Agriculture and Applied Economics Association

Efficiency Gains in Commodity Forecasting with High Volatility in Prices using Different Levels of Data Aggregation

Luis M. Pena-Levano, Octavio Ramirez and Mario Renteria-Pinon

DRAFT

Authors' Affiliation

Luis Pena-Levano is a PhD Candidate in the Department of Agricultural Economics at Purdue University, Octavio Ramirez is a Professor and Department Head in the Department of Agricultural and Applied Economics at the University of Georgia and Mario Renteria-Pinon is a Master student at the University of Washington.

Corresponding Author

Luis M. Pena-Levano
Department of Agricultural Economics
Purdue University
403 West State St.
West Lafayette, IN 47907-2056
E-mail: lpenalev@purdue.edu

Selected Paper prepared for presentation for the 2015 Agricultural & Applied Economics Association and Western Agricultural Economics Association Annual Meeting, San Francisco, CA, July 26-28.

ABSTRACT

This study evaluates the efficiency gains in forecasting three commodity prices (live cattle, coffee and cotton) using time series models. Different leves of temporal aggregations are tested (weekly, monthly, quarterly and annually). The objective is to test whether models based on disaggregated data can produce better price forecasting than the corresponding model using a higher level of aggregation. For example, we test if weekly models can predict better monthly prices than monthly models. Because the high volatility in real prices, we evaluate the possible non-stationarity behavior and heteroskedasticity of each commodity. Then, we use time series methods to model the prices and select the best estimators at each aggregation level and commodity. For the three commodity prices, models based on disaggregated levels effectively provided an efficiency gain in forecasting. Among these levels, the best models were the weekly models. The same behavior was consistent across all possible levels of aggregations.

Keywords: efficiency, forecast, ARMA models, GARCH model, volatility, price, commodity, cotton, livestock, coffee, dissagregation.

JEL codes: C53, E17, C10

SECTION 1 – INTRODUCTION

1.1 Relevance of the topic

An efficient forecast is defined as the prediction from an optimal model that is as close as possible to the true experimental value in a particular period of time (Nordhaus 1987). A way to improve the forecast efficiency that has been overlooked in practice is to select an optimal level of data aggregation for building the model and then making the desired forecast.

Since timing and volatility play important roles in commodity markets, an unexpected change in prices can have significant impact on investors' revenues due to the large volume of these commodities being traded in both the cash and futures markets (Gjølberg and Bengtsson 1997).

Therefore, better forecasts of the cash prices would lead to better decision making by:

- Agribusiness investors, who can compare production costs with future prices and decide on their levels of production of alternative commodities or whether or not to hedge in options and futures markets.
- Banking lenders, because this information could help them assess whether the borrowers would be able to re-pay their production loans and interest.

For approximately four decades, the forecasting efficiency gains that can be obtained by building time series models in which the data are optimally aggregated has been studied sporadically. Nevertheless, comprehensive empirical studies of the efficiency gains that can be achieved through temporal disaggregation have only been conducted in recent years. Interestingly, only few empirical studies have focused on the potential beneficial effects of temporal disaggregation in commodity price forecasting, even though commodities markets are extremely important for the economic performance of the U.S. agricultural sector (Ramirez 2012, Pena-Levano and Ramirez 2014).

Pena-Levano and Ramirez (2014) were the first ones in conducting studies on commodities. They evaluated and showed the efficiency gain using disaggregated models to forecast more aggregated cotton prices, a relative stable commodity in terms of volatility. This paper expands the study by analyzing commodities with high vitality in prices such as coffee and live cattle market; and comparing these results with the previous findings with cotton. The statistics show the relevance of each of those markets:

- U.S. live cattle and calf production was valued at $ 37.0 billion in 2010 and US cattle inventory stood at a total of 92.6 million head in January 1, 2011 (USDA 2012b).

- In fiscal year 2011, The United States's population consumed over half a million pounds of coffee (USDA 2011).
- Cotton's related products and services generated approximately $25.0 billion in annual revenue and were responsible for 200,000 jobs in 2008 (USDA 2012a).

1.2 Contribution and objective of the study

This topic is relevant because most of the commodities present high volatility. Thus, we present in this paper the two cases: commodities with high (e.g. livestock and coffee) and low volatility (e.g. cotton). Thus, this study contributes to the literature by giving a generalization of the forecasting gain of disaggregating data in commodity forecasting.

The primary objective of this study is to evaluate the efficiency gains in forecasting using different levels of temporal aggregations. Specifically, the study assesses: whether the weekly models can produce more accurate monthly, quarterly and annual predictions than the corresponding monthly, quarterly and annual models. We implement the same logic of comparison for the cases of monthly and quarterly versus their respective aggregate. We achieve this through the use of large datasets consisting of approximately 60 years of daily prices and suitable time series models for different levels of data aggregation (weekly, monthly, quarterly and annual).

For agribusiness investors, this result could be very valuable; having an optimal level of aggregation will let them have more efficient forecasting. Thus, better models could be used in investment decisions with respect to (i) the type of crop to be planted as well as (ii) it could help in implementing better hedging strategies. Likewise, as we know, a more efficient model in price forecasting could represent millions of dollars in these commodity sectors because being one more percent accurate could mean significant gains when multiplying by the quantity produced.

SECTION 2 – LITERATURE REVIEW

Forecasting theory posits that the dataset has to be reliable and the model should be as close as possible to the true data-generating process that characterizes the commodity price behavior. An additional factor that this study explores is the effect of the level of aggregation of the data of the model's forecast. The concept is to build models that incorporate the highest amount of about the cyclical behavior or prices into the forecasts.

This section is divided in two sub sections. The first (2.1) draws on the accumulated findings and methods used in previous studies to improve forecasting in commodity markets. The second section (2.2) focuses on previous literature about the effect of temporal aggregation in forecasting efficiency, mainly theoretical econometric models and derivations.

2.1 Previous studies in commodity price forecasting

2.1.i Simultaneous equations and Autoregressive Conditional Heteroskedastic (ARCH) models

The first theoretical models encountered in the literature were simultaneous equations models used for the determination of spot (cash) and futures prices. Initially the models were developed for storable commodities (Peston and Yamey 1960, Stein 1961). Then, some years later, models for non-storable commodities began to appear (Kawai 1983). The main difference between non-storable and storable commodities is that in the first one, the basis (i.e. future price minus spot price) depends on the marginal cost of storage.

Empirical simultaneous models, with no rational expectations and for non-storable commodities were developed by Leuthold and Hartmann. They refined the model forecast approach to semi-strong market efficiency for hogs (Leuthold and Hartmann 1979). Leuthold and Garcia (1992) applied this method to live cattle with a null hypothesis of an efficiency market and no presence of a random walk. They could not reject this hypothesis. Giles et al. (1985) and Goss et al. (1992) built more elaborate simultaneous equation models, with rational expectations. These models were suitable for storable commodities (Giles, Goss, and Chin 1985, Goss and Avsar 1999).

