Universität St.Gallen

Forecasting Copper Prices with Dynamic Averaging and Selection Models

Daniel Buncic, Carlo Moretto

September 2014 Discussion Paper no. 2014-30

School of Economics and Political Science,
Department of Economics

University of St.Gallen

Editor: Martina Flockerzi
University of St.Gallen
School of Economics and Political Science
Department of Economics
Bodanstrasse 8
CH-9000 St. Gallen
Phone +41 71 224 23 25
Fax +41 71 224 31 35
Email seps@unisg.ch

Publisher: School of Economics and Political Science
Department of Economics
University of St.Gallen
Bodanstrasse 8
CH-9000 St. Gallen
Phone +41 71 224 23 25
Fax +41 71 224 31 35

Electronic Publication: http://www.seps.unisg.ch

Forecasting Copper Prices with Dynamic Averaging and Selection Models

Daniel Buncic, Carlo Moretto

Author's address: Daniel Buncic
Institute of Mathematics and Statistics
Bodanstrasse 6
CH-9000 St. Gallen
Phone +41 71 2604
Fax +41 71 2894
Email daniel.buncic@unisg.ch
Website www.danielbuncic.com

Abstract

We use data from the London Metal Exchange (LME) to forecast monthly copper returns using the recently proposed dynamic model averaging and selection (DMA/DMS) framework, which incorporates time varying parameters as well as model averaging and selection into one unifying framework. Using a total of 18 predictor variables that include traditional fundamental indicators such as excess demand, inventories and the convenience yield, as well as indicators related to global risk appetite, momentum, the term spread, and various other financial series, we show that there exists a considerable predictive component in copper returns. Covering an out-of-sample period from May 2002 to June 2014 and employing standard statistical evaluation criteria we show that the out-of-sample R^2 (relative to a random walk benchmark) can be as high as 18.5 percent for the DMA framework. Time series plots of the cumulative mean squared forecast errors and time varying coefficients show further that firstly, a large part of the improvement in the forecasts is realised during the peak of the financial crisis period at the end of 2008, and secondly that the importance of the most relevant predictor variables has changed substantially over the out-of-sample period. The coefficients of the SP500, the VIX, the yield spread, the TED spread, industrial production and the convenience yield predictors are most heavily affected, with the TED spread and yield spread coefficients even changing signs over this period. Our predictability results remain valid for forecast horizons up to 6 months ahead, but are weaker and smaller than at the one month horizon.

Keywords

Copper forecasting, time varying parameter model, state-space modelling, dynamic model and selection models.

JEL Classification

C11, C52, C53, G17.

Table of Contents

1. Introduction

The seminal paper by Chen *et al.* (2010) has shown that the exchange rates from a number of commodity exporting countries — known as commodity currencies — "*have surprisingly robust forecasting power over global commodity prices*" (Chen *et al.*, 2010, page 1145). Using standard regression models and commodity currencies as predictor variables, Chen *et al.* (2010) show further that their results are robust to various control settings, and that they hold in-sample as well as out-of-sample. These encouraging results have lead to a general re-emergence of interest in forecasting commodity prices, in particular with the use of more advanced econometric models and richer predictor sets.

In this study, we implement the recently developed Dynamic Model Averaging and Selection (henceforth DMA/DMS) framework of Raftery *et al.* (2010) and Koop and Korobilis (2012) to forecast monthly London Metal Exchange (LME) copper returns using a large set of 18 predictor variables. Forecasting copper returns with the DMA/DMS framework is particularly appealing in the given context, as it combines time varying parameters and model averaging/selection into one unified framework. These two features substantially increase the flexibility of the prediction environment, as it reduces the commonly encountered problem of over-fitting when too many predictor variables are included in the forecasting model. Since the empirical evidence of parameter instability reported in Chen *et al.* (2010), it is well known that "*addressing parameter instability*" plays a crucial role in uncovering evidence of commodity price predictability. In a more general context, West and Harrison (1997) have argued that incorporating time varying parameters into a model can act as an approximation to neglected non-linearities and also to omitted variables. Such an approximating feature can thus prove advantageous, particularly when the model is used for forecasting. The benefits of forecasting from a weighted average of linear models is widely known since the seminal work of Bates and Granger (1969).

There exist a number of studies that include copper as one of the target commodities of interest in recent forecast evaluations, employing a variety of different modelling approaches (see Chen *et al.* (2010), Groen and Pesenti (2011), Chen and Tsay (2011), Issler *et al.* (2014), Gargano and Timmermann (2014) and many others). For instance, Groen and Pesenti (2011) apply a factor augmented regression approach where the factors are extracted from a large data set of well over 100 regressors using principal component analysis. Gargano and Timmermann (2014) utilize their complete subset regression methodology on a data set that includes the well known Goyal and Welch (2008) predictor variables extensively studied in the equity premium forecasting literature, augmented with a set of variables that capture the state of the U.S. economy, exchange rates, as well as derivative prices. What is common to all of these studies is that a rather general and broad predictor set is used for the various commodity series of interest. There does not seem to exist any recent study that applies a dynamic and flexible forecasting framework where only predictor variables that are relevant for copper are conditioned upon.

The objective of this study is to combine the flexibility of the DMA/DMS modelling framework together with a large set of purposefully selected predictor variables into a forecasting model for copper. More specifically, we construct 18 predictor variables that account for changes in copper fundamentals such as excess demand, inventories, and the convenience yield of holding copper in storage. Due to the recent financialization of commodity prices in general, we also include predictor variables that are designed to capture the global appetite for risk as approximated by the TED spread and the VIX, the equity prices of four large resource based firms which are meant to capture the forward looking (and pricing) behaviour of equity markets, and the Chilean peso and the Australian dollar as the main commodity currencies. In addition to U.S. industrial production as a standard proxy for monthly economic activity, we further add various forward looking measures of economic activity. These are the U.S. term spread, the Baltic Dry shipping index and also the broad S&P500 equity price index. Two other commodities that we add to the above predictor set are gold and oil prices.

Covering an out-of-sample forecast evaluation period from May 2002 to June 2014 we show that the DMA/DMS modelling framework significantly outperforms the random walk benchmark for forecast horizons up to 6 months ahead. We employ two standard statistical evaluation criteria to verify our result. These are the Campbell and Thompson (2008) out-of-sample R^2 and the Clark and West (2007) Mean Squared Forecast Error (MSFE) adjusted $t-$statistic. At the one-step-ahead horizon, the out-of-sample R^2 reaches values as high as 18.5 and 13.7 percent for forecasts from DMA and DMS, respectively, with corresponding MSFE adjusted $p-$values of 0.002 and 0.013. It is interesting to point out here that even a simple expanding window OLS regression using the full predictor set outperforms the random walk benchmark at the one-step-ahead horizon, yielding an out-of-sample R^2 of nearly 10 percent with corresponding MSFE adjusted $p-$value of 0.025. What these results highlight is that the gains in predictability over the random walk model are not solely due to the use of the DMA/DMS framework *per se*, but arise also from the use of predictor variables that were constructed specifically with copper forecasting in mind.

One particularly interesting finding of our paper is that there is a substantial boost in forecast performance at the beginning of the 2008 financial crisis, that is, from September 2008 until the beginning of 2009. After the Lehman Brothers collapse in September 2008, economic and financial uncertainty reached a peak, with many financial assets being heavily affected by this. Despite this uncertainty surrounding the Lehman collapse, we show that there is a large jump in the cumulative MSFEs of the DMA (and also in the simple expanding window OLS regression) model, relative to the random walk benchmark. This means that the biggest improvement in forecast performance (over the random walk benchmark) is realised during a time of extremely high uncertainty, thereby indicating a breakdown of the unpredictability view of asset prices over this period. So far there do not seem to exist any accounts of such a finding in the commodity forecasting literature.

It should be added here that, although the main focus of the paper is an out-of-sample forecast evaluation of the DMA/DMS framework, a positive by-product of the framework is the information that it provides with regards to the time-varying importance of the individual predictor variables that are used to construct the forecasts. This information is contained not only in the filtered estimates of the time varying parameters of the model, but also in the posterior prediction probabilities, which give an indication of the most frequently included variables in the best forecasting models. From time series plots of the posterior inclusion probabilities and the time varying coefficients it is evident that substantial changes occurred in the importance of the 18 predictor variables that we condition upon, particulary over the September 2008 to beginning of 2009 period following the Lehman Brothers collapse. For instance, the coefficients on the SP500, the VIX, the yield spread, the TED spread, industrial production and the convenience yield predictors show sharp upward or downward jumps, with the TED spread and yield spread coefficients even changing signs over this period. The posterior inclusion probability for the convenience yield drops from a value of about 70% to less then 30% and then back up to 70% over a period of 3 months. These abrupt changes highlight the extreme dynamics in the predictive environment around this time. In line with the results reported in Chen *et al.* (2010), we also conclude from our study that *"addressing parameter instability"* plays a crucial role for copper forecasting.

The rest of the paper is structured as follows. Section 2 reviews some of the basic theoretical background to help understand the construction of the relevant predictor variables for copper and provides also a selected recent literature review. In Section 3, we outline the econometric dynamic model averaging and selection framework that we employ in this study. The data set is described in detail in Section 4. The forecast evaluation of the competing models is outlined and discussed in Section 5. In Section 6, we summarise and conclude the study.

2. Theoretical background and existing literature

This section provides a brief overview of the *'fundamental'* factors that influence the dynamics of copper prices. We also discuss here the broad effect of *'financialization'* on copper prices, as well as the specific influence of exchange rates and equity prices of resource based stocks. Lastly, we provide a selected literature review of recent developments in modelling and forecasting commodity prices in general.

2.1. Fundamentals

We discuss two sets of fundamentals that determine the fundamental value of copper prices. These are i) supply and demand dynamics and ii) inventories. In the absence of speculative behaviour, the price of copper should — at least in equilibrium — be determined by these fundamentals.

It should also be noted here that copper can be cheaply and effectively recycled, so as to maintain its original quality. Recycling accounts for around 10% of total global copper supply, and high quality scrap can be considered a nearly perfect substitute for high grade copper. The share of recycled copper is thus expected to play an important role in the determination of copper prices. Nevertheless, the main issue with using information about recycled copper in an econometric model for forecasting purposes is the availability (and reliability) of scrap data. Current available data on recycled copper has an annual frequency, which is much too low to make it feasible to be included in an econometric model. Also, one would expect the price of scrap copper to be closely correlated with that of (high grade) copper, due to their near perfect substitutability. Therefore, from an information content perspective, using the lagged scrap copper price as a predictor for copper is unlikely to add any information that is not already contained in the lagged copper price itself.[1] We thus do not include recycled copper as a viable fundamental variable in the set of predictors that we use.

[1] The results found in Aruga and Managi (2011) confirm the view that the price of high grade copper and scrap copper are contemporaneously determined and contain the same *'information'* from a forecasting perspective.

2.1.1. Supply and demand dynamics

It takes generally several years to build new copper production capacities. Due to this, copper supply is typically very inelastic. Moreover, since it also takes time to switch to other substitute commodities in the production process, copper demand tends to be rather inelastic as well (Fisher *et al.*, 1972). Demand shocks to copper frequently translate into volatility in prices, while supply shocks lead to price cycles (Labys *et al.*, 2000). Evidence of a demand induced price cycle has been found, among others, by Cuddington and Jerrett (2008), who believe that Chinese industrialization may have ignited a super-cycle in the metals market in general, starting in the early 21^{st} century. Issler *et al.* (2014) also find demand induced cyclical features between U.S. industrial production and a number of different metals.

2.1.2. Inventories and the convenience yield

Since copper can be stored without any loss of quality, inventories can interact with demand and supply forces to affect equilibrium prices. Storage theory provides a simple theoretical framework that allows us to model commodity prices by taking inventories into account (Brennan, 1958).[2] When inventories are low, demand shocks can cause big changes in spot prices, while forward prices exhibit lower variation, as demand and supply are expected to adjust (Fama and French, 1988). When inventories are high, the shock to the spot price is dampened by inventories and thus incorporated as a permanent one into the forward price. This means that low inventories lead to higher forward-spot variability, whereas higher inventories should be accompanied by similar variabilities in forward and spot prices. Ng and Pirrong (1994) find that storage theory holds fairly robustly for metal prices.

Another key variable to be included in the set of predictors is the *'convenience yield'*.[3] Fama and French (1987) show that the futures–spot price spread (known as the *'basis'*), should be

[2] Storage theory suggests that equilibrium prices are obtained when the return from buying spot and selling forward is enough to offset the foregone risk-free return and the marginal storage cost net of the convenience yield.

[3] The convenience yield arises also from storage theory.

equalized by the sum of the foregone investment in the risk-free interest rate and the storage costs, net of the advantage accruing from physically holding the commodity. The convenience yield thus cancels arbitrage opportunities that could arise from holding the commodity today, forgoing an equal investment in the risk-free rate, and storing and selling the commodity τ periods in the future. One difficulty with using the convenience yield as a predictor variable is that it is not directly observed and needs to be constructed. For this, one needs to make an assumption about storage costs. We will define explicitly in Section 4 how the convenience yield is constructed.

2.2. Financialization

While the above discussed fundamentals play a central role in the determination of equilibrium prices for copper, they are generally not enough to explain the high level of volatility that is empirically observed in copper prices. Deaton and Laroque (1992, 1996) and Gilbert (2010), have recently documented that speculators and financial arbitrageurs have amplified the variability and the persistence observed in a wide variety of commodity prices. Similarly, Cochran *et al.* (2012) have found that the latest upsurge in commodity financialization is due to changes in legislation. In particular, the *'Commodity Futures Modernization Act of 2000'* has attracted new groups of investors into commodities, as commodities are seen to offer returns similar to equities, are positively correlated with inflation and provide also a hedge against fluctuations in the value of the U.S. Dollar (UNCTAD, 2011). Copper in particular has more recently been found in investment portfolios as an alternative to precious metals such as Gold and Silver, which play the role of safe haven assets during times of uncertainty. Interest to hold also other metals such as platinum and copper in investment portfolios has risen due to the view that gold and silver may be reaching an asset bubble.

Index investment in commodity markets has grown rapidly from 62 billion U.S. Dollars in June 2003 to 492 billion U.S. Dollars in 2008 (Hong and Yogo, 2012, page 482). Moreover, the number of futures and options contracts on commodities markets has increased exponentially since $2004 - 2005$, with notional amounts of commodity over-the-counter (OTC) derivatives increasing sevenfold between 2004 and 2008. Commodity Exchange Traded Products (ETPs) have also been introduced in a large scale. Since ETPs are designed to replicate indices of commodity prices, they are often backed by physical commodities. In this regard, it is clear that an expansion of physically backed ETPs may well cause a tightening of physical commodity supply, because part of the physical commodities available in warehouses of commodity exchanges will be held as collateral (UNCTAD, 2011).[4]

Despite the important effect of financialization on commodity prices, Dwyer *et al.* (2011) note that the overall contribution of financialization to commodity price volatility should not be given an excessive role, as financial volumes are still rather small when compared to world production and export volumes. Dwyer *et al.* (2011) show also that, while the yearly turnover in financial markets is several times higher than yearly production, open interest volume remains much smaller, suggesting further that the recently observed increase in commodity prices and price volatility is no different to what has been observed during past economic crisis periods. Fundamentals, therefore, are still believed to play a fairly dominant role in the determination of equilibrium prices for commodities. Official inventory volumes are small relative to total supply and demand volumes and are thus judged to have an insignificant impact on copper prices.[5]

Nevertheless, irrespective of how we assess the impact of financialization on commodity prices, it should be highlighted that inventories have been accumulated as collateral in financing deals, especially by Chinese banks.[6] There exist media reports that a whole shadow system

[4]See also UNCTAD (2011) on the discussion of how the creation of physically backed ETPs may ignite a speculative bubble. For copper prices, Gilbert (2010) provides evidence of price bubbles occurring approximately between February and April 2004, April and June−July 2006, and December 2008, at least partially induced by index investors. Yet, the measure of Net Index Position that he uses to evaluate the impact of index investors on several commodity prices is built from data regarding index investment in agricultural commodities only.

[5]Our calculations suggest that average monthly LME copper inventories are less than about 15% of the monthly world production of copper.

[6]It is estimated that 90% of warehouse stocks of aluminium are locked in financing deals. The same fig-

of warehouses has been developed alongside the official ones in order to respond to such a high demand in inventories.[7] The lockup of huge amounts of copper and aluminium in financing deals has contributed to generating queues to load off metals. Load-off queues have increased to several months, forcing the LME to introduce new rules to restrict queues to a maximum of 100 days with a mandatory load-off of at least 500 tonnes per day for each metal upon request, irrespective of official queue priorities.

What impact sudden liquidations of inventories have on metal prices is yet to be seen. Currently, data availability of liquidation is limited and questionable, making it difficult to incorporate this information into an econometric model. Nevertheless, it is important to keep this issue in mind, as it suggests that official inventories may serve only as a limited signal for commodity prices.

2.3. Exchange rates and stock prices of resource based firms

Exchange rates of countries that are heavily dependent on exports of one or more commodities are likely to be valid predictors of the prices of those commodities. The reason for this is that these so called *'commodity currencies'* incorporate expectations about future prices of the country's respective commodities. Within a Balassa-Samuelson framework, an increase in the price of an exported commodity increases export revenues from that commodity. Resources are drawn from the non-traded sector towards the traded one, decreasing the supply of non-traded goods. The associated increase in non-traded goods prices leads to a real exchange rate appreciation (see Obstfeld and Rogoff, 1996 for a textbook type treatment of the Balassa-Samuelson effect and Chan *et al.*, 2011 and Aizenman *et al.*, 2012 for its use in a recent commodity forecasting applications). More recently, Makin (2013) has documented empirical evidence that floating exchange rates can act as shock absorbers in the event of export commodity price shocks,

ure is 60% for zinc and only 30% for copper (see, respectively, http://www.reuters.com/article/2013/02/12/warehousing-banks-idUSL5N0BCE4D20130212 and http://www.bloomberg.com/news/print/2013-10-17/lme-aluminum-stocks-jump-most-in-a-year-as-financing-continues.html for further details).

[7] See http://online.wsj.com/news/articles/SB10001424052702304244904579276830893405644.

thereby significantly reducing fluctuations in GDP, especially for small open economies.

Moreover, the seminal paper by Chen *et al.* (2010) has shown that changes in the currencies of commodity exporters such as Australia, Canada, Chile, New Zealand and South Africa '*Granger-cause*' changes in various trade-weighted indexes of world non-fuel commodity prices. Chen *et al.* (2010) show further that all of the exchange rates of the selected countries outperform a random walk (RW) model forecast of world non-fuel commodity prices in an out-of-sample analysis, with the Chilean peso indicating especially strong predictive power for copper prices.[8]

Financial market participants such as traders and quantitative analysts are widely seen to be forward looking and best informed when it comes to the pricing of asset. Since changes in future commodity prices will directly impact upon the revenue generating capacities of copper dependent companies, stock price data from individual mining companies are expected to contain some predictive information about the future movements of copper prices.[9] Although it is clear that equity price data is likely to produce a noisy signal as company specific news and events will react to information about the standing of the firm, its revenue generating capacity as well as the macroeconomic outlook, and not just information about commodity prices. Nevertheless, we do expect some useful information to be contained in the stock prices of big resource firms such as Freeport McMoran, BHP Billiton, Rio Tinto, and Alcoa.

[8] It should also be pointed out here that there exists evidence of commodity prices being a predictor of exchange rates for countries with high commodity export shares. This is documented in Cashin *et al.* (2004), who evaluates the co-movements of commodity prices and trade-weighted exchange rates for 58 countries over the period $1980-2002$. They found that the equilibrium real exchange rate for commodity exporting countries is affected by the respective commodity prices. Since commodity price shocks can be quite persistent, the long-run equilibrium level for the exchange rate is itself time-varying. Cashin *et al.* (2004) suggest that it is the exchange rate that adjusts to these changes when reverting to its equilibrium, rather than commodity prices. Nevertheless, since the focus is here on forecasting LME copper prices, we do not consider this point here any further.

[9] It is expected that companies will likely hedge their price exposures by entering into futures contracts or by buying options. Nevertheless, financial market participants are still likely to look at headline spot price figures and form expectations about future earnings based upon those figures. We therefore do anticipate an impact on stock prices as well, which in turn will be predictive for copper itself.

2.4. Existing literature

The empirical literature on modelling commodity prices has found mixed results regarding predictability. Since the overall positive results found in the seminal paper of Chen *et al.* (2010), there has been a renewed interest in evaluating the forecasting properties of new econometric models as well as new predictor variables. To avoid reviewing this fairly large literature on modelling commodity prices, the review that we present here focuses only on studies that have been undertaken after the results of Chen *et al.* (2010).

Continuing the analysis in Chen *et al.* (2010), Chen *et al.* (2012) extend the causality tests that they utilise to other commodity price indexes, such as agricultural commodities and individual commodities. Using quarterly data from $1980 - 2008$, they forecast agricultural indexes from the Commodity Research Bureau (CRB), the Economist and Standard and Poor's (S&P), as well as rice and wheat prices. Nevertheless, their results are mixed, providing only limited evidence in favor of predictive power of exchange rates. Moreover, their results vary with the agricultural commodity price indexes that are used, with the suggested models sometimes outperforming the forecasts of a random walk benchmark. The interesting insight remains that exchange rates can help predict the prices of those commodities that are more relevant in the export basket of the countries' considered.

Groen and Pesenti (2011) apply a factor augmented model with time-varying factor loadings. They build on a fact noted by Chen *et al.* (2012) that static models may be missing a lot of the possibly evolving relationship between commodity prices and exchange rates, due to evidence of parameter instability. Groen and Pesenti (2011) use monthly averages of broad commodity indexes from CRB, S&P and the IMF from 1973 to 2009 with exchange rates and fundamentals such as industrial production, business and consumer confidence indexes, unemployment rates, core consumer prices, money aggregate and interest rates, inventories and production of metals and oil derivatives, the Baltic Dry Index of transportation costs, and others. Their results are also mixed, with statistical evidence in favor of exchange rates forecasting commodity price indexes being admittedly inconclusive.

Chen and Tsay (2011) employ a mixed-frequency data approach to forecast quarterly changes in commodity price indexes, with daily exchange rates and equity data from 1984 to 2010 as predictor variables. Their findings are encouraging, but are, nevertheless, sensitive to different training sample sizes used in the estimation of the models. The results in Chen and Tsay (2011) also suggest that equity price indexes can have a predictive power similar to exchange rates, which is in line with the broad finding in Rapach *et al.* (2013), that information in U.S. stock prices can be used to forecast stock returns in other international markets. Using daily futures data on the exchange rates of Australia, Canada, New Zealand and South Africa, Chan *et al.* (2011) find no evidence of Granger-causality, and hence no predictability, in either direction and conclude that the futures market, being more liquid than the spot and the forward market for commodities, is more efficient.

