-0.156	0.025	-0.074	0.010	0.113
LRETB10	-	-	-	-	-	-	-	-	1	-0.053	0.126	0.034	-0.065	0.059	0.022
LTY	-	-	-	-	-	-	-	-	-	1	-0.071	-0.505	0.838	-0.354	0.131
RV	-	-	-	-	-	-	-	-	-	-	1	0.295	-0.096	0.320	0.164
S&P500	-	-	-	-	-	-	-	-	-	-	-	1	-0.505	0.955	0.121
TBL	-	-	-	-	-	-	-	-	-	-	-	-	1	-0.435	0.089
USIP	-	-	-	-	-	-	-	-	-	-	-	-	-	1	0.173
VSRETM	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1

Notes: ISRETM Metals Price Index: 1-month excess spot returns; LRETB10: 10-year Treasury bond: 1-month excess returns LTY: Long term yield; RV Realized volatility on S&P 500; TBL:Treasury-bill rates 3-months; USIP: Industrial Production (USA); VSRETM: Metals Price Index: Volatility of 1-month excess spot returns

In order to fit well the cross-sectional asymptotic requirement (large N) regarding

the WLLN, we need to have a large set of diversified forecasts to eliminate the combination of idiosyncratic errors. For this reason, we chose a few classes of different econometric models: ARMA, GARCH, VAR, VECM, with distinct co-variates, and distinct functional forms (levels, logs), and stationarity assumptions (stationarity vs difference-stationarity) for the target variable and co-variates. Considering that some of these models were fairly similar, we discarded a few of those, ending up with N between 109 and 144, i.e., between 109 and 144 distinct models and distinct forecasts for each time horizon (h). Obviously, some of them are nested within each other, and we also have classes of nested models as well. As we argued before, this will not pose a problem to applying the asymptotic techniques as long as we have a large enough number of diverse classes.

For implementing the BCAF and other combining techniques discussed above, we split the sample in three distinct parts, each with a specific purpose: the first one, from 1 to T_1, to estimate the coefficients of each model; the second from $T_1 + 1$ to T_2, to compute the bias; and the third from $T_2 + 1$ to T, to implement truly out-of-sample forecasting, and to assess the forecast accuracy of different forecast strategies and of individual models using the root mean-squared error (RMSE) of forecasts.

To asses forecast accuracy, we constructed an algorithm which is appropriate for the bias-corrected average forecast (BCAF). For alternative forecast combinations or forecasting schemes, slight modifications are required. The algorithm runs as follows:

1. For each model (AR, GARCH or VAR, and a specific set of co-variates), we estimate the coefficients of the regressors using the sub-sample from 1 to T_1.

2. Forecast h-steps ahead the models estimated in step 1 (f_{it}^h) from T_1 to T_2. Each model should be forecasted h-steps ahead $T_2 - T_1 - h + 1$ times.

3. Calculate the bias associated with each h-steps ahead forecast and each model; the bias is the average error between the h-steps ahead forecast and the observed value of the target series (from T_1 to T_2).

4. Forecast h-steps ahead the same models estimated in step 1 for only $T_2 + h$, using the same coefficients estimated in step 1.

5. Store the bias from step 3 and the forecast made in step 4 f^h_{i,T_2+h}.

6. Update $T_1 = T_1 + 1$, $T_2 = T_2 + 1$.

7. Go to step 1 until $T_2 = T$.

8. Adjust the forecasts of each model (made from $T_2 + 1$ to T) by their respective bias.

9. Combine all these adjusted forecasts using equal weights.

10. Compute the RMSE of the BCAF, considering the series of metals price index as the target series.

For the monthly dataset, we took $T_1 = 200$ and $T_2 = 378$. Since $T = 528$, this leaves 150 observations to evaluate out-of-sample performance of different models. For the annual data set, we took $T_1 = 35$ and $T_2 = 70$. Since $T = 111$, this leaves 41 observations for out-of-sample evaluation. In both case, the idea was to keep enough data to estimate the models and two similar-size sub-samples to compute the bias and make the forecasts.

The maximum horizon was set to 6 months for monthly data and to 5 years with annual data.

After computing the average bias for each forecast horizon $(\widehat{B^h})$, we tested the null $H_0 : B^h = 0$, using Issler and Lima's t-ratio test. Tables 5 and 6 present the results, respectively with monthly and annual data.

From the results in Table 5, we conclude that none of the monthly biases were deemed significant by the t-ratio test in Issler and Lima, although with annual data the mean bias for Zinc is nonzero (Table 6). This suggests that, for monthly data, it might not be worth to discard a part of the sample to compute the bias. Since the BCAF collapses to the simple average forecast (AF) when the bias is zero, then, the best strategy is simply to merge the samples E and R into one, where models are estimated. With these models, we can then compute forecast for the sample P, combining them using $\frac{1}{N}\sum_{i=1}^{N} f^h_{i,t}$. With the exception of Zinc for annual data, this is also true for all the other metal prices.

Table 5: Mean Bias Significance Test (Monthly)

Aluminum	Bias	t-statistic	p-value	Copper	Bias	t-statistic	p-value
1 step-ahead	-0.003	-0.014	0.495		0.086	0.155	0.438
2 step-ahead	-0.037	-0.084	0.466		0.228	0.180	0.429
3 step-ahead	-0.057	-0.084	0.466		0.378	0.194	0.423
4 step-ahead	-0.085	-0.095	0.462		0.549	0.213	0.416
5 step-ahead	-0.105	-0.095	0.462		0.747	0.235	0.407
6 step-ahead	-0.114	-0.088	0.465		0.973	0.257	0.398

Lead	Bias	t-statistic	p-value	Nickel	Bias	t-statistic	p-value
1 step-ahead	0.055	0.260	0.397		0.244	0.548	0.292
2 step-ahead	0.129	0.265	0.395		0.462	0.505	0.307
3 step-ahead	0.212	0.273	0.392		0.686	0.502	0.308
4 step-ahead	0.298	0.280	0.390		0.789	0.441	0.330
5 step-ahead	0.389	0.290	0.386		0.923	0.422	0.336
6 step-ahead	0.485	0.299	0.382		1.151	0.448	0.327

Tin	Bias	t-statistic	p-value	Zinc	Bias	t-statistic	p-value
1 step-ahead	1.689	0.420	0.337		-	-	-
2 step-ahead	3.816	0.406	0.342		-	-	-
3 step-ahead	5.982	0.406	0.342		-	-	-
4 step-ahead	8.036	0.406	0.342		-	-	-
5 step-ahead	10.057	0.408	0.342		-	-	-
6 step-ahead	12.061	0.410	0.341		-	-	-

Table 6: Mean Bias Significance Test (Annual)

Aluminum				Copper			
	Bias	t-statistic	p-value		Bias	t-statistic	p-value
1 step-ahead	-276.842	-0.525	0.300		-233.249	-1.114	0.133
2 step-ahead	-444.360	-0.524	0.300		-400.706	-1.114	0.133
3 step-ahead	-486.603	-0.474	0.318		-525.299	-1.212	0.113
4 step-ahead	-487.179	-0.417	0.338		-619.110	-1.309	0.095
5 step-ahead	-450.445	-0.355	0.361		-689.353	-1.349	0.089

Lead				Nickel			
	Bias	t-statistic	p-value		Bias	t-statistic	p-value
1 step-ahead	263.951	0.435	0.332		-1140.074	-1.073	0.142
2 step-ahead	602.373	0.454	0.325		-1834.030	-1.220	0.111
3 step-ahead	783.477	0.458	0.323		-2542.124	-1.500	0.067
4 step-ahead	744.257	0.434	0.332		-3141.645	-1.793	0.037
5 step-ahead	666.113	0.384	0.351		-3495.002	-1.909	0.028

Tin				Zinc			
	Bias	t-statistic	p-value		Bias	t-statistic	p-value
1 step-ahead	16049.680	0.845	0.199		-276.644	-1.277	0.101
2 step-ahead	37310.520	0.751	0.226		-458.376	-1.662	0.048
3 step-ahead	55645.350	0.743	0.229		-586.994	-2.044	0.021
4 step-ahead	53266.960	0.818	0.207		-702.797	-2.652	0.004
5 step-ahead	42927.230	0.983	0.163		-749.982	-2.879	0.002

27

Given the theoretical results in Section 3 for optimality of combined forecasts, we also computed alternative forecast combinations using different strategies: (i) the weighted average forecast (WAF), where weights are based on the inverse of the mean-squared error for each model, normalized to add up to unity; (ii) take the simple average of the 5 best fitting models (by Bayesian Information Criteria); (iii) take the simple average of the 10 best fitting models (by Bayesian Information Criteria); (iv) the median forecast[8].