Barry and Gulay expanded the literature by studying price determination in the Australian live cattle market (a non-storable commodity) using simultaneous rational expectations (RE) models of spot and futures markets. They found that prices in those markets depended on expected increases in consumption. Augmented Dickey-Fuller and Phillips Perron tests for unit roots produced ambiguous results. This fact suggested that, for spot and future prices, the expected real income and consumption of beef have followed a linear polynomial trend. The optimal model for forecasting the futures prices for livestock was an ARIMA (1, 1, 5) model, chosen because it had the lowest Mean Square Error (MSE). Notwithstanding, it was emphasized that there was evidence of ARCH effects in the conditional variance and these effects were represented as an E-GARCH (1, 3) process to capture the variance behavior. Additionally, they concluded that in the Australian live cattle market there was presence of semi-strong efficient market (Goss and Avsar 1999).

2.1.2 Producer expectations

Other studies have focused on actual producer expectations. Heady and Kaldor, in 1954, analyzed the price expectations of 200 farmers in Iowa, using semiannual prices from December 1947 to June 1949. They found that only 52% of individual producer expectations fell within 10% of the actual prices, and that the producer expectations were skewed to the right (i.e. toward higher prices) (Heady and Kaldor 1954). Fisher and Tanner (1978) conducted an economic analysis of 55

Australian wheat farmers to determine the best method to predict their price expectation. Nevertheless, even the expectation model that fit the data best only had an R-square of 20%. They attributed the terrible performance of their models to a possible relation of supply and demand information and the stable wheat price that was observed in the year (Fisher and Tanner 1978). Eales et al. (1990) also studied the producers' expectations in the Australian wheat market from 1987 to 1988, and concluded that expectations were on average very similar to the future price, but there was a notable difference with respect to the variability of the market price (Eales et al. 1990). Schroeder et al. (1990) used surveys to collect information about how the producers say they form their price expectations. Kenyon (2001) analyzed the ability of producers to predict the harvest prices of soybeans and corn using expectation models and monthly data from 1991 to 1998. The producer expectations were elicited though a survey at the annual Virginia Corn-Soybean Conference. The results showed that the price expectations were highly correlated with actual cash and futures prices. The difference between the actual and expected producer price was $0.20-$0.30 per bushel. Interestingly, he also concluded that producer expectations followed a right-skewed distribution for both of those commodities and that producers are significantly underestimating the probability of large soybeans price changes (Kenyon 2001).

Dhuyvetter et al. sought improvements in the forecasting of feeder cattle prices, through hedonic regression models and a composite average forecasting model. His main variables were corn price and weight. The accuracy of these models was inversely related to the weight of the feeder cattle (Dhuyvetter et al. 2008, Tonsor, Dhuyvetter, and Mintert 2004).

2.2 Temporal aggregation studies

The potential efficiency gains in forecasting as a result of the temporal disaggregation of the data have been sporadically explored during the previous four decades. Amemiya and Wu were the first ones in investigating disaggregated series, they formulated a measurement for forecasting efficiency losses due to data aggregation using shot-order autoregressive [AR(1) and AR(2)] processes (Amemiya and Wu 1972).

Tiao conducted similar analyses using shot-order moving average [MA(1) and MA(2)] processes, he showed that for short-term predictions in non-stationary series, the gain in forecasting accuracy could be significant when disaggregated data is used (Tiao 1972). Lutkepohl (1987) analyzed theoretically different aggregation levels using ARMA models assuming that the true parameters were unknown. He compared the Mean Squared Error (MSE) of the predictions from disaggregated versus the aggregated ARMA models. He showed that aggregated forecast from a disaggregated ARMA model was more accurate than the aggregated forecast from an aggregated ARMA model (Lütkepohl 1987).

More recently, in the previous decade, Koreisha and Fang evaluated the forecasting efficiency contrasting also aggregated models and different disaggregation levels. In their theoretical evaluation the data series it was used short-order ARMA [(2,1) and (1,2)] processes. They showed that the disaggregated (monthly) model had a better performance than the aggregate (quarterly) model in forecasting quarterly values. Hence, a quarterly model could be used to improve the performance of the quarterly predictions (Koreisha and Fang 2004).

The first comprehensive empirical study of the forecasting of disaggregation was conducted by Ramirez (2012) who investigated the effects of aggregating weekly observations of four data series (US oil spot price, US Federal fund rate, US/Japan exchange rate and 10-year US bond yield) to forecast monthly, quarterly and annual values using long-order ARMA models. Using the Mean Square Errors of the out-of-sample forecasts (MSE) as a criterion for comparison between models based on different levels of data aggregation, he found consistent and substantial efficiency gains when disaggregated models are used for prediction instead of aggregated models. Specifically he showed that the weekly models produce more accurate monthly, quarterly and annual price predictions than the corresponding monthly, quarterly and annual models; the monthly models yield more precise quarterly and annual predictions than the corresponding quarterly and annual models; and the quarterly models render more accurate annual predictions than the corresponding annual models (Ramirez 2012).

The second study, by Pena-Levano and Ramirez (2014), explored further the effects of the level of aggregations by adding the daily frequency and being the first in evaluating an agricultural commodity (i.e. cotton). Their conclusion is similar to the results of Ramirez (2012): higher level of frequency of aggregation is more efficient when forecasting aggregated prices than using the respective aggregated level. For example, weekly data is more efficient to forecast monthly prices than monthly aggregation.

However, both empirical studies assumed that the volatility of the prices was corrected by a transformation of the variables using the inverse of the squared root of the residuals. For the commodity prices, this is not always true. As many recent empirical researches suggest, the volatility in prices of agricultural commodities are a crucial aspect that is required to be modeled. Thus, the motivation of this research then is to test whether similar efficiency gains can be garnered through the use of time series (i.e. longer-order ARMA) models based on disaggregated data for the forecasting on agricultural commodity prices.

SECTION 3 – THE DATABASE DESCRIPTION

3.1 The Data Series

The data for this study includes nominal daily prices from the USDA Chicago (1948-1971; 1988-2011) and USDA Omaha (1971-1987) databases for three commodities:

- Live Cattle: A total of 15,965 observations, from 01/02/1948 to 08/02/2011 with prices given in dollars per hundred pounds.
- Coffee: The total number of observations includes 15,364 nominal daily prices in dollars per 100 pound, from 01/02/1948 to 04/24/2009.
- Cotton: A total of 8,211 nominal daily prices observations in U.S. dollars per pound from 01/02/1979 to 08/02/2011.

Live cattle nominal prices have followed a positive trend, without major sudden fluctuations, very stable during the first 25 years (1948-1973) and then increasing steadily afterwards. Coffee price is not as stable as live cattle. Specifically, price variability seems to be increasing over time, raising suspicion that this series may not have a constant variance. Three very notorious peaks are observed: In 1977, 1985 and 1999. It is interesting to note that during the period of 1975 – 1993, the price began to gradually decrease with periodic spikes, and then exhibited an abrupt increase during the following six years, achieving an historical peak in 1999. According to ICO composite index, prices fell 22 % at the end of 1999, 25 % in 2000 and 29 % in 2001 reaching a very low nominal price of 52 dollars per 100 pounds. This dramatic fall issue has been attributed to an international collapse in the coffee global market due to a cartel failure. (FAO, 2004). In contrast, the behavior of nominal cotton prices during the last 40 years was relatively stable until 2010, followed by a significant upward spike. Thereafter, in 2011, the prices declined quickly but still settled at a higher level than their historical average (Textile Exchange 2011).