Issler *et al.* (2014) apply forecast combination techniques to predict monthly and quarterly commodity prices from the IMF International Financial Statistics (IFS) database, over the 1965 to 2008 period, using forecasts from random walk, auto-regressive, vector auto-regressive, and vector error-correction models. The averaging techniques include simple averages, averages over the best five or ten models selected by BIC, weighted averages and bias corrected average forecasts. They show that the best way to improve on a RW forecast at monthly frequency appears to be simple forecast averaging.

Gargano and Timmermann (2014) use sets of financial and macroeconomic predictors, including industrial production, unemployment, inflation, the Australian Dollar, the Indian Rupee and futures market open interest. They use both univariate as well as multivariate regressions to predict returns on CRB commodity sub-indices, ranging from industrials to metals, fats/oils, foods, textiles and others. In their multivariate exercise, they use Ridge regressions and forecast averaging techniques where the models are selected from all possible subsets of the full regressor set. Overall forecast improvements over a RW benchmark lie in the range of

$3 - 14\%$ and are found at monthly as well as quarterly frequencies.

3. Econometric model

We model copper returns using the recently proposed Dynamic Model Averaging and Selection (henceforth DMA and DMS, respectively) framework.[10] The DMA/DMS framework is particularly appealing in the given context, as it combines two attractive modelling features. These are: *i*) time varying parameters and *ii*) model averaging and/or model selection. Allowing the parameters of a linear model to be time varying is desirable as economic conditions are frequently thought to evolve over time due to changes in technology, legal environments, agent preferences and learning. In a broader context, West and Harrison (1997) have argued that incorporating time varying parameters into a model can, to some extent, approximate neglected non-linearities and/or omitted variables, which can be beneficial when the model is used for forecasting.

The advantages of averaging over various competing models when forecasting are well known since the seminal work of Bates and Granger (1969).[11] Contrary to a simple time varying parameter model, where one model with the same set of predictors holds for all time periods, the DMA/DMS framework allows in addition to time varying parameters the set of predictors to change at each point in time. These two features substantially increase the flexibility of the model, as the commonly encountered problem of *'over-fitting the data'* from using too many predictors, which is known to lead to poor out-of-sample forecast performance, can be mitigated.

[10]The DMA/DMS framework was first introduced by Raftery *et al.* (2010) and subsequently applied to inflation forecasting in an economic context by Koop and Korobilis (2012).

[11]The popularity of model combination by means of averaging when forecasting has recently resurfaced in the empirical macroeconomic and finance literature (see, among many other studies, Raftery *et al.* (1997), Fernández *et al.* (2001a,b), Cogley and Sargent (2005), Eklund and Karlsson (2007), Wright (2008), and Buncic and Melecky (2014)).

3.1. The DMA/DMS approach

To outline how the DMA/DMS framework is implemented, let y_t denote the variable to be predicted at time period t.[12] Also, let $\mathbf{x}_{t-1}$ be a $(1 \times K)$ vector that contains the full set of k predictors plus an intercept term ($K = k + 1$), and let $m = 1, \ldots, M$ denote the model index, where $M = 2^k$ is the total number of possible (linear) model combinations (including the trivial model with only a constant term in it).[13] The two equations that make up the DMA/DMS framework (for model m) are:

$$\text{Measurement}: \underset{(1\times1)}{y_t} = \underset{(1\times K)}{\mathbf{x}_{t-1}^{(m)}} \underset{(K\times1)}{\boldsymbol{\beta}_t^{(m)}} + \underset{(1\times1)}{u_t^{(m)}} \tag{1a}$$

$$\text{State}: \underset{(K\times1)}{\boldsymbol{\beta}_t^{(m)}} = \underset{(K\times1)}{\boldsymbol{\beta}_{t-1}^{(m)}} + \underset{(K\times1)}{\boldsymbol{\epsilon}_t^{(m)}}, \tag{1b}$$

where (1a) and (1b) are measurement and state equations, respectively. The two disturbance terms $u_t^{(m)}$ and $\boldsymbol{\epsilon}_t^{(m)}$ in (1) are jointly Multivariate Normal (MN) distributed, uncorrelated with each other and over time, that is:

$$\begin{bmatrix} u_t^{(m)} \\ \boldsymbol{\epsilon}_t^{(m)} \end{bmatrix} \sim \text{MN}\left(\begin{bmatrix} 0 \\ \underset{(K\times1)}{\mathbf{0}} \end{bmatrix}, \begin{bmatrix} \underset{(1\times1)}{H_t^{(m)}} & 0 \\ \underset{(K\times K)}{\mathbf{0}} & \underset{(K\times K)}{\boldsymbol{Q}_t^{(m)}} \end{bmatrix} \right), \tag{2}$$

where $H_t^{(m)}$ and $\boldsymbol{Q}_t^{(m)}$ are the variance and covariance matrix of the measurement and state equations, respectively.

Also, let $\mathcal{M}_t$ denote the set of all possible models at time t, so that $\mathcal{M}_t \in \{1, 2, \ldots, M\}$. Given knowledge of $H_t^{(m)}$ and $\boldsymbol{Q}_t^{(m)}$ and by fixing the model set $\mathcal{M}_t = m$, ie., to one particular model,

[12] For reasons of simplicity, we use standard y_t and $\mathbf{x}_t$ notation to denote the left-hand side and predictor variables in the general description of the modelling framework. In our setting, y_t is the monthly copper return computed as $100(S_t/S_{t-1} - 1)$, where S_t is the LME spot price of copper. This will be made explicit in Section 4 and Section 5, where the data and the forecast evaluation results are discussed.

[13] The term model here refers to the different possible linear combinations that can be obtained from using k predictors in a regression context, rather than the more general definition, where a model can be anything, potentially as flexible as non-linear or a non-parametric specification. The use of the term model is standard in the model averaging literature.

the system in (1) takes the form of a standard state-space model, making it thereby possible to extract or *'filter'* the time varying parameters $\boldsymbol{\beta}_t^{(m)}$ as the *'latent states'* using standard Kalman Filter recursions. One-step-ahead forecasts and forecast errors are available as a by product of the Kalman Filter. Given $\mathcal{M}_t = m$, $H_t^{(m)}$ and $\boldsymbol{Q}_t^{(m)}$, the Kalman Filter recursions are:

$$
\begin{aligned}
\text{Prediction}: \quad \hat{\boldsymbol{\beta}}_{t|t-1}^{(m)} &= \hat{\boldsymbol{\beta}}_{t-1|t-1}^{(m)} \\
\boldsymbol{P}_{t|t-1}^{(m)} &= \boldsymbol{P}_{t-1|t-1}^{(m)} + \boldsymbol{Q}_t^{(m)} && \text{(3a)} \\
\hat{y}_{t|t-1}^{(m)} &= \mathbf{x}_{t-1}^{(m)} \hat{\boldsymbol{\beta}}_{t|t-1}^{(m)} && \text{(3b)} \\
\text{Prediction errors}: \quad \hat{u}_t^{(m)} &= (y_t - \hat{y}_{t|t-1}^{(m)}) \\
\text{MSE of prediction errors}: \quad F_t^{(m)} &= \mathbf{x}_{t-1}^{(m)} \boldsymbol{P}_{t|t-1}^{(m)} \mathbf{x}_{t-1}^{\mathsf{T}(m)} + H_t^{(m)} && \text{(3c)} \\
\text{Kalman Gain}: \quad \boldsymbol{G}_t^{(m)} &= \boldsymbol{P}_{t|t-1}^{(m)} \mathbf{x}_{t-1}^{\mathsf{T}(m)} / F_t^{(m)} \\
\text{Updating}: \quad \hat{\boldsymbol{\beta}}_{t|t}^{(m)} &= \hat{\boldsymbol{\beta}}_{t|t-1}^{(m)} + \boldsymbol{G}_t^{(m)} (y_t - \hat{y}_t^{(m)}) && \text{(3d)} \\
\boldsymbol{P}_{t|t}^{(m)} &= \boldsymbol{P}_{t-1|t-1}^{(m)} - \boldsymbol{G}_t^{(m)} \mathbf{x}_{t-1}^{(m)} \boldsymbol{P}_{t-1|t-1}^{(m)}
\end{aligned}
$$

where $\hat{\boldsymbol{\beta}}_{t|t-1}^{(m)} = \mathbb{E}_{t-1}(\boldsymbol{\beta}_t^{(m)})$, $\mathbb{E}_{t-1}(\cdot)$ is the expectation taken with respect to a time $t-1$ information set denoted by $\mathcal{I}_{t-1}$, and $\boldsymbol{P}_{t|t-1}^{(m)}$ is the mean square error (MSE) of $\hat{\boldsymbol{\beta}}_{t|t-1}^{(m)}$. Forecasts from model m using information set $\mathcal{I}_{t-1}$ are denoted by $\hat{y}_{t|t-1}^{(m)}$. The one-step-ahead forecast error is $\hat{u}_t^{(m)}$ and its associated MSE is denoted by $F_t^{(m)}$. The $(K \times 1)$ vector $\boldsymbol{G}_t^{(m)}$ is the Kalman Gain. The terms $\hat{\boldsymbol{\beta}}_{t|t}^{(m)}$ and $\boldsymbol{P}_{t|t}^{(m)}$ are updated (or time t) estimates of the latent states $\boldsymbol{\beta}_t^{(m)}$ and their corresponding MSEs.

The Kalman Filter recursions in (3) are conditional on $H_t^{(m)}$ and $\boldsymbol{Q}_t^{(m)}$ (and model m). To avoid having to estimate $H_t^{(m)}$ and $\boldsymbol{Q}_t^{(m)}$, two simplifying assumptions are used in the literature. The first one, which is due to Raftery *et al.* (2010), is to replace $\boldsymbol{P}_{t|t-1}^{(m)}$ in (3a) by

$$\boldsymbol{P}_{t|t-1}^{(m)} = \frac{1}{\lambda} \boldsymbol{P}_{t-1|t-1}^{(m)}, \tag{4}$$

where $\lambda \in [0,1]$. This approximation implies that $\boldsymbol{Q}_t^{(m)} = \left(\lambda^{-1} - 1\right) \boldsymbol{P}_{t-1|t-1}^{(m)}$. In the given context, the λ parameter is commonly referred to as a *'forgetting factor'*, as it determines how many observations are effectively used for estimation.[14] The second simplifying assumption is to replace the time varying volatility $H_t^{(m)}$ by a simple exponentially weighted moving average (EWMA) estimate, that is, $H_t^{(m)}$ is constructed as:

$$H_t^{(m)} = \kappa H_{t-1}^{(m)} + (1-\kappa)\hat{u}_{t-1}^{2(m)}, \tag{5}$$

where $\kappa \in [0,1]$ is the standard EWMA smoothing parameter. Note here that an EWMA model can be thought of as a special form of a GARCH$(1,1)$ model, ie., a restricted integrated GARCH$(1,1)$, with the restriction being that the intercept term is fixed at 0 and that the weights on the $t-1$ volatility and squared error term sum to unity.[15]

Model averaging or selection in the DMA/DMS framework is achieved by weighting the forecasts by their respective predictive model probabilities. To clarify this, let us define $\pi_{t|t-1}^{(m)}$ to be the probability of model m given information up to time $t-1$, written as:

$$\pi_{t|t-1}^{(m)} = \Pr(\mathcal{M}_t = m|\mathcal{I}_{t-1}). \tag{6}$$

The DMA forecast of y_t, given information up to time $t-1$, denoted as $E(y_t|\mathcal{I}_{t-1})$, is then computed as:

$$\hat{y}_{t|t-1}^{(\mathrm{DMA})} = \sum_{m=1}^{M} \hat{y}_{t|t-1}^{(m)} \pi_{t|t-1}^{(m)}, \tag{7}$$

that is, as a weighted average of the forecasts from all possible models, $\{\hat{y}_{t|t-1}^{(m)}\}_{m=1}^{M}$, with the

[14]This is also known as *'windowing'*. Intuitively, we can think of λ as a weighting function, where observations τ periods in the past receive a weight of λ^{τ}. See the discussion in Section 3.1 in Raftery *et al.* (2010) and pages $872-873$ in Koop and Korobilis (2012) for more background and intuition about the use of forgetting factors in dynamic econometric models and what it implies for the effective sample size.

[15]It is well known in the volatility literature that GARCH$(1,1)$ models are difficult to beat in out-of-sample forecast evaluations (see, for instance, Hansen and Lunde, 2001). Approximating the time varying volatility by EWMA is thus unlikely to create any important loss in accuracy. We discuss later on in the estimation section how κ is calibrated.

averaging weights being the predictive probabilities $\{\pi^{(m)}_{t|t-1}\}^{M}_{m=1}$.

The DMS forecast of y_t is computed by choosing at each point in time the forecast from the model with the highest predictive model probability. Formally, let $\pi^{*}_{t|t-1} = \max\{\pi^{(m)}_{t|t-1}\}^{M}_{m=1}$ and $\mathbb{1}(\mathcal{A})$ be an indicator function that is equal to 1 if statement $\mathcal{A}$ is true, and 0 otherwise. Then

$$\hat{y}^{(\mathrm{DMS})}_{t|t-1} = \sum_{m=1}^{M} \hat{y}^{(m)}_{t|t-1} \mathbb{1}(\pi^{*}_{t|t-1} = \pi^{(m)}_{t|t-1}). \tag{8}$$

To make the construction of the DMA/DMS forecasts in (7) and (8) feasible, model prediction and updating recursions are needed. Let $p_{jm} = \Pr(\mathcal{M}_t = m|\mathcal{M}_{t-1} = j)$ denote the (time invariant) transition probability of moving from model j at time $t-1$ to model m at time t. Also, let $f^{(m)}_{\mathsf{N}}(y_t|\mathcal{I}_{t-1})$ denote the predictive density of y_t given model m and information up to time $t-1$. This predictive density is a Normal density evaluated at y_t with mean and variance given by $\hat{y}^{(m)}_{t|t-1}$ and $F^{(m)}_t$ as computed in (3b) and (3c), respectively. That is, $f^{(m)}_{\mathsf{N}}(y_t|\mathcal{I}_{t-1}) = \mathsf{N}(\hat{y}^{(m)}_{t|t-1}, F^{(m)}_t)$. Given an initial or prior model probability $\pi^{(m)}_{0|0}$, the model prediction and updating equations are then constructed as:

$$\text{Model Prediction}: \quad \pi^{(m)}_{t|t-1} = \sum_{j=1}^{M} \pi^{(j)}_{t-1|t-1} p_{im} \tag{9a}$$

$$\text{Model Updating}: \quad \pi^{(m)}_{t|t} = \frac{\pi^{(m)}_{t|t-1} f^{(m)}_{\mathsf{N}}(y_t|\mathcal{I}_{t-1})}{\sum_{j=1}^{M} \pi^{(j)}_{t|t-1} f^{(j)}_{\mathsf{N}}(y_t|\mathcal{I}_{t-1})}. \tag{9b}$$

A final simplification that is need to make the computation of the predictive model probabilities feasible is to approximate (9a) with

$$\pi^{(m)}_{t|t-1} = \frac{\pi^{\alpha(m)}_{t-1|t-1}}{\sum_{j=1}^{M} \pi^{\alpha(j)}_{t-1|t-1}}, \tag{10}$$

where $\alpha \in [0,1]$.[16] The approximation in (10) has the advantage that one avoids having to

[16]See Raftery *et al.* (2010) for more intuition about this approximation.

specify an $M \times M$ dimensional model probability transition matrix, which would make model prediction computationally infeasible when M is large. The α parameter in (10) can again be interpreted as a '*forgetting factor*'.[17]

The implementation of the DMA/DMS procedure to forecast copper returns requires the calibration of the EWMA smoothing parameter κ as well as the two forgetting factor parameters, λ and α. We follow the guidelines provided in RiskMetrics (1996) for monthly data and fix the κ parameter at 0.97.[18] For the two forgetting factors, Koop and Korobilis (2012) recommend to set the values for λ and α in the $[0.95, 0.99]$ interval, so that the parameters (as well as the model probabilities) evolve reasonably gradually over time.[19] We follow these recommendations and use a common (λ, α) combinations in the empirical evaluation.

4. Data

We analyse the predictive power of the DMA/DMS modelling framework using monthly data over the (full) sample period from June 1996 to June 2014. All data that we use in our analysis were obtained from Bloomberg and the St. Louis Federal Reserve FRED database.[20] With the exception of copper consumption and production data and also U.S. industrial production, which are available only at a monthly frequency, all series were aggregated from daily observations to monthly averages.[21] The choice of the sample period was driven by a trade-off between maximum possible sample size and variability in the Chilean peso, which was offi-

[17] See also Section 3.2 in Raftery *et al.* (2010) and pages $874-875$ in Koop and Korobilis (2012) for additional discussion on this.

[18] See page 97 of the documentation in RiskMetrics (1996). Note here that RiskMetrics (1996) uses λ to denote their EWMA smoothing parameter and not κ as we do.

[19] See the discussion on pages $872-875$ in Koop and Korobilis (2012). The effective window size, ie, how much of a weight observations in the past obtained, is determined from $1/(1-\lambda)$ (or $1/(1-\alpha)$, respectively). Choosing values below say 0.95, would make the window narrow, so that only the very recent past receives non-zero weights, which could result in very noisy forecasts.

[20] FRED Data is available from http://research.stlouisfed.org/fred2/.

[21] Bloomberg and FRED tickers are reported in the description of the variables in Table 1 in parenthesis. We follow Chen *et al.* (2010), Groen and Pesenti (2011), Gargano and Timmermann (2014) and many others and construct monthly averages from daily observations to reduce the volatility that is inherent when using end-of-period data.

cially pegged until September 1999.[22] In 1994, the Chilean central bank moved to an inflation targeting policy regime, setting interest rates to keep inflation rates low, stable, and sustainable over time. One could argue here that the policy focus may have already shifted more towards price stability, with less attention given to the management of the currency, although it is clear that there would have been intervention periods by the central bank to maintain stability in the currency as well. Nevertheless, our view is that with June 1996 as the starting date, we are able to maintain a fairly sizeable sample, while avoiding *'major'* influences of the peg of the Chilean peso on our forecast evaluation results. The full sample period thus consists of 217 monthly observations. We use the first 70 observations as our in-sample or fitting period, and leave the remaining observations for out-of-sample evaluation.[23]

4.1. Copper prices and returns

There exist three major exchanges for copper. These are the London Metal Exchange (LME), the Shanghai Futures Exchange (SHFE) and the New York Mercantile Exchange (NYMEX). In our analysis, we use copper prices from the London Metal Exchange, which is the oldest and most liquid metal exchange in the world and is widely regarded as *the* reference index for world copper prices (see also Watkins and McAleer, 2004 for additional information).[24] To formalise notation, let S_t denote the (LME) spot price of copper at time t. The one period (simple) return for copper is then defined as $r_t = 100(S_t/S_{t-1} - 1)$. To provide a visual overview of the time

[22]Since the end of 1984, the Chilean peso moved from a peg to the U.S. dollar to a crawling peg within some target bands ranging between $2.5\% - 5\%$ that were periodically adjusted. In September 1999, the Chilean peso was official freely allowed to float against the U.S. dollar, nevertheless, with occasional interventions from the Chilean central bank, with the objective to counter excessive depreciation and volatility in the currency.

[23]Note that we are not per se interested in in-sample fitting, but since we are going to compare the forecasts from the DMA/DMS model to simple expanding and rolling window OLS forecasts, an initial fitting period is required. When using the Kalman Filter, it is also advisable to have an initialisation period for the filter to avoid any dependencies on the initial (or *'prior'*) values need to get the filter started.

[24]The London Metal Exchange was founded in 1877. Labys *et al.* (1971) have pointed out that LME prices are highly responsive to world demand and supply, reflecting prices at which bargains are struck throughout the world rather than fixed or administered prices. Nevertheless, it is worth noting here that the relative importance of the LME may have somewhat diminished over recent years due to the increasing importance of the Shanghai Futures Exchange. Although the SHFE accounted for only about 25% of copper price determination in 2007, while the LME accounted for 45%, the Shanghai Futures Exchange has grown substantially since its inception in 1992 (see also Hua *et al.*, 2009, who study the SHFE's growth performance in more detail).

series properties of both, S_t as well as r_t, we show a plot corresponding to monthly LME copper prices and their returns in Panels (a) and (b) of Figure 1.

← Figure 1 about here

We can see from Panel (a) of Figure 1 how copper prices have significantly increased since 2003, peaking at the beginning of 2008, plummeting at the onset of the financial crisis in June 2008, then surging again strongly since the end of 2008, reaching another peak in mid 2011, and gradually declining thereafter to more stable levels at the end of the sample period. From Panel (b) of Figure 1 it is noticeable that volatility has roughly doubled since the beginning of 2006 and has remained higher until approximately the end of 2011. While there exist studies that attribute this increase in volatility entirely to financial crises (see, for instance, Dwyer *et al.*, 2011), Radetzki *et al.* (2008) relate the observed cycles to asymmetries in supply and demand patterns, and argue that the upsurge in copper prices is likely to be reverted as soon as supply adjustments are completed.[25] Nevertheless, irrespective of the sources of the variation, it is clear from the return series that time-varying volatility is an evident feature of the data, albeit being less pertinent than in other financial time series, such as exchange rates or stock prices.

4.2. Predictor variables

Given the theoretical discussion provided in Section 2, it is clear that we expect a mix of fundamental and financial variables to be informative for forecasting copper returns. We therefore select the following three broad groups of predictor variables to match the factors outlined earlier. These are variables relating to *i*) Fundamentals, *ii*) Financialization, and *iii*) Exchange rates and stock prices. In the following sub-sections we describe the predictors in each of these three groups separately.

[25] Several projects are indeed expected to be completed in the next few years, thus predicting a substantial increase in copper production, which may thus potentially lower copper prices if the demand from China does not follow through.

4.2.1. Fundamental variables

We use data from the World Bureau of Metal Statistics (WBMS) on world refined copper production and consumption together with LME warehouse inventories to construct a measure of copper (excess) demand (simply demand henceforth), denoted by Demand_t, which is computed as $\text{Demand}_t = \text{Consumption}_t - (\text{Production}_t + \text{Inventory}_t)$, where Production_t and Consumption_t are defined as world refined copper production and consumption, respectively, and Inventory_t are LME warehouse stocks. Since Demand_t is a strongly trending variable, we use the month-on-month growth rate in demand computed as $\Delta\text{Demand}_t = (\text{Demand}_t - \text{Demand}_{t-1})$ as the predictor variable.[26] Given the availability of inventory data, we also construct monthly growth in inventories (denoted by $\Delta\text{Inventory}_t$) as a predictor.