4.3.1 Comparing forecast accuracy of different models

We now compare the BCAF to other popular forms of forecast combination, such as the simple average forecast $\left(\text{AF}: \frac{1}{N}\sum_{i=1}^{N} f_{i,t}^{h}\right)$ and the weighted average forecast (WAF), in which weights are based on the inverse of the mean-squared error for each model, where weights are normalized to add up to unity. Using data for copper, Figures 1 and 2 present the results for all forecast-combinations strategies for horizons 1- and 6-steps ahead respectively:

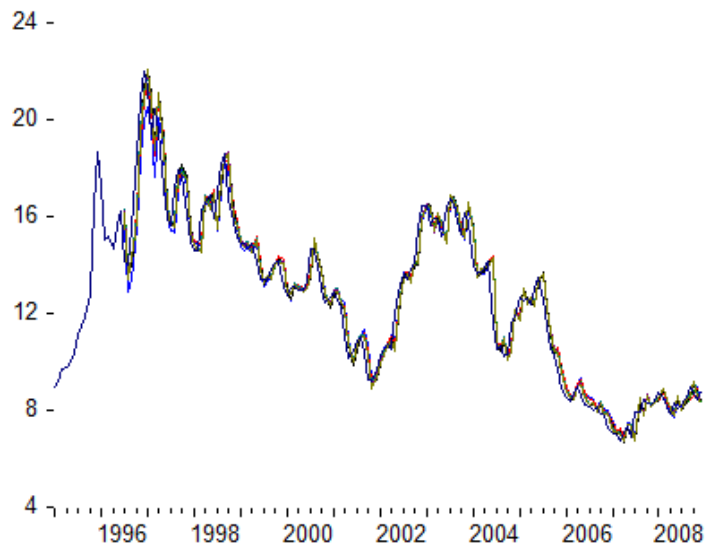[8]For this case we have no optimality result, but we keep it just for reference.

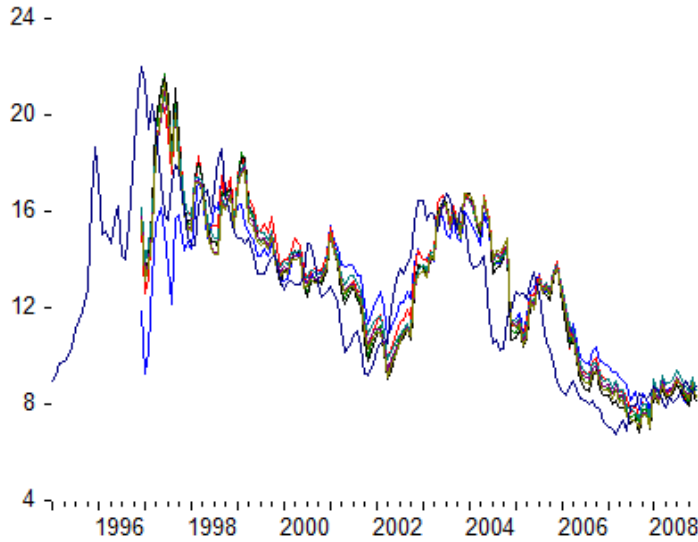Figure 1: Forecast Combination Schemes for Copper
Prices – 1-step ahead

Figure 2: Forecast Combination Schemes for Copper
Prices – 6-steps ahead

Next, we computed the RMSE for the BCAF, the WAF, the AF, the best individual model selected by BIC (Bayesian Information Criterium), and the combinations using the 5-best and 10-best models. Results are presented in Tables 7 and 8[9], respectively for monthly and annual data.

For monthly data, for 1- through 6-steps ahead and across all metal prices, the best performance in terms of RMSE was achieved by the WAF, followed closely by the BCAF and the AF, which are all forecast-combination schemes. On occasion, the median forecast and the best model also performed well.

Regarding annual data, the BCAF outperformed other forecasts strategies, although some individual models performed really well for Copper and Nickel.

The results of this empirical exercise show that combining forecasts proved to be the best strategy to forecast the metal prices: either using the bias-corrected average

_____

[9]In Table 7, we had convergence problems regarding Zinc forecasts, which are not presented in nthis version of the paper.

Aluminum	BCAF	Weighted Average (MSE)	Average Forecast	Median	Best Model (BIC)	5 Best Models	10 Best Models
1 step-ahead	0.619	0.605	0.678	0.663	0.732	0.688	0.684
2 step-ahead	0.926	0.888	1.049	1.005	1.158	1.096	1.057
3 step-ahead	1.402	1.327	1.635	1.570	1.914	1.804	1.699
4 step-ahead	1.674	1.548	1.981	1.911	2.461	2.315	2.129
5 step-ahead	1.833	1.647	2.197	2.161	2.866	2.722	2.436
6 step-ahead	2.125	1.861	2.564	2.530	3.459	3.330	2.912

Copper	BCAF	Weighted Average (MSE)	Average Forecast	Median	Best Model (BIC)	5 Best Models	10 Best Models
1 step-ahead	0.647	0.575	0.571	0.574	0.55	0.566	0.569
2 step-ahead	1.999	1.706	1.716	1.717	1.631	1.695	1.707
3 step-ahead	3.378	2.779	2.817	2.804	2.698	2.795	2.815
4 step-ahead	4.471	3.557	3.587	3.579	3.485	3.604	3.634
5 step-ahead	5.219	4.123	4.091	4.060	4.04	4.157	4.193
6 step-ahead	5.718	4.619	4.592	4.55	4.566	4.672	4.711

Lead	BCAF	Weighted Average (MSE)	Average Forecast	Median	Best Model (BIC)	5 Best Models	10 Best Models
1 step-ahead	0.074	0.074	0.073	0.072	0.076	0.073	0.074
2 step-ahead	0.143	0.145	0.144	0.143	0.150	0.145	0.146
3 step-ahead	0.2	0.206	0.206	0.205	0.214	0.209	0.209
4 step-ahead	0.247	0.256	0.258	0.257	0.272	0.267	0.264
5 step-ahead	0.281	0.293	0.293	0.291	0.310	0.307	0.303
6 step-ahead	0.329	0.341	0.343	0.342	0.367	0.367	0.359

Nickel	BCAF	Weighted Average (MSE)	Average Forecast	Median	Best Model (BIC)	5 Best Models	10 Best Models
1 step-ahead	17.916	17.670	15.688	15.817	16.221	17.095	16.231
2 step-ahead	48.508	47.784	48.316	49.425	50.291	51.964	49.464
3 step-ahead	76.682	75.53	83.621	86.079	88.583	91.630	86.640
4 step-ahead	92.064	91	106.000	109.447	112.618	117.515	110.329
5 step-ahead	100.430	99.401	119.853	123.473	124.583	131.369	122.996
6 step-ahead	105.895	104.082	126.449	131.179	127.222	136.305	126.641