3.2 Real Prices Series

The raw data were adjusted using the Consumer Price Index (CPI) with the base year being 1982. The reasons for choosing the CPI instead of Producer Price Index (PPI) is that the PPI is mainly used to adjust revenues for the measurement of real output growth of finished goods but it does not take into consideration transportation cost. Whereas, the CPI is commonly used to adjust income and expenditures of purchased goods in a particular market. In addition, the CPI does not take into consideration the origin of the goods (Bureau of Labor Statistics 2012). This latter characteristic of the CPI gives it an advantage for evaluating the real value of commodities such as coffee, which are imported from other nations.

Once we make the adjustment with CPI, we observe in terms of real prices that:

- The behavior of the real cattle prices is the opposite of the nominal prices. They exhibit trend and a decreasing level of volatility over time (figure 1).

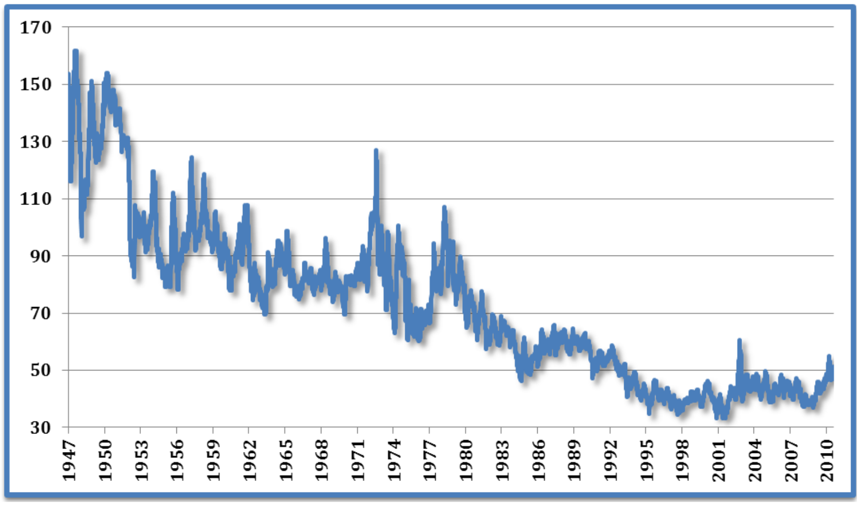

Figure 1. Live Cattle Daily Real Prices (USD per hundred pounds) from 1948 to 2011

Source: USDA, Datastream 2012.

- Coffee also seems to exhibit a negative trend and varying levels of volatility over time. Their highest surge occurred in 1977, but there have been several other less prominent spikes during the last 60 years as well as long periods of depressed prices (figure 2).

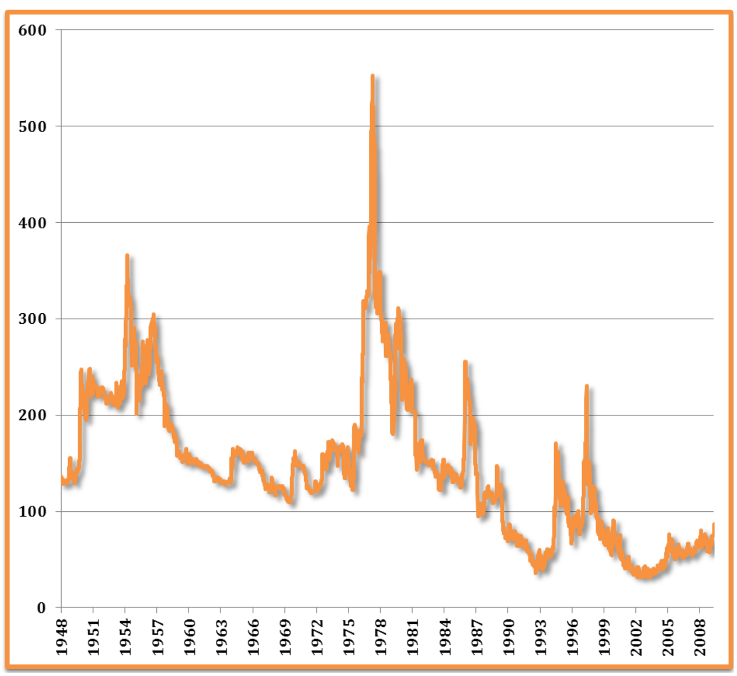

Figure 2. Daily Real Coffee Prices (US dollars per 100 pounds) from 1948 to 2009

Source: USDA, Datastream 2012.

- Real cotton prices follow a negative trend until 2009. However, even after the inflation adjustment, their highest peak and variation occurred in the period of 2010-2011 (figure 3).

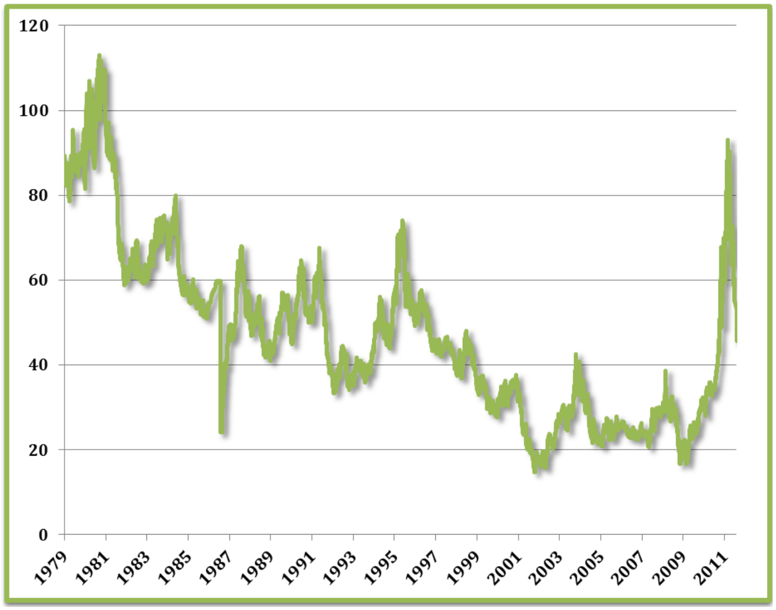

Figure 3. Daily Real Cotton Prices (US dollars per pound) from 1948 to 2011

Source: USDA, Datastream 2012.

3.3 Aggregation into weekly data

The daily data for each commodity was aggregated into weekly data series. Monday was used as the first day of the week for aggregation purposes because cash prices changes more substantially after the weekend.

SECTION 4 - THEORETICAL UNDERPINNING AND THE EMPIRICAL ARMA MODEL SPECIFICATION

In this section, we discuss the data that we are using in our study. We use long-term ARMA models with GLS transformation to obtain the forecasting of the three of them.

