AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

Energy and Food Commodity Prices Linkage:

An Examination with Mixed-Frequency Data

Andres Trujillo-Barrera

(Marketing and Consumer Behaviour Group, Wageningen University)

Joost M.E. Pennings

(Marketing Department, Maastricht University)

(Marketing and Consumer Behaviour Group, Wageningen University)

Selected Paper prepared for presentation at the Agricultural & Applied Economics

Association 2013 AAEA & CAES Joint Annual Meeting, Washington, DC, August 4-6 2013.

Copyright 2013 by Andres Trujillo-Barrera and Joost M.E. Pennings. All rights reserved.
Readers may make verbatim copies of this document for non-commercial purposes by any means, provided this copyright notice appears on all such copies.

Energy and Food Commodity Prices Linkage:

An Examination with Mixed-Frequency Data

(Preliminary draft)

Abstract

Is the relationship between energy and agricultural commodities an important factor in the increasing price variability of food commodities? Findings from the literature appear to be mixed and highly influenced by the data frequency used in those analysis. A recurrent task in time series applied work is to match up data at different frequencies, while macroeconomic variables are often found at monthly or quarterly observations, financial variables are sampled daily or even at higher frequencies. In order to match up time series at different frequencies a common procedure is to aggregate the higher frequency to fit in the low frequency, this has the potential of losing valuable information, and generating misspecification. We study whether the use of mixed frequency estimations with data for the 2006-2011 period helps to improve the out of sample performance of a model that explains grain prices as a function of energy prices, macroeconomic variables such as exchange rate, interest rate, and inflation. Preliminary results suggest that an improvement is feasible, however it is tenuous beyond two months horizons.

Energy and Food Commodity Prices Linkage:

An Examination with Mixed-Frequency Data

Strong co-movements between energy and food commodity prices in recent years led to the notion that energy prices are one of the main contributing factors of the increasing food price variability. This relationship has been prompted by volatile fuel prices coupled with legislative mandates to increase biofuel production, particularly the U.S. Energy Independence and Security act of 2007 (Abbott, 2012). Other related and reinforcing factors are the rapid growth of developing countries, and macroeconomic factors such as monetary policy, and a weak U.S. dollar (Goodwin, 2012).

As identified by Nazlioglu (2011) findings in the literature with regards to the significance and strength of the energy-agricultural link are mixed. Some papers support the neutrality hypothesis where no causation or even strong correlation is found between these sectors (e.g. Zhang and Reed, 2008; Lombardi et al., 2011; Reboredo, 2012). Meanwhile, other branch of the literature support the hypothesis of an increasing linkage where energy prices stimulates price variability and increasing demand of current biofuel feedstocks like corn, sugarcane, and soybeans (e.g. Campiche et al., 2007; Serra et al., 2011). Results are strongly influenced by modeling approaches. Equilibrium models may suffer from arbitrarily price elasticities, and difficulties to capture short run dynamics (Nazlioglu, 2011). While econometric models depend on time series characteristics such as asymmetry, seasonality, linearity, structural changes, and whether macroeconomic or inventory variables are accounted for. Other salient feature found in the literature is that differences in the frequency of the time series data (quarterly, monthly, weekly, daily, etc.) also influence the results of the

energy-agriculture link. For instance, Zhang and Reed (2008) use monthly data rejecting the relationship, while other studies that use weekly data such as Frank and Garcia (2010) support it after the structural break of 2007.

Most empirical time series models involve regressions that relate variables sampled at the same frequency. However, time series are often recorded, collected, or issued at different intervals. For instance, while most macroeconomic variables are reported at monthly or quarterly frequency (i.e. GDP), many other variables including prices or interest rates may have weekly, daily, or even higher frequencies. As a result, matching mixed and irregular sampling frequencies of available data is a common task in applied work that represents a challenge in economic analysis (Chiu et al., 2012). Solutions to the presence of mixed sample frequency include aggregation of the high frequency observation to match the low frequency data. The most common procedure involves taking a simple average, for instance if we have demand data at quarterly frequency, and monthly inventory estimates, then a simple average of the three monthly samples would enter into the regression. In the case we would like to match the low frequency to the high frequency a possible approach is to use interpolation, however this is rarely used. According to Foroni and Marcellino (2013) in pre-filtering the data so that left- and right hand variables are available at the same frequency, a lot of potentially useful information might be destroyed, and mis-specification is inserted in the model. Moreover, the distribution, stationarity, homoscedasticity, potential breaks, and regimes of time series (just to mention some of the most important characteristics), are strongly influenced by data frequency. Therefore, estimation results may differ even when using the same variables.