We further use data on spot and forward rates to construct the *'marginal convenience yield'* (simply convenience yield henceforth and denoted by Convyield_t), which is defined as:

$$\text{Convyield}_t = i_t^{(\tau)} - (F_t^{(\tau)} - S_t)/S_t + W_t^{(\tau)}/S_t, \tag{11}$$

where S_t is the spot price, $i_t^{(\tau)}$ and $F_t^{(\tau)}$ are the risk free interest rate and the futures prices at time t with maturity τ, and $W_t^{(\tau)}$ is the cost of storing the commodity from time t to τ, with $\tau > t$.[27] Fama and French (1988, page 1077) point out that the marginal storage cost is effectively constant over time and thus will only have a levels effect on the convenience yield. Due to the unavailability of reliable storage cost data, we ignore the influence of the marginal storage cost $W_t^{(\tau)}/S_t$ on the convenience yield, and effectively set it to 0 in (11).

Although we expect the demand, inventory and convenience yield based fundamentals to contain important information about global demand and supply imbalances in the market for copper, we also include predictor variables that are meant as proxies for overall economic activ-

[26]Changes are denoted by the difference operator Δ. Also, the change in the monthly demand variable ΔDemand_t is measured in 10 000 metric tonnes. All world production and consumption data from WBMS are available only at a monthly frequency.

[27]See pages $1077 - 1078$ in Fama and French (1988) for further details on the construction.

ity, since the demand for copper is directly related to economic growth because of its primary use in production, construction and manufacturing. We use two different sets of variables to proxy economic activity. The first consists of growth in U.S. Industrial Production (abbreviated by ΔIP_t). Due its high *'correlation'* with real economic activity and availability at monthly frequency, industrial production is one of the most widely used economic performance indices for the U.S.[28] Nevertheless, one weakness of solely using industrial production as a measure of economic activity in a forecasting application is that it is a backward looking measure, because it only accounts for production that has already taken place. To include more forward looking information about economic activity, we use the term spread and the Baltic Dry Index (BDI) in the second set of economic activity proxies.

The information content in the term spread (denoted by Spread_t), defined as the difference between the U.S. 10 year Treasury Bond and the 3 month Treasury Bill rate, is well known to be a real time predictor of economic activity.[29] The Baltic Dry Index has been used as a proxy for global trade flows as well as supply and demand trends in production of finished goods and raw materials. The BDI is made up as a composite index of the Baltic Capesize, Panamax, Handysize and Supramax shipping indices and is designed as the successor to the Baltic Freight Index. The BDI is frequently viewed as a leading indicator of future global economic growth, as the goods that are transported are raw materials as well as final goods, thus give an indication of the demand for primary inputs and overall trade flows. We use the monthly growth rate in the Baltic Dry Index denoted by ΔBDI_t as a predictor variable. In summary, the 6 fundamental

[28] More concretely, the correlation between quarterly growth in industrial production and real GDP growth is 80% for data spanning from 1947:Q2 to 20014:Q2. One drawback of using IP growth as a proxy for economic activity is that it is twice as volatile real GDP growth. Nevertheless, industrial production is also used in, among others, Groen and Pesenti (2011), Gargano and Timmermann (2014) and we thus deem it to be a reasonable indicator of economic activity.

[29] There exists a long and well established literature in monetary economics that links the yield curve to economic activity. One of the seminal papers using only the term spread as we do here, rather then the full set of level, slope and curvature factors, is the study by Estrella and Hardouvelis (1991), who showed that the term spread is predictive for economic output as well as NBER dated recessions. More recently, Ang and Piazzesi (2003) among many others have shown that this holds true even in more rigourous formulated term structure models that account for more than just one factor (see also Brand *et al.* (2010) for the use of yield curve factors in the construction of monetary policy surprises).

variables that we construct are:

$$\mathbf{x}_t^{(\text{fun})} = [\Delta\text{Demand}_t, \Delta\text{Inventory}_t, \text{Convyield}_t, \Delta\text{IP}_t, \text{Spread}_t, \Delta\text{BDI}_t]. \tag{12}$$

4.2.2. Financial variables

To account for the impact of financialization on the evolution of copper prices, we include 5 headline financial variables in the set of predictors for copper. These are the Chicago Board Options Exchange Market Volatility Index (VIX), the TED spread (TED), the Standard and Poor's 500 (SP500) stock price index, the spot price of Gold (Gold) and WTI crude oil spot prices (Oil).

The VIX measures the volatility implied by option prices on the S&P500 (SP500 henceforth) over the coming month. The TED spread is calculated as the difference between the 3 month LIBOR rate (U.S. dollar based) and the 3 month Treasury Bill rate. A higher value in the VIX and/or the TED spread is generally taken as an indication of market participants expecting an overall negative economic or financial outlook and hence an increased (global) aversion to risk.[30] Brunnermeier *et al.* (2009) have shown that the VIX and the TED spread predict higher returns in carry trade strategies used in the foreign exchange market. Given their predictive power in the foreign exchange market, we expect the VIX and the TED spread to also contain some predictive information for copper. We use the (level) VIX and TED spread series (VIX_t and TED_t respectively) as predictors.

We also compute the monthly return on the SP500 index (denoted by ΔSP500_t). Rapach *et al.* (2013) have recently documented that returns in the U.S. stock market are predictive for returns in various other global stock markets. To account for the possible predictive component of the information content in the SP500 index, we include ΔSP500_t as a predictor variable for

[30] The VIX and the TED spread are widely regarded as measures of the *'global appetite for risk'*. This is not only the case for equity markets and equity options markets, but also for corporate credit markets and foreign exchange markets (see for instance the evidence reported in Collin-Dufresne *et al.*, 2001). Also, Pan and Singleton (2008) find that the VIX in particular is strongly related to the variation in risk premiums in sovereign credit default swaps.

copper.[31]

Gold and crude oil are the only two other commodities that we include in the set of predictor variables. The motivation for including gold as a viable predictor variable is twofold. First, as discussed in Section 2, copper has become an alternative asset to precious metals such as gold and silver in investment portfolios that are diversified over equities, bonds, exchange rates and commodities. Second, gold is also considered to be a *'safe haven'* asset and can hence be seen to complement the VIX and the TED spread indicators of risk aversion in financial markets. Gold is further viewed to be a hedge against inflation, deflation, as well as general uncertainties related to economic, financial and political instabilities. We expect, therefore, monthly gold returns (ΔGold_t) to be informative for copper forecasting, especially since the financial crisis in 2008.

The rational for using crude oil prices in the set of predictors is due to oil still being one of the most widely used sources of energy (see for instance, among many other studies, the evidence reported in Lardic and Mignon (2008) and He *et al.* (2010)). Moreover, there is a widely held view that unexpected increases in the price of oil can cause recessions in many oil importing countries (see Kilian (2008), Hamilton (2009) and others). High oil prices are often also linked to periods of higher inflation, thereby directly affecting central bank policy and thus the setting of interest rates (Bhar and Mallik, 2013). Lastly, oil prices, in conjunction with U.S. Energy Information Administration (EIA) inventories are closely monitored by financial market participants and reported in the financial press. These are taken to be early indicators of changes in production and manufacturing demand. We thus include monthly oil (spot price) returns (denoted by ΔOil_t) in the predictor set. To summarise, the five variables that are included under the financial variables heading are:

$$\mathbf{x}_t^{(\text{fin})} = [\text{VIX}_t, \text{TED}_t, \Delta\text{SP500}_t, \Delta\text{Gold}_t, \Delta\text{Oil}_t]\,. \tag{13}$$

[31] The explanation given in Rapach *et al.* (2013) is that the U.S. is a news or information hub, where news about economic and financial developments are most efficiently absorbed and reacted to.

4.2.3. Exchange rate and stock price

The final group of predictor variables consists of two exchange rates, namely, the Chilean peso (CLP) and the Australian Dollar (AUD) and the stock prices of four major commodity/resource based firms.

Following the well known results of Chen *et al.* (2010) that exchange rates of commodity exporting countries can forecast commodity prices, we include monthly returns in the Chilean peso (ΔCLP_t) as one of the predictor variables in the forecasting model. Chile is *the* single largest exporter of copper in the world, providing approximately 39% of world copper in 2008.[32] Copper accounts for about half of Chile's exports. The dependence of Chile's economy on copper was highlighted in the study by Spilimbergo (2002), who showed that copper price cycles almost always preceded business cycles during the 1990s. Since then, the Chilean government has introduced new anti-cyclical fiscal policy rules to counteract at least some of the dependence on copper (Gregorio and Labbé, 2011). Moreover, the free float of the exchange rate in September 1999 has helped to absorb the effect of copper price shocks on the domestic economy. The Chilean central bank has also intervened on a few occasions in the foreign exchange market to counteract strong and unjustified depreciations in the peso.[33]

Apart from the Chilean peso, we also include the monthly return in the Australian dollar (ΔAUD_t) as a predictor variable. Australia accounted for merely 6% of world copper exports in 2008, being nevertheless the third largest exporter after Peru. In addition to copper, Australia also exports other commodities, such as iron ore, aluminium ore (bauxite), steel as well as gold, crude oil, beef and wheat. Any unexpected changes in commodity price will thus affect Australian export earnings, the economic outlook and thereby the value of the Australian dollar. Aside from the well known role of a *'commodity currency'*, the Australian dollar is, due to the high yields offered by Australian bonds, also used as an investment currency in carry

[32] See Table I.4 on page 12 in Meller and Simpasa (2011).

[33] Claro and Soto (2013) give a recent account of foreign exchange interventions by the Chilean central bank.

trades. Since carry trade returns are highly influenced by so called *'crash-risk"*, this strategy is only implemented once the risk appetite of financial market participants has increased to an appropriate level.[34] The role of the Australian dollar in the predictor set is thus twofold: *i*) to complement and broaden the commodity currency information component of the Chilean peso and *ii*) to act as an alternative to the VIX and TED spread based measures of global risk aversion.

We further add monthly stock returns from four of the largest listed commodity/resource based firms in the U.S. to the predictor set. These are monthly returns from Rio Tinto (ΔRio_t), Freeport McMoran (ΔFPM_t), BHBP Billiton (ΔBHP_t), and Alcoa (ΔAlcoa_t). Our main motivation for adding these stock returns is the *'efficiency view'* of the market, that is, any news related to unexpected future commodity price movements will be quickly absorbed and priced accordingly by market participants, providing the markets best view on the stock. Despite this view, we expected there to be also company specific noise that may distort the signal we receive about future copper price movements. Nevertheless, since the DMA/DMS econometric framework that we employ is flexible enough to downweight predictor variables that are less informative for future copper prices, we leave the selection of the most relevant predictors to the model. This last set of predictor variables is summarized as:

$$\mathbf{x}_t^{(\text{other})} = [\Delta\text{CLP}_t, \Delta\text{AUD}_t, \Delta\text{Rio}_t, \Delta\text{FPM}_t, \Delta\text{BHP}_t, \Delta\text{Alcoa}_t]. \tag{14}$$

Note that we also use lagged copper returns as a predictor variable. This is common practice in the asset return forecasting literature, as there can be periods of momentum in asset returns due to some market participants adopting a trend following trading strategy (see, for instance, Allen and Taylor (1990), Menkhoff (1998), and Lui and Mole (1999) for empirical evidence of trend following behaviour in foreign exchange markets and also the theoretical agent

[34]See the discussion on pages 314 to 318 in Brunnermeier *et al.* (2009) for further details on the carry trade, and the link between risk appetite and the Australian dollar.

based models of Brock and Hommes (1997, 1998) which allow for momentum trading). Again, our view is that if lagged copper returns are not informative, the model will downweight its importance in the forecasts.

4.3. Summary statistics and visual overview of predictor variables

In this section we briefly describe some of the basic features of the predictor variables that are evident from Table 1 and the plots shown in Figure 2. The exact transformations that were applied to the series are described in Table 1 and the corresponding Bloomberg and FRED tickers for each of the included variables are given in parenthesis.

← Table 1 about here

← Figure 2 about here

With regards to the transformations that were applied, most of the series are monthly growth rates, computed in the standard way as $100(P_t/P_{t-1} - 1)$ where P_t is the variable of interest at time t so that they can be interpreted as percentage changes. Predictors that were not transformed to growth rates are: the spread between the 10 year U.S. Treasury Bond and the 3 months U.S. Treasury Bill yield, the VIX index, the TED spread and the convenience yield. Changes in copper demand and inventories are computed as first differences.

Looking over the summary statistics that are reported in Table 1 it is evident that the means and medians of the predictors are largely in line with prior expectations in terms of magnitudes. What is interesting to see is that the Chilean peso has declined on average by 0.14 percent per month over the sample period, reflecting the largely depreciating trend since the float of the currency in September 1999. It is further interesting to see that the BDI index is the most volatile series, with a (monthly) standard deviation of 18.32, followed by the stock returns of the resource based companies, as well as oil returns. The relatively high means of the VIX index, the convenience yield and the spread are due to the fact that these variables are captured in levels, rather than growth rates.

Inspecting the plots of the individual predictors shown in Figure 2 highlights the nearly homogenous response of these variables to the September 2008 collapse of Lehman Brothers and hence the beginning of the most intense phase of the financial crisis, with most of the return series dropping to values of around -20 percent and in some extreme cases down to -60 percent for stock prices and the BDI. In fact, the Baltic Dry index seems to have been most affected by the crisis. It is also interesting to note here that, despite the sharp initial increase in changes in inventories after the onset of the crisis and subsequent sharp drop in 2009 and bounce back shortly after, changes in the demand for copper did not seem to be noticeably affected by the crisis period. From Figure 2 it seems that volatility in demand was in fact lower over the end of 2008 to beginning of 2009 period than during most other periods in the sample, spiking upwards only towards the end of 2009 and in mid 2013.

5. Forecast evaluation

In this section we evaluate the forecast performance of the dynamic model averaging and selection framework. Since we are primarily interested in the real time predictive ability of the the model, we focus here on evaluating the models' out-of-sample performance only. Before we outline in detail the statistical evaluation criteria that we use as well as the results of the forecast evaluation, we initially describe the prediction setting that we employ in our evaluation.

5.1. Prediction setting

5.1.1. Model set-up

Following the general description of the DMA/DMS framework in Section 3, the forecasting model for copper returns r_{t+1} takes the form

$$r_{t+1} = \mathbf{x}_t^{(m)} \boldsymbol{\beta}_{t+1}^{(m)} + u_{t+1}^{(m)} \tag{15a}$$

$$\boldsymbol{\beta}_{t+1}^{(m)} = \boldsymbol{\beta}_t^{(m)} + \boldsymbol{\epsilon}_{t+1}^{(m)} \tag{15b}$$

where $m = 1, \dots, M$ denotes the model index, $r_{t+1} = 100(S_{t+1}/S_t - 1)$, S_t is the LME spot price of copper, and the full $(1 \times K)$ predictor set $\mathbf{x}_t$ (including the intercept term) is defined as:

$$\mathbf{x}_t = \left[1, r_t, \mathbf{x}_t^{(\text{fun})}, \mathbf{x}_t^{(\text{fin})}, \mathbf{x}_t^{(\text{other})}\right]. \tag{16}$$

In (16), the intercept term is denoted by 1, r_t is the lagged copper return, and $\mathbf{x}_t^{(i)}$ for all $i =$\{fun, fin, other\} is as defined in (12), (13) and (14), respectively. The number of predictor variables (excluding the intercept term) is $k = (1 + 6 + 5 + 6) = 18$, thus a total of $M = 2^{18} = 262144$ models are available, at each point in time.[35]

To compute return forecasts for copper from the recursions in (15), time t filtered estimates $\hat{\boldsymbol{\beta}}_{t|t}^{(m)}$ are needed to construct the optimal forecast of $\boldsymbol{\beta}_{t+1}^{(m)}$. Given our random walk specification of the state dynamics in (1b), this forecast is given by $\hat{\boldsymbol{\beta}}_{t|t}^{(m)}$, for all $m = 1, \dots, M$. The sequence of $\hat{\boldsymbol{\beta}}_{t|t}^{(m)}$ are obtained from the Kalman filter recursions outlined in (3). To implement the Kalman filter, we need to specify initial values. We follow Koop and Korobilis (2012) and use a diffuse prior for $\hat{\boldsymbol{\beta}}_{0|0}^{(m)} \sim \mathsf{MN}(\mathbf{0}_{K_m}, 100\boldsymbol{I}_{K_m})$, where K_m denotes the dimension of the m^{th} model, $\mathbf{0}_{K_m}$ is a $(K_m \times 1)$ dimensional vector of zeros and $\boldsymbol{I}_{K_m}$ is $(K_m \times K_m)$ dimensional identity matrix. The model updating probabilities $\pi_{t|t}^{(m)}$ in (9b) are initialised with an uninformative prior $\pi_{0|0}^{(m)} = \frac{1}{M}$, so that all models are assumed to be equally likely. The α and λ parameters are set to $(0.95, 0.99)$.[36] The κ term in the EWMA specification is fixed at 0.97, in line with current RiskMetrics (1996) recommendations.

Given the model updating probabilities $\pi_{t|t}^{(m)}$ and $\alpha = 0.95$, forecasts of the model probabilities are computed as $\pi_{t+1|t}^{(m)} = \pi_{t|t}^{\alpha(m)} / \sum_{j=1}^{M} \pi_{t|t}^{\alpha(j)}$, yielding the DMA and DMS based forecasts

[35] Note that the null or base model is the one with only an intercept term in it.

[36] This combination is also used in Koop and Korobilis (2012) (see the results in Tables 4 and 5). We set the forgetting factor for the model probability updating somewhat lower to allow for a more frequent updating in the model probabilities. The reason for this is that we have a fairly large number of regressors and want to allow for the possibility of fast changes over time in terms of which ones are included in the model. One could experiment here with a few different values as well, nevertheless, we want to abstract from a search for the best values and use this simple parameter setting instead.

of copper returns:

$$\hat{r}^{(\mathrm{DMA})}_{t+1|t} = \sum_{m=1}^{M} \mathbf{x}^{(m)}_{t} \hat{\boldsymbol{\beta}}^{(m)}_{t|t} \pi^{(m)}_{t+1|t} \tag{17a}$$

$$\hat{r}^{(\mathrm{DMS})}_{t+1|t} = \sum_{m=1}^{M} \mathbf{x}^{(m)}_{t} \hat{\boldsymbol{\beta}}^{(m)}_{t|t} \mathbb{1}\left(\pi^{*}_{t+1|t} = \pi^{(m)}_{t+1|t}\right) \tag{17b}$$

respectively, where $\mathbb{1}\left(\pi^{*}_{t+1|t} = \pi^{(m)}_{t+1|t}\right)$ is again as defined above in (8), ie., an indicator variable that is equal to one for the most probable model.

5.1.2. Fitting and evaluation periods

Our entire available data set consists of $T = 216$ observations, covering the period from June 1996 to June 2014.[37] We follow common practice in the out-of-sample forecast evaluation literature and split the full sample into roughly 1/3 in-sample and 2/3 out-of-sample portions. More specifically, we use the first 70 observations for '*in-sample*' fitting and the remaining 146 for out-of-sample evaluation (the out-of-sample period thus spans from May 2002 to June 2014). The reason for the 1/3 and 2/3 split here is to have a relatively large number of out-of-sample forecasts available to be able to conduct statistically meaningful inference. Moreover, since we are working with a model that allows for time varying dynamics in the model parameters as well as the selected predictors, it will be interesting to see how the model performs over a sample that includes the period before, during and after the 2008 financial crisis.

5.2. Evaluation criteria

We assess the out-of-sample forecast performance of the DMA/DMS framework by following the recent literature on forecasting the equity premium. That is, we follow the approach of Rapach *et al.* (2013), Neely *et al.* (2014) and many others and evaluate the forecasts in terms of the Campbell and Thompson (2008) out-of-sample R^2 (denoted by R^2_{os} henceforth) and the Clark and West (2007) Mean Squared Forecast Error (MSFE) adjusted $t-$statistic, which we

[37] One observations is lost due to lagging.

denote by CW − statistic. The forecast errors from the various competing models are defined as:

$$\hat{e}^{(n)}_{t+1|t} = \left(r_{t+1} - \hat{r}^{(n)}_{t+1|t}\right) \tag{18}$$

with corresponding MSFEs being

$$\text{MSFE}_{(n)} = \frac{1}{T_{os}} \sum_{t=T_{is}}^{T} \hat{e}^{2(n)}_{t+1|t}, \tag{19}$$

for all $n = \{\text{DMA}, \text{DMS}, \text{BM}\}$, where BM is the benchmark model that is used. T_{os} and T_{is} denote, respectively, the number of out-of-sample and in-sample observations, so that $T_{is} + T_{os} = T$.[38]

The Campbell and Thompson (2008) R^2_{os} is computed as follows. Let $\text{MSFE}_{(\text{DMA})}$ be the MSFE from the DMA model and let $\text{MSFE}_{(\text{BM})}$ denote the MSFE from some alternative benchmark model. Then, the R^2_{os} comparing the performance of the DMA model to the BM is defined as:

$$R^2_{os} = 1 - \frac{\text{MSFE}_{(\text{DMA})}}{\text{MSFE}_{(\text{BM})}}. \tag{20}$$

Intuitively, the R^2_{os} statistic in (20) measures the reduction in the MSFE of the proposed DMA model relative to the benchmark model. When $R^2_{os} > 0$, then this is an indication that the proposed DMA model (or other competing model that is used) performs better than the benchmark model in terms of MSFE, while $R^2_{os} < 0$ suggests that the benchmark model performs better.

The Clark and West (2007) MSFE adjusted $t-$statistic is computed as (again assessing the performance of the DMA model relative to the BM):

$$\text{CW} - \text{statistic} = -\frac{2}{T_{os}} \sum_{t=T_{is}}^{T} \hat{e}^{(\text{BM})}_{t+1|t} \left(\hat{e}^{(\text{BM})}_{t+1|t} - \hat{e}^{(\text{DMA})}_{t+1|t}\right) \tag{21}$$

[38]For longer forecast horizons, ie., when $h > 1$, where h denotes the forecast horizon, this means that the number of out-of-sample observations is $T_{os} = T - T_{is} - h + 1$.

