Fundamentals, Speculation or Macroeconomic Conditions? Modelling and Forecasting Arabica Coffee Prices*

Jesus Crespo Cuaresma[1,2,3,4], Jaroslava Hlouskova[1,5,6], and Michael Obersteiner[1]

[1]International Institute of Applied Systems Analysis (IIASA)
[2]Vienna University of Economics and Business (WU)
[3]Wittgenstein Center for Demography and Global Human Capital (IIASA,VID/OEAW,WU)
[4]Austrian Institute of Economic Research (WIFO)
[5]Institute for Advanced Studies (IHS), Vienna
[6]Thompson Rivers University, Kamloops, Canada

Abstract

We analyse the role played by market fundamentals, speculation and macroeconomic conditions as empirical determinants of price changes in Arabica coffee. We combine model averaging techniques to explain historical patterns with an in-depth analysis of out-of-sample predictability of Arabica coffee prices using fundamentals as well as macroeconomic and financial variables. Our results indicate that variables related to global macroeconomic and financial developments contain valuable information to explain the historical pattern of coffee price developments, as well as to improve out-of-sample predictions.

Keywords: Commodity prices, coffee, forecasting, forecast combination, vector autoregressive models, model uncertainty, model averaging.

*The authors would like to thank two anonymous referees for very helpful comments on an earlier draft of this paper and Olha Danylo for research assistance. Funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 633692: "Metrics, Models and Foresight for European SUStainable Food And Nutrition Security" (SUSFANS) is gratefully acknowledged.

1 Introduction

The world coffee market has undergone a significant transformation over the last 50 years. The coffee market was regulated, up until 1989, by a series of International Coffee Agreements which were intended to manage supply and maintain price stability. Price levels during the regulated market period (1965 to 1989) were relatively high since both upward and downward trends were corrected through the application of export quotas. This system subsequently collapsed, and since 1990 the coffee market has been subject to free market forces. The free market period beginning in 1990 had two sub-periods of markedly low price levels: 1990 to 1993 and 1999 to 2004 (see Figure 1). The latter sub-period (known as the *coffee crisis* period) was the longest period of low prices ever recorded with severe negative consequences on the economies of exporting countries. Prices recovered strongly after 2004, reaching a 34-year high in mid-2011. However, there has subsequently been a severe deterioration in prices while costs of coffee production inputs, particularly fertilizers and labor, continued to rise. These price increases were in part driven by higher expenditures for pesticides to combat emerging large scale diseases attacking coffee plantations and increasing fertilizer prices both squeezing the margins for labor inputs.

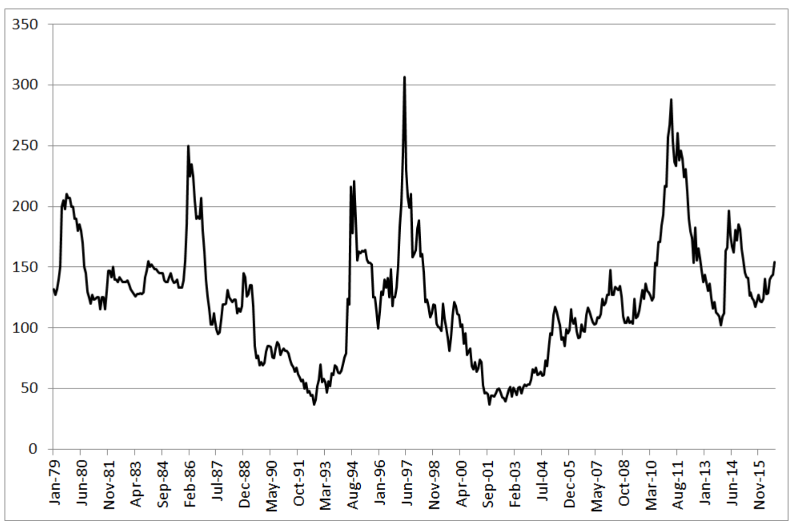

Figure 1: Arabica (Brazilian) coffee price (Cents/lb).

During the regulated market period the highest volatility was recorded in years following severe climate shocks recorded in exporting countries, notably in Brazil in 1975 and 1985. The highest volatility levels are generally recorded for the months of May, June, July and August, since they cover the period of possible frosts in Brazil, thought to fuel speculative activity. During the free market high volatility was recorded in 1994 and 1997 (see Figure 2) where in 1994 a climate shock was recorded in Brazil.

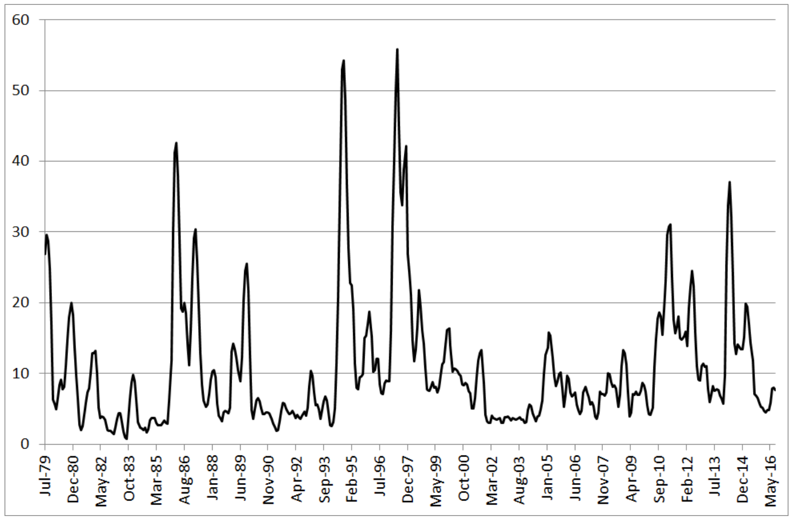

Figure 2: Volatility of Arabica (Brazilian) coffee price (based on 6-months rolling window).

In this paper, we assess empirically the role that different theoretical driving factors of coffee prices play as predictors of their dynamics. By carrying out a systematic assessment of the drivers of Arabica coffee price dynamics in the context of specification uncertainty, we provide a novel methodological framework to model commodity price changes and aim at improving both the in-sample fit and and out-of-sample forecasting ability of existing approaches. Our method explicitly addresses specification uncertainty both in terms of explaining historical trends and using forecast combination methods which aggregate predictions from different models using diverse techniques to integrate the uncertainty over specification choice in out-of-sample forecasts. We entertain individual models that contain information about climate, coffee market fundamentals, global macroeconomic developments and speculation. Making use of forecast pooling techniques that account for model uncertainty, we are also able to assess quantitatively the differences in predictive ability of competing explanatory factors. In addition to homoskedastic and heteroskedastic univariate time series models, we include vector autoregressive (VAR) and vector error correction (VEC) specifications aimed at exploiting the relationship between the price of Arabica coffee and its potential determinants. We tackle the issue of specification uncertainty by using Bayesian Model Averaging methods and assessing the potential improvements in predictive ability that can be gained from both frequentist and Bayesian forecast averaging methods. For this purpose, we employ a large number of forecast combination techniques that have been proposed in the literature on exchange rate forecasting (Costantini et al., 2016; Crespo Cuaresma et al., 2018) but have not yet been used for commodity price prediction specifications.

Our results indicate that variables related to global macroeconomic and financial developments contain valuable information to explain the historical pattern of coffee price developments, as well as to improve out-of-sample predictions of Arabica coffee prices. The paper is structured as follows. Section 2 presents the econometric framework used to explain historical data on Arabica coffee prices, which is based on Bayesian Model Averaging techniques, and discusses the results based on in-sample fit. Section 3 shows the results of the out-of-sample predictability analysis under specification uncertainty and section 4 concludes.

2 Explaining historical Arabica coffee price dynamics

2.1 Modelling and predicting commodity prices

A large number of studies evaluate the predictive power of commodity futures prices for actually realized spot prices, and evaluate the information content of macroeconomic and financial variables as leading indicators of commodity price dynamics. While early studies tend to show that futures prices contain useful information for predicting commodity price developments (Just and Rausser, 1981), more modern studies concentrate on the potential predictive content of other variables for commodity price forecasting. Gargano and Timmermann (2014) present evidence that the out-of-sample predictability of commodity prices depends on the state of the economy and that the information contained in macroeconomic variables improves forecast accuracy in models of commodity prices. Husain and Bowman (2004), on the other hand, analyse 15 different commodities and show that statistical models based on futures tend to yield better results in terms of predictive ability than those based exclusively on spot price dynamics or on judgement. The predictive models used to obtain commodity forecasts are also very diverse in terms of their methodological background, ranging from artificial neural networks (Kohzadi et al., 1996) to specifications aimed at modelling the dynamics of the second moment of the commodity price time series (Bernard et al., 2008).

In spite of the fact that many specifications, estimation methods and conceptual modelling settings have been employed in the empirical literature in order to obtain forecasts of commodity prices, the literature hitherto has not yet assessed model uncertainty in a systematic manner when it comes to creating commodity price predictions. The literature on modelling and forecasting commodity prices tends to rely on particular model specifications to draw conclusions concerning in-sample and out-of-sample predictive accuracy. Egelkraut et al. (2003) examine the accuracy of corn and soybean production forecasts provided by the USDA, while Brockhaus et al. (2016) analyze the response of different factors (such as prices) on production of rice, wheat and corn in China. Algieri (2014) investigates the main drivers of wheat price using vector error correction models. Ahumada and Cornejo (2015), Ahumada and Cornejo (2016a) and Ahumada and Cornejo (2016b) are also relevant examples of the use of single vector error correction specifications to model and predict food prices using similar theoretical settings as those in Deaton and Laroque (1992) or Deaton and Laroque (2003). By explicitly addressing model uncertainty using forecast pooling, Gargano and Timmermann (2014) is a notable exception in the literature. In our application, we concentrate exclusively on Arabica coffee price dynamics and analyse a much larger model space than that entertained in Gargano and Timmermann (2014), as well as a broader set of specification averaging techniques.