Overcoming the potential problems of combining unbalanced frequency datasets has recently attracted attention in the macroeconomics and forecasting literature and is currently a fast growing area of research (Andreou et al., 2011; Kuzin et al., 2011; Armesto 2010;

Clements and Galvão 2008; Ghysels, Santa-Clara, and Valkanov 2006). Several alternatives to model mixed frequency data have been proposed, the most promising developments include the Mi(xed) Da(ta) S(ampling) MIDAS (Andreou, Ghysels, and Kourtellos (2011); Ghysels, Sinko, and Valkanov, (2007) and the mixed frequency vector autoregression (mixed frequency VAR) ((Kuzin, Marcellino, and Schumacher (2011); Ghysels (2012); Kuzin et al. (2011); Qian (2010); and Schorfheide and Song (2012)). MIDAS models are essentially extensions of augmented distributed lags, and offer a balance between the parsimony of simple time aggregations and the flexibility of space state models. Mixed frequency VAR models incorporate multiple time series, and with the use of structural model it offer a rich analysis environment. Both techniques look for a parsimonious model that reflects properly the dynamics of the data (Foroni and Marcellino, 2013).

In this paper we focus our attention to the use of MIDAS as a way of evaluating whether the use of mixed frequency data improves the out of sample performance of a model of agricultural prices as a function of energy prices, and other macroeconomic components. Our approach follows the logical positivism of Bayesian methods, that states that a model is as good as its predictions (Geweke and Amisano, 2010). The data used in this analysis is a subset of the one used in the standard VAR of Enders and Holt (2012). These data include an index of grain prices, an index of energy prices, exchange rates, interest rate, and inflation for the period 2006-2011

Our study is also motivated by recent work in the commodity prices literature that identifies the importance of accounting for data frequency. In the context of the energy-agricultural link, Vacha et al. (2012) estimates correlations between biofuels and feedstock at different frequencies by using wavelets, finding that those correlations vary in time and across frequencies. Karali and Power (2013) decomposed realized commodity price volatility into its high- and low-frequency components, identifying the role of macroeconomic forces, and

higher frequency news. However, to the best of our knowledge no study attempts the use of mixed frequencies. We generate forecasts using different functional forms for the weighting scheme of the MIDAS regressions. Results suggest that improvement on short-term predictive ability of grain prices can be achieved by incorporating mixed frequency data. Further work is needed to develop frameworks that incorporate multiple time series such as the mixed frequency VAR. We leave this for future research.

The Mi(xed) Da(ta) S(ampling) MIDAS Model

As with most time series, variables related to the energy and agricultural markets exhibit mixed frequency. For instance, futures prices can be obtained at daily or weekly frequency, while, macroeconomic variables, or reports from agencies such as USDA are usually released monthly or quarterly. Findings from the literature suggest that data frequency influences the estimated results of the energy-agricultural markets relationship. Moreover, potential misspecification and loss of information exist when filtering the data to match either lower or higher common frequencies.

MIDAS regressions estimate time series sampled at different frequencies by specifying conditional expectations as distributed lags of regressors recorded at higher frequencies According to Ghysels (2012b) MIDAS regressions can be viewed as a reduced form of a linear projection from a state space model. Therefore, it can be represented as an approximation of a Kalman filter where the full state space system of equation is not required. Although the Kalman filter provides a good tool to deal with aggregation or interpolation of mixed time frequencies it also exhibit disadvantages such as the "curse of dimensionality" since it produces a large amount of parameters. As a result, computational complexities and specification errors increase. On the other hand MIDAS regressions are easier to estimate and less prone to specification errors. Several surveys of econometric analysis of MIDAS

regressions are worth mentioning. (Andreou et al., 2011) and (Foroni and Marcellino, 2013) provide a comprehensive review of the model, while (Armesto et al., 2010) gives a brief introduction to the topic.