(see equation 4.1 on page 297 in Clark and West (2007)). Following the suggestion in Clark and West (2007, page 294) , the easiest way to compute the CW − statistic is to form the sequence

$$cw_{t+1} = dm_{t+1} + adj_{t+1} \tag{22}$$

where

$$dm_{t+1} = \hat{e}^{2(\text{RW})}_{t+1|t} - \hat{e}^{2(\text{DMA})}_{t+1|t} \tag{23}$$

and

$$adj_{t+1} = \left[\hat{r}^{(\text{RW})}_{t+1|t} - \hat{r}^{(\text{DMA})}_{t+1|t}\right]^2. \tag{24}$$

The dm_t term is the standard Diebold and Mariano (1995) sequence that is computed to test for (unconditional) superior predictive ability. The adjustment term adj_t arises due to the nested nature of the models being compared and performs a bias correction (see Clark and West (2007) for more details). The CW − statistic is then computed as

$$\text{CW} - \text{statistic} = \frac{\overline{cw}}{\sqrt{\text{Var}(\overline{cw})}} \tag{25}$$

where $\overline{cw} = T_{os}^{-1} \sum_{t=T_{is}}^{T} cw_{t+1}$ and $\text{Var}(\overline{cw})$ is the variance of the sample mean, which can simply be obtained as the heteroskedasticity and autocorrelation (HAC) robust $t-$statistic on the intercept term of a regression of cw_{t+1} on a constant.[39]

The CW − statistic implements a test of the null hypothesis that the MSFE of the benchmark model is equal to the MSFE of the DMA model, against the one sided alternative hypothesis that the benchmark's MSFE is greater than that of the DMA. A rejection of the null hypothesis hence suggests that DMA forecasts are (on average) *significantly* better than BM forecasts. It should be highlighted here that the CW − statistic is particularly suitable in the given context, as it is designed for a comparison of nested forecasting models. The benchmark model is frequently

[39] See also the discussion in section 2.1 in Diebold (2014) for more background on this in the context of the traditional Diebold-Mariano (DM) statistic.

assumed to be a RW model, which can be obtained from the DMA/DMS framework by setting (or restricting) $\boldsymbol{\beta}_t^{(m)} = 0$.

5.3. Forecast evaluation results

Table 2 presents the one-step-ahead forecast evaluation results for the out-of-sample period from May 2002 to June 2014. We use 70 observations from June 1996 to April 2002 as the in-sample fitting period for the OLS based benchmarks.[40] Following the large literature on out-of-sample forecast evaluation, we use the random walk (RW) model as the benchmark model in the statistical tests. The first column in Table 2 shows the models that are fitted, the second column shows the mean squared forecast errors (MSFEs), the third column the MSFEs relative to the RW benchmark, the fourth column the Campbell and Thompson (2008) R^2_{os} (in percent) as defined in (20), and the fifth and sixth columns display the Clark and West (2007) MSFE adjusted $t-$statistic (CW$-$statistic) and its corresponding one-sided $p-$value.

← Table 2 about here

We include a number of alternative models in our forecast evaluation exercise. These are the Historical Average (HA) and a standard OLS regression on the full $(1 \times K)$ dimensional predictor set $\mathbf{x}_t$ as defined in (16). The historical average forecast has recently gained popularity in the equity premium forecasting literature (see for instance Campbell and Thompson (2008), Rapach *et al.* (2013), Neely *et al.* (2014) and others). The historical average is computed simply by taking the arithmetic mean for data up to time t, and using that as the forecast for the next time period $t+1$. More intuitively, it can be easily implemented as an OLS regression of copper returns on an intercept term. For both, the OLS and HA based alternative models, we construct the forecasts with two different schemes: $i)$ a rolling window scheme and $ii)$ an expanding window scheme.[41] Apart from the DMA/DMS models with α and λ calibrated at 0.95

[40] We should note here also that since the Kalman filter recursions in (3) depend on the initial conditions (or priors), the in-sample fitting period can also be thought of as a burn-in period for the Kalman filter. After this initial period, the influence of the initial conditions becomes smaller as the recursions proceed.

[41] Under a rolling window scheme, we keep the estimation sample fixed and '*roll*' through the remaining out-of-sample data points to construct the forecast. The parameter estimates thus change at each point in time over the out-of-sample data, in the same way that a '*real time*' forecast would be constructed. That is, we use observations 1 to 70 for fitting, forecast observation 71, and then move forward to observations 2 to 71 (keeping the sample size

and 0.99 respectively, we also include forecasts from a simple time-varying parameter (TVP) model. Note here that the TVP model (with $\lambda = 0.99$) arises as a by-product of the DMA/DMS procedure, as it is readily computed and simply corresponds to the model which contains the full set of predictors.[42] Also, as there is no model averaging or selection, the value of the α parameter is irrelevant.

Looking over the results reported in Table 2, one can notice that both rolling and expending window historical averages perform poorly in forecasting monthly LME copper returns, in fact, much worse than a simple RW benchmark. It is further evident that the rolling window OLS regression on the full predictor set forecasts also rather poorly, noticeably worse than the two historical average forecasts. All three models produce negative R^2_{os} and rather small and insignificant CW−statistics. Nevertheless, all remaining four models produce statistically superior forecasts to the RW benchmark with sizeable R^2_{os} values, ranging from 9.6% for the TVP model to about 18.5% for the DMA model. The forecast performance of the DMS model is marginally worse than that of the DMA model in terms of MSFE, the R^2_{os}, as well as statistical significance. What is interesting to observe from the results that are reported in Table 2 is that even the simple OLS model using the full regressor set on an expanding window size does remarkably well in forecasting, producing an R^2_{os} of nearly 10%, and a CW−statistic of 1.96, which is significant at the 5% level. From these results we can thus see that the (rather large) set of predictor variables that we use contains valuable information for forecasting copper returns. Comparing these results to the ones from the TVP model seems to suggest that it is not solely the time-varying parameter part that leads to improved forecasts when using DMA/DMS, but a combination of the selected predictor set, time varying parameters, and most importantly,

fixed at 70), re-estimate the model parameters again and then forecast observation 72, etc. Under an expanding scheme, the first step is the same as under the rolling scheme, but all of the forecasts use the full in-sample data for fitting, ie., from observations 1 to 71 (increasing sample size of 71 now), re-estimating and then forecasting observation 72 and so forth. One can argue that both have their advantages and disadvantages. A rolling window scheme should adapt faster to parameter changes, should these occur, while the expanding window scheme would be preferred if there is no change in parameters because more precise estimates are obtained with a growing sample as opposed to a fixed one.

[42] Note that there is no model averaging or switching in the TVP model, since the model with the full set of predictors is used to forecast.

the averaging over the various models that are estimated. In summary, the statistical results presented here are overall positive and promising.

To facilitate our understanding of the statistical results that are summarised in Table 2, we show various plots of interest from the forecast evaluation exercise in Figure 3. In Panel (a) of Figure 3, we show a plot of the actual monthly copper returns (blue line) together with predicted values from DMA (red line) and expanding window OLS (green line), to get a visual impression of the overall out-of-sample fit of these two models. It is noticeable from the plots in Panel (a) that both models follow the actual copper return series rather well, with the two models producing broadly similar forecasts. There are a few important differences though. The DMA model seems to forecast somewhat worse at the end of 2006, and produces a marginally stronger crisis drop in copper returns towards the end of 2008. Nevertheless, DMA produces noticeably better forecasts than the expanding window OLS counterpart from early 2009 onwards for about 4 consecutive months, with the forecasts thereafter being again fairly similar. Looking at the squared forecast errors plotted in Panel (b) of Figure 3, one gets the same overall impression regarding the performance of these two models.[43]

← Figure 3 about here

In Panel (c) of Figure 3 we show plots of the time series evolution of the cumulative sum of the differences of the squared forecast errors of the DMA (red line) and expanding window OLS (green line) models, relative to those of the RW benchmark. This cumulative sum of differences is commonly used in the equity premium forecasting literature as a tool to highlight the predictive performance of the model relative to the benchmark over time (see Goyal and Welch, 2008). For DMA, with the benchmark being the RW model, this difference is computed as:[44]

$$\mathrm{CSD}_t^{(\mathrm{RW,DMA})} = \sum_{t=T_{is}}^{T_{os}} \left(\hat{e}_{t+1|t}^{2(\mathrm{RW})} - \hat{e}_{t+1|t}^{2(\mathrm{DMA})} \right). \tag{26}$$

A value above zero of $\mathrm{CSD}_t^{(\mathrm{RW,DMA})}$ indicates that the cumulative squared forecast errors of

[43]The plot in Panel (b) is fairly self explanatory. To conserve space, we do not discuss this plot any further and include it simply to provide additional information regarding the forecasting performance of these two models.
[44]Analogously for the expanding window OLS model.

the RW model are larger than those corresponding to the DMA model, suggesting that DMA produces better forecasts whenever $\mathrm{CSD}_t^{(\mathrm{RW,DMA})} > 0$. As is evident from Panel (c) of Figure 3, for both, DMA as well as expanding window OLS forecasts, $\mathrm{CSD}_t^{(\mathrm{RW},\bullet)}$ is uniformly greater than 0, apart from a short time period at the beginning of the out-of-sample evaluation period in 2003. What is interesting to observe is that the biggest boost in $\mathrm{CSD}_t^{(\mathrm{RW},\bullet)}$ is obtained during a short window around the peak of the financial crisis in September 2008. This suggests that, apart from the weaker predictability of LME copper returns during *normal times*, there was a period from September 2008 to the beginning of 2009 where substantial forecast gains could be realised. This is an interesting finding that has not been documented in the copper forecasting literature thus far.

In the last panel in Figure 3, we show a scatter plot of predicted copper returns from the DMA model on the $x-$axis against actual copper returns on the $y-$axis. Scatter plots of actual against predict values are informative, as they can shed light on the overall performance of the model. Since we can think of the out-of-sample R^2 as being the standard regression R^2 from an OLS fit of actual copper returns on predicted values from the DMA model, examining a scatter plot of predicted against actual can reveal possible abnormalities in this relationship such as outliers, which can influence the statistical results. A scatter plot can further reveal features that may be lost or hidden away in the R^2_{os} and/or CW$-$statistic.[45] To help us identify relational patterns in the data, we superimpose two simple model fits on the scatter in Panel (d). These are fits from i) a linear OLS regression of actual on predicted returns (solid blue line) and ii) a flexible non-parametric local linear regression (LOESS) of actual on predicted returns. The non-parametric LOESS fit is evaluated at the DMA predicted ($x-$axis) values (marked by red circles), as well as over an evenly spaced grid from min(DMA predicted) to max(DMA predicted) (drawn as a solid red line). Grey shading around the LOESS fit mark the

[45]The objective here is to learn about models and their fit to data in the spirit of Pagan (2002) and Breunig *et al.* (2003), which was recently also applied in Buncic (2012) in the context of exchange rate forecasting.

95% confidence interval computed from asymptotic standard errors.[46]

Looking over the plots in Panel (d) of Figure 3, one can immediately identify the general positive and mostly linear relationship between the predicted and actual return series. It is further noticeable that the linear OLS and the flexible non-parametric LOESS fits are reasonably similar over the range of the data where the bulk of the observations lie, ie., broadly the $[-5, 10]$ $x-$axis interval. There are evidently some observations on either end of the tails of the return distribution that are different from a linear fit, especially the handful of observations that are in the $[-15, -5]$ interval. Over this range, the LOESS estimates is rather flat, indicating no relationship between DMA predictions and actual returns, while the linear OLS fit is still upward sloping. Another feature that should be pointed out from the non-parametric fit is the very mild (inverted) $S-$shape in the predicted against actual relation over the $[-5, 10]$ interval. Nevertheless, one needs to be careful with the interpretation of this *visual* non-linearity in the fit, as the OLS regression line that crosses this interval stays mostly within the 95% confidence interval of the LOESS fit, indicating that the uncertainty surrounding the non-parametric point estimate is large enough to include the (linear) OLS regression line. This could be seen as *'insignificant'* or redundant non-linearity and therefore we do not discuss this any further.[47]

In summary, we can conclude from the visual inspection of the plots in Figure 3 that the DMA model performs reasonably well overall as a predictive model for copper returns. Interestingly, the plots of the cumulative difference of the squared forecast errors relative to the RW benchmark reveal that there is a boost in copper predictability after the Lehman Brothers collapse in September 2008, which is a time period of high economic and financial uncertainty. Over this period, a considerable gain in predictive ability over a simple RW model is realised.

[46]We used a simple Silverman (1986) plug in bandwidth and a first order polynomial in the LOESS regression (see Pagan and Ullah, 1999, page 104 for details).
[47]A visual indication of non-linearity is clearly not equivalent to a statistical test, however, the empirical finance literature has used visual evidence to argue for the need to fit non-linear models to the data. We do not pursue this issue any further as the objective here is purely to provide a visual overview of how the actual and fitted values relate to each other.

5.3.1. Which predictors perform best?

Given our flexible model specification with regards to the predictor variables that are included, as well as the time varying nature of the modelling setup that we implement, we can now look at which predictors are the most influential, and more importantly, how this influence evolves over time. Since we include various fundamental as well as financial predictor variables that we deem important for copper return forecasting, we expect to see changes, particularly during the peak of the financial crisis period from September 2008 to the beginning of 2009, where asset prices world wide went through a substantial re-valuation phase.

To obtain a visual impression of the time varying influence of our predictor variables on copper return forecasts, we examine the dynamic evolution of two quantities of interest. The first one is the posterior inclusion probability (PIP_t for short) of the predictor variable. The second one is the weighted average of the updated estimates of the latent state vector, which we denote by $\hat{\boldsymbol{\beta}}_{t|t}^{(\text{DMA})}$. The $(k \times 1)$ dimensional PIP vector at time t is computed as:[48]

$$\text{PIP}_t = \sum_{m=1}^{M} \pi_{t|t}^{(m)} \mathbb{1}\left(\mathbf{x}_t \in \mathcal{M} = m\right), \tag{27}$$

where $\pi_{t|t}^{(m)}$ is the updated model probability as defined in (10) and $\mathbb{1}\left(\mathbf{x}_t \in \mathcal{M} = m\right)$ is an indicator variable that is equal to 1 if any of the regressors in $\mathbf{x}_t$ are included in the m^{th} model. To construct the PIP, one thus simply sums over all updated model probabilities $\pi_{t|t}^{(m)}$ which contain the i^{th} predictor variable $\{x_{i,t}\}_{i=1}^{k}$. The weighted average of the time varying parameter estimates is obtained as:

$$\hat{\boldsymbol{\beta}}_{t|t}^{(\text{DMA})} = \sum_{m=1}^{M} \pi_{t|t}^{(m)} \hat{\boldsymbol{\beta}}_{t|t}^{(m)} \mathbb{1}\left(\mathbf{x}_t \in \mathcal{M} = m\right), \tag{28}$$

where $\hat{\boldsymbol{\beta}}_{t|t}^{(m)}$ is as defined in (3d) and $\mathbb{1}\left(\mathbf{x}_t \in \mathcal{M} = m\right)$ is as above. Intuitively, to calculate

[48] It is k here, rather $K = (k+1)$. This excludes the intercept term as it is by default included in all models so by construction has a PIP of 1.

$\hat{\boldsymbol{\beta}}_{t|t}^{(\mathrm{DMA})}$, we simply average each $\hat{\boldsymbol{\beta}}_{t|t}^{(m)}$ estimate over the updated model probabilities $\pi_{t|t}^{(m)}$, at each point in time for the regressors that are included in the model.

In Figures 4 and 5 we show plots of the time series evolution of PIP_t and $\hat{\boldsymbol{\beta}}_{t|t}^{(\mathrm{DMA})}$ over the out-of-sample period from May 2002 to June 2014. Let us first discuss the plots of the PIPs shown in Figure 4. Looking over the bottom row of plots, it is evident that three of the four resource based equity return series have very low probabilities of being included in the forecasting models. There is some mild variation over time, but most of the PIPs remain well below 20%. The influence of the Spread, the VIX, the TED spread, and Gold prices shown in the middle of Figure 4 have fairly stable PIPs over time (at least since March 2006), indicating that, on average, these variables are included about 50% of the time in the forecasting models. This stability is particularly interesting to see for the VIX and the TED spread, as our prior expectation would have been an increasing influence with an overall jump in risk aversion during the financial crisis period from September 2008 onwards, due to these two variables being market proxies for global risk appetite. ← Figure 4 about here

The SP500, Oil prices, the Chilean Peso, Australian Dollar and the stock price of Alcoa are also overall fairly stable (broadly in the 50% PIP region), nevertheless, with some adjustments taking place in March 2006 and May 2009. In March 2006, the PIP of ΔOil changed from around 20% to nearly 80% by June 2006 and declined gradually thereafter, stabilising at 40%. The importance of the SP500 jumped from around 30% to just under 60% in March 2006, while the Australian Dollar and Alcoa equity returns experienced a marginal increase in their inclusion probabilities from March 2006 until the end of 2008, remaining fairly stable at around 45% over the rest of the sample.

The biggest PIP changes are visible for lagged copper returns, the convenience yield, industrial production, and to a lesser extent, the Baltic dry index. The variability in the PIP of the lagged copper return series is particularly pronounced, with the posterior inclusion probability changing from rather high values of over 80% at the beginning of the evaluation period, to low values of around 40% in March 2006, increasing back to 80% by September 2008 and then dropping sharply to around $20 - 30\%$ and remaining at this level until May 2009. The importance of lagged copper returns has increased rapidly again since May 2009, reaching levels well above 80% by mid 2010 and staying at this level until the end of the evaluation period. What is interesting to see from the PIP of the lagged copper return series (and also from the $\hat{\boldsymbol{\beta}}_{t|t}^{(\mathrm{DMA})}$ plots in Figure 5) is the varying degree of momentum that the series exhibits, particularly over the January 2006 to January 2009 period, and also from January 2009 to February 2011 (see also Panel (a) of Figure 1 for the build up, drop and built up cycle of LME copper prices).

The inclusion probability of the convenience yield has been reasonably stable since 2006 at around $50 - 60\%$, but then spiked sharply down and back up again in September 2008, indicating that the predictive content of the convenience yield was much lower during this period of high uncertainty. A broadly similar picture is visible for the PIP of industrial production, which was also fairly constant at a level of $40 - 60\%$ from the beginning of the out-of-sample period in May 2002 until September 2008. In September 2008, the PIP then dropped from a level of around 40% to nearly zero, moving gradually back towards a level of 40% by June 2014. The sudden and sizable drops in the PIPs of these two variables give an indication of how normally informative '*fundamental*' variables were irrelevant for the determination of copper prices over this time period of heightened economic and financial uncertainty. ← Figure 5 about here

Given our knowledge of how the inclusion probabilities of the different predictor variables have changed over time, we can now examine the evolution of the $\hat{\boldsymbol{\beta}}_{t|t}^{(\mathrm{DMA})}$ coefficients shown in Figure 5 to complement the PIP results. Due to the very low inclusion probabilities of the 3 resource based equity return series shown in the last row of Figure 5, we do not discuss their time varying $\hat{\boldsymbol{\beta}}_{t|t}^{(\mathrm{DMA})}$ plots. Looking over the time variation of the model averaged coefficients for the Chilean peso, the Australian dollar, Alcoa stock, Oil and Gold we can identify a number of interesting features. Initially, we can notice the fairly stable and positive coefficient on the Chilean peso from April 2004 until the end of the sample. The coefficient was close to zero from

May 2002 until January 2004, but then increased fairly rapidly to a value of around 0.20 by July 2004 and staying at that level. In September 2008, the value of the coefficient increased further to about 0.30 until the middle of 2010, suggesting an heightened influence of the Chilean peso on copper returns over the crisis period.[49]

The coefficients on the Australian dollar, Alcoa stock , Oil and Gold were small or close to zero for most of the out-of-sample period, with important exceptions being the two time periods March 2006 and September 2008. All four predictor variables had positive and increasing coefficients over these two periods suggesting a positive relationship with copper returns. For instance, Oil, Alcoa stock and also Gold went through a similar initial rapid increase and then subsequent fall over the March 2006 to December 2006 period. This is reflected in the magnitudes of the coefficients over this time frame. The Australian dollar grew in importance over the resource based run up until mid 2008. What is different for the Australian dollar is the impact over the September 2008 to May 2009 period. While the coefficients on Oil, Alcoa stock and Gold jumped up, the one on the Australian dollar dropped down towards zero. Although all four predictor variables were affected by the September 2008 period, the intensity and magnitude of the responses varied. Alcoa's equity price was most heavily affected by this period, with its stock dropping from approximately 38USD just before September 2008 to about 6USD by February 2009. Oil prices tumbled from 140USD a barrel to under 40USD.[50] Over the same time frame, the Australian dollar was heading towards parity with the USD before September 2008, but then depreciated sharply to 0.65USD. Gold prices also dropped in response to the Lehman Brothers collapse, but the drop was rather small from around 1000USD to 800USD (around 20%) when compared to the magnitudes of decline in the other three predictors, and also when measured against the sharp increase in Gold prices that followed.

[49] A coefficient value of 0.3 translates into a copper return increase of 0.3% for every 1% increase in ΔCLP, provided all else is the same.

[50] Oil had an immense run up before September 2008 and had peaked in fact in July 2008, then dropped down from 143USD to around 91USD, increased once more to 122USD a barrel, before collapsing to 33USD in December 2008.

Looking over the evolution of the coefficients on the two risk aversion proxies, the VIX and the TED spread, we can clearly identify the impact of the September 2008 crisis period. What is interesting to observe here is that the coefficient on the TED spread was positive, suggesting that a rise in the TED spread projected an increase in copper returns, *ceteris paribus*, for the first half of the out-of-sample period, that is, before September 2008. There is some variation in the magnitude before the financial crisis, but the coefficient stayed positive until September 2008. Thus, during non-crisis (or normal) times, an increase in the TED spread gave a substantially different prediction signal for copper. Following September 2008, the coefficient changed abruptly to a negative value, dropping as low as -0.25 by the end of 2008. Although the TED spread had stabilised fairly quickly at a value of around 0.5, which is below its long term average of 0.61 after September 2009 (see the plot of the predictors in Figure 2), the coefficient value on the TED spread stayed negative at about -0.18. We can hence see that even months after the September 2008 collapse of Lehman Brothers, marginal increases in the TED spread had a negative impact on copper prices, which is in contrast with the first half of the out-of-sample period.

Such a change in sign of the coefficient did not occur for the VIX, which remained negative for the entire out-of-sample period, with values ranging from -0.01 to 0, being closer to 0 from March 2006 to December 2007.[51] In September 2008, there was a substantial shift down in the coefficient, dropping as low as -0.02 and stabilising at around -0.01 by January 2012. Despite being negative over the whole forecast evaluation period, the sensitivity of copper to changes in the VIX more than doubled from the time before the Lehman collapse until May 2009, suggesting that market participants were extremely responsive to risk aversion. Looking over the time series plot of the coefficient on the SP500 one can see two interesting features. The first is the obvious jump during the September 2008 to May 2009 period. The second is the steady positive trend in the coefficient, particularly from June 2007 until the end of the sample

[51]Note here the that coefficient looks rather small when compared to the TED spread one, but it needs to be kept in mind that scale of the VIX predictor variable is 10 times that of the TED spread.

period. Both of these results are consistent with the recent finding by Rapach *et al.* (2013), that the U.S. is an information center when it comes to news related to global economic and financial activity. The persistence in the upward trend after September 2008 is particularly interesting, as it suggests that the U.S. played the role of an information hub not only during the peak of the financial crisis period, but also afterwards.[52]

The three predictor variables that are meant to capture economic activity, namely, industrial production, the Baltic dry index and the term spread also show a number of interesting features with regards to the time series evolution of their coefficients. Industrial production, for instance, had a fairly large coefficient over the first half of the evaluation period, varying broadly between 0.8 and 1.3, suggesting that changes in industrial production (our proxy for U.S. economic activity) had a positive impact on copper prices. This changed dramatically in September 2008, with the coefficient on ΔIP dropping to less than 0.05 by December 2008. So not only did the PIP of ΔIP go towards zero (as discussed before and shown in Figure 4), but also the magnitude of its coefficient, indicating that the usual predictive content of industrial production for copper had disappeared after the Lehman Brothers collapse. The coefficient did recover gradually after May 2009, which is in line with the findings for the PIP values, but remained subdued at values well below 0.5.