2.2 Model uncertainty and coffee price dynamics: An in-sample analysis

In order to analyse the relative importance of different potential explanatory factors of the dynamics of Arabica coffee prices, we start by entertaining models of the class of autoregressive distributed lag specifications. The models we consider are of the form

$$\Delta P_t = \alpha + \sum_{k=1}^{q_0} \theta \Delta P_{t-k} + \sum_{i=1}^{v} \sum_{j=1}^{q_v} \phi_{ij} x_{i,t-j} + \varepsilon_t, \tag{1}$$

where ΔP_t is the (annual) log-change in the price of Arabica coffee, which is assumed to be explained by its own lags and by lags of a set of variables $\{x_{it}\}_{i=1}^{v}$, as well as by a random normally distributed shock, ε_t, assumed to fulfil the standard assumptions of linear regression specifications.

The specifications nested in equation (1) do not contain contemporaneous covariates for two reasons. On the one hand, we aim at imposing a Granger-causality structure between the explanatory variables and changes in coffee prices, which in addition should reduce the potential problem of correlation between coffee price shocks and these covariates.[1] On the other hand, the out-of-sample predictive analysis carried out in section 3 is based on models that include exclusively lagged regressors (in order to allow for forecasting without particular assumptions on the behaviour of the explanatory variables), so the use of the specification without contemporaneous effects allows for a consistent modelling framework within the analysis carried out. Theoretical models of commodity prices such as those in Deaton and Laroque (1992) and Deaton and Laroque (2003) predict autoregressive dynamics in price behaviour and lagged adjustment to deviations from the supply-demand equilibrium, thus justifying specifications of the type put forward by equation (1). The individual model structures entertained by Ahumada and Cornejo (2016a), aimed at forecasting food prices, can be reconciled with specifications such as those nested in equation (1) and in the vector autoregressive structures used in our forecasting exercise. In addition, the use of an autoregressive term in equation (1) accounts for the persistence in coffee price changes and, to the extent that such persistence is related to the effect of other slow-moving variables (such as costs of coffee production inputs), also control for other potential omitted determinants.

[1]Such a specification is in contrast to those used in other empirical studies whose main aim is not forecasting but understanding the contemporaneous reaction of commodity prices to supply and demand shocks, such as the models estimated by Borensztein and Reinhart (1994).

We address uncertainty about the particular specification of the model (in the sense of covariate inclusion) by using Bayesian model averaging (BMA) techniques to carry out inference in the class of models given by equation (1).[2] The extensive number of candidate variables proposed in the literature as candidates to enter the model presented in equation (1) implies that model uncertainty may lead to flawed inference if it is not explicitly assessed. Instead of basing our inference on a particular selected model, we learn about the drivers of commodity prices using a weighted average of single regressions. In the Bayesian framework, the natural weighting scheme is based on the corresponding *posterior model probabilities* (PMP) of the individual specifications. In particular, if we are interested in performing inference for a quantity χ, which could be a parameter of the model, a combination of parameters or a predicted value of the dependent variable, the posterior probability over χ can be obtained as

$$p(\chi|y) = \sum_{l=1}^{2^K} p(\chi|M_l, y)p(M_l|y), \tag{2}$$

with $p(\cdot|y)$ denoting posterior distributions (that is, conditional on the data, y) and $p(\cdot|M_l, y)$ denoting posterior distributions conditional on the choice of covariates implied by model M_l. Assuming that v potential independent variables are available and that up to q lags are allowed to enter the specification, the cardinality of the model space based on equation (1) is given by $K = 2^{(v+1)q}$, which corresponds to the number of models that can be built by combining these covariates and lags in addition to the autoregressive terms.

Bayesian reasoning allows us to write the posterior model probabilities in equation (2) as proportional to the product of the marginal likelihood of the corresponding model, $p(y|M_l)$ and the model prior $p(M_l)$,

$$p(M_l|y) \propto p(y|M_l)p(M_l). \tag{3}$$

In order to obtain PMPs, prior distributions need to be elicited on the parameters of the models that can be formed by combining the covariates and lags, as well as on the variance of the error term, σ^2. Following the literature on BMA for linear models, improper priors are placed on the intercept $p(\alpha) \propto 1$ and variance $p(\sigma) \propto \sigma^{-1}$, reflecting lack of prior subjective information about these quantities. For the rest of the parameters in a given specification within the class of models described by equation (1), we follow the standard convention in BMA and use Zellner's g prior (Zellner, 1986),

$$\phi_{ij}| \left(\sigma^2, M_l, g\right) \; \sim \; N\left(0, \sigma^2 g(X_l' X_l)^{-1}\right), \tag{4}$$

where X_l is the matrix of observations of the independent variables included in model M_l. Characteristic choices of the parameter g are T, the number of observations (unit information prior, UIP), proposed

[2] See Koop (2003) for an introduction to BMA.

by Kass and Wasserman (1995) and K^2, (the risk inflation criterion, RIC) put forward by Foster and George (1994). Fernández et al. (2001a) propose $g = \max(T, K^2)$ (BRIC prior) after comparing the performance of the UIP and RIC priors in simulated settings.[3]

The prior probabilities assigned to individual models, $p(M_l)$, allow the researcher to include prior beliefs about the relative adequacy of the different specifications nested in the class of models given by 1. Following Ley and Steel (2009), in our application we use a beta-binomial prior for inclusion of a given variable with a prior expected model size of $K/2$ regressors. Such a prior over the model space is uninformative about model size.

2.3 The determinants of coffee price dynamics

We assess the nature of the factors robustly affecting changes in Arabica coffee prices by applying BMA to a dataset on potential explanatory variables which we divide into four thematic groups: (i) *fundamental variables* (coffee production in Brazil, y^{BR}_{coffee}, world coffee production, y^{world}_{coffee}), (ii) *macroeconomic variables* (output for Brazil, y^{BR}, output for the EU, y^{EU}, output for the US, y^{US}, leading indicator for Germany, li^{EU}, leading indicator for the US, li^{US}, real effective exchange rate, REER), (iii) *financial variables* (stock market index for the EU, $stock^{EU}$, stock market index for the US, $stock^{US}$, S&P Goldman Sachs commodity index, GSCI) and (iv) *other climatic and meteorological variables* (precipitation, temperature of the area in Brazil where Arabica coffee is grown). We employ monthly data spanning the period from January 1985 to March 2016. The description of the variables and source of the data can be found in Table 1.

The set of potential determinants is chosen to strike a balance between covering the most important theoretical drivers of coffee price dynamics and the existence of data for such proxies at a monthly frequency. From a theoretical point of view, coffee prices are expected to react to supply changes, here proxied by Arabica coffee production both for the whole world and for Brazil, which are in turn affected by climatic conditions. In addition, coffee demand dynamics are expected to be driven on the one hand by income developments worldwide, which are proxied using industrial production indices for Brazil, the EU and the US. Since expectations on developments of macroeconomic variables have also been proposed as determinants of changes in commodity prices (see, for example, Ahumada and Cornejo, 2016b), we include leading indicator indices for both the US and Germany (with the latter as a proxy of overall macroeconomic expectations in Europe). Exchange rates are also proposed as a relevant macroeconomic factor affecting commodity prices in the existing empirical literature (Ahumada and Cornejo, 2015, 2016a), while financial speculation has been recently proposed as an important

[3]Alternatively, a hierarchical structure can be imposed by defining a prior on g, as put forward by Liang et al. (2008), Feldkircher and Zeugner (2009) or Ley and Steel (2012).

determinant of food price dynamics (Headey and Fan, 2008; Gilbert, 2010), which leads us to include the set of financial variables described above as potential covariates in our models.

Table 1: Variable description

Variable	Description	Source	Start date
Coffee price	Coffee-Brazilian (Arabica), (NY) Cents/lb	Datastream: COFBRAZ	1980:1
Fundamental variables			
Coffee production in Brazil	In thousand 60kg bags	International Coffee Organization	1965:4
Coffee production in the world	In thousand 60kg bags	International Coffee Organization	1965:4
Marcoeconomic variables			
Output for Brazil	Industrial production index. Indexed 2000:1=100. Seasonally adjusted.	Datastream: BRIPTOT.G	1985:1
Output for EU	Industrial production index for Eurozone. Indexed 2000:1=100. Seasonally adjusted.	Datastream: EKESIMANG	1980:1
Output for US	Industrial production index for the US Indexed 2000:1=100. Seasonally adjusted.	Datastream: USIPTOT.G	1980:1
Leading indicator for Germany	IFO: Business climate index	Datastream: BDIFOIDXE	1980:1
Leading indicator for the US	ISM: Manufacturing index	Datastream: USCNFBUSQ	1980:1
Real effective exchange rate	with respect to Brazil	Bloomberg	1980:1
Financial variables			
Stock market index for EU	Index covers at leats 80% of the market capitalization in the EMU	Datastream: TOTMKEM	1980:1
Stock market index for the US	Index covers at leats 80% of the market capitalization in the US	Datastream: TOTMKUS	1980:1
S&P Goldman Sachs commodity index	A composite index of commodity sector returns representing an unleveraged, long-only investment in commodity futures that is broadly diversified across a spectrum of commodities	Datastream: GSCITOT	1980:1
Climatic variables			
Precipitation	Mean precipitation of the Arabica coffee region (mm/pentad)	Google Earth Engine for Brazil (Arabica coffee region)	1981:1
Temperature	Mean value of daytime land surface temperature for each aggregated pixel, in degrees Celsius	Google Earth Engine for Brazil (Arabica coffee region)	1980:1

We apply BMA using the class of models defined by equation (1) and the set of variables described in Table 1, after transforming trending and seasonal variables when necessary by using annual changes. We employ the BRIC prior for the parameters corresponding to the covariates and their lags and the beta-binomial in Ley and Steel (2009) to define prior model probabilities. In order to overcome the computational constraints that are given by the large cardinality of the model space, we employ Markov Chain Monte Carlo (MCMC) methods proposed by Kass and Wasserman (1995) to explore the model space. The results presented in this section are based on two million model draws after a burn-in of one million draws.