Tin	BCAF	Weighted Average (MSE)	Average Forecast	Median	Best Model (BIC)	5 Best Models	10 Best Models
1 step-ahead	3.497	3.374	2.614	2.535	2.479	2.497	2.502
2 step-ahead	11.160	9.501	7.130	7.002	6.853	6.778	6.757
3 step-ahead	23.028	18.181	13.257	13.039	13.128	12.894	12.853
4 step-ahead	37.227	27.824	20.298	19.901	20.564	20.184	20.094
5 step-ahead	51.951	36.476	26.372	25.575	26.988	26.381	26.214
6 step-ahead	67.825	44.722	31.727	30.62	32.705	31.877	31.612

Zinc	BCAF	Weighted Average (MSE)	Average Forecast	Median	Best Model (BIC)	5 Best Models	10 Best Models
1 step-ahead	-	-	-	-	-	-	-
2 step-ahead	-	-	-	-	-	-	-
3 step-ahead	-	-	-	-	-	-	-
4 step-ahead	-	-	-	-	-	-	-
5 step-ahead	-	-	-	-	-	-	-
6 step-ahead	-	-	-	-	-	-	-

$T = 528$, $T_1 = 200$, $T_2 = 378$ and the out-of-sample range is 150

Table 7: Forecast Root-Mean-Squarred-Error (Monthly)

forecast (BCAF), the weighted average forecast (WAF), or the average forecast (AF). This was true for most metal prices and horizons, although some individual models performed well on occasion.

5 Conclusion and Further Research

The objective of this paper was to study (understand and forecast) spot metal price levels and changes at monthly, quarterly, and annual horizons. The data to be used consists of metal-commodity prices in a monthly and quarterly frequencies from 1957 to 2012 from the IFS. At annual frequency we use USGS data from 1900 through 2010. We will also employ the (relatively large) list of co-variates used in Welch and Goyal (2008) and in Hong and Yogo (2009) , which are available for download.

Regarding short- and long-run comovement, we tested and confirmed their existence using the techniques proposed in the *common-feature* literature at different frequencies. These findings entail the estimation of parsimonious VARs with cointegration restrictions and *quasi-structural* relationships. Cointegration restrictions were found with long-span annual data: using pairwise analysis, we found that, from all possible 15 cases, we found cointegration among 10 pairs of metal-commodity prices. One interesting issue is the long-run behavior of real metal-commodity prices: while three of them displayed an obvious increase in prices (Copper, Nickel, and Zinc) in 110 years – more than double from 1900 to 2010 – the other three displayed an obvious decrease over time (Aluminium, Lead, and Tin) of about 70%-90%. Common cycles are also observed at all frequencies, but their existence is more useful at the monthly frequency, where we showed how to use them to *nowcast* the current state of industrial production using copper price data. At annual frequency common-cycle evidence is more abundant than at higher frequencies – quarterly or monthly.

Finally, in a forecasting exercise, we combined forecasting models to build metal-commodity prices optimal forecasts. Regarding out-of-sample forecasts, we used a variety of models (linear and non-linear, single equation and multivariate) and a variety of co-variates to forecast the returns and prices of six metal commodities. With the forecasts of a large number of models (N large) and a large number of

Aluminum	BCAF	Weighted Average (MSE)	Average Forecast	Median	Best Model (BIC)	5 Best Models	10 Best Models
1 step-ahead	37.206	40.657	38.358	40.861	37.900	36.715	36.116
2 step-ahead	59.374	70.052	63.762	66.324	67.485	65.130	63.431
3 step-ahead	65.961	82.476	70.684	75.429	77.922	74.876	72.269
4 step-ahead	69.016	93.859	75.082	81.393	83.052	79.635	76.129
5 step-ahead	69.13	99.694	75.681	84.285	87.665	83.348	79.164

Copper	BCAF	Weighted Average (MSE)	Average Forecast	Median	Best Model (BIC)	5 Best Models	10 Best Models
1 step-ahead	17.665	17.128	13.821	12.165	12.414	15.894	15.021
2 step-ahead	44.467	46.675	35.102	34.609	31.649	32.732	32.260
3 step-ahead	75.580	83.605	59.676	62.879	55.337	55.212	55.486
4 step-ahead	106.519	121.296	100.515	100.425	81.601	82.790	84.118
5 step-ahead	128.250	150.484	128.840	128.84	100.425	102.647	105.021

Lead	BCAF	Weighted Average (MSE)	Average Forecast	Median	Best Model (BIC)	5 Best Models	10 Best Models
1 step-ahead	87.215	88.310	95.419	102.963	98.490	96.390	96.919
2 step-ahead	164.499	161.849	182.150	199.458	188.621	180.429	180.717
3 step-ahead	194.459	199.206	231.145	247.535	225.020	218.144	219.826
4 step-ahead	197.945	217.125	261.908	280.107	231.141	228.625	232.594
5 step-ahead	204.004	223.123	274.640	288.786	226.445	230.735	237.098

Nickel	BCAF	Weighted Average (MSE)	Average Forecast	Median	Best Model (BIC)	5 Best Models	10 Best Models
1 step-ahead	780.106	791.635	784.654	760.113	728.444	739.280	747.302
2 step-ahead	1727.239	1840.379	1825.805	1898.553	1475.888	1421.853	1441.278
3 step-ahead	1967.854	2148.107	2060.348	2160.290	1696.273	1587.752	1607.805
4 step-ahead	2263.931	2663.476	2317.876	2342.069	2065.623	1971.324	2000.096
5 step-ahead	2678.195	3306.581	2872.925	2839.961	2335.586	2306.738	2352.873

Tin	BCAF	Weighted Average (MSE)	Average Forecast	Median	Best Model (BIC)	5 Best Models	10 Best Models
1 step-ahead	64.457	72.979	64.111	67.698	74.204	72.228	72.420
2 step-ahead	159.763	168.255	141.708	147.906	161.558	148.089	148.736
3 step-ahead	221.401	216.056	176.578	183.942	192.313	179.855	179.121
4 step-ahead	203.960	258.841	204.587	210.945	201.923	199.095	197.25
5 step-ahead	201.165	313.839	226.033	229.935	204.341	214.522	211.547

Zinc	BCAF	Weighted Average (MSE)	Average Forecast	Median	Best Model (BIC)	5 Best Models	10 Best Models
1 step-ahead	141.114	143.832	195.797	161.538	150.502	149.399	153.070
2 step-ahead	244.172	251.799	479.166	292.524	254.273	255.094	267.004
3 step-ahead	245.351	266.658	875.957	291.309	247.958	248.965	260.206
4 step-ahead	242.105	279.292	250.617	253.082	233.096	236.309	241.427
5 step-ahead	250.851	301.163	264.693	269.355	235.248	242.127	248.668

$T = 111$, $T_1 = 35$, $T_2 = 70$ and the out-of-sample range is 41

Table 8: Forecast Root-Mean-Squarred-Error (Annual)

time periods (T large), we showed that combining forecasts proved to be the best strategy to forecast the metals prices. For metal prices, regarding possible optimal combination schemes, we found that the best performances in terms of RMSE was achieved by the bias-corrected average forecast (BCAF), the weighted average forecast (WAF) using the reciprocal of the MSE, or the average forecast (AF). These are all forecast-combination schemes, which achieve optimality by eliminating individual-model forecast error by means of the use of a weak law-of-large-numbers. These empirical results are true for most metal prices, frequencies, and horizons, although some individual models performed well on occasion.