4.1 Stationary behavior of the time series

4.1.1 The Unit Root Test

It was observed that commodities prices could exhibit non-stationary behavior (figure 4). Therefore, it was necessary to verify these properties. For the analysis of stationary behavior we tested the series using two test for Unit Root: Dickey-Fuller test (Dickey and Fuller 1979) and Phillips-Perron test (Phillips and Perron 1988) without an intercept or a time trend. The null hypothesis of stationarity of the unit root was tested using a level of significance (α) of 0.10 in all the cases.

When the null hypothesis was not rejected, we proceeded to compute different polynomial degrees to find an adequate systematic trend of the series with respect to time. Specifically first, second and third degree polynomials were evaluated:

Linear: $Y(t) = \alpha_0 + \alpha_1 t + \hat{\omega}_t$ (1a)

Quadratic: $Y(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \hat{\omega}_t$ (1b)

Cubic: $Y(t) = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 t^3 + \hat{\omega}_t$ (1c)

Where Y is weekly prices; t represents time, $\{\boldsymbol{t}$=1,..., n$\}$ ($\boldsymbol{n}$ is the number of observations), and $\boldsymbol{\hat{\omega}_t}$ is the model's residual. Models were estimated using OLS once an appropriate polynomial is identified, the series was de-trended and the residuals $\{\boldsymbol{\hat{\omega}_t}\}$ were tested using Dickey-Fuller and Phillips-Perron tests to make sure that the de-trended data is stationary.

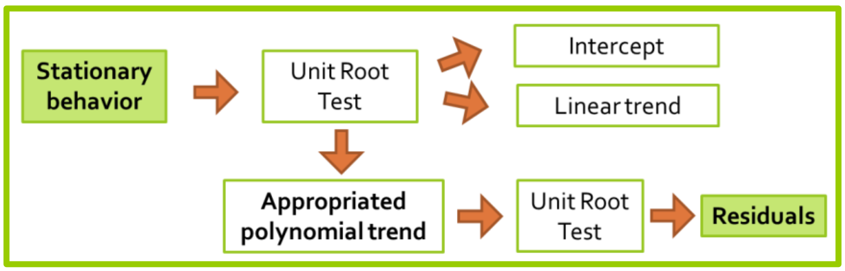

Figure 4. Analysis of the stationary behavior of the data series

Table 1 summarizes the results of the Dickey-Fuller unit root tests. Without any allowance for a time-trend, all commodities appear to have a non-stationary behavior (Phillips-Perron gave similar results). However, when an appropriate polynomial trend is incorporated for each of the three commodities (Ouliaris, Park, and Phillips 1989), all of them show a stationary behavior according to both the unit root tests.

Table 1. Dickey-Fuller Unit Root Test Results from

Commodity	Dickey-Fuller test before the transformation	Adequate polynomial trend	Dickey-Fuller test after detrending
LIVE CATTLE	-1.98	QUADRATIC	-5.73*
COFFEE	-1.28	CUBIC	-3.40*
COTTON	-1.52	LINEAR	-3.02*

* Reject H_0 of a unit root at a significance level of less than 1%

4.2 Heteroskedasticity in the time series data

4.2.1 Detection of Heteroskedasticity

The procedure of the analysis of heteroskedasticity is described in figure 5. First, white tests were conducted to evaluate whether the de-trended residuals $(\boldsymbol{\hat{\omega}_1}, \dots, \boldsymbol{\hat{\omega}_t})$ fulfill the constant variance assumption for each commodity. An auxiliary regression was computed:

$$\widehat{\omega}_i^2 = \beta_0 + f(t) + \widehat{v}_i \quad (2)$$

in which the dependent variable was the squared of the de-trended residuals $\widehat{\omega}_i^2$ ***(i=1,…,n),*** β_0 was the intercept**,** the systematic component was the time trend {$f(t)$**;** i.e., linear, quadratic or cubic}, and $\widehat{v}_i$ was the residual of the auxiliary regression. The time trend depended on the behavior of the weekly prices.

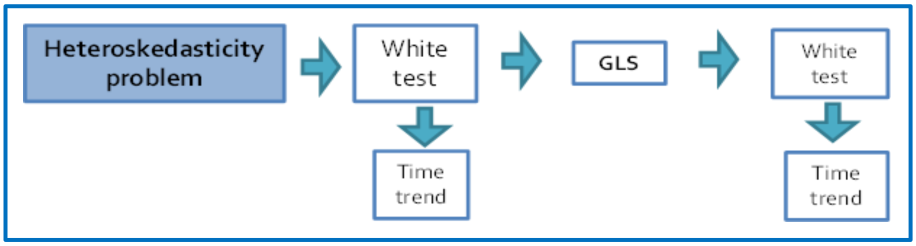

Figure 5. Heteroskedasticity analyses of the data series

The R^2 from this auxiliary regression was used to compute the Lagrange Multiplier **(LM):**

$$LM = n * R^2 \sim \chi^2_{p-1} \quad (3)$$

The null hypothesis (***H₀***) in this test is homoscedasticity. H_0 is rejected when ***LM*** is greater than the Chi-Square table value at a particular level of significance (in this case 0.10) and ***(p-1)*** degrees of freedom (White 1980).

4.1.2 Addressing Heteroskedascity

If H_0 is rejected, a transformation using GLS was required to address the problem:

$$p_t = \psi * y_t = \psi * g(t) + \psi * \widehat{\omega}_t \quad (4a)$$

$$w_t = \psi * \widehat{\omega}_t \quad (4b)$$

where ψ is the appropriate transformation**,** y_t is the original data price, $g(t)$ is the appropriate polynomial time trend found in section 4.1, p_t is the transformed price and w_t is the transformed residual. The White Test was computed again on w_t to verify that the behavior of the transformed series fulfill the assumption of constant variance over time. The appropriate transformation for each commodity is found and described in table 2.

Table 2. White Test Results and GLS Transformations

Commodity	Lagrange Multiplier: Before GLS Transformation	Behavior	Lagrange Multiplier: After GLS transformation
LIVE CATTLE	480.11	QUADRATIC	0.26*
COFFEE	13.78	LINEAR	4.31**
COTTON	6.59	LINEAR	1.59*

* Ho of homoscedasticity cannot be rejected at the 20% significance level or less.
** Ho of homokedasticity cannot be rejected at the 1% significance or less.

Table 2 suggests that the three commodities suffer from inter-temporal heteroskedastic problems, with the Lagrange Multiplier test values being very high in all cases. The most likely patterns of heteroskedasticty for live cattle, coffee and cotton prices were determined to be quandratic, linear and linear, respectively.

Interestingly, for all the commodities, the GLS transformation necessary to achieve homoscedasticity was the same. That is, to divide the detrended price values by the reciprocals of the squared residuals from the artificial regression used in the White Test. After this transformation is incorporated, the residuals exhibit a constant variance over time for all of commodities if constant variance is no longer rejected.