We start by presenting results that correspond to models without an autoregressive term, that is, based on specifications of the form

$$\Delta P_t = \alpha + \sum_{i=1}^{v} \phi_i x_{i,t-1} + \varepsilon_t, \tag{5}$$

By abstracting away from modelling the persistence of the commodity price series, the analysis may be polluted by identifying partial correlation structures which are driven by common persistence patterns across variables. However, it serves as a first approach to pinpoint the robustness of partial correlations between Arabica coffee price changes and the lagged variables entertained in the analysis. The first column of Table 2 shows the posterior inclusion probability (PIP) of each one of the covariates considered within the set of potential determinants. The PIP is defined as the sum of posterior model probabilities of specifications containing a particular variable, and is routinely interpreted as a measure of the importance of that covariate as a robust determinant of changes in the dependent variable under model uncertainty (see for example Fernández et al., 2001b; Ley and Steel, 2009). Our prior elicitation implies an expected prior inclusion probability of 0.5 for the variables considered, so we will label variables with PIP above 0.5 as *robust covariates*, since the confidence on model inclusion increases after observing the data.

While the results of the BMA exercise which explores models without an autoregressive component unveil robust partial correlations for covariates belonging to the group of fundamental, macroeconomic and financial variables (see first column in Table 2), most of these factors lose their robustness once the persistence of coffee prices is explicitly included in the model. In the second column of Table 2 we present the PIPs of the variables based on entertaining models that include the lagged dependent variable (with a one-month lag) in addition to the rest of the lagged covariates. Modelling the persistence of commodity prices leads to only two variables besides the lag of the dependent variable having PIP above 0.5. These covariates (US leading indicator and the yearly log-change in the US stock market index) reflect global macroeconomic and financial developments. The corresponding commodity market fundamentals (global coffee production or production in Brazil) lose their relevance as determinants of coffee price dynamics under model uncertainty once the persistence of this variable is modelled through a lag.

Table 2: BMA analysis: Posterior inclusion probabilities

	$q_0 = 0$ $q_v = 1$	$q_0 = 1$ $q_v = 1$	$q_0 = 2$ $q_v = 2$
Coffee production in Brazil	**1.00**	0.02	0.01
Coffee production in the world	**0.81**	0.02	0.01
Output for EU	**0.51**	0.02	0.01
Output for US	0.14	0.02	0.23
Output for Brazil	**0.99**	0.03	0.14
Leading indicator, Germany	**0.64**	0.06	0.01
Leading indicator, US	**1.00**	**0.73**	0.01
Real effective exchange rate	**1.00**	0.03	0.01
Stock market index, EU	0.10	0.04	0.01
Stock market index, US	**1.00**	**0.59**	0.07
S&P GS commodity index	0.27	0.03	0.01
Precipitation	0.10	0.02	0.03
Temperature	0.12	0.02	0.01
Lagged dependent variable	–	**1.00**	**1.00**

PIPs based on the class of models defined by equation (1). Bold figures if PIP>0.5
PIPs in column "$q_{m0} = 2$ and $q_{mv} = 2$" correspond to the lag with maximum PIP.
Dependent variable is the annual log-change in Arabica coffee price.
Results based on two million MCMC replications after one million burn-in draws.

Finally, in the third column of Table 2 we enlarge the number of potential lags of the dependent and independent variables included in the model to two, thus allowing for more complex dynamic relationships between changes in the commodity price and its determinants. For each variable, the figures in Table 2 present the maximum PIP for the two lags included. Once more complex autoregressive dynamics are allowed for in the specification, no single covariate besides the lagged dependent variables achieve PIP above the implied prior expectation. Such a result is a reflection of the difficulty of finding individual drivers that are able to explain historical changes in commodity prices once the autoregressive structure of price changes is accounted for.

On the one hand, our results emphasize the role of global macroeconomic developments and financial markets as drivers of coffee prices. However, the large degree of model uncertainty renders this result unrobust once slightly more complex autoregressive dynamics are assumed for Arabica coffee price changes. Although historical in-sample dynamics may be difficult to assess with the covariates proposed, these may still contain information which is useful for out-of-sample forecasting of coffee prices. In the following section we provide a comprehensive assessment of the out-of-sample predictive power of fundamental, macroeconomic, financial and climatic variables for commodity prices, applying the analysis to Arabica coffee prices.

3 Out-of-sample predictability analysis

3.1 The econometric setting: Prediction models and forecast averaging techniques

In order to provide a comprehensive analysis of the predictive power of the variables put forward above for Arabica coffee prices, we consider a battery of univariate and multivariate model structures as potential prediction models and perform a systematic comparison of their out-of-sample forecasting power. We consider a large number of univariate and multivariate models as well as forecast combination methods for variables corresponding to the different categories described above (fundamentals, macroeconomic, financial and climatic variables). The particular models and combination methods used for the analysis are presented in Table 3.

Table 3: Models and combination methods

Abbreviations	Description
	Individual models
AR(p)	Autoregression in levels with p lags
DAR(p)	Autoregression in first differences with p lags
s-AR(p)	Subset autoregression in levels with p lags
s-DAR(p)	Subset autoregression in first differences with p lags
ARCH(p,q)	Autoregression conditional heteroskedasticity in levels with p lags in mean equation and q lags in variance equation
DARCH(p,q)	Autoregression conditional heteroskedasticity in first differences with p lags in mean equation and q lags in variance equation
GARCH(p,q)	Generalized autoregression conditional heteroskedasticity in levels with p lags in mean equation and q lags in variance equation
DGARCH(p,q)	Generalized autoregression conditional heteroskedasticity in first differences with p lags in mean equation and q lags in variance equation
VAR(p)	Vector autoregression in levels with p lags
DVAR(p)	Vector autoregression in first differences with p lags
VEC(c,p)	Vector error correction model with c cointegration relationships and p lags
s-VAR(p)	Subset vector autoregression in levels with p lags
s-DVAR(p)	Subset vector autoregression in first differences with p lags
BVAR(p)	Bayesian vector autoregression in levels with p lags
BDVAR(p)	Bayesian vector autoregression in first differences with p lags
	Forecast combination methods
mean	Forecasting combination based on mean of individual predictions
tmean	Forecasting combination based on trimmed mean of individual predictions
median	Forecasting combination based on median of individual predictions
OLS	Forecasting combination based on pooling using OLS
PC	Forecasting combination based on principal components
DMSFE	Forecasting combination based on discounted mean square forecast errors
HR	Forecasting combination based on hit rates
EHR	Forecasting combination based on exponential of hit rates
EEDF	Forecasting combination based on the economic evaluation of directional forecasts
BMA	Forecasting combination based on Bayesian model averaging weights using the predictive likelihood
FMA-aic	Forecasting combination based on AIC weights
FMA-bic	Forecasting combination based on BIC weights
FMA-hq	Forecasting combination based on Hannan-Quinn weights

Within the class of linear univariate time series models, we consider autoregressive (AR) models, where the price of Arabica coffee is assumed to depend on its own p lags and a random white noise shocks. Alternatively, autoregressive models in first differences are also considered as potential model structures for the price variable, as well as specifications whose dynamics are driven by heteroskedastic disturbances in the form of autoregressive conditional heteroskedastic (ARCH) and generalized autoregressive conditional heteroskedastic (GARCH) errors. A group of linear multivariate time series specifications are also entertained in our analysis. In such specifications, we consider the Arabica coffee price, P_t as an element of the vector x_t, which includes other fundamental, macroeconomic, financial or climatic variables. The vector x_t is assumed to depend on its past values and on a multivariate normal random shock, so that

$$x_t = \Psi_0 + \sum_{l=1}^{p} \Psi_l x_{t-l} + \varepsilon_t, \qquad \varepsilon_t \sim \textbf{NID}(\mathbf{0}, \Sigma_\varepsilon), \tag{6}$$

where Ψ_l for $l = 1, \ldots, p$ are matrices of coefficients and Ψ_0 is a vector of intercept terms. Instead of assuming a relationship in levels, it can be assumed that the linear linkage is among first differences of the variables, so that the corresponding model would be given by

$$\Delta x_t = \chi_0 + \sum_{l=1}^{p} \chi_l \Delta x_{t-l} + \mu_t, \qquad \mu_t \sim \textbf{NID}(\mathbf{0}, \Sigma_\mu). \tag{7}$$

Alternatively, if the elements of x_t are integrated of order one and linked by a cointegration relationship, the vector error correction representation would be used and is given by

$$\Delta x_t = \delta_0 + \lambda\beta' x_{t-1} + \sum_{l=1}^{p-1} \delta_l \Delta x_{t-l} + u_t, \qquad u_t \sim \textbf{NID}(\mathbf{0}, \Sigma_u). \tag{8}$$

Following a similar systematic approach to out-of-sample prediction as that put forward in Costantini et al. (2016) and Crespo Cuaresma et al. (2018), combinations of out-of-sample predictions of individual specifications of these classes of univariate and multivariate models will be considered in addition to the forecasts of each individual model. The assessment of predictive ability is based on a series of profit/cost measures. Denoting $\hat{P}_{c,t+h|t}$ the forecast of the price of Arabica coffee for time $t+h$ conditional on the information available at time t obtained by model or forecast combination method c, $c = 1, \ldots, M$, the loss measures we evaluate include the standard square forecast error,

$$SE_{c,t,h} = \left(\hat{P}_{c,t|t-h} - P_t\right)^2 \tag{9}$$

and the absolute error

$$AE_{c,t,h} = \left| \hat{P}_{c,t|t-h} - P_t \right|, \tag{10}$$

which are standard loss measures in assessments of forecasting models for continuous variables.