References

Athanasopoulos, G., O. T. de Carvalho Guillén, J. V. Issler, and F. Vahid (2011). Model selection, estimation and forecasting in var models with short-run and long-run restrictions. *Journal of Econometrics 164*(1), 116 – 129.

Batchelor, R. (2007). Bias in macroeconomic forecasts. *International Journal of Forecasting 23*(2), 189 – 203.

Bates, J. M. and C. W. J. Granger (1969). The combination of forecasts. *OR 20*(4), pp. 451–468.

Beveridge, S. and C. Nelson (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'. *Journal of Monetary economics 7*(2), 151–174.

Cashin, P., H. Liang, and C. McDermott (2000). How persistent are shocks to world commodity prices? *Working Paper International Monetary Fund*.

Cashin, P., C. McDermott, and A. Scott (1999). The myth of comoving commodity prices. *Working Paper International Monetary Fund 99*.

Cashin, P., C. McDermott, and A. Scott (2002). Booms and slumps in world commodity prices. *Journal of Development Economics 69*(1), 277–296.

Chambers, M. J. and R. E. Bailey (1996). A theory of commodity price fluctuations. *Journal of Political Economy 104*(5), pp. 924–957.

Clements, M. P. and D. F. Hendry (2006). Chapter 12 forecasting with breaks. Volume 1 of *Handbook of Economic Forecasting*, pp. 605 – 657. Elsevier.

Cuddington, J. (1992). Long-run trends in 26 primary commodity prices: A disaggregated look at the prebisch-singer hypothesis. *Journal of Development Economics 39*(2), 207–227.

Cuddington, J. and C. Urzúa (1989). Trends and cycles in the net barter terms of trade: a new approach. *The Economic Journal 99*(396), 426–442.

Davies, A. and K. Lahiri (1995). A new framework for analyzing survey forecasts using three-dimensional panel data. *Journal of Econometrics 68*(1), 205 – 227.

Deaton, A. (1999). Commodity prices and growth in africa. *The Journal of Economic Perspectives 13*(3), 23–40.

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

Deaton, A. and G. Laroque (1996). Competitive storage and commodity price dynamics. *Journal of Political Economy 104*(5), pp. 896–923.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica 50*(4), pp. 987–1007.

Engle, R. F. and C. W. J. Granger (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica 55*(2), pp. 251–276.

Engle, R. F. and J. V. Issler (1995). Estimating common sectoral cycles. *Journal of Monetary Economics 35*(1), 83 – 113.

Engle, R. F. and S. Kozicki (1993). Testing for common features. *Journal of Business and Economic Statistics 11*(4), pp. 369–380.

Fuller, W. A. and G. E. Battese (1974). Estimation of linear models with crossed-error structure. *Journal of Econometrics 2*(1), 67 – 78.

Gargano, A. and A. Timmermann (2012). Predictive dynamics in commodity prices. *Working Paper*.

Grilli, E. and M. Yang (1988). Primary commodity prices, manufactured goods prices, and the terms of trade of developing countries: what the long run shows. *The World Bank Economic Review 2*(1), 1–47.

Hamilton, J. D. (2011). Nonlinearities and the macroeconomic effects of oil prices. *Macroeconomic Dynamics 15*, 364–378.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50*(4), pp. 1029–1054.

Hecq, A., F. Palm, and J. Urbain (2006). Common cyclical features analysis in var models with cointegration. *Journal of Econometrics 132*(1), 117–141.

Hendry, D. F. and M. P. Clements (2004). Pooling of forecasts. *Econometrics Journal 7*(1), 1–31.

Hong, H. and M. Yogo (2009). Digging into commodities. *Unpublished Working Paper, Princeton University and Wharton of University of Pennsylvania*.

Hong, H. and M. Yogo (2012). What does futures market interest tell us about the macroeconomy and asset prices? *Journal of Financial Economics 105*(3), 473 – 490.

IMF (2012, April). *World Economic Outlook: Growth resuming, dangers remain*, Chapter 1, pp. 1–47. International Monetary Fund.

Issler, J. V. and L. R. Lima (2009). A panel data approach to economic forecasting: The bias-corrected average forecast. *Journal of Econometrics 152*(2), 153 – 164.

Issler, J. V. and F. Vahid (2001). Common cycles and the importance of transitory shocks to macroeconomic aggregates. *Journal of Monetary Economics 47*(3), 449 – 475.

Issler, J. V. and F. Vahid (2006). The missing link: using the nber recession indicator to construct coincident and leading indices of economic activity. *Journal of Econometrics 132*(1), 281 – 303.

Jerrett, D. and J. Cuddington (2008). Broadening the statistical search for metal price super cycles to steel and related metals. *Resources Policy 33*(4), 188–195.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica 59*(6), pp. 1551–1580.

Labys, W., A. Achouch, and M. Terraza (1999). Metal prices and the business cycle. *Resources Policy 25*(4), 229–238.

Laster, D., P. Bennett, and I. Geoum (1999). Rational bias in macroeconomic forecasts. *The Quarterly Journal of Economics 114*(1), 293–318.

Lettau, M. and S. Ludvigson (2001). Consumption, aggregate wealth, and expected stock returns. *The Journal of Finance 56*(3), 815–849.

Lettau, M. and S. C. Ludvigson (2004). Understanding trend and cycle in asset values: Reevaluating the wealth effect on consumption. *The American Economic Review 94*(1), pp. 276–299.

Newey, W. and K. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica: Journal of the Econometric Society*, 703–708.

Palm, F. C. and A. Zellner (1992). To combine or not to combine? issues of combining forecasts. *Journal of Forecasting 11*(8), 687–701.

Patton, A. and A. Timmermann (2007). Testing forecast optimality under unknown loss. *Journal of the American Statistical Association 102*(480), 1172–1184.

Phillips, P. C. B. and H. R. Moon (1999). Linear regression limit theory for nonstationary panel data. *Econometrica 67*(5), pp. 1057–1111.

Pindyck, R. S. and J. J. Rotemberg (1990). The excess co-movement of commodity prices. *The Economic Journal 100*(403), pp. 1173–1189.

Stock, J. and M. Watson (2006). Forecasting with many predictors. *Handbook of economic forecasting 1*, 515–554.

Timmermann, A. (2006). Forecast combinations. In C. G. G. Elliott and A. Timmermann (Eds.), *Handbook of Economic Forecasting*, Volume 1 of *Handbook of Economic Forecasting*, pp. 135 – 196. Elsevier.

Vahid, F. and R. F. Engle (1993). Common trends and common cycles. *Journal of Applied Econometrics 8*(4), pp. 341–360.

Vahid, F. and R. F. Engle (1997). Codependent cycles. *Journal of Econometrics 80*(2), 199 – 221.

Vahid, F. and J. V. Issler (2002). The importance of common cyclical features in var analysis: a monte-carlo study. *Journal of Econometrics 109*(2), 341 – 363.

Wallace, T. D. and A. Hussain (1969). The use of error components models in combining cross section with time series data. *Econometrica 37*(1), pp. 55–72.

Welch, I. and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies 21*(4), pp. 1455–1508.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica 64*(5), pp. 1067–1084.

A Joint Short- and Long-Run Restriction for Metal-Price Dynamics using VAR Models

Before discussing the dynamic representation of metal-price data, we present useful definitions of the concepts used to measure the degree of comovement in them. These include long-run comovement (cointegration) and short-run comovement (common cycles for the growth rate of these prices). Engle and Issler (1995) and Issler and Vahid (2001) present previous applications of the techniques discussed in this section. For an in-depth theoretical discussion of these issues see Engle and Granger (1987) and Vahid and Engle (1993) respectively for cointegration and common cycles. Theoretical extensions of the standard common-cycle case can be found in Hecq et al. (2006) and Athanasopoulos et al. (2011).