MIDAS regressions are distributed lag models with variables sampled at different frequencies. In a general form the distributed lag model is:

$$y_{t_q} = \alpha + B(L)\, x_{t_q} + \varepsilon_{t_q} \tag{1}$$

where B(L) is a lag polynomial operator, and y_{t_q} and x_{t_q} are the dependent and independent variables sampled at time t, and for illustrative purposes let's assume that it is quarterly frequency data (q). Suppose we are interested on estimating quarterly food price changes y_{t_q} as a function of monthly interest rates x_{t_m}. The conventional approach is to aggregate the monthly frequency data to quarterly frequency by computing a simple average $x_{t_q} = \frac{1}{3}\sum_{i=1}^{3} x_{t_m}$ and proceed with regression at equation (1). Other possibility would be to include the high frequency data as explanatory variables as seen in equation (2). However, parameter proliferation makes this alternative unappealing (Andreou et al., 2011)

$$y_{t_q} = \alpha + B(L_m)\, x_{t_{m,q}} + \varepsilon_{t_q} \tag{2}$$

The alternative offered by MIDAS is to incorporate a functional form to the aggregation scheme, allowing a parsimonious representation based on a data driven linear projection of the high frequency data onto the lower frequency dependent variable, represented by polynomial $W(\theta) = \sum_{i}^{N} w_i(\theta)\, x_{t_{m,q}-i}$, as seen in equation (3) (Ghysels, 2012b).

$$y_{t_q} = \alpha + B(L)W(\theta)\, x_{t_{m,q}} + \varepsilon_{t_q} \tag{3}$$

The weighting function $W(\theta)$ can have any kind of functional form, its objective it to provide flexibility while maintaining parsimony (Armesto et al., 2010). Various functional forms of suggested in the literature meet that goal, in our empirical estimation we use some of the most

popular alternatives in the literature including the Exponential Almon Lag, the Beta lag, and the unrestricted MIDAS model.

Ghysels et al. (2007) proposed the used of the Exponential Almon Lag, its name comes from its resemblance with the smooth polynomial Almon lag functions used to reduce multicollinearity in the distributed lag literature (Foroni and Marcellino, 2013). This function is expressed as:

$$w_i(\theta_j) = \frac{\exp\,(\theta_1 k + \cdots + \theta_n k^N)}{\sum_{n=1}^{N} \exp\,(\theta_1 k + \cdots + \theta_n k^N)} \tag{4}$$

Foroni and Marcellino (2013) argue that this function is quite flexible and can take various shapes with only a few parameters.

Another possible parameterization is based in the Beta function, therefore is known as Beta Lag. The Beta distribution provides as well a function that is flexible and only requires two parameters. It is expressed as follows:

$$w_i(\theta_{1,}\theta_2) = \frac{\mathrm{f}\left(\frac{\mathrm{k}}{\mathrm{K}}, \theta_1, \theta_2\right)}{\sum_{n=1}^{N} \mathrm{f}\left(\frac{\mathrm{k}}{\mathrm{K}}, \theta_1, \theta_2\right)} \tag{5}$$

where $f(i, \theta_1, \theta_2) = \frac{i^{\theta_1 - 1}(1-i)^{\theta_2 - 1}\,\Gamma(\theta_1 + \theta_2))}{\Gamma(\theta_1)\Gamma(\theta_2)}$, and $\Gamma(\theta_p)$ is the gamma function.

We also consider the Unrestricted MIDAS which was proposed by Foroni et al. (2011). This weighting scheme does not depend on the functional form of the distributed lag polynomials but is derived from a linear dynamic framework. The model is based on a linear lag polynomial as follows:

$$B(L)w_i(\theta) = \delta_1(L)x_1 + \cdots + \delta_N(L)x_N + \epsilon_{tm} \tag{6}$$

Where $\delta(L) = (\delta_{j,0} + \delta_{j,1}L + \cdots + \delta_{j,v}L$. In this study we will evaluate the out of sample performance of our data using the alternative MIDAS procedures.

Data and Modeling Approach

Our data corresponds to a subsample of the dataset used by Enders and Holt (2012) for their unrestricted VAR. We do our analysis for the period 2006-2011 that has been identified as the period of strong links between energy and agricultural commodities. Data start in September 2006 (following the identification of this period by Frank and Garcia, (2010)), and finishes in February 2011.

The dataset is includes a grain price index, constructed by the World Bank, as a weighted average of monthly world prices of corn and sorghum (40.8 percent), rice (30.2-percent), wheat (25.3-percent), and barley (3.7-percent). Similarly the energy price index also constructed by the World Bank, is a composite of monthly world prices for coal (4.7-percent), crude oil (84.6-percent), and natural gas (10.8-percent). The nominal exchange rate is obtained from the Federal Reserve, and is the broad exchange trade-weighted exchange rate, which is a weighted average of the foreign exchange values of the U.S. dollar against the currencies of a large group of major U.S. trading partners. The interest rate is the three-month Treasury bill secondary market, both exchange rates and interest rates are sampled at weekly frequency. Finally, the inflation rate is constructed as: $infl = 400 * (\frac{CCPI_t}{CCPI_{t-3}} - 1)$, where $CCPI_t$ is the core consumer price index, that is, the CPI adjusted by deleting prices for food and energy, this variable is calculated monthly.