Denoting T_3 the end of the available data and T_2 the beginning of the out-of-sample period, the statistics of interest based on these two measures of predictive error are the *mean square error* (MSE) at horizon h

$$MSE_{c,h} = \frac{1}{T_3 - T_2 + 1} \sum_{j=0}^{T_3 - T_2} SE_{c,T_2+j,h} \tag{11}$$

and the *mean absolute error* (MAE) at horizon h,

$$MAE_{c,h} = \frac{1}{T_3 - T_2 + 1} \sum_{j=0}^{T_3 - T_2} AE_{c,T_2+j,h}. \tag{12}$$

In our forecast analysis we also use composite forecasts based on the relative performance of particular methods over certain out-of-sample periods. In particular, for this technique at each time point t we choose the model or forecast combination method (and thus also the forecast for time point $t+h$) with the best performance (i.e. minimum MSE and/or MAE) over a certain time window ending at time point t. Namely,

$$\hat{P}^{MSE,l}_{t+h|t} = \hat{P}_{c^{MSE}_{mlth},t+h|t} \quad \text{where} \quad c^{MSE}_{mlth} = \text{argmin}_c \sum_{j=l}^{t} SE_{c,j,h}. \tag{13}$$

Time point l, such that $T_2 \leq l \leq t$, defines the beginning of the window over which the performance is evaluated, i.e., the evaluation window is $[l, t]$ where $l \leq t \leq T_3$. In a similar way

$$\hat{P}^{MAE,l}_{t+h|t} = \hat{P}_{c^{MAE}_{mlth},t+h|t} \quad \text{where} \quad c^{MAE}_{mlth} = \text{argmin}_c \sum_{j=l}^{t} AE_{c,j,h}. \tag{14}$$

In terms of profit measures, we use directional accuracy (*DA*), directional value (*DV*), the returns from a trading strategy generated by our forecasts and a risk adjusted performance measure given by the Sharpe ratio.

The DA measure is given by

$$DA_{c,t,h} = I\left(\text{sgn}(P_t - P_{t-h}) = \text{sgn}(\hat{P}_{c,t|t-h} - P_{t-h})\right) \tag{15}$$

where $I(\cdot)$ is the indicator function. $DA_{c,th}$ is thus a binary variable indicating whether the direction of the price change was correctly forecast at horizon h ($DA_{c,t,h} = 1$) or not ($DA_{c,t,h} = 0$). The economic value of directional forecasts is better captured by assigning to each correctly predicted change its magnitude (see Blaskowitz and Herwartz, 2011). We use the directional value (DV) statistic for this purpose,

$$DV_{c,t,h} = |P_t - P_{t-h}| DA_{c,t,h}. \tag{16}$$

We entertain composite forecasts based on forecasts from all models and forecast combination methods. At each time point t, we choose the model or forecast combination method, and thus also the forecast for time point $t+h$, with the largest DA or DV over certain time window ending at time point t. That is,

$$\hat{P}^{DA,l}_{t+h|t} = \hat{P}_{c^{DA}_{mlth},t+h|t} \qquad \text{where} \qquad c^{DA}_{mlth} = \text{argmax}_c \sum_{j=l}^{t} DA_{c,j,h}, \tag{17}$$

where l, $T_2 \leq l \leq t$, defines the beginning of the window over which is the performance evaluated, i.e., the evaluation window is $[l, t]$ where $l \leq t \leq T_3$. In a similar way

$$\hat{P}^{DV,l}_{t+h|t} = \hat{P}_{c^{DV}_{mlth},t+h|t} \qquad \text{where} \qquad c^{DV}_{mlth} = \text{argmax}_c \sum_{j=l}^{t} DV_{c,j,h}. \tag{18}$$

The performance of Arabica coffee price forecasts based on their profitability is also evaluated by the returns or Sharpe ratios implied by a simple trading strategy that is based on predictions. Selling/buying signals are based on the difference between the current spot price and the forecast for horizon h. Positive returns are executed as long positions while negative returns are executed as short positions (see for example Gencay, 1998). The (discrete) return of the spot price at time t over period h is $r_{th} = P_t/P_{t-h} - 1$. If the trading signal implied my model or model combination c at time t is given by

$$y_{c,t-h,h} = \begin{cases} -1, & \text{for selling signal (forecast downward movement for horizon } h) \\ & \hat{P}_{c,t|t-h} < P_{t-h}, \\ & \\ 1, & \text{for buying signal (forecast upward movement for horizon } h) \\ & \hat{P}_{c,t|t-h} > P_{t-h}, \end{cases} \tag{19}$$

then the return of the trading strategy (at time t over period h) implied by model c is $R_{c,th} = y_{c,t-h,h} r_{th}$ for $t = 1, \ldots, n$, and the total return of the trading strategy over n periods, i.e., over interval $[t, t+n]$, implied by model c and with respect to all realized $h-$period returns ($h \leq n$), is given by

$$R_{c,h,[t,t+n]} = \frac{1}{h} \sum_{j=0}^{h-1} \Pi_{i=0}^{n_j} \left(R_{c,t+j+ih,h} + 1 \right) - 1 \tag{20}$$

where n_j, $j = 1, \ldots, h-1$, is the largest integer such that $t + j + n_j h \leq n$.[4]

As in the previous cases, we create an aggregate/composite forecast with the maximum averaged or realized return – based on forecasts from all models and forecast combination methods. I.e., at each time point t we choose the model or forecast combination method, and thus also the forecast for time point $t+h$, with the largest average return over time window $[l,t]$, namely

$$\hat{P}^{TS,l}_{t+h|t} = \hat{P}_{c^{TS}_{mlth},t+h|t} \qquad \text{where} \qquad c^{TS}_{mlth} = \text{argmax}_c \sum_{k=l}^{t} R_{c,k,h} \tag{21}$$

and the largest total realized return until time point t, namely

$$\hat{P}^{TS}_{t+h|t} = \hat{P}_{c^{TS}_{mth},t+h|t} \qquad \text{where} \qquad c^{TS}_{mth} = \text{argmax}_c R_{c,h,[1,t]}. \tag{22}$$

We also perform comparisons based on *Sharpe ratios* - the excess return per unit of deviation generated by a trading strategy. In our application we take zero return as a benchmark return in the definition of the Sharpe ratio.

The forecast averaging methods employed use different weights, with some of the schemes using the predictive ability of each one of the specifications to compute them. Starting with the simplest methods, forecast pooling based on the *mean* uses the average of the forecasts of the individual models. The *trimmed mean* method uses the same type of weighting after discarding the lowest and highest forecast generated by the set of models considered. The *median* combination method uses the median of the predictions produced by the battery of specifications entertained.

Granger and Ramanathan (1984) propose to use weights based on the parameter estimates obtained from regressing the actual realizations in a hold-out sample on the corresponding forecasts from the individual models. We denote this combination method *OLS*. To avoid potential problems caused by multicolinearity, we also use a similar forecast pooling method based on building OLS weights based on the principal components of the model-specific forecasts instead of the individual set of predictions (*PC*). Stock and Watson (2004) put forward a forecast combination technique based on discount mean square forecast errors (*DMSFE*) which corresponds to using weights in equation (23) which depend

[4] Note that for $h = 1$ is the total return over $[t, t+n]$ given by $\Pi_{i=0}^{n} \left(R_{c,t+i,h} + 1 \right) - 1$. Note in addition that the total return given by equation (20) is the average of all possible $h-$period returns. We decided to proceed this way so as to take into account all $h-$step ahead forecasts.

inversely on the discounted squared forecast errors obtained in the hold-out sample for each model. Such a discounting scheme implies that the recent predictive performance of the individual models is considered more relevant for this weighting strategy.

We also use a combination method based on the proportion of correctly predicted directions of change in the Arabica coffee price by model i (the hit rate, *HR*), as well as a pooling strategy based on the exponential of the hit rate (*EHR*), a method put forward, for instance, in Bacchini et al. (2010). While these methods base the weight of the individual specifications on their ability to predict direction of change, we can also construct weights based on the economic evaluation of directional forecasts (*EEDF*), that is, taking into account the magnitude of the realized change in the price. In this case, the weights are built using the relative performance of the individual models in terms of the variable created by multiplying the absolute change in the price by a variable that takes value one if the direction of change was forecast adequately and zero otherwise.

Bayesian model averaging (*BMA*) techniques provide a framework which can be used to construct weights for pooling forecasts. In the spirit of weighting based on posterior model probabilities, weights for the individual models can be obtained making use of the Laplace approximation of the marginal likelihood of each model evaluated using the out-of-sample forecast errors, as proposed by Kapetanios et al. (2006). While the Laplace approximation of the marginal likelihood relies on the use of the Bayesian Information Criterion (BIC), frequentist approaches also propose the use of the Akaike Information Criterion (*AIC*) or the Hannan-Quinn Information Criterion (HQ) as alternatives to the BIC when building the model averaging weights see (see, for instance, Claeskens et al., 2008).

The pooled forecast methods considered in this analysis build linear combination of the predictions of individual specifications,

$$\hat{P}_{c,t+h|t} = w^h_{c,0t} + \sum_{i=1}^{F} w^h_{c,it}\hat{P}_{i,t+h|t}, \tag{23}$$

where c is the combination method, F is the number of individual forecasts and the weights are given by $\{w^h_{c,it}\}_{i=0}^{F}$. Table 4 presents the exact definition of the weights corresponding to each one of the methods entertained.[5]

3.2 Out-of-sample results on Arabica coffee price

We base our comparisons on monthly data spanning the period from January 1985 until March 2016 for Arabica coffee. The beginning of the hold-out forecasting sample for individual models used in order to obtain weights based on predictive accuracy is given by January 2000. The beginning of the actual out-of-sample forecasting sample is January 2005, and the end of the data sample is March 2016. In

[5] We use also the median of forecasts, i.e., $\hat{P}_{\text{median},t+h|t} = \text{median}\{\hat{P}_{c,t+h|t}\}_{c=1}^{M}$, which can not be expressed by (23).