Assume that y_t is a n-vector of $I(1)$ metal prices (or log metal prices), with the stationary $(MA(\infty))$ Wold representation given by:

$$\Delta y_t = C(L)\epsilon_t, \tag{11}$$

where $C(L)$ is a matrix polynomial in the lag operator, L, with $C(0) = I_n$, $\sum_{j=1}^{\infty}\|C_j\| < \infty$. The vector ϵ_t is a $n \times 1$ a multivariate white noise process. We can rewrite equation (11) as:

$$\Delta y_t = C(1)\epsilon_t + \Delta C^*(L)\epsilon_t \tag{12}$$

where $C^*(L) = C_0^* + C_1^* L + C_2^* L^2 + \cdots$, with $C_i^* = \sum_{j>i} -C_j$ for all $i \geq 0$, and, in particular, $C_0^* = I_n - C(1)$.

Integrating both sides of equation (12), given an initial condition y_0, we get:

$$\begin{aligned} y_t &= C(1)\sum_{s=0}^{t-1}\epsilon_{t-s} + C^*(L)\epsilon_t + y_0 \\ &= T_t + C_t \end{aligned} \tag{13}$$

Equation (13) is the multivariate version of the Beveridge-Nelson trend-cycle rep-

resentation (Beveridge and Nelson (1981)). Apart from an initial condition y_0, the series y_t are represented as sum of a Martingale part $T_t = C(1) \sum_{s=0}^{t-1} \epsilon_{t-s}$, which is called the "trend," and a stationary and ergodic part $C_t = C^*(L)\,\epsilon_t$, which is called the "cycle."

Definition 1. The variables in y_t are said to have common trends (or cointegrate) if there are r linearly independent vectors, $r < n$, stacked in an $r \times n$ matrix α', with the following property[10]:

$$\underset{r \times n}{\alpha'}\, C(1) = 0.$$

Definition 2. The variables in y_t are said to have common cycles if there are s linearly independent vectors, $s \leq n - r$, stacked in an $s \times n$ matrix $\tilde{\alpha}'$, with the property that:

$$\underset{s \times n}{\tilde{\alpha}'}\, C^*(L) = 0.$$

Thus, cointegration and common cycles represent restrictions on the elements of $C(1)$ and $C^*(L)$ respectively.

We now discuss what role these restrictions play on the dynamic autoregressive representation of y_t. We assume that y_t is generated by a Vector Autoregression (VAR). Note that VARs are the working horses of time-series econometric analysis. They have been applied extensively for reduced-form and structural-form estimation and forecasting, since they fit most macroeconomic and financial data fairly well:

$$y_t = \Gamma_1 y_{t-1} + \ldots + \Gamma_p y_{t-p} + \epsilon_t, \tag{14}$$

---

[10]This definition could alternatively be expressed in terms of an $n \times r$ matrix γ, such that:

$$C(1)\,\gamma = 0.$$

The Granger-Representation Theorem (Engle and Granger (1987)) shows that if the series in y_t are cointegrated, α and γ in equation (2) below satisfy:

$$
\begin{aligned}
C(1)\,\gamma &= 0, \text{and,}\\
\alpha' C(1) &= 0.
\end{aligned}
$$

where the autoregressive matrix polynomial is $\Phi(L) = I - \Gamma_1 L - \Gamma_2 L^2 - \ldots - \Gamma_p L^p$.

If elements of y_t cointegrate, then the matrix $\Phi(1) = I - \sum_{i=1}^{p} \Gamma_i$ must have less than full rank. In this case, Engle and Granger (1987) showed that the system (14) can be written as a Vector Error-Correction model (VECM) as:

$$\Delta y_t = \Gamma_1^* \Delta y_{t-1} + \ldots + \Gamma_{p-1}^* \Delta y_{t-p+1} + \gamma \alpha' y_{t-1} + \epsilon_t \tag{15}$$

where γ and α are full rank matrices of order $n \times r$, r is the rank of the cointegrating space, $-\left(I - \sum_{i=1}^{p} \Gamma_i\right) = \gamma \alpha'$, and $\Gamma_j^* = -\sum_{i=j+1}^{p} \Gamma_i$, $j = 1, \ldots, p-1$.

For our purposes, testing for cointegration will be used to verify whether metal-price data share common trends (or have long-run comovement). As is well known, metals are an important input in industrial processes, and thus it is expected that most metal commodities would have their long-run prices linked to global industrial factors. Testing for common trends among y_t will use the maximum-likelihood approach in Johansen (1991).

A key issue to assure that inference is done properly in this case is to estimate the lag length of the VAR (14) consistently, i.e., to estimate p consistently. When data have common cycles as well as common trends, Athanasopoulos et al. (2011) showed that some popular information criteria do not have an appropriate small-sample behavior, and that a combination of traditional information criteria and criteria with data-dependent penalties can estimate the lag length consistently for VARs with common trends and cycles. An alternative to way to infer p is to perform diagnostic testing to rule out the risk of underestimation of p, which leads to inconsistent estimates for the parameters in (15).

Vahid and Engle (1993) show that the dynamic representation of y_t may be further restricted if there exist white noise independent linear combinations of the series Δy_t, i.e., that the Δy_t share common cycles. To see this, recall that the cofeature vectors $\tilde{\alpha}_i'$, stacked in an $s \times n$ matrix $\tilde{\alpha}'$, eliminate all serial correlation in Δy_t, i.e. lead to

$\tilde{\alpha}' \Delta y_t = \tilde{\alpha}' \epsilon_t$. Therefore, they should restrict the elements of (15) as follows:

$$\tilde{\alpha}' \Gamma_1^* = \tilde{\alpha}' \Gamma_2^* = \ldots = \tilde{\alpha}' \Gamma_{p-1}^* = 0, \text{ and} \tag{16}$$

$$\tilde{\alpha}' \gamma = 0. \tag{17}$$

Hecq et al. (2006) have labelled the joint restrictions (16) and (17) as strong-form *serial-correlation common features* (SCCF), whereas they call only imposing restrictions (16) as weak-form SCCF. For the latter, notice that we only inherit an unpredictable linear combination of Δy_t once we *control* for the long-run deviations $\alpha' y_{t-1}$. Hence,

$$\tilde{\alpha}' [\Delta y_t - \gamma \alpha' y_{t-1}] = \tilde{\alpha}' \epsilon_t. \tag{18}$$

We now continue the discussion of common cycles in the case of strong-form serial-correlation common features, since the weak-form case can be immediately inferred from it. As is well known, cofeature vectors are identified only up to an invertible transformation[11]. Without loss of generality, consider $\tilde{\alpha}$ to have an s dimensional identity sub-matrix:

$$\tilde{\alpha} = \left[\begin{array}{c} I_s \\ \tilde{\alpha}_{(n-s) \times s}^* \end{array} \right]$$

Now, $\tilde{\alpha}' \Delta y_t = \tilde{\alpha}' \epsilon_t$ constitute s equations in a system. Completing the system by adding the unconstrained VECM equations for the remaining $n - s$ elements of Δy_t, we obtain,

$$\left[\begin{array}{cc} I_s & \tilde{\alpha}^{*\prime} \\ \mathbf{0}_{(n-s) \times s} & I_{n-s} \end{array} \right] \Delta y_t = \left[\begin{array}{c} \mathbf{0}_{s \times (np+r)} \\ \Gamma_1^{**} \ldots \Gamma_{p-1}^{**} \gamma^* \end{array} \right] \left[\begin{array}{c} \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-p+1} \\ \alpha' y_{t-1} \end{array} \right] + v_t, \tag{19}$$

where Γ_i^{**} and γ^* represent the partitions of Γ_i^* and γ respectively, corresponding to

---

[11] The same is true regarding cointegrating vectors. We are only able to identify a subspace of $\mathbb{R}^n$ of dimension r.

the bottom $n - s$ reduced form VECM equations, and $v_t = \begin{bmatrix} I_s & \tilde{\alpha}^{*\prime} \\ \underset{(n-s)\times s}{\mathbf{0}} & I_{n-s} \end{bmatrix} \epsilon_t.$

It is easy to show that (19) parsimoniously encompasses (15). Since $\begin{bmatrix} I_s & \tilde{\alpha}^{*\prime} \\ \underset{(n-s)\times s}{\mathbf{0}} & I_{n-s} \end{bmatrix}$ is invertible, it is possible to recover (15) from (19). Notice however that the latter has $s \cdot (np + r) - s \cdot (n - s)$ fewer parameters.