The coefficients on ΔBDI and Spread jumped in September 2008. For the Spread variable, the coefficient was close to zero for a few years before the crisis, but then dropped down to -0.3 by November 2008, before gradually increasing back to -0.1. Note here the interpretation. The yield curve is said to be inverted when the spread is negative. An inverted yield curve is predictive for U.S. recessions. A decreasing value thus signals that markets are anticipating a worsening economic outlook. As is visible from the plot of the term spread in Figure 2, the yield curve was inverted for two episodes: from July 2000 to January 2001 and from July 2006 to May

[52]It might seem obvious to think that the U.S. was the source of all news related to, and around, the Lehman Brothers collapse due to the news actually originating there. Nevertheless, it is astounding to see that this importance, as captured by the coefficient size, did not decrease after news related to the financial crisis diminished.

2007. Although the term spread was positive from May 2007 onwards and thus not suggestive of recessionary expectations building up, there were substantial jumps up and down in the term spread around the time of the Lehman Brothers collapse in September 2008, due to the increased economic and financial uncertainty. For example, in August 2008, the term spread was 2.14%, then jumped to 3.30% by November 2008, moved down to 2.35% by January 2009, before increasing fairly steadily afterwards until May 2010. This period of uncertainty was accompanied by a sharp drop in the coefficient on the spread.

The spread had rather negligible predictive power in the two years preceding the Lehman Brothers collapse, as is evident from its coefficient being close to zero in June 2008. Nevertheless, by November 2008, the coefficient had dropped to -0.29, highlighting the increased predictive power over this time frame. What is interesting to note here is the reversal in the sign of the coefficient. For the first half of the out-of-sample period, the spread coefficient was small but mainly positive (only turning negative in March 2009), so that an increasing yield spread predicted positive copper returns. After September 2008, this relationship reversed so that increases in the yield spread were predicting negative returns. Changes in the Baltic dry shipping index were largely of less importance as is visible from the close to zero coefficient value for a substantial part of the out-of-sample period. Nevertheless, noticeable increases in the coefficient to values just below 0.1 are evident for the March 2006 and September 2008 periods, emphasising the increased predictive content of the BDI for copper over these two particular periods.

The three predictors related to copper fundamentals, that is, (excess) demand, inventory and the convenience yield also portray considerable time variation in their coefficients. Changes in inventories had initially a fairly constant negative impact on copper returns, but then dropped noticeably lower in September 2008 (and to a lesser extent in March 2006) before recovering to a value of -0.1 by the end of the sample period. The coefficient on ΔDemand jumped during the September 2008 period from around 0.04 to a value of 0.10, highlighting the increased sensitivity to expectations about demand pressures over this time period. After September 2008, the coefficient adjusted again to a considerably smaller value, dropping below 0.02 by the end of the sample. The coefficient on the convenience yield seems to be the most volatile of the three fundamental predictors, starting off at a value close to -0.4, increasing steadily to -0.1 until March 2006, then dropping initially rapidly and finally more slowly towards the September 2008 period. In September 2008, the coefficient on the convenience yield jumped from -0.27 to -0.15, and then dropped as low as -0.4 by May 2009, indicative of substantial changes in predictive power over this time frame. Since then, the coefficient has been increasing towards 0, suggesting a more subdued influence on copper returns.

The variation in the coefficient on lagged copper returns shows similar patterns to the ones seen in the posterior inclusion probabilities that are plotted in Figure 4. Initially, there was a strong build up in momentum for copper, with the coefficient on lagged returns increasing from just below 0.4 to 0.6 over the period from May 2002 to March 2006. This strong momentum phase is also evident in LME spot prices plotted in Panel (a) of Figure 1. After March 2006, a drop in momentum is visible, with the coefficient on lagged copper returns falling to 0.22 by August 2006, but then increasing steadily towards 0.45 by May 2008. The final drop in momentum occurred in two phases after September 2008, with a noticeable rebound phase from November 2008 to January 2009, bottoming out in February 2009 at a coefficient value of 0.12. Inline with the strong increases in copper prices that are visible in Panel (a) of Figure 1, momentum began once again to increase steadily until the end of the sample in June 2014.

In summary, the results presented in this section show that there is considerable time variation not only in the set of predictor variables that are most informative for forecasting copper returns, but also in the magnitude of the model averaged coefficients attached to the predictors. Moreover, the effect of the Lehman Brothers collapse in September 2008 is visible in nearly all the predictor variables through jumps and spikes in the inclusion probabilities as well as in the magnitudes of the coefficients. This highlights the fast changing nature of the prediction environment and information flow during that time. The SP500, the VIX, the yield spread, the TED spread and the convenience yield predictors were particulary heavily influenced during the September 2008 to May 2009 period, re-emphasising the need for a flexible forecasting model.

5.4. Forecasting performance at longer horizons

Given the overall positive results at the one month ahead horizon, we now turn to assess the forecasting performance of our model and predictor set at horizons greater than one month. It is well known from the literature on exchange rate forecasting (see, for instance, Mark, 1995) that some predictor variables, especially fundamentals, can contain more predictive information when forecasting multiple periods into the future, as the influence of *'noise traders'* can drive the evolution of the return process at shorter horizon.[53] To assess the multiple steps ahead forecast performance of our framework, we construct and evaluate copper return forecasts at horizons of $2, 3, 6, 9$, and 12 months into the future.

5.4.1. Computing multiple period ahead out-of-sample forecasts

We implement the so-called *'direct'* forecasting approach to construct multiple-step-ahead out-of-sample forecasts form the DMA/DMS framework.[54] That is, we re-formulate the relation in (1) (again using the general y_t and $\mathbf{x}_t$ notation as in Section 3.1) as

$$y_t = \mathbf{x}_{t-h}^{(m)}\boldsymbol{\beta}_{t,h}^{(m)} + u_t^{(m)} \tag{29a}$$

$$\boldsymbol{\beta}_{t,h}^{(m)} = \boldsymbol{\beta}_{t-1,h}^{(m)} + \boldsymbol{\epsilon}_t^{(m)}, \tag{29b}$$

[53]The influence of noise traders or chartists, that is, traders that use simple technical indicators as trading signals is well known in the exchange rate and equity premium forecasting literature (see, among many other studies, Allen and Taylor, 1990 and Neely *et al.*, 2014 for two widely cited studies in the exchange rate and equity forecasting literature).

[54]See Clements and Hendry (1996), Chevillon and Hendry (2005), Marcellino *et al.* (2006), Chevillon (2007), and Pesaran *et al.* (2011), among others, for a motivation, evaluation and comparison of the direct forecasting approach to iterated forecasts.

where the h subscript in $\boldsymbol{\beta}_{t,h}^{(m)}$ signifies the relation to the $h-$period lagged value of $\mathbf{x}_t$. Using the same Kalman Filter recursions as in (3), but now on the $h-$period lagged relation as specified in (29) yields filtered estimates of the latent states (for each model), that is, $\hat{\boldsymbol{\beta}}_{t|t,h}^{(m)}$. Analogues to the forecasting equations in (15), the $h-$period copper return forecast relation becomes:

$$r_{t+h}^{h} = \mathbf{x}_t^{(m)} \boldsymbol{\beta}_{t+h,h}^{(m)} + u_{t+h}^{(m)} \tag{30a}$$

$$\boldsymbol{\beta}_{t+h,h}^{(m)} = \boldsymbol{\beta}_{t+h-1,h}^{(m)} + \boldsymbol{\epsilon}_{t+h}^{(m)} \tag{30b}$$

where $r_{t+h}^{h} = 100(S_{t+h}/S_t - 1)$, that is, the $h-$period holding return from buying LME spot copper at time t and selling it at time $t+h$.

Given (30), the DMA/DMS based $h-$step ahead forecasts are then computed as:

$$\hat{r}_{t+h|t}^{h(\mathrm{DMA})} = \sum_{m=1}^{M} \mathbf{x}_t^{(m)} \hat{\boldsymbol{\beta}}_{t|t,h}^{(m)} \pi_{t+h|t}^{(m)} \tag{31a}$$

$$\hat{r}_{t+h|t}^{h(\mathrm{DMS})} = \sum_{m=1}^{M} \mathbf{x}_t^{(m)} \hat{\boldsymbol{\beta}}_{t|t,h}^{(m)} \mathbb{1}\left(\pi_{t+h|t}^{*} = \pi_{t+h|t}^{(m)}\right) \tag{31b}$$

for all $t = T_{is}, \ldots, T$, where, again due to the random walk evolution of the latent state vector $\boldsymbol{\beta}_{t+h,h}^{(m)}$, the best forecast of $\boldsymbol{\beta}_{t+h,h}^{(m)}$ is its last observed filtered estimate, that is, $\mathbb{E}_t(\boldsymbol{\beta}_{t+h,h}^{(m)}) = \hat{\boldsymbol{\beta}}_{t|t,h}^{(m)}$. As defined earlier, the term $\mathbb{1}\left(\pi_{t+h|t}^{*} = \pi_{t+h|t}^{(m)}\right)$ is an indicator that is equal to one for the model with the highest predictive probability and $\pi_{t+h|t}^{*} = \max\{\pi_{t+h|t}^{(m)}\}_{m=1}^{M}$. The $h-$step ahead predictive model probabilities at time t are computed from

$$\pi_{t+h|t}^{(m)} = \frac{\pi_{t|t}^{\alpha(m)}}{\sum_{j=1}^{M} \pi_{t|t}^{(j)}}, \tag{32}$$

where $\pi_{t|t}^{(m)}$ is as defined in (9b).

5.4.2. Multiple-step-ahead forecast evaluation

We use the same calibration for the λ, α and κ parameters that were used in the one-step-ahead forecast setting to implement the Kalman Filter recursions. The results for forecast horizons $h = 2, 3, 6, 9$, and 12 months are reported in Table 3. The content of the columns of Table 3 is the same as described earlier in the one-step-ahead forecast evaluation in Table 2, with the only exception being that the second column now reports MSFEs deflated by the forecast horizon h. Also, since $h-$step ahead forecast errors will be MA($h-1$) processes in general (ie., will be autocorrelated of order $h-1$) which affects also the CW$-$statistic, we use a HAC robust variance for $\text{Var}(\overline{cw})$ in (25). More specifically, we follow the recommendation of Andrews and Monahan (1992) and employ a data driven bandwidth using a Quadratic Spectral (QS) Kernel with a *'pre-whitening'* step, where we set the (optimal) bandwidth parameter with an AR(1) as the approximating model (see equation 3.5 in Andrews and Monahan (1992)).[55] ← Table 3 about here

From the results reported in Table 3 it is evident that the forecasting performance of the DMA model (relative to a simple random walk forecast) remains in tact for forecast horizons up to 6 months ahead. The overall predictive ability diminished fairly consistently as the forecast horizon increases, with the out-of-sample R^2 dropping from close to 8% at forecast horizon 2, to 6.8% at $h = 3$, to 5.2% at $h = 6$ months ahead, yielding corresponding CW$-$statistics that are significant at the 5% for $h = 2, 3$ and 10% level for $h = 6$. What is interesting to see from the results reported in Table 3 is that the performance of DMS, TVP and OLS on the full regressor set (using an expanding window) perform noticeably worse than DMA. Also, the performance of the simple historic average on an expanding window seems to keep its overall predictive performance, producing marginally lower out-of-sample MSFEs than the random walk benchmark.

[55] That is, to pre-whiten the cw_{t+1} series, we first fit an ARMA$(1,1)$ to cw_{t+1} and then use the QS Kernel with the bandwidth parameter set to $1.3221\,(\hat{\alpha}(2)T_{os})^{1/5}$, where $\hat{\alpha}\,(2) = 4\hat{\rho}^2/(1-\hat{\rho})^4$ and $\hat{\rho}$ is the AR(1) parameter estimate obtained from an AR(1) regression of the (pre-whitened) residual series obtained from the ARMA$(1,1)$ model fitted to cw_{t+1}. To obtained the HAC variance, we then *'re-colour'* again with the ratio of the square of the ARMA lag polynomials (see Andrews and Monahan, 1992 for more details on the exact computations).

To visualise how the actual long-horizon returns as well as the predicted values from the DMA model compare, we plot again the same contents that were shown in Figure 3 in Figures 6 to 10. The main aim here is once again to learn about the forecasts from the different models. From Figures 6 and 7 we can broadly identify the same pattern in the plots that we found in the corresponding plots shown in Figure 3. These are the reasonably close performance of DMA and OLS on an expanding window up to the September 2008 crisis period. What is different at the 2 and 3 step ahead horizons is that the expanding window OLS regression does not produce the rebound in copper returns following the January 2009 period. The DMA model is much better able to adapt to the strong appreciation in copper prices that followed the end of 2008 collapse. From about January 2010, the forecasts from the two models are quite similar again, producing forecasts that are visually not different to a random walk model. The reasonably good out-of-sample performance of DMA at horizons 2 and 3 is also mirrored in the predicted against actual scatters shown in Panel (d) of the two figures. The predicted against actual relationship seems again fairly linear with little influence from aberrant observations.

← Fig.6-10 about here

Looking over the plots for the 6 months forecast horizon shown in Figure 8 it is interesting to see that OLS on an expanding window performed uniformly better than DMA (and also the RW benchmark) for roughly one half of the out-of-sample period, as is evident from the positive cumulative sum of squared forecast errors shown in Panel (c). From May 2009, the performance dropped of substantially and rapidly, again due to (expanding window) OLS not being able to generate the rebound after the September 2008 slump. The DMA model gets most of its gain in predictability over the time frame from April 2009 to February 2010. From around February 2010, we can see from Panel (a) of Figure 8 that the forecasts from DMA and OLS were remarkably similar again. The weaker predictive power for 6 months ahead forecasts that is evident from the statistical results reported in Table 3 is also visible from the predicted vs. actual scatter shown in Panel (d) of Figure 8. Although the (linear) OLS fit as marked by the blue line in the plot indicates an overall positive relationship between predicted and actual values, the non-parametric fit shows that this positive relation holds only mildly over the $[-10, 10]$ $x-$axis range, while it is flat over (nearly) the entire remaining $x-$axis range.[56]

The plots summarising the performance of 9 and 12 months ahead forecasts shown in Figure 9 and Figure 10, respectively, are fairly similar in terms of highlighting the breakdown of the expanding OLS based forecasts. From the actual vs predicted plots in Panel (a) of the figures it is evident that the models performed fairly similarly, producing rather *flat* forecasts. Nevertheless, conditioning on the values of the predictor variables at their September 2008 crisis period values produces extremely poor and exaggerated negative 9 and 12 months ahead forecasts from expanding window OLS. All the parameter estimates that are obtained during *'normal times'* are now used to forecast with the same sensitivities for predictor variables that have a fundamentally different relation to copper, producing a strong and wrongly signed forecast. The DMA framework seems to be able to better adjust to these differences, forecasting at least broadly in the same direction as the target variable. The two plots showing predicted against actual values in Panel (d) of Figure 9 and Figure 10 signify the poor forecasts.

Overall, we can conclude here that forecasting copper returns over a holding period of longer than 6 months using the DMA framework is no better than a random walk forecast, with forecasts from an OLS based model on the full predictor set being substantially worse. Also, given the long-horizon forecast results, the simple historical average computed over an expanding window could be a viable alternative forecasting model to use when an alternative to the random walk forecast is desirable.

6. Conclusion

We use the recently proposed dynamic model averaging and selection (DMA/DMS) framework to forecast monthly LME copper returns using a purposefully selected large set of 18 predictor variables. The DMA/DMS modelling framework is particularly appealing for forecasting cop-

[56]The exception being the few points in the right hand side of the scatter, ie., for $x-$axis values greater than 30.

per, as it combines a time varying parameter and model selection approach under one unified setting. Letting not only the model parameters but also the set of predictors change over time is important in the given context, as the role of copper has evolved over time from being a *'simple commodity'* that is used as a primary input in the production process of final goods, to a financial asset that is held and traded for speculative purposes. This changing role of copper makes it necessary to condition on a large and diverse set of predictor variables that control not only for standard demand, inventory and convenience yield factors, but also for the impact that financialization has had on copper, since the importance of these particular factors has shifted over the last few decades.

Covering an out-of-sample period from May 2002 to June 2014, we show in our empirical forecast evaluation exercise that the DMA/DMS modelling framework significantly outperforms the random walk benchmark for predicting monthly copper returns. The out-of-sample R^2 can be as high as 18.5% for the DMA/DMS framework and is nearly 10% for even a much simpler (expanding window) OLS regression model on the full predictor set. Moreover, these forecast gains are statistically significant. What these results highlight is that the attained forecast gains are not purely due to the flexibility of the DMA/DMS specification that we implement. A substantial part is also attributable to the set of highly relevant predictor variables that were specifically selected for copper forecasting. Using visualisation techniques to learn about the fit of the models to the data, our results highlight further that the biggest improvement in the forecast performance over the benchmark model is realised over the September 2008 to beginning of 2009 period, which seems to have been the worst phase of the 2008 financial crisis. During this time period, copper return predictability is larger than during any other time period in our sample. This finding has not been documented so far in the commodity forecasting literature.

From plots of the time varying parameter estimates as well as the posterior inclusion probabilities (PIPs) we show further that the influence of the most relevant predictor variables for copper has changed over time, and substantially so over the period following the Lehman Brothers collapse in September 2008. The coefficients and/or the PIPs of nearly all predictor variables show either a sizeable jump or a drop around the time of the Lehman collapse. The coefficients of the SP500, the VIX, the yield spread, the TED spread, industrial production and the convenience yield predictors were most heavily affected, with the TED spread and yield spread coefficients even changing signs over this period. The time series plots highlight further that momentum in copper has varied considerably and abruptly, most visibly around March 2006 and September 2008.

Results from our multiple-step-ahead forecast evaluation indicate that copper predictability holds for horizons up to 6 months into the future. Predictability is strongest at the one month forecast horizon and decreases as the horizon increases. At $2, 3$ and 6 months ahead horizons, only forecasts from DMA produce statistically superior forecasts over a random walk benchmark. The performance of the expanding window OLS regression model on the full predictor set deteriorates quickly, producing substantially worse forecasts than a random walk model. DMS and a simple time varying parameter (TVP) model without predictor selection perform also notably worse than the random walk benchmark.

References

Aizenman, Joshua, Sebastian Edwards and Daniel Riera-Crichton (2012): "Adjustment patterns to commodity terms of trade shocks: the role of exchange rate and international reserves policies," *Journal of International Money and Finance*, **31**(8), 1990–2016.

Allen, Helen and Mark P. Taylor (1990): "Charts, Noise and Fundamentals in the Foreign Exchange Market," *The Economic Journal*, **100**(1), 4959.

Andrews, Donald W. K. and J. Christopher Monahan (1992): "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," *Econometrica*, **60**(4), 953–966.

Ang, Andrew and Monika Piazzesi (2003): "A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables," *Journal of Monetary Economics*, **50**(4), 745–787.

Aruga, Kentaka and Shunsuke Managi (2011): "Price Linkages in the Copper Futures, Primary, and Scrap Markets," *Resources, Conservation and Recycling*, **56**, 43–47.

Bates, John M. and Clive W. J. Granger (1969): "The Combination of Forecasts," *Operations Research Quarterly*, **20**(4), 451–468.

Bhar, Ramprasad and Girijasankar Mallik (2013): "Inflation uncertainty, growth uncertainty, oil prices, and output growth in the UK," *Empirical Economics*, **45**(3), 1333–1350.

Brand, Claus, Daniel Buncic and Jarkko Turunen (2010): "The Impact of ECB Monetary Policy Decisions and Communication on the Yield Curve," *Journal of the European Economic Association*, **8**(6), 1266–1298.

Brennan, Michael J. (1958): "The Supply of Storage," *The American Economic Review*, **48**(1), 50–72.

Breunig, Robert V., Serineh Najarian and Adrian R. Pagan (2003): "Specification Testing of Markov Switching Models," *Oxford Bulletin of Economics and Statistics*, **65**(S1), 703–725.

Brock, William A. and Cars H. Hommes (1997): "A Rational Route to Randomness," *Econometrica*, **65**(5), 1059–1095.

——— (1998): "Heterogeneous Beliefs and Routes to Chaos in a Simple Asset Pricing Model," *Journal of Economic Dynamics and Control*, **22**(8-9), 1235–1274.

Brunnermeier, Markus K., Stefan Nagel and Lasse H. Pedersen (2009): "Carry Trades and Currency Crashes," in *NBER Macroeconomics Annual 2008*, edited by Kenneth Rogoff Daron Acemoglu and Michael Woodford, University of Chicago Press, Volume 23, 313–347s.

Buncic, Daniel (2012): "Understanding forecast failure of ESTAR models of real exchange rates," *Empirical Economics*, **34**(1), 399–426.

Buncic, Daniel and Martin Melecky (2014): "Equilibrium credit: The reference point for macroprudential supervisors," *Journal of Banking and Finance*, **41**(April), 135–154.

Campbell, John Y. and Samuel B. Thompson (2008): "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?" *Review of Financial Studies*, **21**(4), 1509–1531.

Cashin, Paul, Luis Felipe Céspedes and Ratna Sahay (2004): "Commodity Currencies And The Real Exchange Rate," *Journal of Development Economics*, **75**, 239–268.

Chan, Kalok, Yiuman Tse and Michael Williams (2011): "The Relationship between Commodity Prices and Currency Exchange Rates: Evidence from the Futures Markets," in *Commodity Prices and Markets, NBER – East Asia Seminar on Economics*, edited by Takatoshi Ito and Andrew K. Rose, University of Chicago Press, Volume 20, 47–71.

Chen, Yu, Kenneth Rogoff and Barbara Rossi (2010): "Can exchange rates forecast commodity prices?" *Quarterly Journal of Economics*, **125**(3), 1145–1194.

——— (2012): "Predicting Agri-Commodity Prices: An Asset Pricing Approach," in *Global Uncertainty and the Volatility of Agricultural Commodities Prices*, edited by Bertrand R. Munier, IOS Press, 45–71.

Chen, Yu and Wen-Jen Tsay (2011): "Forecasting Commodity Prices with Mixed-Frequency Data: An OLS-based generalized ADL Approach," *SSRN Working Paper*. Available from: http://ssrn.com/abstract=1782214.