Table 4: Weights of forecast combination methods

Method	Weights, w_{it}^h
Mean	$\frac{1}{k}$
Trimmed mean	$\frac{1}{k-2}$ where the smallest and largest forecasts are discarded
OLS	coefficients from regressing actual values on forecasted values
PC	coefficients from regressing actual values on factors
DMSFE	$\sum_{s=T_1-1+h}^{t} \theta^{T-h-s} \left(P_{s+h} - \hat{P}_{i,s+h\vert s}\right)^2$ where $\theta = 0.95$ is a discount factor
HR	$\frac{\sum_{j=T_1+h-1}^{t} DA_{i,jh}}{\sum_{c=1}^{M}\left(\sum_{j=T_1+h-1}^{t} DA_{c,jh}\right)}$ where $DA_{c,jh} = I\left(\text{sgn}(P_j - P_{j-h}) = \text{sgn}(\hat{P}_{c,j\vert j-h} - P_{j-h})\right)$ and $I(\cdot)$ is the indicator function
EHR	$\frac{\exp\left(\sum_{j=T_1+h-1}^{t}(DA_{i,jh}-1)\right)}{\sum_{c=1}^{M}\exp\left(\sum_{j=T_1+h-1}^{t}(DA_{c,jh}-1)\right)}$
EEDF	$\frac{\sum_{j=T_1+h-1}^{t} DV_{i,jh}}{\sum_{c=1}^{M}\left(\sum_{j=T_1+h-1}^{t} DV_{c,jh}\right)}$ where $DV_{c,th} = \vert P_t - P_{t-h}\vert DA_{c,th}$
BMA	$\frac{(t-T_1-h+2)^{\frac{p_1-p_i}{2}}\left(\frac{\sum_{j=T_1+h-1}^{t} SE_{1,jh}}{\sum_{j=T_1+h-1}^{t} SE_{i,jh}}\right)^{\frac{t-T_1-h+2}{2}}}{\sum_{c=1}^{M}(t-T_1-h+2)^{\frac{p_1-p_l}{2}}\left(\frac{\sum_{j=T_1+h-1}^{t} SE_{1,jh}}{\sum_{j=T_1+h-1}^{t} SE_{c,jh}}\right)^{\frac{t-T_1-h+2}{2}}}$ where $SE_{c,th} = \left(\hat{P}_{c,t\vert t-h} - P_t\right)^2$
FMA	$\frac{\exp\left(-\frac{1}{2}IC_{it}\right)}{\sum_{c=1}^{M}\exp\left(-\frac{1}{2}IC_{ct}\right)}$ where IC_{ct} is the information criterion of model c and t is the last time point of the data over which are models estimated

a first stage, the forecasting exercise is performed for groups of variables corresponding to each one of the groups (as introduced in section 2.3, see Table 1), in order to assess the relative performance of each one of the potential types of determinants of coffee price dynamics. In particular, multivariate time series models are estimated for each one of the possible combinations of variables within a group. The lag length of each multivariate model specifications under consideration is selected using the AIC criterion for potential lag lengths ranging from 1 to 6 lags. For the case of VEC models, selection of the lag length and the number of cointegration relationships is carried out simultaneously using the AIC as a model selection criterion. That is, for a given choice of variables in a group, VEC models for all possible cointegration relationships are estimated and the specification corresponding to the best AIC is chosen. We also estimate subset-VAR specifications, where individual (insignificant) parameters of the VAR specification are set equal to zero recursively using t-tests.

Tables 5 and 6 summarizes the results of the forecast performance analysis for each individual group and forecast horizons of one, three, six, nine and twelve months. When comparing among the forecast horizons, one can see that the MAE and the MSE increase with increasing horizon. Thus, for each group the smallest mean square and mean absolute errors are obtained when forecasting one month ahead where, when comparing among the groups for one month forecast horizon, the smallest forecast errors were obtained for the group of fundamental variables for the MSE and the group of financial variables for the MAE. The similar feature is observed when the forecast performance is measured by the average annual returns implied by the trading strategy where for each group the return decreases with increasing horizon and the highest return is obtained for the group of macroeconomic variables and forecast horizon of one month ahead. On the other hand, the highest profit-based performance measures like the directional accuracy, directional value and the Sharpe ratio are obtained for the highest forecast horizon of twelve months and the group of macroeconomic variables (namely, 74.8% for directional accuracy, 83.2% for the directional value and 0.64 for the Sharpe ratio). These profit-based measures (with exception of the return) increase with increasing horizon for the groups of fundamental[6] and macroeconomic variables and for all groups when the performance is measured by the Sharpe ratio. In contrast to this, for the group of financial variables the profit-based performance (based on DA and DV) increases till forecast horizon of three months and then decreases. Regarding the group of climatic variables the directional value decreases with increasing forecast horizon.

Among all forecast horizons, the best performance is by far achieved by the group of macroeconomic variables. The group of macroeconomic variables outperforms other groups for all forecast horizons and all performance measures except for the loss measures for the one month forecast horizon, where the group of fundamental variables achieved the smallest mean square forecast error and the group of financial variables achieved the smallest mean absolute error.

[6]The only exception is the forecast horizon of three months for the group of fundamental variables and the directional value performance measure.

With respect to the *loss measures*, the smallest MSE is obtained for the group of fundamental variables implied by composite forecasts[7] and one month forecast horizon and the smallest MAE is obtained for the group of financial variables implied by the VEC model and one month forecast horizon. For forecast horizons larger than one month the smallest loss measures are always obtained for the group of macroeconomic variables with the best model being VAR or s-VAR with variables such as output for the US and Brazil, y^{US}, y^{BR} and the leading indicator for Germany, li^{EU}.[8]

The highest *directional accuracy* (or the hit rate) is always given by the group of macroeconomic variables for all forecast horizons where the largest hit rate of 74.8% is achieved for forecast the horizon of twelve months. The best models are always multivariate models (s-DVAR, VEC and VAR) where the following variables occur the most often: leading indicator for Germany (for all forecast horizons but $h = 3$), leading indicator for the US (for all forecast horizons but $h = 9$) and the real effective exchange rate (for $h > 1$). In addition, the output variables appear in the best models for the following horizons: $h > 6$ for the EU, $h = 1, 6$ for the US and $h = 1, 9$ for Brazil.

The highest *directional value* for all horizons is again obtained for the group of macroeconomic variables, with the largest DV of 83.2% achieved for a forecast horizon of twelve months. The best models are always multivariate models (VAR and s-VAR), in which the leading indicator for the US appears in specifications for all forecast horizons, the real effective exchange rate is included in the best models for all horizons but twelve months ahead, while the output for Brazil enters the best models for all horizons but one month ahead. The leading indicator for Germany enters models for forecast horizons of three, six and nine months while the output for the EU does not appear in any of the best models. Note that models for six and nine months forecast horizons are the same, namely s-VAR(2) where all macroeconomic variables appear except for the leading indicator for Germany.

The largest average annualized *return* implied by the trading strategy for all horizons is again obtained for the group of macroeconomic variables, where the largest return of 27.9% is achieved for the forecast horizon of one month (returns decrease with increasing horizons) where only output variables for the EU and Brazil are included in the DVAR model. The best models (based on the highest realized return with chosen variables from the macroeconomic group) are: VEC for $h = 3$, s-VAR for $h = 6, 9$ (with the same variables, namely output for the US and Brazil, leading indicators for Germany and the US and the real effective exchange rate) and VAR for $h = 12$.

Finally, the best performance with respect to the *Sharpe ratio* (SR) is achieved again for the group of macroeconomic variables, with the Sharpe ratio increasing with longer horizons and the highest SR

[7] In this case 64% of the time a univariate model is picked up as the best model, 30% of the time a multivariate model is picked up that includes the Arabica coffee production in Brazil and 30% of the time a multivariate model is picked up that includes the world Arabica coffee production.

[8] There are two exceptions: MAE for forecast horizon of three months where the only variable, except for the coffee price, was the real effective exchange rate and both MAE and MSE for horizon of twelve months where the leading indicator for Germany was replaced by the leading indicator for the US.

obtained for the forecast horizon of twelve months. Regarding the models and chosen variables, the results are identical to the ones for returns.

In the second stage, we undertake the forecast performance analysis based on models created out of combinations of variables from all categories under consideration (see Table 1), following the same logic as in the case of specifications built from group-specific covariates. In particular, we entertain the specifications that can be obtained from all possible combinations of variables from the different categories that were included in the models with the best predictive performance depicted in Tables 5 and 6. The findings are summarized in Table 7. The performance results improve with respect to the results for specifications based on group-specific variables (as presented in Tables 5 and 6), for all forecast horizons and performance measures (there are two exceptions: the forecast horizon of one month and the mean absolute error, where the smallest MAE is based solely on the group of financial variables and the forecast horizon of three months and the mean square error, where the smallest MSE is based only on some macroeconomic variables, namely the output variables for the US and Brazil, y^{US}, y^{BR}, and the leading indicator for Germany, li^{EU}). Thus, the combination of variables from all individual groups helps to improve the performance for almost all forecast horizons. Another general observations we can extract from the results based on the full set of explanatory factors are:

- Macroeconomic and/or financial variables appear in all best models for all performance measures and forecast horizons. This indicates, on the one hand, that information that goes beyond that included in the dynamics of fundamentals is relevant for predicting future coffee price changes. To a certain extent, present and past changes in the price of coffee appears to contain the relevant predictive information concerning market fundamentals and further information on these variables have little systematic impact on out-of-sample predictive ability.