4.3 The aggregation levels of the commodity data

As Figure 6 deploys, the weekly series of each of the three commodities were aggregated into four different levels: Weekly, monthly, quarterly and annual average prices. For the purpose of comparison, it was assumed that a month was equivalent to four weeks, a quarter equal to three months and a year equal to four quarters. Thus, a total of 12 series were analyzed in the study.

Figure 6. Aggregation levels for the data series

4.4 Criteria of selection for the best ARMA models

As shown in the figure 6, a total of 600 ARMA (p, q) models were tested and levels of aggregation (weekly, monthly, quarterly and annual). This number results from a combination of p x q orders (p=0,1,…,24; q=0,1,…,24). As the values of p and q increases, the time series process becomes more complex (Box and Jenkins 1976). Because of this reason, it is anticipated that the weekly models could follow complex ARMA models that could approach p=24 and q=24.

When the data is aggregated into weekly or monthly, it is likely that ARMA models with lower orders in p and q can be used. In regard to the selection of the appropriated ARMA orders (Figure 7), two criteria were combined while also checking for independence of the error terms: ***Akaike Information Criterion*** (Akaike 1998) and ***Parsimony criterion*** (e.g. models with the least number of parameters). Generally the model with the lowest AIC was selected unless there was another much more parsimonious model with only slightly higher AIC and independent residuals.

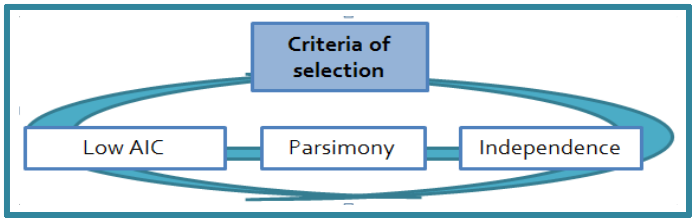

Figure 7. Criteria of selection of the true models

Independence was evaluated by testing for autocorrelation using Ljung-Box Pagan test. The null hypothesis (H_0) assumes independence. If its residuals were not independent according to this test, the model couldn't be selected as the best specification (Ljung and Box 1978, Box and Pierce 1970). Thus, the best model was selected according to have one of the lowest AIC value while being parsimonious and exhibiting independent errors.

4.5 Out-of-Sample Test and MSE as measurement of forecasting efficiency

The comparison of forecasting efficiencies between two ARMA models was based on the minimum squared errors (MSE) of the models. The MSE of a model is computed as follows:

$$MSE = \frac{\sum_{i=1}^{m}(w_i - \widehat{w}_i)^2}{m} \qquad (5)$$

where $\boldsymbol{m}$ is the number of predictions, $\boldsymbol{w_i}$ is the empirical value and $\boldsymbol{\widehat{w}_i}$ is the predicted value from the ARMA model (Amemiya and Wu 1972, Nijman and Palm 1990).

Two tests were used for the comparison in forecasting efficiency between two ARMA models (Figure 8): out-of-sample forecasting and out-of-sample backcasting. For both tests, 960 observations were taken out of the data series. From these 960 observations, a total of 240 monthly, 80 quarterly and 20 annual comparisons were made (Figure 9).

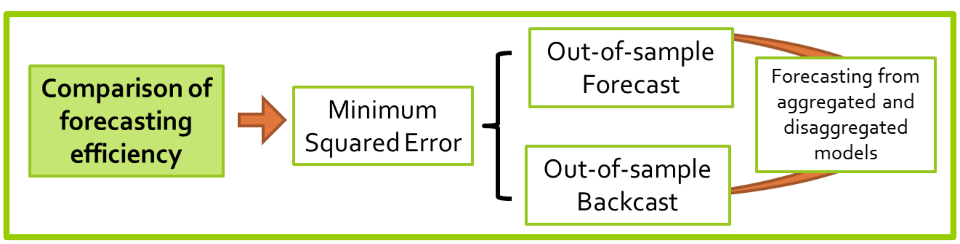

Figure 8 Tests for the comparison of forecasting efficiency

Figure 9. Forecasting and backcasting comparisons

In the comparison in forecasting efficiency of the two models, one of them was an aggregated model used to forecast an aggregated value, and the other one was a less aggregated ARMA model used to forecast the same aggregated value. Figure 10 illustrates the procedure of how to get the aggregated forecast values from a disaggregated model. In this case, four forecast weeks were averaged to obtain the monthly forecast; the average of three and twelve monthly and weekly predictions respectively resulted in a quarterly forecast; and an average of four, twelve and forty eight quarterly, monthly and weekly predictions respectively were equal to an annual forecast.

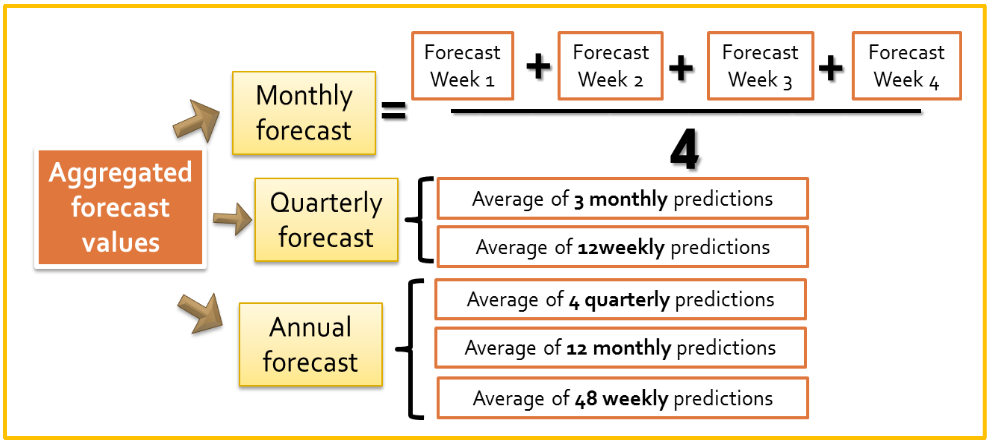

Figure 10. Procedure to obtain the aggregated forecast values

Figure 11 displays the six comparisons made in the out-of-sample forecasting and backcasting tests. The first group of comparisons was made based on the monthly forecasts, where the MSE of the weekly model was compared to the MSE of the monthly model to predict monthly values. The second and third groups of comparisons were based on the efficiency in forecasting quarterly and annual values using the same process.