Assuming that y_t share common trends and cycles leading to (19), we test for common cycles using a canonical-correlation approach once we determine what is the lag length of the VAR, i.e., p. We proceed as follows:

1. Compute the sample squared canonical correlations between $\{\Delta y_t\}$ and $\{\alpha' y_{t-1}, \Delta y_{t-1}, \Delta y_{t-2}, \cdots, \Delta y_{t-p+1}\}$, labelled λ_i, $i = 1, \cdots, n$, where n is the number of variables in the system.

2. Test whether the first smallest s canonical correlations are jointly zero by computing the test statistic:

$$-T \sum_{i=1}^{s} \log (1 - \lambda_i),$$

which has a limiting χ^2 distribution with $s(np + r) - s(n - s)$ degrees of freedom under the null, where r is the number of cointegrating relationships. The maximum number of zero canonical correlations that can possibly exist is $n - r$.

3. Suppose that s zero canonical correlations were found in the previous step. Use these s contemporaneous relationships between the first differences as s pseudo-structural equations in a system of simultaneous equations. Augment them with $n - s$ equations from the VECM and estimate the system using full information maximum likelihood (FIML). The restricted VECM will be the reduced form of this pseudo-structural system.

4. The case where we have weak-form restrictions, in step 1 above, we Compute the sample squared canonical correlations between $\left\{ \left(\Delta y_t', (\alpha' y_{t-1})' \right)' \right\}$

43

and $\{\alpha' y_{t-1}, \Delta\, y_{t-1}, \Delta\, y_{t-2}, \cdots, \Delta\, y_{t-p+1}\}$. The rest of the steps remain identical.

B GMM Based Common-Cycle Tests

One possible drawback of the canonical-correlation approach is that it assumes homoskedastic data, and that may not hold for metal-price (and other macroeconomic and financial data) collected at high frequency. In this case, a GMM approach is more robust, since inference can be conducted with Heteroskedastic and Auto-correlation (HAC) robust estimates of variance-covariance matrices of parameter estimates. Regarding common cycles, the system with n-equations in (19) can be estimated by GMM. The vector of instruments comprise the series in $\{\alpha' y_{t-1}, \Delta\, y_{t-1}, \Delta\, y_{t-2}, \cdots, \Delta\, y_{t-p+1}\}$ which are collected in a vector Z_{t-1}. Indeed, GMM estimation exploits the orthogonality between all the elements in v_t and all the elements in Z_{t-1}, where v_t is defined in equation (19):

$$\mathbb{E}\left[v_t \otimes Z_{t-1}\right] = \mathbf{0}. \tag{20}$$

The test for common cycles is the over-identifying restriction test – the J test – proposed in Hansen (1982). It has an asymptotic χ^2 distribution with degrees of freedom equal to the number of over-identifying restrictions. As usual, over-identifying restriction tests verify whether or not errors are orthogonal to instruments in an instrumental-variable setup. Thus, it checks whether the exclusions of the elements of Z_{t-1} in the first s equations is appropriate. Heuristically, since the cyclical behavior (serial correlation behavior) of the data $\Delta\, y_t$ is captured by Z_{t-1}, this test verifies whether the linear combinations in these s equations have no serial correlation, i.e., are unpredictable. Therefore, it is a test of common serial correlation or *common cycles*. If two series have a *common cycle*, their impulse response functions are colinear, making their response to shocks proportional and therefore similar. Here, contrary to the canonical correlation approach above, we can deal with heteroskedasticity of unknown form by employing the HAC robust estimates for the variance-covariance matrix of sample means counterparts of (19) $\left(\widehat{S_T}\right)$ using the Newey and West (1987)

procedure. The parameters estimated by GMM, stacked in a vector θ, comprise all parameters in $\tilde{\alpha}^{*\prime}$ and all parameters in the matrices $\Gamma_1^{**}, \Gamma_2^{**}, \ldots \Gamma_{p-1}^{**}$, and in γ^*.

If we want to test for weak-form SCCF, the only additional twist if that now v_t takes the form:

$$
v_t = \left[\begin{array}{cc} I_s & \tilde{\alpha}^{*\prime} \\ \mathbf{0} \\ {}_{(n-s)\times s} & I_{n-s} \end{array} \right] \Delta\, y_t - \left[\begin{array}{cccc} \mathbf{0} & \cdots & \mathbf{0} & \gamma_1 \\ \Gamma_1^{**} & \cdots & \Gamma_{p-1}^{**} & \gamma_2 \end{array} \right] \left[\begin{array}{c} \Delta\, y_{t-1} \\ \vdots \\ \Delta\, y_{t-p+1} \\ \alpha' y_{t-1} \end{array} \right],
$$

where $\gamma = \left[\begin{array}{c} \gamma_1 \\ \gamma_2 \end{array} \right]$.

C   Optimal Forecast Combinations

In this section, we list the set of assumptions needed to obtain optimal forecast combinations and discuss some details on how to construct these optimal forecast under mean-squared error (MSE) risk function. For an in-depth theoretical discussion of these issues see Issler and Lima (2009) and the references therein.

We are interested in forecasting y_t, stationary and ergodic, using information up to h periods prior to t, then, under a MSE risk function, the optimal forecast is the conditional expectation using information available up to $t - h$: $\mathbb{E}_{t-h}(y_t)$. We label forecasts of y_t, computed using conditioning sets lagged h periods, by $f_{i,t}^h$, $i = 1, 2, \ldots, N$. Therefore, $f_{i,t}^h$ are h-step-ahead forecasts and N is the number of models estimated to forecast $f_{i,t}^h$.

Hendry and Clements (2004) argue that the fact that the simple forecast average $\frac{1}{N} \sum_{i=1}^{N} f_{i,t}^h$ usually outperforms individual forecasts $f_{i,t}^h$ shows our inability to approximate $\mathbb{E}_{t-h}(y_t)$ reasonably well with individual models. However, since $\mathbb{E}_{t-h}(y_t)$ is optimal, this is exactly what these individual models should be doing.

With this motivation, our setup writes the $f_{i,t}^h$'s as approximations to the optimal forecast as follows:

$$
f_{i,t}^h = \mathbb{E}_{t-h}(y_t) + k_i^h + \varepsilon_{i,t}^h, \tag{21}
$$

where k_i^h is the individual model time-invariant bias for h-step-ahead prediction and $\varepsilon_{i,t}^h$ is the individual model error term in approximating $\mathbb{E}_{t-h}(y_t)$, where $\mathbb{E}(\varepsilon_{i,t}^h) = 0$ for all i, t, and h. Here, the optimal forecast is a *common feature* of all individual forecasts and k_i^h and $\varepsilon_{i,t}^h$ arise because of forecast misspecification.