Plots for these four monthly series, 2006-2011, are presented in Figure 1. The grain index and the energy index exhibit strong variability, and some co-movements are identified during certain periods. They share a strong increase from 2007 until mid 2008, followed by a sharp decrease. However, after 2009 some differences are found, the grain index exhibits a positive trend after the financial crisis of 2009, meanwhile the energy index shows a relatively flat price index from 2009 until 2011, followed by a sharp increase. Exchange rates show a

negative trend until the financial crisis followed by a strong increase in 2009, and showing a negative trend since then. The interest rate declined during the first part of the period and became close to zero since at least 2009.

After plotting the data we tested the stationarity of the series by performing an ADF test. As seen in table 1, grain price index, energy price index and exchange rates are integrated of order 1, I(1). Therefore we perform a transformation on those variables by taking the first difference.

Using these data we evaluate the out of sample performance of the monthly grain index return as a function of monthly energy index return, weekly exchange rate, weekly interest rates and monthly inflation. We generate recursive forecast that use as a initial trading period observations from September 2006 until January 2008, afterwards each observation at t is added to the training period in order to predict observation t+1.

We calculate the root mean square error of the predictions at horizons h=1 month, h=2 months, h=3 months, and h=6 months ahead. We evaluate the performance of different functional forms used in MIDAS such as the Exponential almon polynomial, Beta, MIDAS U, and a simple time averaging to forecast the monthly grain returns.

Preliminary Results and Conclusions

We generate our calculations with modified versions of the scripts in the mixed frequency toolbox in MATLAB of Ghysels (2012b), that estimate MIDAS regressions and generates out of sample root mean square error. In our case the out of sample period corresponds to observations from February 2008 until February 2011 for a total of 37 observations. Results from Table 2 suggest that including mixed frequency data improves prediction ability at least in the short run. The results of alternative MIDAS schemes outperform the simple average at

h=1 and h=2. However for h=3 and h=6 there is almost no difference between simple average and U-MIDAS, and by h=6 the Beta calibration produces inferior forecasts in terms of RMSE. As a result, we claim that the use of mixed frequency data can improve the estimation and prediction of the agricultural commodities explained by energy and macroeconomic factors.

As a summary, mixed results from the literature evaluating the energy-agricultural prices link suggest that is still not clear whether this relationship has strongly influenced food commodity prices. Results appear to be influenced by modeling, and the account of control variables, structural changes. Another aspect influencing results, in particular for econometric models, is the issue of data frequency, results from using weekly or daily data may not concur with the results of using monthly or quarterly data, furthermore misspecification and loss of information is introduce by aggregating or interpolating the data. Several recent papers such as Vacha et al., (2012) and Karali and Power (2013) identify the importance of different frequencies in the economic modeling of commodities. However, the use of mixed frequency data in this context has not been explored. We make use of recent advances in econometric estimation that can help to overcome the potential shortcomings of either too simple time aggregation or too complicated space state models by finding a parsimonious and efficient middle point offered by mixed frequency data models, in particular the Mi(xed) Da(ta) S(ampling) MIDAS model. Results show that different functional forms of the MIDAS weighting schemes improve the predictive performance of the series compared to a simple average time aggregation, but this is only noteworthy for short-term horizons. Test of predictive accuracy including Mariano-Diebold are currently being developed.

Further work will be oriented to the development of structural VAR models with mixed frequency, which can incorporate multiple equation modeling and offer a more comprehensive analysis of the influence of mixed frequency in the measurement of the relationship between agricultural and energy markets.

References

Abbott, P., 2012. Biofuels, Binding Constraints and Agricultural Commodity Price Volatility, in: The Economics of Food Price Volatility. University of Chicago Press.

Andreou, E., Ghysels, E., Kourtellos, A., 2011. Forecasting with mixed-frequency data. Oxford Handbook of Economic Forecasting 225–245.

Armesto, M., Engemann, K., M, O., 2010. Forecasting with mixed frequencies. Federal Reserve Bank of Saint Louis 92, 521–536.