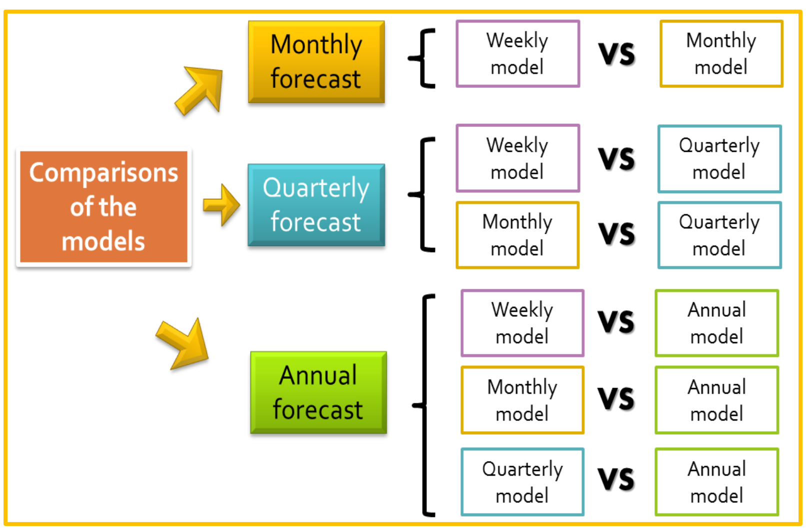

Figure 11 Comparisons between two ARMA models based on the MSE

The procedure of the MSE comparison was based on the method developed by Ramirez (2012) and Koreisha and Fang (2004):

$$\%\Delta MSE = \frac{(MSE_{dm} - MSE_{am})}{MSE_{am}} \quad (6)$$

where $\boldsymbol{MSE_{dm}}$ is the MSE of the disaggregated model, $\boldsymbol{MSE_{am}}$ is the MSE of the aggregated model and $\%\Delta\boldsymbol{MSE}$ is the percent difference between $\boldsymbol{MSE_{dm}}$ and $\boldsymbol{MSE_{am}}$. Therefore, $\%\Delta\boldsymbol{MSE}$ represents the percentage gain or loss in efficiency when an aggregated value is forecasted using a disaggregated model, as opposed to using the corresponding aggregated model.

SECTION 5 – RESULTS AND DISCUSSION

In this section, we provide the results from the estimations explained in section 4. Sub-section 5.1 shows the best models selected for each of the level of aggregations. Sub-section 5.2 compares the efficiency forecasting of these models.

5.1 Best models

Table set 3 presents the best model orders selected for each of the three commodities according to the AIC and parsimony criteria for the weekly, monthly, quarterly and annual aggregation levels. Specifically, they show the order which achieved very low AICs while being reasonably parsimonious and exhibiting independently distributed residuals.

Table 3.a Best ARMA weekly model order selected for each commodity

Commodity	AIC[1]	P[2]	Q[3]	24[4]	48[4]	96[4]	240[4]	600[4]	P+Q[5]
Live Cattle	-3.0133	12	18	0.9995	0.9987	0.8582	0.2519	0.9604	**30**
Coffee	-6.6678	18	16	0.9999	0.9919	0.9600	0.0553	0.9997	**34**
Cotton	-2.0367	6	4	0.9871	0.9051	0.9414	0.9990	1.0000	**10**

Table 3.b Best ARMA monthly model order selected for each commodity

Commodity	AIC[1]	P[2]	Q[3]	12[4]	24[4]	72[4]	120[4]	144[4]	P+Q[5]
Live Cattle	0.2356	12	6	0.9998	1.0000	0.8510	0.9821	0.9919	**18**
Coffee	-0.5235	11	7	0.9999	0.9917	0.8772	0.9918	0.9863	**18**
Cotton	0.0046	2	4	0.9071	0.9772	0.9839	0.9845	0.9996	**6**

Table 3.c Best ARMA quarterly model order selected for each commodity

Commodity	AIC[1]	P[2]	Q[3]	4[4]	8[4]	12[4]	24[4]	60[4]	P+Q[5]
Live Cattle	0.4181	5	0	0.9979	0.9979	0.9975	0.7145	0.9071	5
Coffee	0.1401	1	11	0.9975	1.0000	0.9998	0.9973	0.9999	12
Cotton	0.1510	3	0	0.8875	0.5708	0.3967	0.7675	0.9319	3

Table 3.d Best ARMA annual model order selected for each commodity

Commodity	AIC[1]	P[2]	Q[3]	2[4]	4[4]	8[4]	10[4]	12[4]	P+Q[5]
Live Cattle	0.1362	2	0	0.9687	0.9268	0.9731	0.9292	0.9503	2
Coffee	0.1110	1	1	0.9421	0.5766	0.6554	0.6133	0.5968	2
Cotton	0.0743	1	1	0.9977	0.9847	0.9965	0.9884	0.9851	2

[1] AIC = Akaike Criterion

[2] P = Autoregressive component

[3] Q = Moving average component

[4] P-values presented for Box-Pierce autocorrelation test at the correspondent lag.

[5] P+Q = Parsimony criterion

The best-model results for each commodity can be summarized as follows:

- ***Live Cattle:*** ARMA (12, 18) weekly model, ARMA (12,6) monthly model, AR(5) quarterly model and AR(2) annual model.
- ***Coffee:*** ARMA (18, 16) weekly model, ARMA (11,7) monthly model, ARMA(1,11) quarterly model and ARMA(2,1) annual model.
- ***Cotton:*** ARMA (6, 4) weekly model, ARMA (6,1) monthly model, AR(3) quarterly model and ARMA(1,1) annual model.

From the above results, it is evident that as the data are more aggregated, the orders of the ARMA models become shorter, which is similar to what was observed in Ramirez (2012). Coffee is the commodity with the largest ARMA orders in all the cases closely followed by live cattle. In contrast, cotton prices exhibited much shorted ARMA orders in the weekly, monthly and quarterly models. Likely reasons for these differences can be attributed to: the period of storage that every commodity requires and proportion produced in competitive versus concentrated industries and black swam events that can affect the commodities in a particular period of time.

All the selected ARMA models fulfilled the three previously outlined criteria: having one of the lowest possible AICs as well as low number of parameters (parsimony criteria) and independent error terms.

5.2 The comparison between models: The out-of-sample test results and MSE

Tables 4-6 display the results of the one-period-ahead forecasting/backcasting contests for live cattle, cotton and coffee prices, respectively. Table 7 averages these results per commodity.

5.2.1 Live cattle

The comparisons show that there are substantial efficiency gain when using the ARMA (12, 18) weekly model to forecast and backcast the quarterly (25% and 46%, respectively) and monthly

(15.3% and 28.2%, respectively) prices relative to the quarterly AR (5) and monthly ARMA (12, 6) models. Likewise, there's a significant gain in using the ARMA (12, 6) monthly model over AR (5) quarterly model for forecasting quarterly prices (18.6% and 38.5%, respectively), which corroborates the results of Ramirez (2012). In short, when forecasting monthly or quarterly prices, the model based on the lowest level of aggregation performs best by a wide margin (table 4).

Table 4. Out-of-sample forecasting/backcasting results for live cattle prices (in %)

Model 1 v.s. Model 2	Prediction Target	% MSE difference Forecasting method	% MSE difference Backcasting method
Weekly vs Monthly	Monthly	15.25%	28.21%
Weekly vs Quarterly	Quarterly	25.01%	45.95%
Weekly vs Annual	Annual	16.32%	-8.95%
Monthly vs Quarterly	Quarterly	18.63%	38.50%
Monthly vs Annual	Annual	28.61%	-5.66%
Quarterly vs Annual	Annual	11.65%	-8.54%

In the case of annual forecasts, the evidence from live cattle prices is not that conclusive. When looking at the average between the forecasting and backcasting efficiency comparisons (Table 4), the disaggregated model perform better (4% higher efficiency on weekly versus annual, 12% higher on monthly versus annual, and 2% higher on quarterly versus annual), the annual forecast from the annual model exhibit a lower MSE than those from the weekly, monthly and quarterly in the backcasting comparisons.