Chevillon, Guillaume (2007): "Direct multi-step Estimation and Forecasting," *Journal of Economic Surveys*, **21**(4), 746–785.

Chevillon, Guillaume and David F. Hendry (2005): "Non-parametric direct multi-step estimation for forecasting economic processes," *International Journal of Forecasting*, **21**(2), 201–218.

Clark, Todd E. and Kenneth D. West (2007): "Approximately normal tests for equal predictive accuracy in nested models," *Journal of Econometrics*, **138**(1), 291–311.

Claro, Sebastian and Claudio Soto (2013): *"Exchange rate policy and exchange rate interventions: the Chilean experience,"* Bank for International Settlement. Available from: www.bis.org/publ/bppdf/bispap73g.pdf.

Clements, Michael P. and David F. Hendry (1996): "Multi-step Estimation for Forecasting," *Oxford Bulletin of Economics and Statistics*, **58**(4), 657–84.

Cochran, Steven J., Iqbal Mansur and Babatunde Odusami (2012): "Volatility persistence in metal returns: A FIGARCH approach," *Journal of Economics and Business*, **64**, 287–305.

Cogley, Timothy and Thomas J. Sargent (2005): "The conquest of US inflation: Learning and robustness to model uncertainty," *Review of Economic Dynamics*, **8**(2), 528–563.

Collin-Dufresne, Pierre, Robert S. Goldstein and J. Spencer Martin (2001): "The Determinants of Credit Spread Changes," *The Journal of Finance*, **56**(6), 2177–2207.

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

Deaton, Angus and Guy Laroque (1992): "On the behaviour of commodity prices," *The Review of Economic Studies*, **59**(1), 1–23.

——— (1996): "Competitive Storage and Commodity Price Dynamics," *Journal of Political Economy*, **104**(5), 896–923.

Diebold, Francis X. (2014): "Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests," *Journal of Business and Economic Statistics*, **Forthcoming**.

Diebold, Francis X. and Roberto S. Mariano (1995): "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, **13**(1), 253–263.

Dwyer, Alexandra, George Gardner and Thomas Williams (2011): "Global Commodity Markets - Price Volatility and Financialisation," *RBA Bulletin*, **June Quarter**, 49–58.

Eklund, Jana and Sune Karlsson (2007): "Forecast Combination and Model Averaging Using Predictive Measures," *Econometric Reviews*, **26**(2-4), 329–363.

Estrella, Arturo and Gikas A. Hardouvelis (1991): "The Term Structure as a Predictor of Real Economic Activity," *The Journal of Finance*, **46**(2), 555–576.

Fama, Eugene F. and Kenneth R. French (1987): "Commodity Futures Prices: Some Evidence on Forecast Power, Premiums, and the Theory of Storage," *The Journal of Business*, **60**(1), 55–73.

——— (1988): "Business Cycles and the Behavior of Metals Prices," *The Journal of Finance*, **43**(5), 1075–1093.

Fernández, Carmen, Eduardo Ley and Mark F. J. Steel (2001a): "Benchmark Priors For Bayesian Model Averaging," *Journal of Econometrics*, **100**(2), 381–427.

——— (2001b): "Model uncertainty in cross-country growth regressions," *Journal of Applied Econometrics*, **16**(5), 563–576.

Fisher, Franklin M., Paul H. Cootner and Martin Neil Baily (1972): "An Econometric Model of the World Copper Industry," *The Bell Journal of Economics and Management Science*, **3**(2), 568–609.

Gargano, Antonio and Allan Timmermann (2014): "Forecasting commodity price indexes using macroeconomic and financial predictors," *International Journal of Forecasting*, **30**(3), 825–843.

Gilbert, Christopher Leslie (2010): "Speculative influences on commodity futures prices 2006-08," UNCTAD, no. 197 in Discussion papers.

Goyal, Amit and Ivo Welch (2008): "A Comprehensive Look at The Empirical Performance of Equity Premium Prediction," *Review of Financial Studies*, **21**(4), 1455–1508.

Gregorio, José De and Felipe Labbé (2011): "Copper, the real exchange rate and macroeconomic fluctuations in Chile," *Working Papers No. 640*, Central Bank of Chile.

Groen, Jan J. J. and Paolo A. Pesenti (2011): "Commodity Prices, Commodity Currencies, and Global Economic Developments," in *Commodity Prices and Markets, NBER – East Asia Seminar on Economics*, edited by Takatoshi Ito and Andrew K. Rose, University of Chicago Press, Volume 20, 15–42.

Hamilton, James D. (2009): "Causes and Consequences of the Oil Shock of 2007-08," *Brookings Papers on Economic Activity*, **40**(1), 215–283.

Hansen, Peter Reinhard and Asger Lunde (2001): *"A Forecast Comparison of Volatility Models: Does Anything Beat a Garch(1,1)?"* Working Paper.

He, Yanan, Shouyang Wang and Kin Keung Lai (2010): "Global economic activity and crude oil prices: A cointegration analysis," *Energy Economics*, **32**(4), 868–876.

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

Hua, Renhai, Bing Lu and Baizhu Chen (2009): "Price Discovery Process in the Copper Markets: Is Shanghai Futures Market Relevant?" .

Issler, João Victor, Claudia Rodrigues and Rafael Burjack (2014): "Using common features to understand the behavior of metal-commodity prices and Forecast them at different horizons," *Journal of International Money and Finance*, **42**(April), 310–335.

Kilian, Lutz (2008): "The Economic Effects of Energy Price Shocks," *Journal of Economic Literature*, **46**(4), 871–909.

Koop, Gary and Dimitris Korobilis (2012): "Forecasting Inflation Using Dynamic Model Averaging," *International Economic Review*, **53**(3), 867–886.

Labys, Walter, Eugene Kouassi and Michel Terrazza (2000): "Short-Term Cycles in Primary Commodity Prices," *The Developing Economies*, **38**(3), 330–342.

Labys, Walter C., H. J. B. Rees and C. M. Elliott (1971): "Copper price behavior and the London Metal Exchange," *Applied Economics*, **3**(2), 99–113.

Lardic, Sandrine and Valerie Mignon (2008): "Oil prices and economic activity: An asymmetric cointegration approach," *Energy Economics*, **30**(3), 847–855.

Lui, Yu-Hon and David Mole (1999): "The Use of Fundamental and Technical Analysis by Foreign Exchange Dealers: Hong Kong Evidence," *Journal of International Money and Finance*, **17**(3), 535–545.

Makin, Anthony J. (2013): "Commodity prices and the macroeconomy: An extended dependent economic approach," *Journal of Asian Economics*, **24**, 80–88.

Marcellino, Massimiliano, James H. Stock and Mark W. Watson (2006): "A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series," *Journal of Econometrics*, **135**(1-2), 499–526.

Mark, Nelson C. (1995): "Exchange Rates and Fundamentals: Evidence on Long-Horizon Predictability," *The American Economic Review*, **85**(1), 201–218.

Meller, Patricio and Anthony M. Simpasa (2011): "Role of Copper in the Chilean & Zambian Economies: Main Economic and Policy Issues," *GDN Working Paper No. 43*, Global Development Network.

Menkhoff, Lukas (1998): "The Noise Trading Approach Questionnaire Evidence from Foreign Exchange," *Journal of International Money and Finance*, **17**(3), 547–564.

Neely, Christopher J., David E. Rapach, Jun Tu and Guofu Zhou (2014): "Forecasting the Equity Risk Premium: The Role of Technical Indicators," *Management Science*, **60**(7), 1772–1791.

Ng, Victor K. and Stephen Craig Pirrong (1994): "Fundamentals and Volatility: Storage, Spreads, and the Dynamics of Metals Prices," *Journal of Business*, **67**(2), 203–230.

Obstfeld, Maurice and Kenneth Rogoff (1996): *Foundations of International Macroeconomics*, MIT Press.

Pagan, Adrian R. (2002): "Learning About Models and Their Fit to Data," *International Economic Journal*, **16**(2), 118.

Pagan, Adrian R. and Aman Ullah (1999): *Nonparametric Econometrics*, New York: Cambridge University Press.

Pan, Jun and Kenneth J. Singleton (2008): "Default and Recovery Implicit in the Term Structure of Sovereign CDS Spreads," *The Journal of Finance*, **63**(5), 2345–2384.

Pesaran, M. Hashem, Andreas Pick and Allan Timmermann (2011): "Variable selection, estimation and inference for multi-period forecasting problems," *Journal of Econometrics*, **164**(1), 173–187.

Radetzki, Marian, Roderick G. Eggert, Gustavo Lagos, Marcos Lima and John E. Tilton (2008): "The boom in mineral markets: How long might it last?" *Resource Policy*, **33**(3), 125–128.

Raftery, Adrian E., Miroslav Kárný and Pavel Ettler (2010): "Online Prediction Under Model Uncertainty via Dynamic Model Averaging: Application to a Cold Rolling Mill," *Technometrics*, **52**(1), 52–66.

Raftery, Adrian E., David Madigan and Jennifer A. Hoeting (1997): "Bayesian Model Averaging for Linear Regression Models," *Journal of the American Statistical Association*, **92**(437), 179–191.

Rapach, David E., Jack K. Strauss and Guofu Zhou (2013): "International Stock Return Predictability: What Is the Role of the United States?" *The Journal of Finance*, **68**(4), 1633–1662.

RiskMetrics (1996): *"RiskMetrics: Technical Document, 4th Edition,"* J.P.Morgan/Reuters. Available from: http://www.riskmetrics.com/system/files/private/td4e.pdf.

Silverman, B. W. (1986): *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.

Spilimbergo, Antonio (2002): "Copper and the Chilean Economy: 1960-98," *The Journal of Policy Reform*, **5**(2), 115–126.

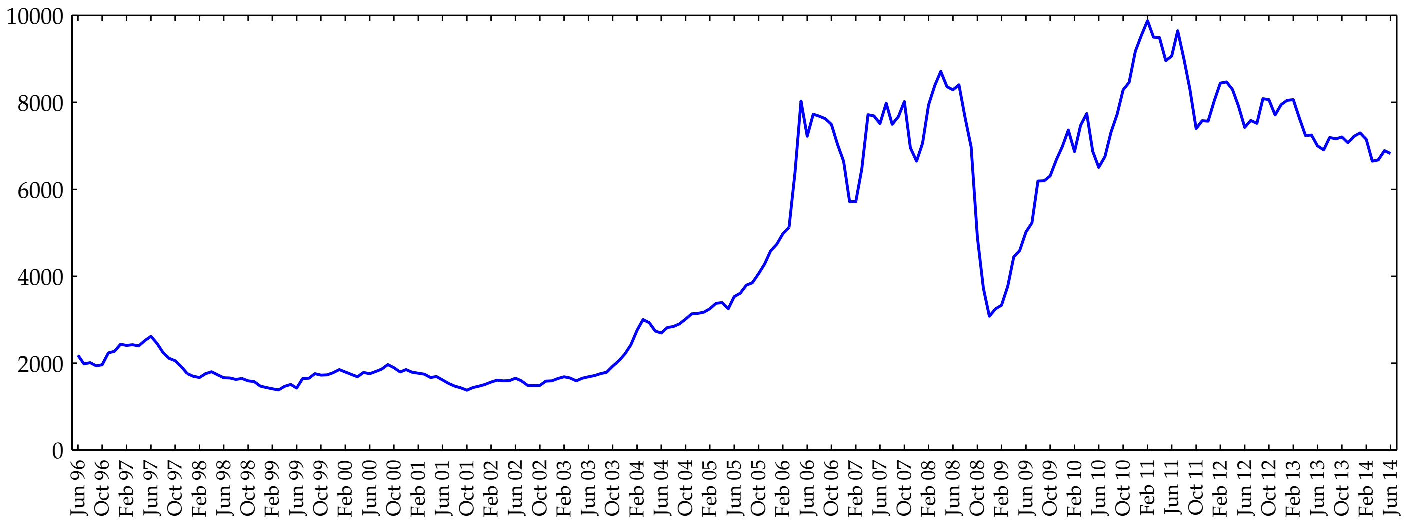
UNCTAD (2011): *"Price formation in financialized commodity markets: The role of information,"* United Nations Conference on Trade and Development. Available from: http://unctad.org/en/docs/gds20111_en.pdf.

Watkins, Clinton and Michael McAleer (2004): "Econometric modeling of non-ferrous metal prices," *Journal of Economic Surveys*, **18**(5), 654–701.

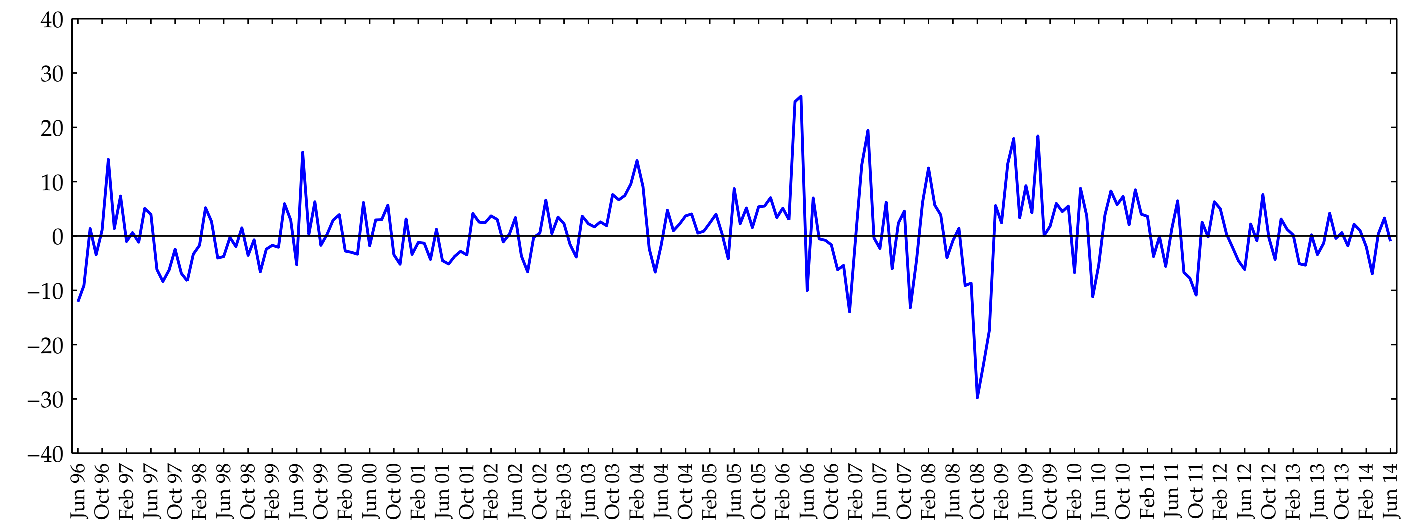
West, M. and J. Harrison (1997): *Bayesian Forecasting and Dynamic Models*, Springer Verlag.

Wright, Jonathan H. (2008): "Bayesian Model Averaging and exchange rate forecasts," *Journal of Econometrics*, **146**(2), 329–341.

Figures and Tables



(a) Level series



(b) Return series

Figure 1: Time series evolution of London Metal Exchange (LME) Copper from June 1996 to June 2014. Spot prices (S_t) are shown in Panel (a) and returns (r_t) are shown in Panel (b).

Table 1: Summary statistics of predictor variables

Variable	Variable description (Bloomberg and FRED tickers in parenthesis)	Mean	Med	Stdev	Skew	Kurt	Min	Max
r_t	Copper spot return: $100(\text{Spot}_t/\text{Spot}_{t-1} - 1)$, Copper Spot (LMCADY)	0.67	0.61	6.78	−0.14	6.52	−29.82	25.70
ΔDemand_t	Change in demand (10 000 metric tonnes): Consumption (WBMURCWT)− [Production (WBMURPWT) + Inventories (LSCA)] .	0.01	−0.27	8.39	0.28	2.86	−18.64	24.06
$\Delta\text{Inventories}_t$	Change in Inventories (10 000 metric tonnes): $\text{Inventories}_t - \text{Inventories}_{t-1}$.	−0.07	−0.37	4.00	0.40	3.96	−10.94	14.87
Convyield_t	Marginal Convenience Yield: Risk free rate (USGG3M) − [3 month Forward (LMAHDS03) − Spot (LMCADY)] / Spot	2.81	2.54	3.02	0.83	3.01	−1.02	13.38
ΔIP_t	Growth in Industrial Production (INDPRO from FRED database):	0.15	0.20	0.69	−1.94	12.37	−4.30	2.06
Spread_t	Term spread: U.S. 10 year Treasury Bond (USGG10YR) - 3 month Treasury Bill (USGG3M).	1.73	1.71	1.20	−0.17	1.87	−0.68	3.67
ΔBDI_t	Growth rate of Baltic Dry Index (BDIY): $100(\text{BDIY}_t/\text{BDIY}_{t-1} - 1)$.	1.45	0.21	18.32	0.89	8.40	−63.67	102.08
VIX_t	Chicago Board Options Exchange Market Volatility Index (VIX)	21.49	20.19	8.17	1.78	8.26	10.79	62.25
TED_t	Ted Spread (TEDRATE from FRED database)	0.52	0.41	0.42	2.65	14.22	0.12	3.41
ΔSP500_t	Return on SP500 (SPX): $100(\text{SPX}_t/\text{SPX}_{t-1} - 1)$.	0.50	1.09	3.97	−1.36	8.64	−23.05	11.40
ΔGold_t	Growth rate of GOLD (XAU): $100(\text{XAU}_t/\text{XAU}_{t-1} - 1)$.	0.55	0.02	3.85	0.31	4.27	−12.14	15.98
ΔOil_t	Growth rate of OIL (WTI): $100(\text{WTI}_t/\text{WTI}_{t-1} - 1)$.	0.74	1.72	8.29	−0.84	5.07	−34.07	20.54
ΔCLP_t	Return on Chilean peso (CLP, [CLP per USD]): $100(\text{CLP}_t^{-1}/\text{CLP}_{t-1}^{-1} - 1)$.	−0.14	−0.05	2.54	−1.03	9.31	−16.07	7.53
ΔAUD_t	Return on Australian dollar (AUD, [USD per AUD]): $100(\text{AUD}_t/\text{AUD}_{t-1} - 1)$.	0.07	0.22	2.94	−1.10	8.07	−17.43	7.19
ΔAlcoa_t	Return on Alcoa (ALCOA): $100(\text{ALCOA}_t/\text{ALCOA}_{t-1} - 1)$.	0.53	0.75	8.95	−0.55	8.41	−50.66	36.28
ΔRio_t	Return on Rio Tinto (RIO): $100(\text{RIO}_t/\text{RIO}_{t-1} - 1)$.	1.29	1.97	8.61	−1.20	7.92	−42.95	24.71
ΔFPM_t	Return on Freeport McMoran (FCX): $100(\text{FCX}_t/\text{FCX}_{t-1} - 1)$.	1.15	1.52	10.82	−0.50	4.69	−48.97	30.47
ΔBHP_t	Return on BHP Billiton (BHP): $100(\text{BHP}_t/\text{BHP}_{t-1} - 1)$.	1.22	1.56	7.27	−0.44	5.26	−34.94	21.32

Notes: This table reports standard summary statistics of the predictor variables that are used in the forecasting model. It also includes summary statistics of copper spot return, which is the target variable in the forecast evaluation.

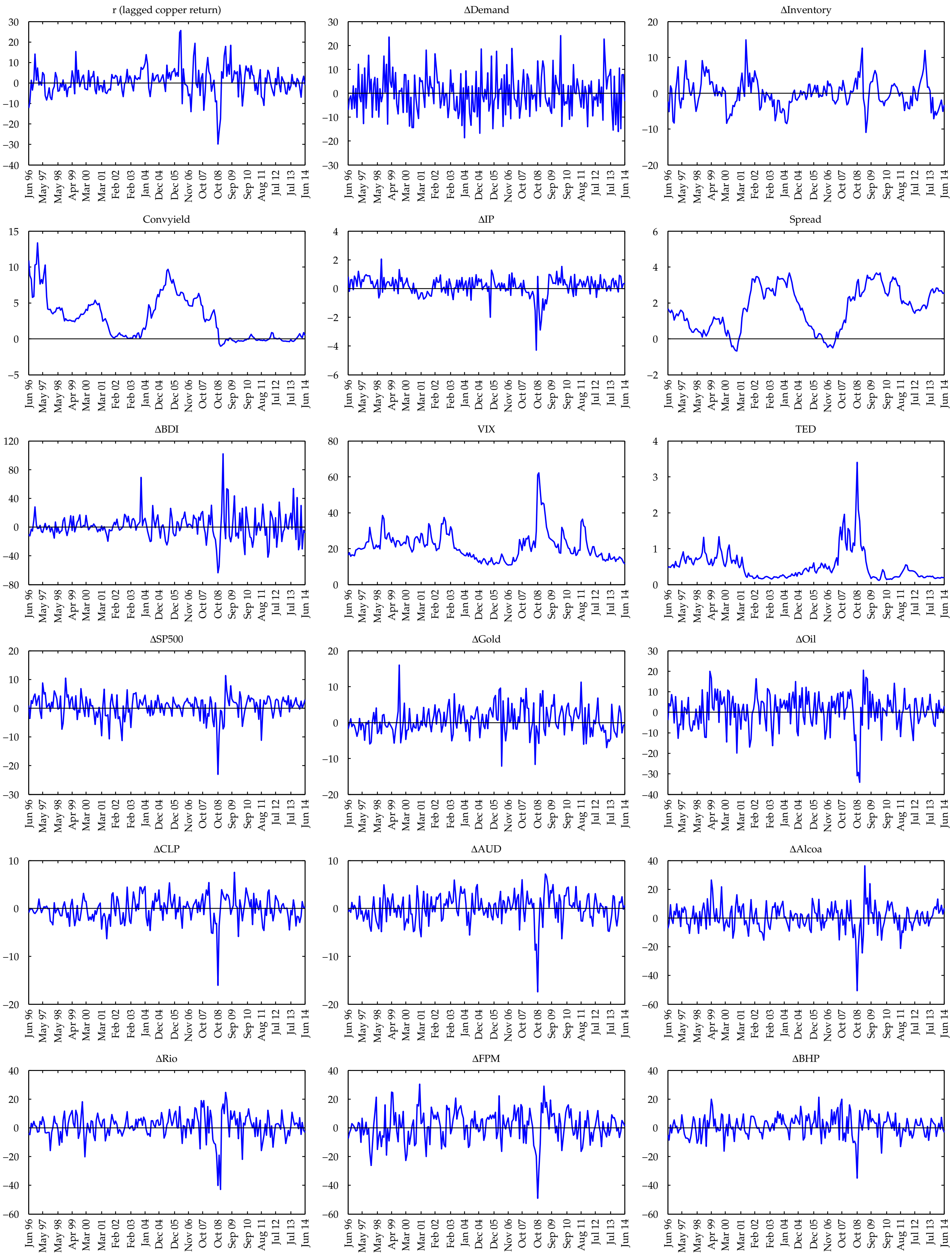
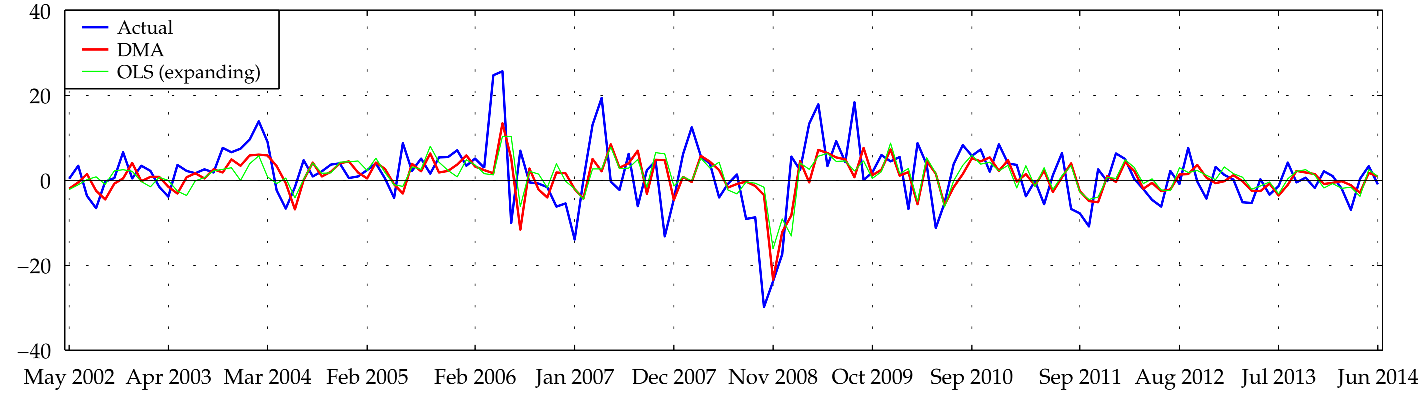


Figure 2: Time series plot of the set of predictor variables (June 1996 to June 2014).