We can always decompose the series y_t into $\mathbb{E}_{t-h}(y_t)$ and an unforecastable component ζ_t, such that $\mathbb{E}_{t-h}(\zeta_t) = 0$ in:

$$y_t = \mathbb{E}_{t-h}(y_t) + \zeta_t. \tag{22}$$

Combining (7) and (8) yields,

$$
\begin{aligned}
f_{i,t}^h &= y_t - \zeta_t + k_i^h + \varepsilon_{i,t}^h, \qquad \text{or,} \\
f_{i,t}^h &= y_t - \eta_t^h + k_i^h + \varepsilon_{i,t}^h, \text{ where, } \eta_t^h = -\zeta_t.
\end{aligned} \tag{23}
$$

This yields the well known two-way decomposition, or error-component decomposition, of the forecast error $f_{i,t}^h - y_t$:

$$f_{i,t}^h = y_t + \mu_{i,t}^h, \qquad i = 1, 2, ..., N \text{ and } t > T_1 \tag{24}$$
$$\mu_{i,t}^h = k_i^h + \eta_t^h + \varepsilon_{i,t}^h.$$

By construction, the framework in (9) specifies explicit sources of forecast errors that are found in both y_t and $f_{i,t}^h$; see also the discussion in Palm and Zellner (1992) and Davies and Lahiri (1995). The term k_i^h is the time-invariant forecast bias of model i. It captures the long-run effect of forecast-bias of model i. Its source is $f_{i,t}^h$. The term η_t^h arises because forecasters do not have future information on y_t between $t - h + 1$ and t. Hence, the source of η_t^h is y_t , and it is an additive aggregate zero-mean shock affecting equally all forecasts. The term $\varepsilon_{i,t}^h$ captures all the remaining errors affecting forecasts, such as those of idiosyncratic nature and others that affect some but not all the forecasts (a group effect). Its source is $f_{i,t}^h$.

From the perspective of combining forecasts, the components k_i^h, $\varepsilon_{i,t}^h$ and η_t^h play very different roles. If we regard the problem of forecast combination as one aimed at

diversifying risk, i.e., a finance approach, then, on the one hand, the risk associated with $\varepsilon_{i,t}^h$ can be diversified, while that associated with η_t^h cannot. On the other hand, in principle, diversifying the risk associated with k_i^h can only be achieved if a bias-correction term is introduced in the forecast combination, which reinforces its usefulness.

We now list the assumptions needed to construct optimal forecast combinations under this framework.

Assumption 1 We assume that k_i, $\varepsilon_{i,t}$ and η_t are independent of each other for all i and t.

Independence is an algebraically convenient assumption used throughout the literature on two-way decompositions; see Wallace and Hussain (1969) and Fuller and Battese (1974) for example. At the cost of unnecessary complexity, it could be relaxed to use orthogonal components, something we avoid here.

Assumption 2 k_i is an identically distributed random variable in the cross-sectional dimension, but not necessarily independent, i.e.,

$$k_i^h \sim \text{i.d.}(B, \sigma_{k^h}^2), \tag{25}$$

where B^h and $\sigma_{k^h}^2$ are respectively the mean and variance of k_i^h. In the time-series dimension, k_i^h has no variation, therefore, it is a fixed parameter.

The idea of dependence is consistent with the fact that forecasters learn from each other by meeting, discussing, debating, etc. Through their ongoing interactions, they maintain a current collective understanding of where their target variable is most likely heading to, and of its upside and downside risks. Given the assumption of identical distribution for k_i^h, B^h represents the market (or collective) bias. Since we focus on combining forecasts, a pure idiosyncratic bias does not matter but a collective bias does. In principle, we could allow for heterogeneity in the distribution of k_i – means and variances to differ across i. However, that will be a problem in testing the hypothesis that forecast combinations are biased.

It is desirable to discuss the nature of the term k_i^h, which is related to the question of why we cannot focus solely on unbiased forecasts, for which $k_i^h = 0$. The role of k_i^h is to capture the long-run effect, in the time dimension, of the bias of econometric models of y_t. The model-based forecasts bias results from model misspecification. Here, it is important to distinguish between in-sample and out-of-sample model fitting. The fact that, in sample, a model approximates well the data-generating process (DGP) of y_t does not guarantee that it will in out-of-sample forecasting; see the discussion in Clements and Hendry (2006) and in Hendry and Clements (2004). Notice that bias correction is a form of intercept correction. We now discuss survey-based forecasts. In this case, a relevant question to ask is: why would forecasters introduce bias under a MSE risk function? Laster et al. (1999), Patton and Timmermann (2007) , and Batchelor (2007) list different arguments consistent with forecasters having a non-quadratic loss function. Following their discussion, we assume that all forecasters employ a combination of quadratic loss and a secondary loss function. Bias is simply a consequence of this secondary loss function and of the intensity in which the forecaster cares for it. The first example is that of a bank selling an investment fund. In this case, the bank's forecast of the fund return may be upward-biased simply because it may use this forecast as a marketing strategy to attract new clients for that fund. Although the bank is penalized by deviating from $\mathbb{E}_{t-h}(y_t)$, it also cares for selling the shares of its fund. The second example introduces bias when there is a market for *pessimism* or *optimism* in forecasting. Forecasters want to be labeled as optimists or pessimists in a "branding" strategy to be experts on "worst-" or on "best-case scenarios," respectively. Batchelor lists governments as examples of experts on the latter.

Assumption 3 The aggregate shock η_t^h is a stationary and ergodic MA process of order at most $h-1$, with zero mean and variance $\sigma_{\eta^h}^2 < \infty$.

Since h is a bounded constant in our setup, η_t^h is the result of a cumulation of shocks to y_t that occurred between $t-h+1$ and t. Being an $MA(\cdot)$ is a consequence of the wold representation for y_t. If y_t is already an $MA(\cdot)$ process, of order smaller than $h-1$, then, its order will be the same of that of η_t^h. Otherwise, the order is $h-1$.

In any case, it must be stressed that η_t^h is unpredictable, i.e., that $\mathbb{E}_{t-h}\left(\eta_t^h\right) = 0$. This a consequence of (8) and of the law of iterated expectations, simply showing that, from the perspective of the forecast horizon h, unless the forecaster has superior information, the aggregate shock η_t^h cannot be predicted.

Assumption 4: Let $\varepsilon_t^h = \left(\varepsilon_{1,t}^h, \varepsilon_{2,t}^h, \dots \varepsilon_{N,t}^h\right)'$ be a $N \times 1$ vector stacking the errors $\varepsilon_{i,t}^h$ associated with all possible forecasts, where $\mathbb{E}\left(\varepsilon_{i,t}^h\right) = 0$ for all i and t. Then, the vector process $\left\{\varepsilon_t^h\right\}$ is assumed to be covariance-stationary and ergodic for the first and second moments, uniformly on N. Further, defining as $\xi_{i,t}^h = \varepsilon_{i,t}^h - \mathbb{E}_{t-1}\left(\varepsilon_{i,t}^h\right)$, the innovation of $\varepsilon_{i,t}^h$, we assume that

$$\lim_{N\to\infty} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left|\mathbb{E}\left(\xi_{i,t}^h \xi_{j,t}^h\right)\right| = 0. \tag{26}$$

Because the forecasts are computed h-steps ahead, forecast errors $\varepsilon_{i,t}$ can be serially correlated. Assuming that $\varepsilon_{i,t}^h$ is weakly stationary is a way of controlling its time-series dependence. It does not rule out errors displaying conditional heteroskedasticity, since the latter can coexist with the assumption of weak stationarity; see Engle (1982).

The techniques discussed in this section are appropriate for forecasting a weakly stationary and ergodic univariate process $\{y_t\}$ using a large number of forecasts that will be combined to yield an optimal forecast in the mean-squared error (MSE) sense. These forecasts are the result of several econometric models that need to be estimated prior to forecasting[12].