A possible explanation of this lack of conclusiveness is likely due to an insufficient number of observations of annual comparisons. That is, there are only 20 observations available for the annual forecasting and 20 for the back casting comparisons, versus 80 for the quarterly, 240 for the monthly and 960 for the weekly comparisons.

5.2.2 Coffee

Significant gains in efficiency are observed in both the forecasting and backcasting (table 5) comparisons between the ARMA (18, 16) weekly model and to the ARMA (11, 7) monthly model (33.5% and 44.2%, respectively), the ARMA (1, 11) quarterly model (26.7% and 55.6%, respectively), and the ARMA (1, 1) annual model (29.5% and 36.9% respectively). That is, the weekly models are much more precise in predicting monthly, quarterly and annual coffee prices than the monthly, quarterly and annual models, respectively.

Table 5. Out-of-sample forecasting/backcasting results for live cattle prices (in %)

Model 1 v.s. Model 2	Prediction Target	% MSE difference Forecasting method	% MSE difference Backcasting method
Weekly vs Monthly	Monthly	**33.50%**	**42.30%**
Weekly vs Quarterly	Quarterly	**26.72%**	**55.61%**
Weekly vs Annual	Annual	**29.58%**	**36.90%**
Monthly vs Quarterly	Quarterly	**24.57%**	**28.32%**
Monthly vs Annual	Annual	**-10.64%**	**32.54%**
Quarterly vs Annual	Annual	**24.84%**	**30.90%**

Similarly high gains are found when comparing the MSE of the forecasts/backcasts from the ARMA (11, 7) monthly model with those from the ARMA (1, 11) quarterly model to predict quarterly prices (24.6% and 28.3%, respectively) and the ARMA (1, 11) quarterly model versus the ARMA (1, 1) annual model when forecasting annual prices (24.8% and 30.9%, respectively). Still, there is one inconsistent result when comparing the efficiency of the annual model forecast, where the MSE of the forecast from the annual model is 10% smaller than that of the monthly model. However, this is reversed on the case of backcasting, where the monthly model is 32% more efficient than the annual model. On average, forecasting and backcasting, the monthly model showed a gain in efficiency by 11% compared to the annual model.

5.2.3 Cotton

In the case of cotton prices, the efficiency gains using models based on disaggregated versus aggregated data were consistent and substantial in both out-of-sample forecasting and out-sample backcasting contests (table 6).

Table 6. Out-of-sample forecasting/backcasting results for cotton prices (in %)

Model 1 v.s. Model 2	Prediction Target	% MSE difference Forecasting method	% MSE difference Backcasting method
Weekly vs Monthly	Monthly	**19.83%**	**34.22%**
Weekly vs Quarterly	Quarterly	**35.79%**	**49.03%**
Weekly vs Annual	Annual	**65.75%**	**36.48%**
Monthly vs Quarterly	Quarterly	**43.21**	**41.26%**
Monthly vs Annual	Annual	**59.80%**	**28.59%**
Quarterly vs Annual	Annual	**54.49%**	**12.98%**

For monthly forecasting, the ARMA (6, 4) weekly model showed an efficiency gain of 19.8% (forecasting) and 34.2%, (backcasting) in comparison to the ARMA (2, 4) monthly model. For quarterly forecasting, the said weekly model was 35.8% and 49.0% more efficient than the AR (3) quarterly model and it was 65.7% and 36.5% more accurate than the ARMA (1, 1) annual model to annual forecasting. The results are similarly striking when comparing the quarterly forecasting and backcasting accuracy of the ARMA (2, 4) monthly model with that of the AR (3) quarterly model (43.3% and 41.2% higher) and its annual forecasting precision with that of the ARMA (1, 1) annual model (59.8% and 28.6% higher). Finally the efficiency of the annual forecasts/backcasts from the AR (3) quarterly model is 55% and 13% higher than the ARMA (1, 1) annual cotton price model.

5.2.4 Average gains in efficiency

Table 7 displays the average gain of forecasting and backcasting results for all three commodities. For the three cases, the efficiency gains are substantial when using weekly models to forecast the monthly and quarterly prices relative to quarterly and monthly models. Similarly high gains are found when comparing on average the MSE of the forecasts from the monthly models with those from the quarterly to predict quarterly prices. Interestingly, on average, there is evidence of efficiency gains in annual forecasts from the monthly and quarterly models with those from the annual model prices. The only case of annual forecast in which the evidence is not that conclusive is the gain in annual forecast when comparing on average the MSE of the weekly models and the annual models; as previously outlined, this is likely due to the insufficient number of observations for the annual comparisons.

Table 7 Average of efficiency gain/loss for all the three commodities.

Model 1 v.s. Model 2	Prediction target	Live Cattle	Coffee	Cotton
Weekly vs Monthly	Monthly	21.73%	37.90%	27.02%
Weekly vs Quarterly	Quarterly	35.48%	41.16%	42.41%
Weekly vs Annual	Annual	-8.95%	33.24%	51.11%
Monthly vs Quarterly	Quarterly	38.50%	26.44%	42.24%
Monthly vs Annual	Annual	11.48%	10.95%	44.19%
Quarterly vs Annual	Annual	1.55%	27.87%	33.73%

In summary, the results for the three commodities are consistent with each other in the fact that a disaggregated model, on average, it is on average preferred because it gives a more efficient forecast, which is consistent with Tiao (1972), Amemiya and Wu (1972), Koreisha and Fang (2004) and Ramirez (2012) studies.

SECTION 6 - CONCLUSIONS

Empirical data of three different commodity prices were evaluated in this study: Live cattle, coffee and cotton at closing daily prices. Under the three scenarios, the commodities were subjected to a transformation due to their non-stationary behavior (dependence to the time period) incorporating quadratic, cubic and linear polynomials respectively. Likewise, all the commodities required transformations for ameliorating heteroskedasticity, illustrating the presence of volatility.

The three commodity prices were aggregated into four different levels of aggregation: weekly, monthly, quarterly and annual, with very large samples based on 60 years of historical data in which it was assumed that the true model parameters are known. The models were selected according to AIC and parsimonious criteria in which it was verified that the residuals were independent and identically distributed.

Under the three different scenarios, disaggregation levels effectively provided an efficiency gain in forecasting, and the best models for this, in the three commodities (live cattle, cotton and coffee) were always the weekly models [ARMA (12, 18), ARMA (18, 16) and ARMA (6, 4), respectively]. The same behavior was consistent across all possible levels of aggregations [i.e., monthly models ARMA (12, 6), ARMA (11, 7) and ARMA (2, 4), respectively] over the quarterly models [AR (5), ARMA (1, 11) and AR (3), respectively].