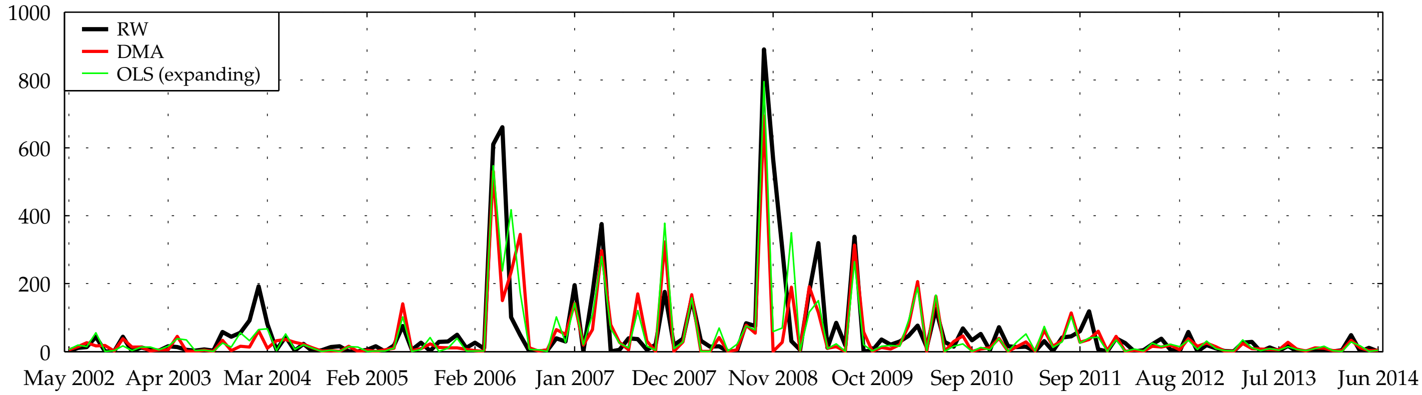
Table 2: One-step-ahead out-of-sample forecast results (May 2002 to June 2014)

Model	MSFE	Relative−MSFE	$R^2_{os}(\%)$	CW−statistic	p−value
Random Walk (RW)	56.9500	1.0000	–	–	–
HA (rolling window)	57.4338	1.0085	−0.8494	0.5884	0.2781
HA (expanding window)	57.0555	1.0019	−0.1852	0.3693	0.3559
OLS (rolling window)	59.6750	1.0478	−4.7849	0.0767	0.4694
OLS (expanding window)	51.2692	0.9002	9.9751	1.9639	0.0248
DMA ($\alpha = 0.95, \lambda = 0.99$)	45.6190	0.8150	18.5009	2.8551	0.0022
DMS ($\alpha = 0.95, \lambda = 0.99$)	49.1047	0.8622	13.7759	2.2284	0.0129
TVP ($\lambda = 0.99$)	51.4550	0.9035	9.6488	1.9136	0.0278

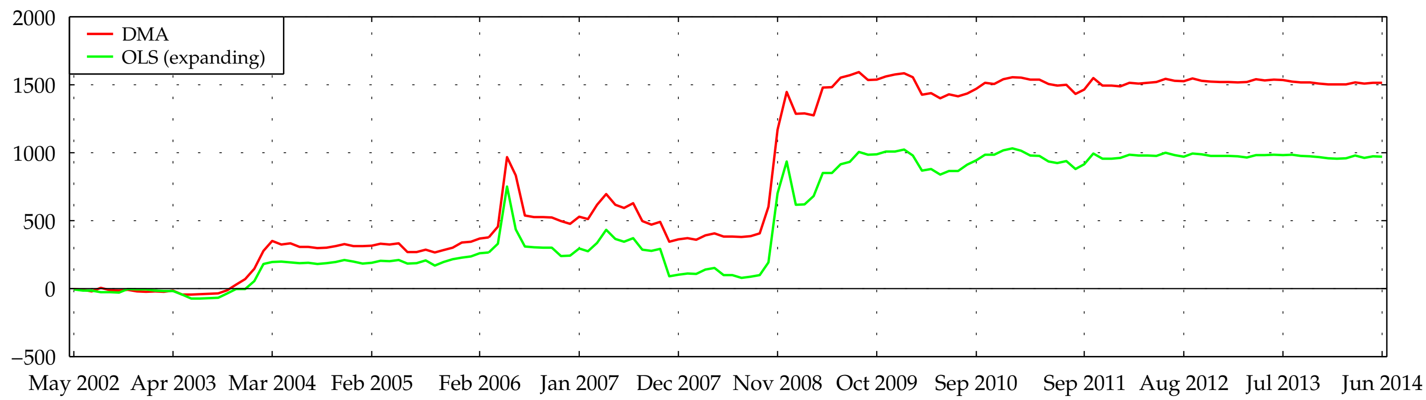
Notes: This table reports the one-step-ahead out-of-sample forecast evaluation results over the time period from May 2002 to June 2014. 70 observations (ie., from June 1996 to April 2002) were used for in-sample fitting. The first column shows the models that are fitted, the second column the mean squared forecast error (MSFE), the third column shows the MSFEs relative to the MSFE of a random walk (RW) model, the fourth column provides the Campbell and Thompson (2008) out-of-sample R^2_{os} (in percent), the fifth column is the Clark and West (2007) MSFE-adjusted t−statistic, and the sixth column is the corresponding (one-sided) p−value.



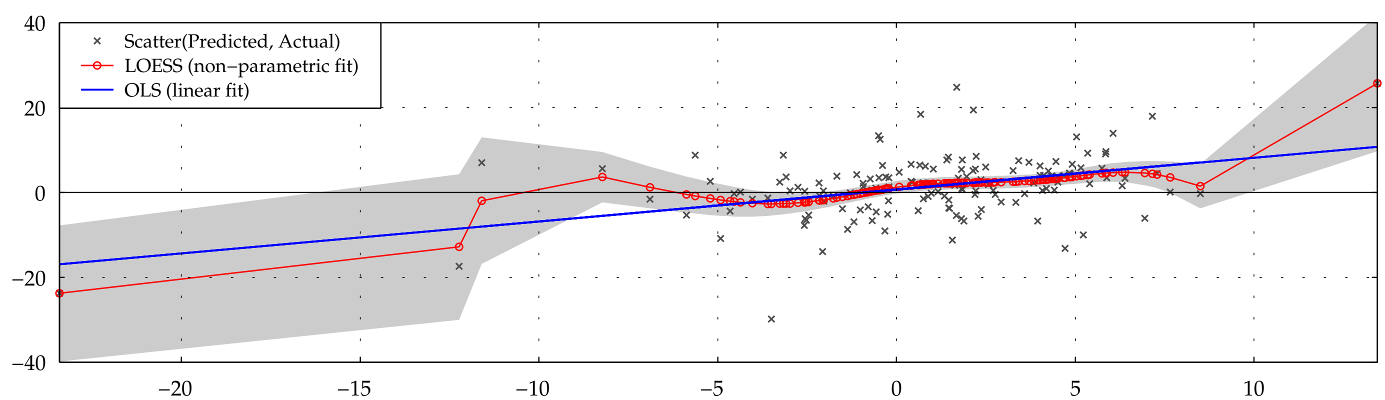
(a) Actual and Predicted



(b) Squared Forecast Errors



(c) Cumulative Difference of Squared Forecast Errors



(d) Scatter Plot of Predicted and Actual (with fits)

Figure 3: 1-step-ahead out-of-sample forecasts. Panel (a) shows the evolution of the actual copper return together with the predicted values from DMA and expanding window OLS predictions. Panel (b) plots the squared forecast errors from random walk (RW), DMA and expanding window OLS. Panel (c) shows plots of the cumulative difference between the squared forecast errors of the DMA and expanding window OLS forecasts, relative to the RW model. Panel (d) shows a scatter plot of predicted copper returns from the DMA model ($x-$axis) and actual returns ($y-$axis), together with an OLS based (linear) and a LOESS non-parametric fit of this relationship. 95% confidence intervals for the LOESS fit, based upon asymptotic standard errors are marked in grey shading.

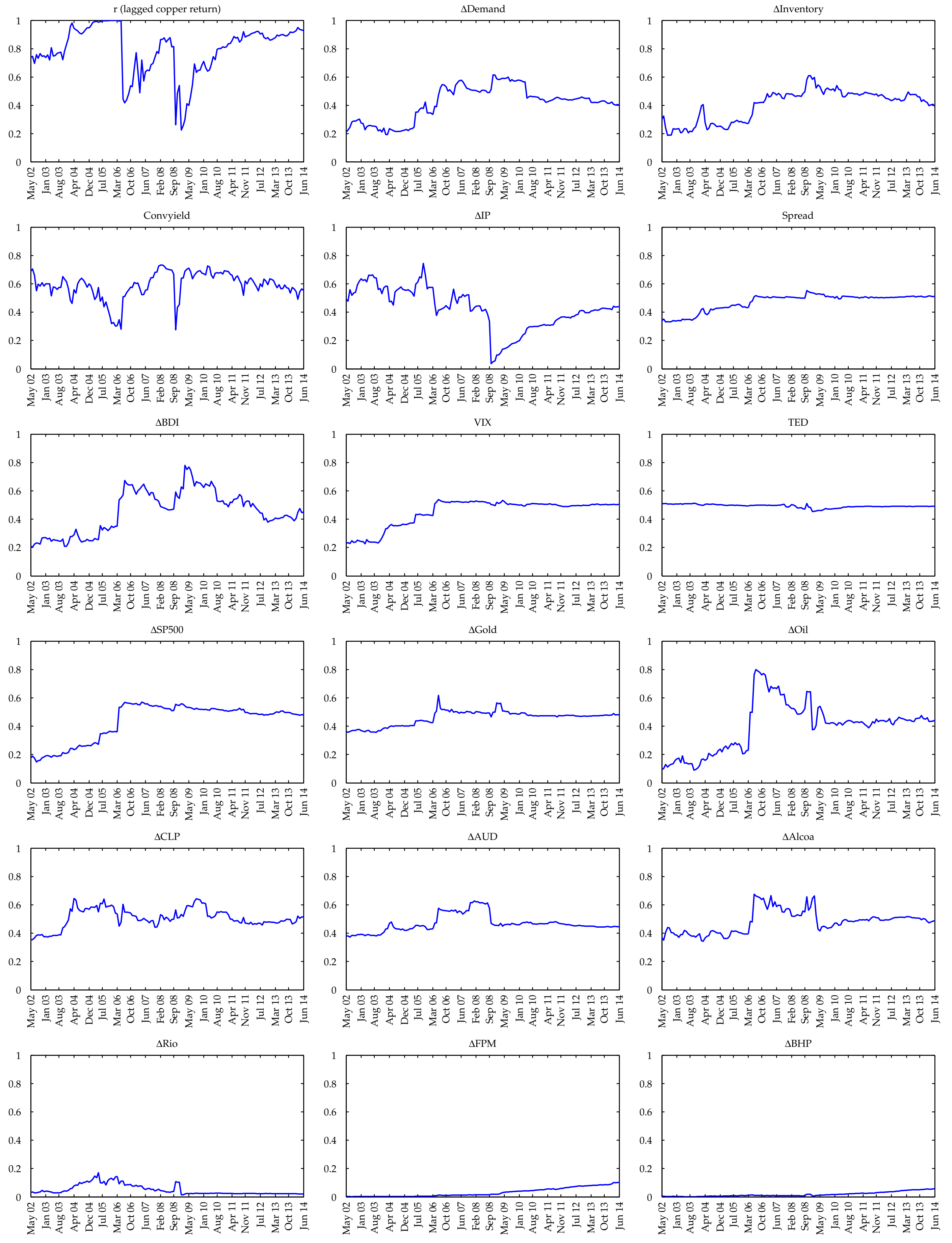


Figure 4: Time series plots of the posterior inclusion probability (PIP_t) over the out-of-sample evaluation period from May 2002 to June 2014.

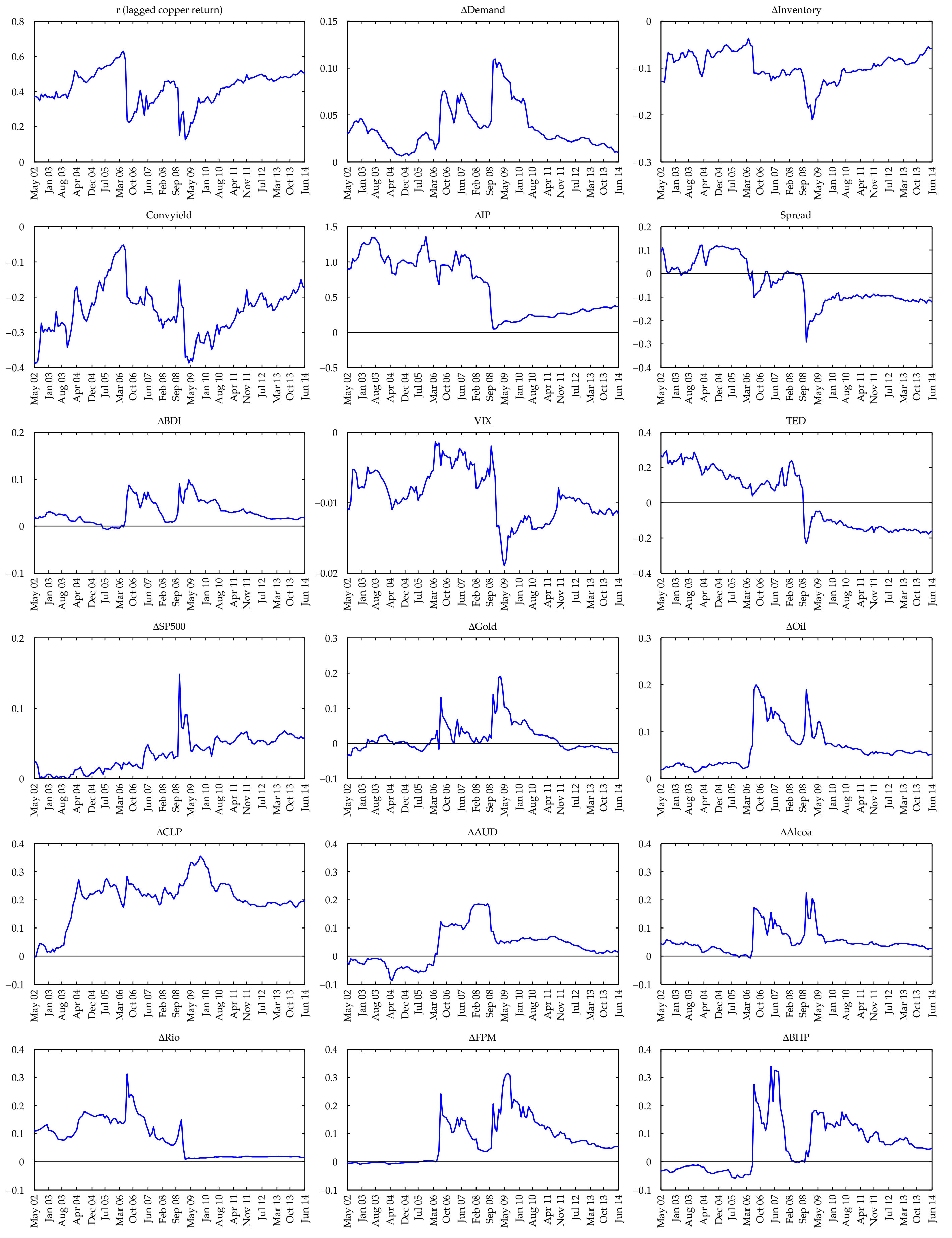


Figure 5: Time series plots of the DMA aggregated time varying parameter estimates ($\hat{\boldsymbol{\beta}}_{t|t}^{(\mathrm{DMA})}$) over the out-of-sample period from May 2002 to June 2014.

Table 3: Multiple-step-ahead out-of-sample forecast results

Model	MSFE/h	Relative−MSFE	$R^2_{os}(\%)$	CW−statistic	p−value
			$h = 2$		
Random Walk (RW)	82.1507	1.0000	–	–	–
HA (rolling window)	82.3034	1.0019	−0.1859	0.7231	0.2348
HA (expanding window)	81.4063	0.9909	0.9062	0.7271	0.2336
OLS (rolling window)	101.0579	1.2302	−23.0153	0.3921	0.3475
OLS (expanding window)	86.1708	1.0489	−4.8936	0.6041	0.2729
DMA ($\alpha = 0.95, \lambda = 0.99$)	75.6150	0.9204	7.9557	2.1961	0.0140
DMS ($\alpha = 0.95, \lambda = 0.99$)	81.4557	0.9915	0.8459	0.7247	0.2343
TVP ($\lambda = 0.99$)	88.3177	1.0751	−7.5070	0.7523	0.2259
			$h = 3$		
Random Walk (RW)	97.6540	1.0000	–	–	–
HA (rolling window)	98.1034	1.0046	−0.4602	0.5363	0.2959
HA (expanding window)	96.4411	0.9876	1.2420	0.5588	0.2881
OLS (rolling window)	126.3863	1.2942	−29.4226	0.1310	0.4479
OLS (expanding window)	108.2863	1.1089	−10.8877	0.4798	0.3157
DMA ($\alpha = 0.95, \lambda = 0.99$)	91.0030	0.9319	6.8108	1.9307	0.0268
DMS ($\alpha = 0.95, \lambda = 0.99$)	101.6393	1.0408	−4.0811	0.4982	0.3092
TVP ($\lambda = 0.99$)	98.3304	1.0069	−0.6927	0.6649	0.2531
			$h = 6$		
Random Walk (RW)	134.8147	1.0000	–	–	–
HA (rolling window)	135.2127	1.0030	−0.2952	0.5507	0.2909
HA (expanding window)	131.4130	0.9748	2.5232	0.7900	0.2148
OLS (rolling window)	172.4352	1.2791	−27.9053	0.2166	0.4142
OLS (expanding window)	167.2238	1.2404	−24.0398	0.3990	0.3450
DMA ($\alpha = 0.95, \lambda = 0.99$)	127.7949	0.9479	5.2070	1.5404	0.0617
DMS ($\alpha = 0.95, \lambda = 0.99$)	143.5110	1.0645	−6.4506	0.7343	0.2314
TVP ($\lambda = 0.99$)	135.9420	1.0084	−0.8362	0.4345	0.3320
			$h = 9$		
Random Walk (RW)	160.0248	1.0000	–	–	–
HA (rolling window)	160.2217	1.0012	−0.1230	0.5925	0.2767
HA (expanding window)	155.7709	0.9734	2.6583	0.8151	0.2075
OLS (rolling window)	265.6328	1.6599	−65.9948	−0.1554	0.5618
OLS (expanding window)	224.4761	1.4028	−40.2758	0.2217	0.4123
DMA ($\alpha = 0.95, \lambda = 0.99$)	171.4907	1.0717	−7.1651	0.5588	0.2882
DMS ($\alpha = 0.95, \lambda = 0.99$)	191.7606	1.1983	−19.8318	0.4548	0.3246
TVP ($\lambda = 0.99$)	249.1862	1.5572	−55.7173	−0.1041	0.5414
			$h = 12$		
Random Walk (RW)	176.1519	1.0000	–	–	–
HA (rolling window)	178.3301	1.0124	−1.2365	0.5092	0.3053
HA (expanding window)	172.7305	0.9806	1.9423	0.4516	0.3258
OLS (rolling window)	319.8395	1.8157	−81.5703	−0.2590	0.6022
OLS (expanding window)	251.5167	1.4278	−42.7839	0.1339	0.4467
DMA ($\alpha = 0.95, \lambda = 0.99$)	184.0053	1.0446	−4.4583	0.4608	0.3225
DMS ($\alpha = 0.95, \lambda = 0.99$)	196.0132	1.1128	−11.2751	0.1923	0.4238
TVP ($\lambda = 0.99$)	286.9439	1.6290	−62.8957	−0.1898	0.5753

Notes: This table reports the multiple-step-ahead out-of-sample forecast evaluation results. All entries in this table are the same as in Table 2, with the only exception being the second column which shows the MSFE deflated by the forecast horizon h. The CW−statistic in the 5^{th} column uses HAC standard errors computed with an ARMA(1,1) pre-whitening step and a Quadratic Spectral Kernel where the bandwidth parameter was chosen optimally with the data driven approach outlined in Andrews and Monahan (1992).