- In case of the loss-based performance measures, climatic variables (namely precipitation) tend to be included in the best models for longer forecast horizons (starting from six months ahead), where the same model, namely, s-VAR(2), with y^{US}, y^{BR}, li^{EU} and precipitation, was chosen as the one generating the smallest forecast error;[9]

- Fundamental variables, namely Arabica coffee production in Brazil, y^{BR}_{coffee}, appear only in best models for shorter forecast horizons (one and three months);

- Models that appear most often among the best are: VEC with one cointegration relationship, followed by s-VAR and VAR models. The same VEC model (with two lags and the combination of the following macroeconomic and financial variables: y^{EU}, y^{BR}, li^{EU}, li^{US}, $stock^{US}$ and GSCI) gives the highest profit-base measures (DA, DV, returns and Sharpe ratio) for the forecast horizon

[9]The only exception is the MAE with forecast horizon of six months where also $stock^{US}$ was included among the set of variables.

of six months and also for the return and the Sharpe ratio for $h = 3$. In addition, another VEC model (again with two lags) with macroeconomic and financial variables (y^{EU}, y^{US} y^{BR}, li^{EU}, li^{US} and GSCI) performs the best in terms of profit-based measures for $h = 9$ and $h = 12$ (except for the DV measures and $h = 12$).

- Forecasts implied by univariate models, forecast combination methods or composite forecasts never appear as the ones that generate the best performance for any loss- or profit-based measures (for any forecast horizon).

In more detail, for one month forecast horizon the improvement in performance (with respect to the results based on variables from individual groups, see Table 5) is given by the combination of: fundamental, macroeconomic and financial variables for the MSE and the DV measures; macroeconomic, financial and climatic variables for the DA measure; and the macroeconomic and financial variables for the return and the SR. For three months forecast horizon the performance improvement is implied by the combination of: fundamental, macroeconomic and financial variables for the MAE; macroeconomic and financial variables for the directional accuracy, return and the Sharpe ratio; and fundamental and macroeconomic variables for the directional value. For six months forecast horizon the performance is improved by the combination of: macroeconomic, financial and climatic variables for the MAE; macroeconomic and climatic variables for the MSE; and macroeconomic and financial variables for all profit-based measures. For forecast horizons of nine and twelve months, the combination of the following variables improves the performance: macroeconomic and climatic variables for loss-based measures and for the DV (and $h = 12$); and macroeconomic and financial variables for all profit-based measures (except for the DV and $h = 12$).

The smallest forecast errors, MSE and MAE, are obtained for one month forecast horizon (similar result as in the group based analysis, see Tables 5), the largest directional accuracy (82.2%) is reached for twelve months forecast horizon (the largest DA value in the group based analysis was 74.8% for twelve month forecast horizon) as well as the largest Sharpe ratio of 0.7. The largest directional value of 87.5% was achieved for six months forecast horizon. Finally, the biggest annual return of 31.9%, implied by the trading strategy, was observed for one month forecast horizon.[10]

3.3 Data snooping bias free test for equal performance

To assess whether the performance superiority of the "best" models with respect to the simple benchmark model such as the random walk model[11] is systematic and not due to luck, we perform the bootstrap

[10]Note that the largest DV in the group based analysis was 83.2%, the largest annualized return was 27.9% and the largest Sharpe ratio reached 0.6.

[11]More precisely the benchmark model for MAE, MSE measures and returns implied by the trading strategy is the random walk model, while for DA and DV measures it is the random walk with an intercept.

stepwise multiple superior predictive ability test (stepM-SPA) by Hsu et al. (2010). The test is based on the bootstrap method of Politis and Romano (1994), the stepwise test of multiple check by Romano and Wolf (2005) and the test for superior predictive ability of Hansen (2005).

The following relative performance measures, $d_{c,th}$, implied by model c, $t =$ January 2005 to March 2016, $h = 1, 3, 6, 12$ months are computed and the tests are defined based on them

$$d_{c,th} = \begin{cases} SE_{RW,th} & - & SE_{c,th} \\ AE_{RW,th} & - & AE_{c,th} \\ DA_{c,th} & - & DA_{RW,th} \\ DV_{c,th} & - & DV_{RW,th} \\ R_{c,th} & - & R_{RW,th} \end{cases} \tag{24}$$

RW stands for the random walk. Note that the bootstrap test cannot be performed for the Sharpe ratio, as for negative values it is not true that larger values are associated with a better performance (the test involves the ordering of calculated statistics).

The bootstrap stepM-SPA test is a comprehensive test across all models and five performance measures under consideration and directly quantifies the effect of data snooping by testing the null hypothesis that the performance of the best model is no better than the performance of the benchmark model. The following individual testing problems are considered

$$H_0^c: \ \mathbb{E}\left(d_{c,th} \leq 0\right), \quad \text{versus} \quad H_A^c: \ \mathbb{E}\left(d_{c,th} > 0\right). \tag{25}$$

This multiple testing method yields a decision for each individual testing problem (by either rejecting H_0^c or not)[12] and is implemented as follows. Without loss of generality we assume that the sequence of averages od $d_{c,th}$ across the time, $\left\{\bar{d}_{ch}\right\}_c$, are arranged in a descending order. Top j_1 null hypotheses are rejected (i.e. top j_1 models outperform the benchmark) if $\sqrt{n}\,\bar{d}_{lh}$, $l = 1, \ldots, j_1$, where n is the number of observations, is greater than the bootstrapped critical value calculated using the stationary bootstrap method of Politis and Romano (1994). If none of the null hypotheses is rejected, the procedure terminates. Otherwise, $d_{1,th}, \ldots, d_{j_1,th}$ are removed from the data and the bootstrap simulation is applied to the rest of the data to obtain the new critical value. The procedure continues until no more null hypotheses are rejected. In our analysis we use significance levels of 5% and 10%.[13]

Results are presented in Table 7 where two stars (**) indicate that the null hypothesis that the best model does not outperform the benchmark model is rejected at the 5% significance level. We can

[12] The individual decisions are made such that the familywise error rate is asymptotically achieved at the significance level α, which is achieved by constructing a joint confidence region with a nominal joint coverage probability of $1 - \alpha$. The familywise error rate is defined as the probability of rejecting at least one true null hypothesis. For more details, see Romano and Wolf (2005).

[13] For more details on the test, see Hsu et al. (2010).

observe that the (significant) outperformance of best models with respect to all performance measures (except for the Sharpe ratio) is increased with increasing forecast horizon. The results indicate that there is no significant outperformance of the best models with respect to the random walk for forecast horizon of one month. For the forecast horizon of three months, the best models outperform the random walk model for directional accuracy and directional value measures (profit based measures DA and DV) while for the forecast horizon of six months the best models outperform the random walk for all performance measures under consideration but MAE. Finally, for the forecast horizon of nine and twelve months the best models (as presented in Table 7) for all performance measures under consideration outperform the random walk.

3.4 Is the forecasting performance of models for Arabica coffee prices asymmetric along the business cycle?

There are theoretical arguments that imply that the predictive power of econometric models may depend on the particular phase of the business cycle in which they are performed (Jurado et al., 2015). For the particular case of commodity prices, the results in (Gargano and Timmermann, 2014), for example, show that the forecast quality of econometric models for commodity prices differs between expansions and recessions. We evaluate the statistical significance of differences in predictive power in expansions relative to recessions using regressions of differences of forecast performance measures (M_{th}) between the benchmark model (the random walk, RW) and the best model in our battery of specifications on the recession indicator (D_t). We thus estimate the regression model

$$M_{RW,t,h} - M_{best,t,h} = c_0 + c_1 D_t + \varepsilon_t, \tag{26}$$

where the performance measure M_{th} is alternatively the absolute error AE_{th}, see equation (10); square forecast error SE_{th}, see equation (9); directional accuracy measure DA_{th}, see equation (15); directional value measure DV_{th}, see equation (16); the return R_{th} implied by the trading strategy as given by equation (20) and the Sharpe ratio implied by this return. The dummy variable D_t represents periods of recession when $D_t = 1$ and expansion $D_t = 0$ and is calculated based on the turning points of the growth cycle as captured by the corresponding OECD composite leading indicator. We perform the analysis based alternatively on recessions and expansions for the euro area, US, OECD and Brazil and the recession period is identified as the period following peak through trough. Negative and significant values of c_1 for loss measures (AE and SE) suggest that the best model is more accurate relative to the benchmark during expansions than during recessions. Regarding the profit measures (DA, DV, return and SR), positive and significant values of c_1 suggest that the best model is more accurate relative to the benchmark during expansions than during recessions while negative and significant values of c_1 suggest that the best model is more accurate during recessions the during expansions. Table 8 presents the results of the analysis by forecasting horizon and recession indicator. The figures in Table 8 without an asterisk show the forecasting horizons at which the best model performs significantly better than the benchmark in expansions, while figures with asterisk refer to significantly better performance in recessions. All in all, our results are not in line with those in Gargano and Timmermann (2014), who tend to find that commodity prices are more predictable in recessions than in expansions using exclusively standard forecast loss error measures.

Our results suggest that for forecast horizons of six months ahead and above (except for the DV and return measures, where even lower forecast horizons are relevant), the best models outperform the benchmark model in expansions for both loss and profit based measures for the euro area, US and OECD, while for Brazil this is the case (for less forecast horizons) only for profit based measures. In case of loss measures, the best models outperform the benchmark model in recessions for short-term horizons (one and three months ahead for absolute error and three months ahead for square error). The opposite occurs for the forecast horizon of twelve months ahead, where the absolute error of the best model is significantly below the absolute error of the benchmark model and the expansion period. These results hint at the existence of differences in the predictive accuracy of models over time, thus calling for the use of time-varying weights in forecast averaging exercise, in the spirit of the methods put forward by Onorante and Raftery (2016). Dynamic model averaging methodologies take explicitly into account such variation in weights across specifications and could be an important building block of potential future methodological developments in coffee price prediction models.