Campiche, J., Bryant, H., Richardson, J., Outlaw, J., 2007. Examining the Evolving Correspondence between Petroleum Prices and Agricultural Commodity Prices., in: Selected Paper Prepared for Presentation at the American Agricultural Economics Association Annual Meeting, Portland, OR, July 29-August1,.

Chiu, C.W.J., Eraker, B., Foerster, A.T., Kim, T.B., Seoane, H.D., 2012. Estimating VAR's Sampled at Mixed or Irregular Spaced Frequencies: A Bayesian Approach.

Enders, W., Holt, M.T., 2012. The Evolving Relationships between Agricultural and Energy Commodity Prices: A Shifting-Mean Vector Autoregressive Analysis, in: Chavas, J.P. (Ed.), The Economics of Food Price Volatility. National Bureau of Economic Research, Inc.

Foroni, C., Marcellino, M., Schumacher, C., 2011. U-MIDAS: MIDAS regressions with unrestricted lag polynomials.

Foroni, C., Marcellino, M.G., 2013. A Survey of Econometric Methods for Mixed-Frequency Data. SSRN Electronic Journal.

Frank, J., Garcia, P., 2010. How Strong are the Linkages among Agricultural, Oil, and Exchange Rate Markets?, in: Proceedings of the NCCC- 134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management. St. Louis, MO.

Geweke, J., Amisano, G., 2010. Comparing and Evaluating Bayesian Predictive Distributions of Asset Returns. International Journal of Forecasting 26, 216–230.

Ghysels, E., 2012a. Mixed Frequency Vector Autoregressive Models.

Ghysels, E., 2012b. Matlab Toolbox for Mixed Sampling Frequency Data Analysis using MIDAS Regression Models.

Ghysels, E., Sinko, A., Valkanov, R., 2007. MIDAS Regressions: Further Results and New Directions. Econometric Reviews 26, 53–90.

Goodwin, B.K., 2012. Comment on“ The Evolving Relationships between Agricultural and Energy Commodity Prices: A Shifting-Mean Vector Autoregressive Analysis”, in: The Economics of Food Price Volatility. University of Chicago Press.

Karali, B., Power, G.J., 2013. Short- and Long-Run Determinants of Commodity Price Volatility. American Journal of Agricultural Economics 95, 724–738.

Kuzin, V., Marcellino, M., Schumacher, C., 2011. MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the Euro Area. International Journal of Forecasting 27, 529–542.

Lombardi, M.J., Osbat, C., Schnatz, B., 2011. Global commodity cycles and linkages: a FAVAR approach. Empirical Economics 43, 651–670.

Nazlioglu, S., 2011. World oil and agricultural commodity prices: Evidence from nonlinear causality. Energy Policy 39, 2935–2943.

Qian, H., 2010. Vector autoregression with varied frequency data. MPRA Paper 88.

Reboredo, J.C., 2012. Do food and oil prices co-move? Energy Policy 49, 456–467.

Schorfheide, F., Song, D., 2012. Real-Time Forecasting with a Mixed Frequency VAR.

Serra, T., Zilberman, D., Gil, J.M., Goodwin, B.K., 2011. Nonlinearities in the U.S. corn-ethanol-oil-gasoline price system. Agricultural Economics 42, 35–45.

Vacha, L., Janda, K., Kristoufek, L., Zilberman, D., 2012. Time-Frequency Dynamics of Biofuels-Fuels-Food System. Arxiv 16.

Zhang, Q., Reed, M.R., 2008. Examining the Impact of the World Crude Oil Price on China’s Agricultural Commodity Prices: The Case of Corn, Soybean, and Pork, in: 2008 Annual Meeting, February 2-6, 2008, Dallas, Texas.