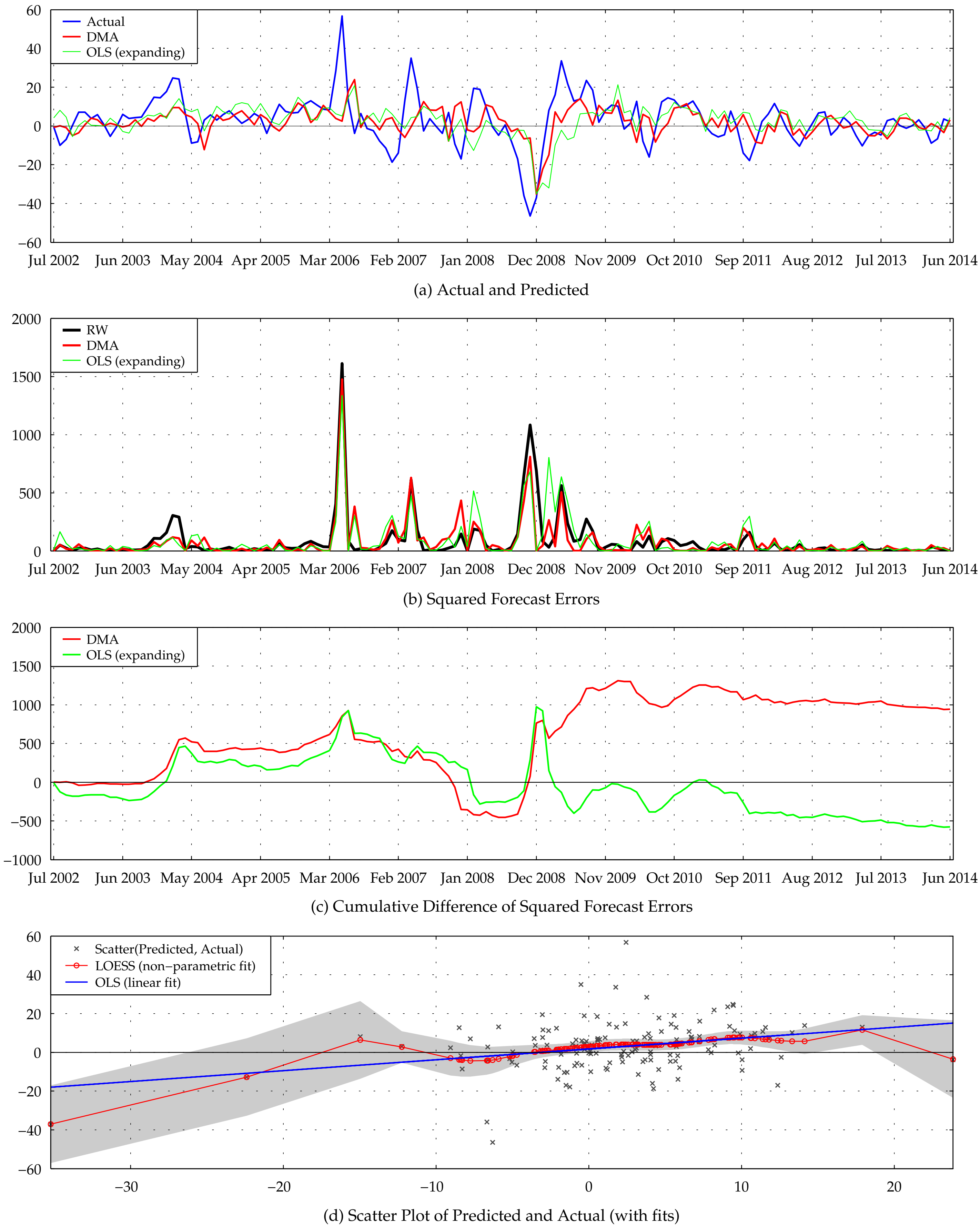


Figure 6: 2-step-ahead out-of-sample forecasts. The contents of the plot are the same as in Figure 3.

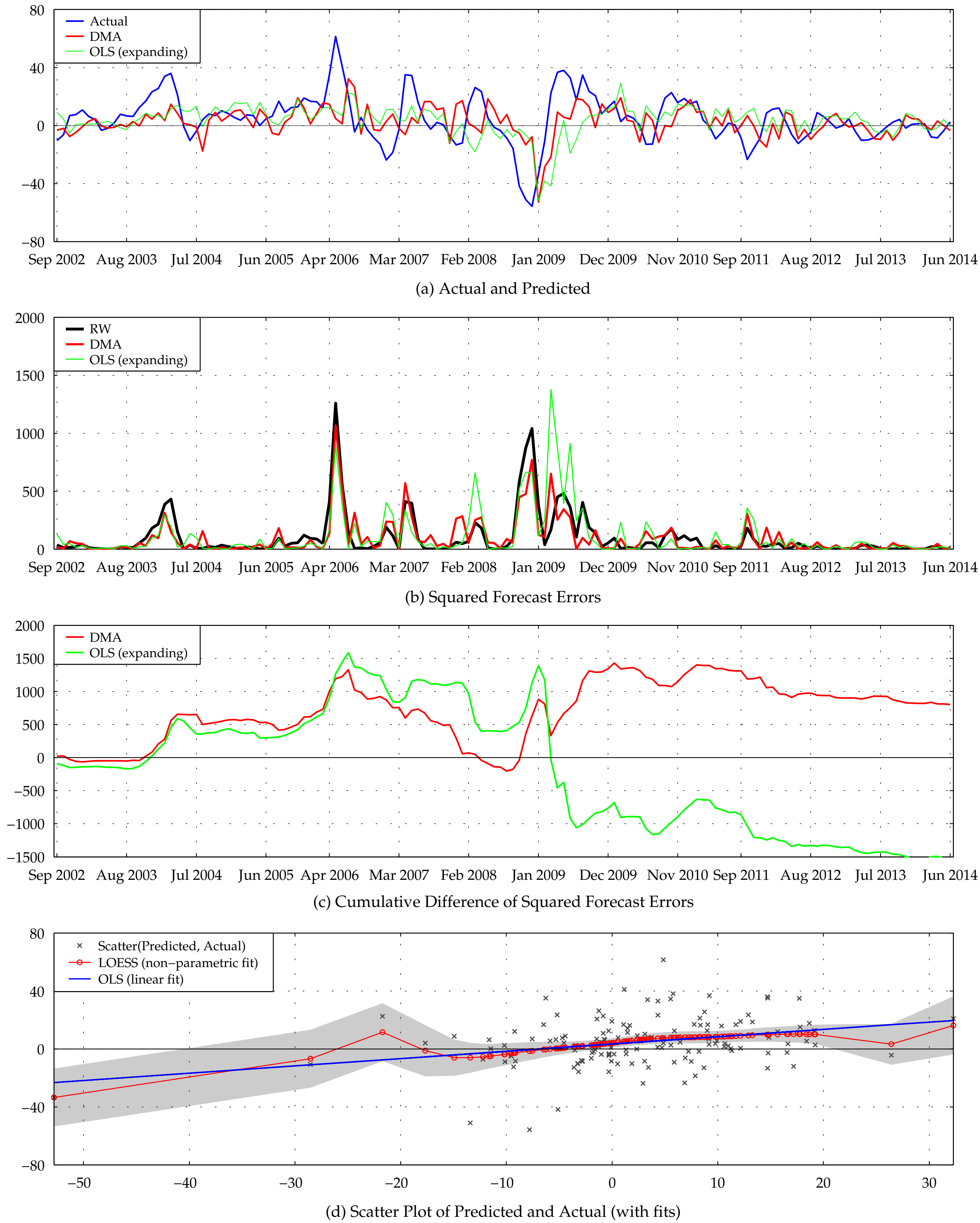


Figure 7: 3-step-ahead out-of-sample forecasts. The contents of the plot are the same as in Figure 3.

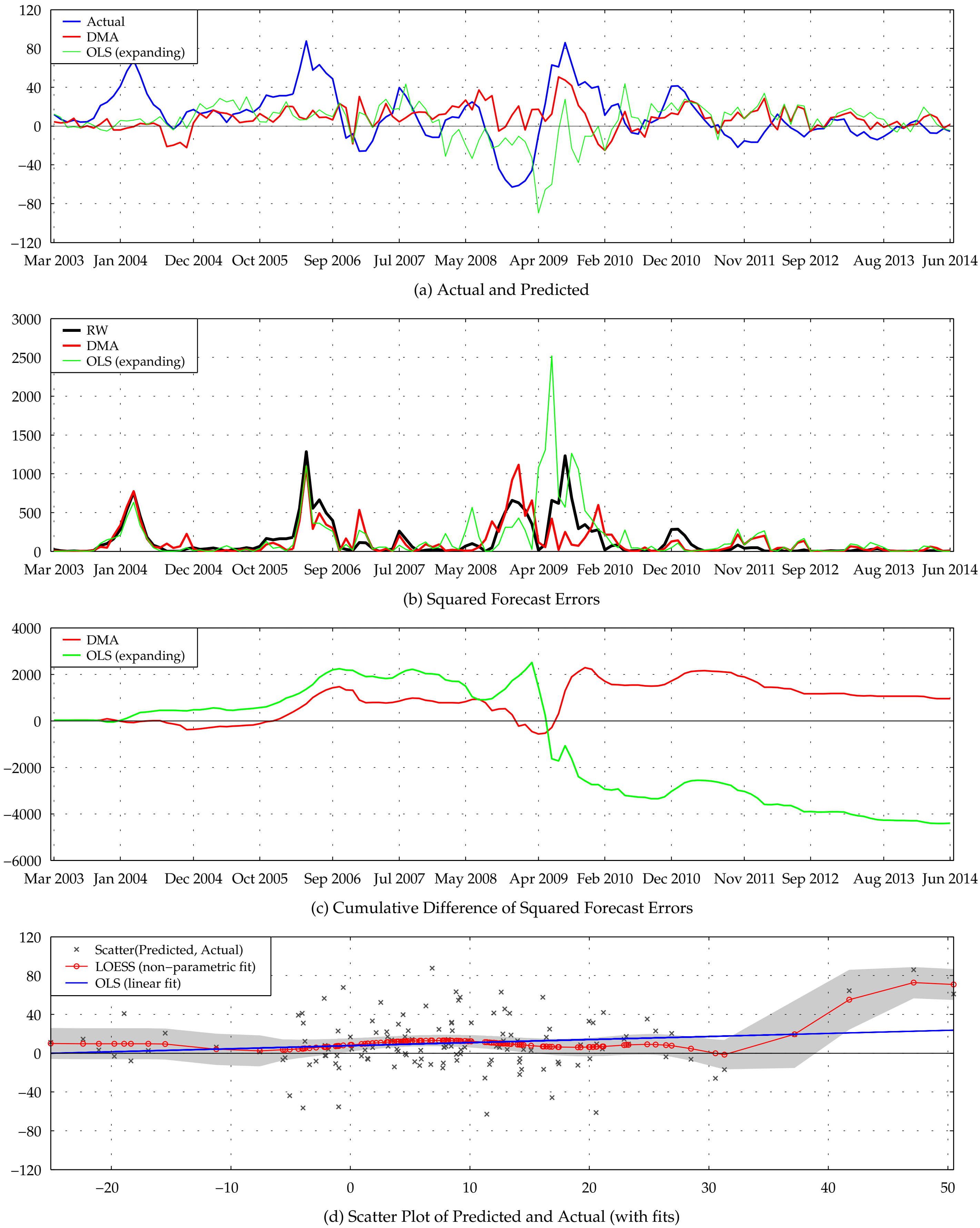
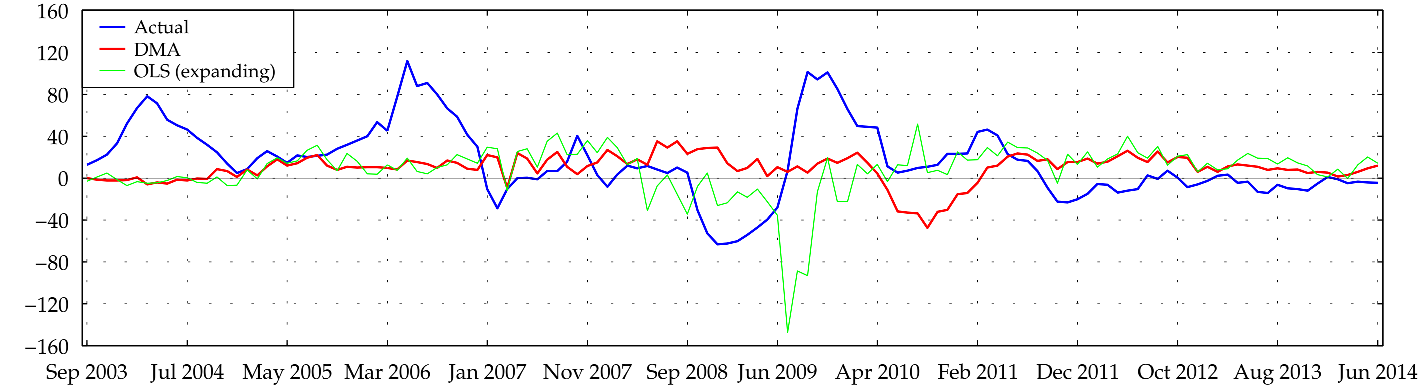
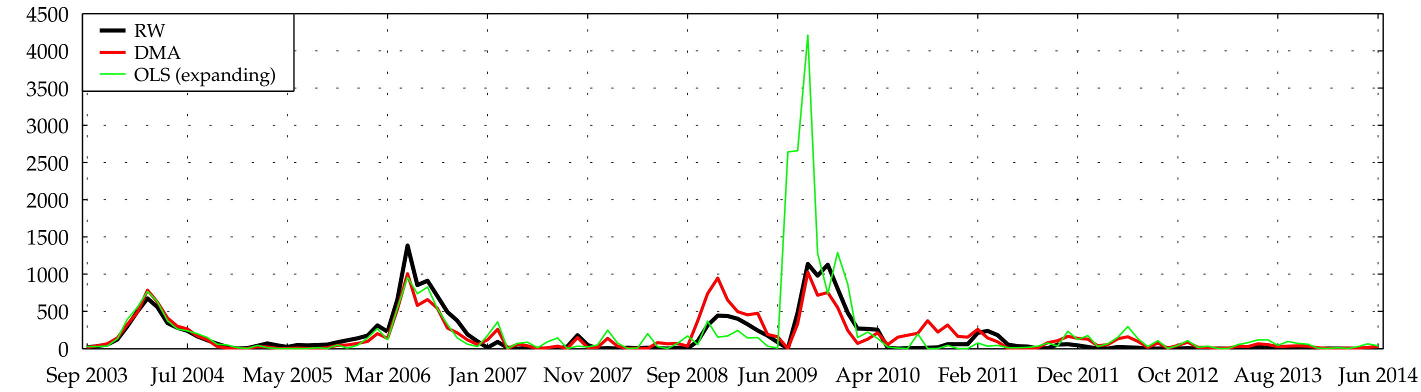


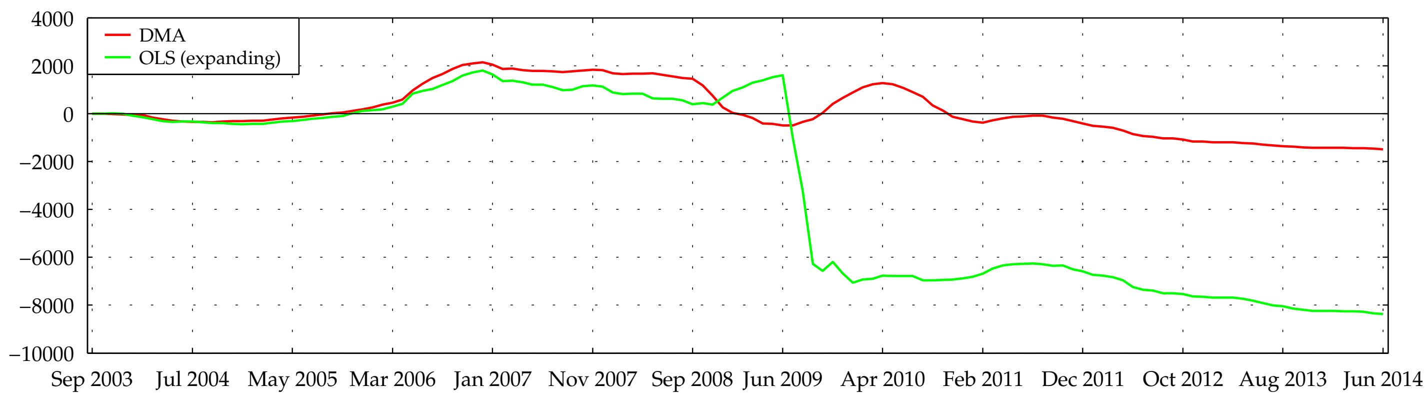
Figure 8: 6-step-ahead out-of-sample forecasts. The contents of the plot are the same as in Figure 3.



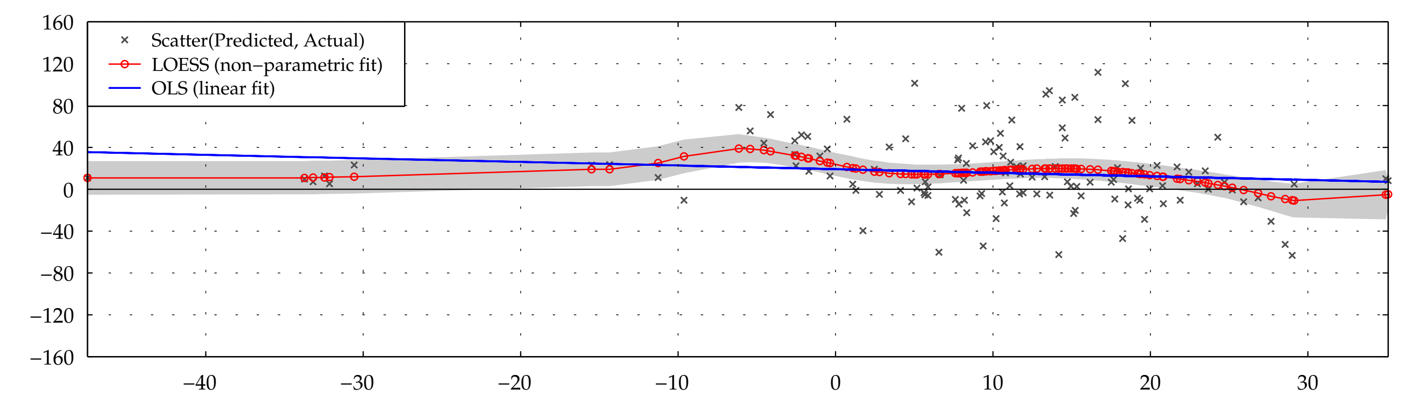
(a) Actual and Predicted



(b) Squared Forecast Errors

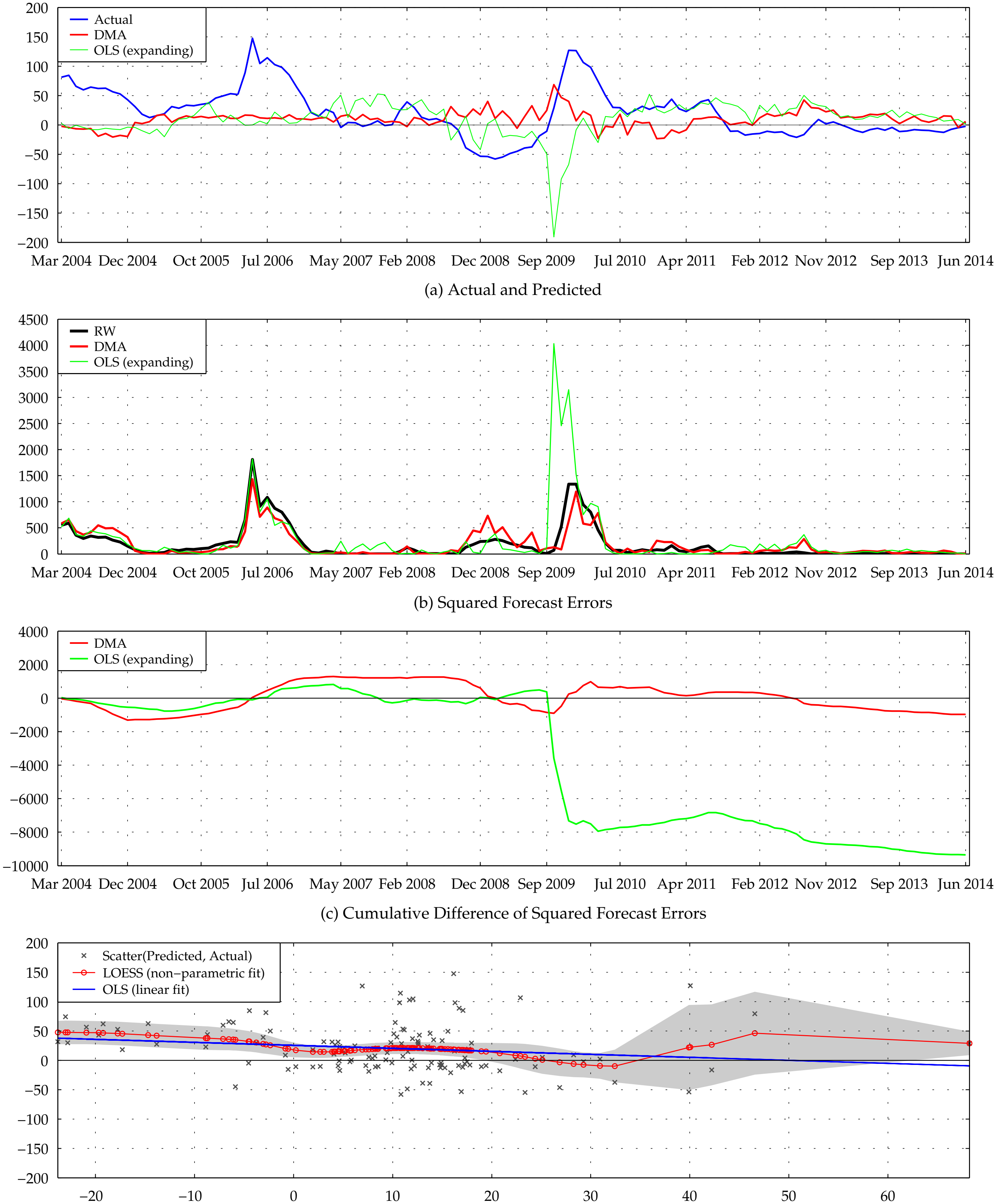


(c) Cumulative Difference of Squared Forecast Errors



(d) Scatter Plot of Predicted and Actual (with fits)

Figure 9: 9-step-ahead out-of-sample forecasts. The contents of the plot are the same as in Figure 3.



(a) Actual and Predicted

(b) Squared Forecast Errors

(c) Cumulative Difference of Squared Forecast Errors

(d) Scatter Plot of Predicted and Actual (with fits)

Figure 10: 12-step-ahead out-of-sample forecasts. The contents of the plot are the same as in Figure 3.