4 Conclusions

As is the case of many other commodities, price trends and volatility is a major concern for stakeholders in the coffee market. In exporting countries, the price volatility is a source of uncertainty in relation to export earnings and tax revenues, as well as instability in producer incomes, many of which are smallholders. Sustained low coffee prices can imply considerable social hardship in many export dependent countries. In importing countries, price volatility makes it difficult for roasters to control processing costs and affects profit margins along the supply chain. In the free market period since 1990, smallholder farmers in many countries have been more exposed to fluctuations in coffee prices, as the internal regulatory mechanisms in producing countries were predominantly dismantled. These price fluctuations have increased rural poverty as it became difficult for small producers to efficiently plan their resource allocations. As a result, risk management strategies are becoming increasingly recommended to producers in developing countries. However, the scope and applicability of these instruments can vary significantly depending on the nature of the underlying and direct drivers of price trends and volatility. The development of early warning mechanisms and prediction models for coffee prices is thus of particular importance for both the supply and the demand side of the market. This paper provides a first quantitative assessment of a comprehensive set of forward-looking drivers of Arabica coffee price formation, which can be employed when designing complementary early warning systems to those developed in the framework of the G20 Action Plan on Food Price Volatility and Agriculture, such as the Agricultural Market Information System (AMIS) and the Rapid Response Forum, GEO Global Agricultural Monitoring Initiative (GEOGLAM) for market and production international monitoring.

Our results indicate that information on global macroeconomic and financial developments is valuable for explaining the historical pattern of Arabica coffee price developments, as well as to improve out-of-sample predictions. The forecasting horizon plays an important role when it comes to choosing the adequate econometric specification both in terms of the covariates included in the model and the particular model structure. In addition, the performance of the best forecasting models varies significantly in recessions as compared to expansions for most prediction horizons.

Our results suggest that macroeconomic and financial market variables are more important to understand and predict coffee prices than previously assumed. This has important implications for how individual producers, including smallholder coffee producers and producer countries, should manage the consequences of commodity price risks. Predictive tools such as the ones presented in this paper appear to be key for the implementation of such a risk management systems.

Table 5: Summary of forecast performance of best models for Brazilian Arabica coffee over different variable groups: fundamentals, macro financial and other for time horizons one, three and six months ahead.

1-month horizon	MAE	MSE	DA	DV	return	Sharpe ratio
Fundamentals	6.983 last 9 months	**84.956** last month	57.463 last 9 months	71.453 last 6 months	25.390 last 3 months	0.289 last 3 months
Macroeconomic	7.045 VAR(1) y^{US} y^{BR} li^{EU} REER	87.566 VAR(1) y^{US} y^{BR} li^{EU} REER	**61.481** VAR(2) y^{US} y^{BR} li^{EU} li^{US}	**72.026** VAR(2) li^{US} REER	**27.912** DVAR(1) y^{EU} y^{BR}	**0.317** DVAR(1) y^{EU} y^{BR}
Financial	**6.974** VEC(1,1) $stock^{EU}$ $stock^{US}$ GSCI	85.750 last 6 months	60.000 s-DVAR(1) $stock^{EU}$ GSCI	70.108 last 6 months	24.861 last 12 months, DV	0.283 last 12 months, DV
Climatic	7.177 ARCH(2,4)	87.755 last 9 months	57.037 DVAR(1) temperature	71.733 last 6 months	25.079 last 6 months	0.285 last 6 months
3-months horizon	MAE	MSE	DA	DV	return	Sharpe ratio
Fundamentals	14.916 last 3 months	412.162 last 3 months	58.519 s-DAR(2)	70.929 s-DAR(2)	21.085 s-DAR(2)	0.327 s-DAR(2)
Macroeconomic	**14.837** s-VAR(3) REER	**361.876** s-VAR(2) y^{US} y^{BR} li^{EU}	**67.407** VEC(1,1) li^{US} REER	**79.476** VAR(1) y^{BR} li^{EU} li^{US} REER	**27.854** VEC(1,2) y^{EU} y^{US} li^{EU} li^{US} REER	**0.440** VEC(1,2) y^{EU} y^{US} li^{EU} li^{US} REER
Financial	14.978 last month	427.363 s-VAR(3) GSCI	61.364 last 3 months	73.403 last 3 months	21.085 s-DAR(2)	0.327 s-DAR(2)
Climatic	15.158 last month	418.840 last 9 months	61.364 last 3 months	71.431 last 3 months	21.085 s-DAR(2)	0.327 s-DAR(2)
6-months horizon	MAE	MSE	DA	DV	return	Sharpe ratio
Fundamentals	23.668 AR(3)	1006.104 whole	63.704 s-VAR(2) y^{BR}_{coffee}	76.588 s-VAR(2) y^{BR}_{coffee}	19.711 s-VAR(2) y^{BR}_{coffee}	0.399 s-VAR(2) y^{BR}_{coffee}
Macroeconomic	**22.619** s-VAR(2) y^{US} y^{BR} li^{EU}	**772.865** s-VAR(2) y^{US} y^{BR} li^{EU}	**72.593** VEC(2,2) y^{US} li^{EU} li^{US} REER	**82.489** s-VAR(2) y^{US} y^{BR} li^{EU} li^{US} REER	**24.961** s-VAR(2) y^{US} y^{BR} li^{EU} li^{US} REER	**0.524** s-VAR(2) y^{US} y^{BR} li^{EU} li^{US} REER
Financial	23.206 s-VAR(3) GSCI	1024.283 s-VAR(3) GSCI	60.741 s-DAR(2)	63.823 s-DAR(2)	7.865 s-DAR(2)	0.154 s-DAR(2)
Climatic	23.668 AR(3)	1050.843 last month	60.741 s-DAR(2)	63.823 s-DAR(2)	7.865 s-DAR(2)	0.154 s-DAR(2)

See Table 3 for the abbreviation of the models. **Bold** figures indicate the best performance among all groups but within certain forecast horizon and **bold** figures indicate the best performance among all groups and forecast horizons.

Table 6: Summary of forecast performance of best models for Brazilian Arabica coffee over different variable groups: fundamentals, macro financial and other for time horizons nine and twelve months ahead.

9-months horizon	MAE	MSE	DA	DV	return	Sharpe ratio
Fundamentals	31.581 AR(3)	1673.914 s-VAR(2) y^{BR}_{coffee}	65.185 s-VAR(2) y^{BR}_{coffee}	79.291 s-VAR(2) y^{BR}_{coffee}	19.587 s-VAR(2) y^{BR}_{coffee}	0.486 s-VAR(2) y^{BR}_{coffee}
Macroeconomic	**27.771** s-VAR(2) y^{US} y^{BR} li^{EU}	**1245.128** s-VAR(2) y^{US} y^{BR} li^{EU}	**74.074** s-DVAR(2) y^{EU} y^{BR} li^{EU} REER	**82.818** s-VAR(2) y^{US} y^{BR} li^{EU} li^{US} REER	**22.535** s-VAR(2) y^{US} y^{BR} li^{EU} li^{US} REER	**0.579** s-VAR(2) y^{US} y^{BR} li^{EU} li^{US} REER
Financial	30.986 s-VAR(3) GSCI	1712.960 s-VAR(3) GSCI	57.778 s-DAR(2)	60.274 s-DAR(2)	4.340 s-DVAR(1) GSCI	0.099 s-DVAR(2) GSCI
Climatic	31.463 VAR(2) precipitation	1787.762 AR(3)	61.481 s-DVAR(2) precipitation	60.274 s-DAR(2)	4.482 s-DVAR(1) temperature	0.102 s-DVAR(1) temperature
12-months horizon	MAE	MSE	DA	DV	return	Sharpe ratio
Fundamentals	37.992 s-VAR(2) y^{BR}_{coffee}	2365.755 s-VAR(2) y^{BR}_{coffee}	**74.815** s-VAR(2) y^{BR}_{coffee}	80.359 s-VAR(2) y^{BR}_{coffee}	18.716 s-VAR(2) y^{BR}_{coffee}	0.559 s-VAR(2) y^{BR}_{coffee}
Macroeconomic	**32.615** s-VAR(2) y^{US} y^{BR} li^{US}	**1760.559** s-VAR(2) y^{US} y^{BR} li^{US}	**74.815** s-DVAR(2) y^{EU} li^{EU} li^{US} REER	**83.169** VAR(3) y^{US} y^{BR} li^{US}	**20.655** VAR(1) y^{BR} li^{US} REER	**0.640** VAR(1) y^{BR} li^{US} REER
Financial	37.217 s-VAR(3) GSCI	2366.536 s-VAR(3) GSCI	57.037 s-DAR(2)	57.623 s-DAR(2)	3.524 s-DAR(2)	0.092 s-DAR(2)
Climatic	38.035 AR(3)	2487.673 AR(3)	58.519 VAR(2) precipitation temperature	57.623 s-DAR(2)	3.524 s-DAR(2)	0.092 s-DAR(2)

See Table 3 for the abbreviation of the models. **Bold** figures indicate the best performance among all groups but within certain forecast horizon and **bold** figures indicate the best performance among all groups and forecast horizons.

Table 7: Summary of forecast performance of best models for Brazilian Arabica coffee over variables with highest predictive power.