Interestingly, in the case of annual forecast the evidence from the commodities prices [from AR (2, 0); ARMA (1, 1) and ARMA (1, 1) respectively] is not that conclusive. One explanation of this lack of conclusiveness is an insufficient number of observations of annual comparisons. That is, there are only 20 observations available for the annual forecasting and 20 for the back casting comparisons, versus 80 for the quarterly, 240 for the monthly and 960 for the weekly comparisons.

The results were consistent with the previous theoretical studies of Tiao (1972), Amemiya and Wu (1972), Koreisha and Fang (2004) using short-order ARMA models. The results were also consistent with the empirical study of Ramirez (2012) in oil prices, bond yields, exchange and federal fund rates. Finally, it can be concluded the efficiency gain for each commodity is markedly different, perhaps due to the specific cyclical behavior and volatility of their prices. That is, each commodity must be tested for its best polynomial trend and aggregation traits, depending on the requirements of the decision(s) to be made from the forecast.

REFERENCES

Akaike, Hirotogu. 1998. "Information theory and an extension of the maximum likelihood principle." In *Selected Papers of Hirotugu Akaike*, 199-213. Springer.

Amemiya, Takeshi, and Roland Y Wu. 1972. "The effect of aggregation on prediction in the autoregressive model." *Journal of the American Statistical Association* 67 (339):628-632.

Box, George E, and GM Jenkins. 1976. "Time Series Analysis: Forecasting and Control." *Time Series and Digital Processing*.

Box, George EP, and David A Pierce. 1970. "Distribution of residual autocorrelations in autoregressive-integrated moving average time series models." *Journal of the American statistical Association* 65 (332):1509-1526.

Bureau of Labor Statistics, BLS. 2012. "How does the Producer Price Index differ from the Consumer Price Index." http://www.bls.gov/ppi/ppicpippi.htm

Dhuyvetter, Kevin C, Kole Swanser, Terry Kastens, James Mintert, and Brett Crosby. 2008. "Improving Feeder Cattle Basis Forecasts." Selected paper, Western Agricultural Economics Association Annual Meetings, Big Sky, Montana, June.

Dickey, David A, and Wayne A Fuller. 1979. "Distribution of the estimators for autoregressive time series with a unit root." *Journal of the American statistical association* 74 (366a):427-431.

Eales, James S, Brian K Engel, Robert J Hauser, and Sarahelen R Thompson. 1990. "Grain price expectations of Illinois farmers and grain merchandisers." *American Journal of Agricultural Economics* 72 (3):701-708.

Fisher, Brian S, and Carolyn Tanner. 1978. "The Formulation of price expectations: an empirical test of theoretical models." *American Journal of Agricultural Economics* 60 (2):245-248.

Giles, David EA, Barry A Goss, and Olive PL Chin. 1985. "Intertemporal allocation in the corn and soybean markets with rational expectations." *American Journal of Agricultural Economics* 67 (4):749-760.

Gjølberg, Ole, and Berth-Arne Bengtsson. 1997. "Forecasting quarterly hog prices: Simple autoregressive models vs. naive predictions." *Agribusiness* 13 (6):673-679.

Goss, Barry A, and S Gulay Avsar. 1999. "Non-storables, Simultaneity and Price Determination: The Australian (Finished) Live Cattle Market." *Australian Economic Papers* 38 (4):461-480.

Heady, Earl O, and Donald R Kaldor. 1954. "Expectations and errors in forecasting agricultural prices." *The Journal of Political Economy*:34-47.

Kawai, Masahiro. 1983. "Spot and Futures Prices of Nonstorable Commodities Under Rational Expectations*." *The Quarterly Journal of Economics*:235-254.

Kenyon, David E. 2001. "Producer ability to forecast harvest corn and soybean prices." *Review of Agricultural Economics* 23 (1):151-162.

Koreisha, Sergio G, and Yue Fang. 2004. "Updating ARMA predictions for temporal aggregates." *Journal of Forecasting* 23 (4):275-296.

Leuthold, Raymond M, and Peter A Hartmann. 1979. "A semi-strong form evaluation of the efficiency of the hog futures market." *American Journal of Agricultural Economics* 61 (3):482-489.

Ljung, Greta M, and George EP Box. 1978. "On a measure of lack of fit in time series models." *Biometrika* 65 (2):297-303.

Lütkepohl, Helmut. 1987. *Forecasting aggregated vector ARMA processes*. Vol. 284: Springer Science & Business Media.

Nijman, Theo E, and Franz C Palm. 1990. "Predictive accuracy gain from disaggregate sampling in ARIMA models." *Journal of Business & Economic Statistics* 8 (4):405-415.

Nordhaus, William D. 1987. "Forecasting efficiency: concepts and applications." *The Review of Economics and Statistics*:667-674.

Ouliaris, Sam, Joon Y Park, and Peter CB Phillips. 1989. "Testing for a unit root in the presence of a maintained trend." In *Advances in econometrics and modelling*, 7-28. Springer.

Pena-Levano, Luis, and Octavio A Ramirez. 2014. "Efficiency gains in cotton price forecasting using different levels of data aggregation." American Agricultural Economics Association 2014, Minneapolis, MN, July 26th-28, 2014.

Peston, MH, and BS Yamey. 1960. "Inter-Temporal Price Relationships with Forward Markets: A Method of Analysis." *Economica*:355-367.

Phillips, Peter CB, and Pierre Perron. 1988. "Testing for a unit root in time series regression." *Biometrika* 75 (2):335-346.

Ramirez, Octavio A. 2012. "Conclusive Evidence on the Benefits of Temporal Disaggregation to Improve the Precision of Time Series Model Forecasts." 2012 Annual Meeting, August 12-14, 2012, Seattle, Washington.

Stein, Jerome L. 1961. "The simultaneous determination of spot and futures prices." *The American Economic Review*:1012-1025.

Textile Exchange, TE. 2011. "Global Organic Cotton Market grows percent, Hits $5.61 Billion in 2010." http://www.prweb.com/releases/2011/9/prweb8770173.htm

Tiao, George C. 1972. "Asymptotic behaviour of temporal aggregates of time series." *Biometrika* 59 (3):525-531.

Tonsor, Glynn T, Kevin C Dhuyvetter, and James R Mintert. 2004. "Improving cattle basis forecasting." *Journal of Agricultural and Resource Economics*:228-241.

USDA, Unite States Department of Agriculture. 2011. "Coffee: World Markets and Trade." Accessed January 15th, 2012. http://www.fas.usda.gov/htp/2011_June_coffee.pdf

USDA, Unite States Department of Agriculture. 2012a. "Cotton and Wool Outlook." http://usda01.library.cornell.edu/usda/current/CWS/CWS-05-11-2012.pdf

USDA, Unite States Department of Agriculture. 2012b. "Livestock, Dairy and Poultry Outlook." http://www.ers.usda.gov/publications/ldp/2012/01Jan/LDPM211.pdf.

White, Halbert. 1980. "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity." *Econometrica: Journal of the Econometric Society*:817-838.