We consider 3 consecutive distinct time sub-periods, where time is indexed by $t = 1, 2, \dots, T_1, \dots, T_2, \dots, T$. The first sub-period E is labeled the "estimation sample", where models are usually fitted to forecast y_t subsequently. The number of observations in it is $E = T_1 = \kappa_1 \cdot T$, comprising $(t = 1, 2, \dots, T_1)$. For the other two, we follow the standard notation in West (1996). The sub-period R (for regression) is

---

[12]In this setting, we can also imagine that some (or all) responses use no formal econometric model at all, e.g., just the result of an opinion poll on the variable in question using a large number of individual responses.

49

labeled the post-model-estimation or "training sample", where realizations of y_t are usually confronted with forecasts produced in the estimation sample, and weights and bias-correction terms are estimated. It has $R = T_2 - T_1 = \kappa_2 \cdot T$ observations in it, comprising $(t = T_1 + 1, \ldots, T_2)$. The final sub-period is P (for prediction), where genuine out- of-sample forecast is entertained. It has $P = T - T_2 = \kappa_3 \cdot T$ observations in it, comprising $(t = T_2 + 1, \ldots, T)$. Notice that $0 < \kappa_1, \kappa_2, \kappa_3 < 1$, $\kappa_1 + \kappa_2 + \kappa_3 = 1$, and that the number of observations in these three sub-periods keep a fixed proportion with T - respectively, κ_1, κ_2 and κ_3 - being all $O(T)$. This is an important ingredient in our asymptotic results for $T \to \infty$.

Issler and Lima (2009) propose a non-parametric estimator of k_i^h, which exploits the fact that k_i^h represents the fixed effect of a panel of forecasts:

$$(f_{i,t}^h - y_t) = k_i^h + \eta_t^h + \varepsilon_{i,t}^h, \qquad i = 1, 2, ..., N \qquad (27)$$
$$t = T_1 + 1, ..., T_2$$

It does not depend on any distributional assumption on $k_i^h \sim i.d.(B^h, \sigma_{k^h}^2)$ and it does not depend on any knowledge of the models used to compute the forecasts $f_{i,t}^h - y_t$.

Issler and Lima propose the following consistent estimators for the components k_i^h, B^h, η_t^h, and $\varepsilon_{i,t}^h$:

$$\widehat{k_i^h} = \frac{1}{R} \sum_{t=T_1+2}^{T_2} f_{i,t}^h - \frac{1}{R} \sum_{t=T_1+2}^{T_2} y_t \qquad (28)$$

$$\widehat{B^h} = \frac{1}{N} \sum_{i=1}^{N} \widehat{k_i^h} \qquad (29)$$

$$\widehat{\eta^h}_t = \frac{1}{N} \sum_{i=1}^{N} f_{i,t}^h - \widehat{B^h} - y_t \qquad (30)$$

$$\widehat{\varepsilon}_{i,t}^h = f_{i,t}^h - y_t - \widehat{k_i^h} - \widehat{\eta^h}_t \qquad (31)$$

Now we state the two most important results from Issler and Lima (2009), which

asserts that the (feasible) bias-corrected average forecast (BCAF) is an optimal forecasting device.

Theorem 1. *If Assumptions 1-4 hold, the feasible bias-corrected average forecast* $\sum_{i=1}^{N} \omega_i f_{i,t}^h - \widehat{B^h}$ *obeys*

$$\operatorname*{plim}_{(T,N\to\infty)_{seq}} \left(\frac{1}{N} \sum_{i=1}^{N} f_{i,t}^h - \widehat{B^h} \right) = y_t + \eta_t^h = \mathbb{E}(y_t)$$

and has a mean-squared error as follows:

$$\mathbb{E}\left[\operatorname*{plim}_{(T,N\to\infty)_{seq}} \left(\frac{1}{N} \sum_{i=1}^{N} f_{i,t}^h - \widehat{B^h} \right) - y_t \right]^2 = \sigma_{\eta^h}^2$$

Therefore it is an optimal forecasting device.

Indeed, there are infinite ways of combining forecasts. The next corollary presents alternative weighting schemes.

Corollary 1. *Consider the sequence of deterministic weights* $\{\omega_i\}_{i=1}^{N}$, *such that* $\omega_i \neq 0$, $\omega_i = O(N^{-1})$ *uniformly, with* $\sum_{i=1}^{N} \omega_i = 1$ *and* $\lim_{N\to\infty} \sum_{i=1}^{N} \omega_i = 1$. *Then, under Assumptions 1-4, an optimal forecasting device is:*

$$\mathbb{E}\left[\operatorname*{plim}_{(T,N\to\infty)_{seq}} \left(\sum_{i=1}^{N} \omega_i f_{i,t}^h - \sum_{i=1}^{N} \omega_i \widehat{k_i^h} \right) \right]^2 = \sigma_{\eta^h}^2$$

Optimal population weights, constructed from the variance-covariance structure of models with stationary data, will obey the structure in Corollary 1 and cannot perform better than $\frac{1}{N}$ coupled with bias correction. Therefore, there is no forecast-combination puzzle in the context of populational weights.

Theorem 1 shows that the feasible BCAF is asymptotically equivalent to the optimal weighted forecast. Its advantage is that it employs equal weights. As $N \to \infty$, the number of estimated parameters is kept at unity: $\widehat{B^h}$. This is a very attractive feature of the BCAF compared to devices that combine forecasts using estimated

weights. Our answer to the curse of dimensionality is parsimony, implied by estimating only one parameter- $\widehat{B^h}$. One additional advantage is that we need not limit the asymptotic path of N and T, which is the case of forecasts based on estimated weights.

Finally, there is one interesting case in which we can dispense with estimation in combining forecasts: when the mean bias is zero, i.e., $B^h = 0$, there is no need to estimate B^h and the BCAF is simply equal to $\frac{1}{N} \sum_{i=1}^{N} f_{i,t}^h$, the sample average of all forecasts. This is the ultimate level of parsimony. To be able to test the null that $B^h = 0$, Issler and Lima developed a robust t-ratio test that takes into account the cross-sectional dependence in k_i^h.

C.1 Forecast Combinations for Nested Models

It is important to discuss whether the techniques above are applicable to the situation where some (or all) of the models we combine are nested. The potential problem is that the innovations from nested models can exhibit high cross-sectional dependence. In what follows, we introduce nested models into our framework in the following way. Consider a continuous set of models and split the total number of models N into M classes (or blocks), each of them containing m nested models, so that $N = mM$. In the index of forecasts, $i = 1, \ldots, N$, we group nested models contiguously. Hence, models within each class are nested but models across classes are non-nested. We make the number of classes and the number of models within each class to be functions of N, respectively as follows: $M = N^{1-d}$ and $m = N^d$, where $0 \leq d \leq 1$. Notice that this setup considers all the relevant cases: (i) $d = 0$ corresponds to the case in which all models are non-nested; $d = 1$ corresponds to the case in which all models are nested and; (iii) the intermediate case $0 < d < 1$ gives rise to N^{1-d} blocks of nested models, all with size N^d.

For each block of nested models, Assumption 4 may not hold because the innovations from that block can exhibit high cross-sectional dependence. Regarding the interaction across blocks of nested models, it is natural to impose that the correlation structure of innovations across classes is such that Assumption 4 holds, since we

should expect that the cross-sectional dependence of forecast errors across classes is weak.

Here, keeping some nested models poses no problem at all, since the mixture of models will still deliver the optimal forecast. From a practical point of view, the choice of $0 \leq d < 1$ seems to be superior. Here, we are back to the main theorem in finance about risk diversification: do not put all your eggs in the same basket, choosing a large enough number of diversified (classes of) models.