Forecast horizon	MAE	MSE	DA	DV	return	Sharpe ratio
1-month	**6.974** VEC(1,1) $stock^{EU}$ $stock^{US}$ GSCI	**83.609** VEC(1,2) y^{BR}_{coffee} y^{US} li^{EU} REER $stock^{US}$	68.148 s-VAR(2) y^{EU} y^{BR} li^{EU} $stock^{US}$ precipitation temperature	74.467 VAR(2) y^{BR}_{coffee} y^{US} li^{EU} li^{US} REER GSCI	**31.850** VEC(1,2) y^{EU} y^{BR} li^{EU} li^{US} $stock^{EU}$ GSCI	0.361 VEC(1,2) y^{EU} y^{BR} li^{EU} li^{US} $stock^{EU}$ GSCI
3-months	13.883 s-VAR(2) y^{BR}_{coffee} li^{EU} REER $stock^{US}$	361.876 s-VAR(2) y^{US} y^{BR} li^{EU}	70.370** VAR(1) y^{BR} li^{US} REER $stock^{US}$	81.275** s-VAR(3) y^{BR}_{coffee} li^{US} REER	31.808 VEC(1,2) y^{EU} y^{BR} li^{EU} li^{US} $stock^{US}$ GSCI	0.510 VEC(1,2) y^{EU} y^{BR} li^{EU} li^{US} $stock^{US}$ GSCI
6-months	21.697 s-VAR(2) y^{US} y^{BR} li^{EU} $stock^{US}$ precipitation	762.386** s-VAR(2) y^{US} y^{BR} li^{EU} precipitation	77.778** VEC(1,2) y^{EU} y^{BR} li^{EU} li^{US} $stock^{US}$ GSCI	**87.506**** VEC(1,2) y^{EU} y^{BR} li^{EU} li^{US} $stock^{US}$ GSCI	30.767** VEC(1,2) y^{EU} y^{BR} li^{EU} li^{US} $stock^{US}$ GSCI	0.685 VEC(1,2) y^{EU} y^{BR} li^{EU} li^{US} $stock^{US}$ GSCI
9-months	27.345** s-VAR(2) y^{US} y^{BR} li^{EU} precipitation	1217.996** s-VAR(2) y^{US} y^{BR} li^{EU} precipitation	79.259** VEC(1,2) y^{EU} y^{US} y^{BR} li^{EU} li^{US} GSCI	85.328** VEC(1,2) y^{EU} y^{US} y^{BR} li^{EU} li^{US} GSCI	25.879** VEC(1,2) y^{EU} y^{US} y^{BR} li^{EU} li^{US} GSCI	0.700 VEC(1,2) y^{EU} y^{US} y^{BR} li^{EU} li^{US} GSCI
12-months	32.098** s-VAR(2) y^{US} y^{BR} li^{EU} precipitation	1702.567** s-VAR(2) y^{US} y^{BR} li^{EU} precipitation	**82.222**** VEC(1,2) y^{EU} y^{US} y^{BR} li^{EU} li^{US} GSCI	84.965** VAR(1) y^{BR} li^{US} REER temperature	22.441** VEC(1,2) y^{EU} y^{US} y^{BR} li^{EU} li^{US} GSCI	**0.722** VEC(1,2) y^{EU} y^{US} y^{BR} li^{EU} li^{US} GSCI

See Table 3 for the abbreviation of the models. **Bold** figures indicate the best performance among all forecast horizons. Two stars (**) indicate that the null hypothesis that the best model does not outperform the benchmark random walk model is rejected at the 5% significance level

Table 8: Forecast horizons when the best models significantly outperform the benchmark model (RW) in either expansion or recession times.

	Euro area	USA	OECD	Brazil
AE	6, 9, 12	9, 12	6, 9, 12	1*, 3*, 12
SE	6, 9, 12	6, 9	6, 9, 12	3*
DA	6, 9, 12	3, 6, 9	3, 6, 9	3, 6
DV	3, 6, 9, 12	1, 6, 9	1, 3, 6, 9, 12	1, 3, 6, 12
return	3, 6, 9, 12	6, 9	6, 9	3, 6
Sharpe ratio	6, 9, 12	6, 9	6, 9	6

No star values indicate horizons when best models outperform the benchmark model in expansion times while star values indicate horizons when best models outperform the benchmark model in recession times.

References

Ahumada, H. and Cornejo, M. (2015). Explaining commodity prices by a cointegrated time series-cross section model. *Empirical Economics*, 48(4):1667–1690.

Ahumada, H. and Cornejo, M. (2016a). Forecasting food prices: The case of corn, soybeans and wheat. *International Journal of Forecasting*, 32(3):838–848.

Ahumada, H. and Cornejo, M. (2016b). Out-of-sample testing price discovery in commodity markets: the case of soybeans. *Agricultural Economics*, 47(6):709–718.

Algieri, B. (2014). A roller coaster ride: an empirical investigation of the main drivers of the international wheat price. *Agricultural Economics*, 45(4):459–475.

Bacchini, F., Ciammola, A., Iannaccone, R., and Marini, M. (2010). Combining forecasts for a flash estimate of euro area gdp. *Contributi ISTAT*, (3).

Bernard, J.-T., Khalaf, L., Kichian, M., McMahon, S., et al. (2008). Forecasting commodity prices: Garch, jumps, and mean reversion. *Journal of Forecasting*, 27(4):279.

Blaskowitz, O. and Herwartz, H. (2011). On economic evaluation of directional forecasts. *International Journal of Forecasting*, 27(4):1058–1065.

Borensztein, E. and Reinhart, C. M. (1994). The macroeconomic determinants of commodity prices. *Staff Papers (International Monetary Fund)*, 41(2):236–261.

Brockhaus, J., Huang, J., Hu, J., Kalkuhl, M., von Braun, J., and Yang, G. (2016). When do prices matter most? Rice, wheat, and corn supply response in China. In *Food Price Volatility and Its Implications for Food Security and Policy*, pages 435–456. Springer.

Claeskens, G., Hjort, N. L., et al. (2008). *Model selection and model averaging*, volume 330. Cambridge University Press Cambridge.

Costantini, M., Crespo Cuaresma, J., and Hlouskova, J. (2016). Forecasting errors, directional accuracy and profitability of currency trading: The case of eur/usd exchange rate. *Journal of Forecasting*, 35(7):652–668.

Crespo Cuaresma, J., Fortin, I., and Hlouskova, J. (2018). Exchange rate forecasting and the performance of currency portfolios. *Journal of Forecasting*, forthcoming.

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

Deaton, A. and Laroque, G. (2003). A model of commodity prices after sir arthur lewis. *Journal of Development Economics*, 71(2):289–310.

Egelkraut, T. M., Garcia, P., Irwin, S. H., and Good, D. L. (2003). An evaluation of crop forecast accuracy for corn and soybeans: Usda and private information agencies. *Journal of Agricultural and Applied Economics*, 35(01):79–95.

Feldkircher, M. and Zeugner, S. (2009). Benchmark Priors Revisited: On Adaptive Shrinkage and the Supermodel Effect in Bayesian Model Averaging. *IMF Working Paper*, WP/09/202.

Fernández, C., Ley, E., and Steel, M. F. (2001a). Benchmark Priors for Bayesian Model Averaging. *Journal of Econometrics*, 100:381–427.

Fernández, C., Ley, E., and Steel, M. F. (2001b). Model Uncertainty in Cross-Country Growth Regressions. *Journal of Applied Econometrics*, 16:563–576.

Foster, D. P. and George, E. I. (1994). The Risk Inflation Criterion for Multiple Regression. *The Annals of Statistics*, 22:1947–1975.

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

Gencay, R. (1998). The predictability of security returns with simple technical trading rules. *Journal of Empirical Finance*, 5(4):347–359.

Gilbert, C. L. (2010). How to understand high food prices. *Journal of Agricultural Economics*, 61(2):398–425.

Granger, C. W. and Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting*, 3(2):197–204.

Headey, D. and Fan, S. (2008). Anatomy of a crisis: the causes and consequences of surging food prices. *Agricultural economics*, 39(s1):375–391.

Husain, M. A. M. and Bowman, C. (2004). *Forecasting Commodity Prices: Futures Versus Judgment*. Number 4-41. International Monetary Fund Working Paper.

Jurado, K., Ludvigson, S., and Ng, S. (2015). Measuring uncertainty. *American Economic Review*, 105(3):1177–1216.

Just, R. E. and Rausser, G. C. (1981). Commodity price forecasting with large-scale econometric models and the futures market. *American Journal of Agricultural Economics*, 63(2):197–208.

Kapetanios, G., Labhard, V., and Price, S. (2006). Forecasting using predictive likelihood model averaging. *Economics Letters*, 91(3):373–379.

Kass, R. and Wasserman, L. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association*, pages 928–934.

Kohzadi, N., Boyd, M. S., Kermanshahi, B., and Kaastra, I. (1996). A comparison of artificial neural network and time series models for forecasting commodity prices. *Neurocomputing*, 10(2):169 – 181. Financial Applications, Part I.

Koop, G. (2003). Bayesian Econometrics. *John Wiley & Sons*.

Ley, E. and Steel, M. F. (2009). On the Effect of Prior Assumptions in Bayesian Model Averaging with Applications to Growth Regressions. *Journal of Applied Econometrics*, 24:4:651–674.

Ley, E. and Steel, M. F. (2012). Mixtures of g-priors for bayesian model averaging with economic applications. *Journal of Econometrics*, 171(2):251–266.

Liang, F., Paulo, R., Molina, G., Clyde, M. A., and Berger, J. O. (2008). Mixtures of g Priors for Bayesian Variable Selection. *Journal of the American Statistical Association*, 103:410–423.

Onorante, L. and Raftery, A. E. (2016). Dynamic model averaging in large model spaces using dynamic occam's window. *European Economic Review*, 81:2–14.

Stock, J. H. and Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23(6):405–430.

Zellner, A. (1986). *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti*, chapter On Assessing Prior Distributions and Bayesian Regression Analysis with g-Prior Distributions. North-Holland: Amsterdam.