Table 1

Augmented Dickey-Fuller Tests

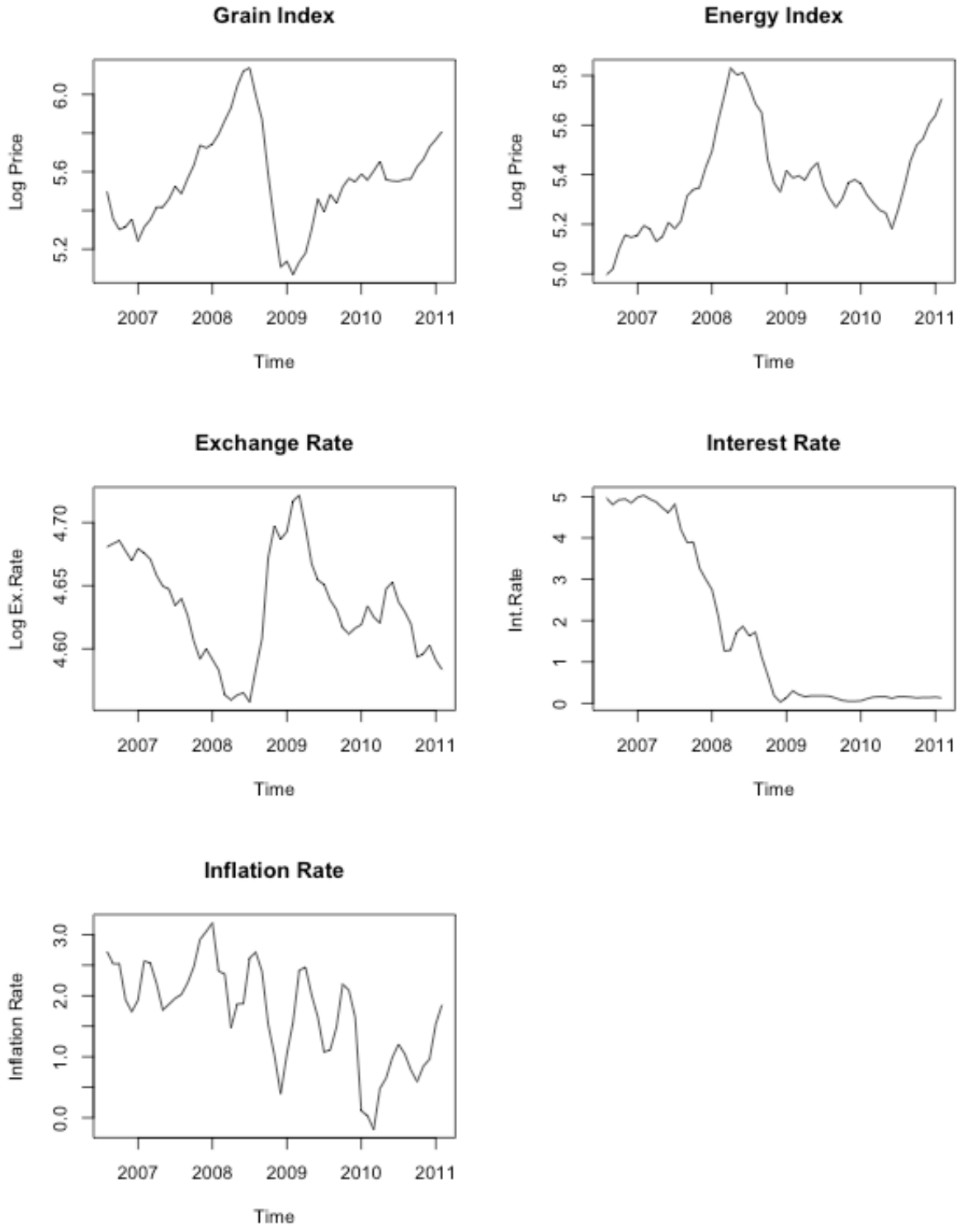
Variable	Deterministic Term	Number of Lags (AIC)	Test Statistic	5% Critical Value
Grain Price	trend	2	-3.14	-3.45
Energy Price	trend	9	-2.02	-3.45
Exchange Rate	trend	4	-2.91	-3.45
Int Rate	trend	19	-4.30	-3.45
Grain returns	constant	2	-3.11	-2.89
Energy Returns	constant	1	-3.38	-2.89
Exch. rates Ret.	constant	6	-3.51	-2.89

Table 2

Out of sample Root Mean Square errors of

Weighting Scheme	RMSE h=1	RMSE h=2	RMSE h=3	RMSE h=4
Exponential Almon	4.01	5.71	6.91	10.94
Beta	3.78	5.64	7.04	11.24
U-MIDAS	3.88	5.62	6.93	10.88
Simple Average	4.21	5.81	7.14	11.21

Figure 1


Grain Index
Log Price
5.2
5.6
6.0
2007
2008
2009
2010
2011
Time
Energy Index
Log Price
5.0
5.2
5.4
5.6
5.8
2007
2008
2009
2010
2011
Time
Exchange Rate
Log Ex.Rate
4.60
4.65
4.70
2007
2008
2009
2010
2011
Time
Interest Rate
Int.Rate
0
1
2
3
4
5
2007
2008
2009
2010
2011
Time
Inflation Rate
Inflation Rate
0.0
1.0
2.0
3.0
2007
2008
2009
2010
2011
Time