Efficiency gains in commodity forecasting using disaggregated levels versus more aggregated predictions

By

Luis Moisés Peña-Lévano and Kenneth Foster

DRAFT

Authors' Affiliation

Luis Peña-Lévano is a PhD Candidate, and Kenneth Foster is a Professor and Department Head in the Department of Agricultural Economics at Purdue University

Corresponding Author

Luis M. Peña-Lévano

Department of Agricultural Economics Purdue University

403 West State St., West Lafayette, IN 47907-2056

E-mail: lpenalev@purdue.edu

ABSTRACT

This study evaluates the efficiency gains in forecasting of using disaggregated data in time series modeling compared to high levels of aggregation. This represents an important topic for agribusinessmen and farmers because it could provide them with insights on how to obtain more accurate predictions. This information then can be use to improve their hedging and negotiation strategies. In our research, we simulated commodity prices and evaluated them under different leves of temporal aggregations are tested (weekly, monthly 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. Then, we use time series methods to model the prices and select the best estimators at each aggregation level and commodity. For the commodity prices under different sample sizes and long time series, 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.

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 forecasting 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.

Important potential areas of application of improved forecasting efficiency are commodity and financial markets which are extremely important for the economic performance of the agricultural sector in many countries. 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). Hence, better predictions 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 and also can help them decide which type of crop is better to be planted at that time period.

- Banking lenders, because this information could help them assess whether the borrowers would be able to re-pay their production loans and interest.

In other words, 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.

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. Several theoretical papers have (i) evaluated short-order univariate and/or multiple time series model (Amemiya and Wu 1972, Tiao 1972, Lütkepohl 1987, Koreisha and Fang 2004) and (ii) suggested new measurements and formulations for the evaluation of prediction efficiency (Lütkepohl 1987). These studies showed forecasting efficiency losses due to data aggregation.

Recently, two studies have covered empirical applications. The first one, by Ramirez (2012), investigated the effects of aggregating observations in US oil spot price and US Federal fund rate. The second one, by Pena-Levano and Ramirez (2012), explored the effects of the level of aggregation on agricultural commodities (i.e. cotton, corn and livestock). Their conclusions were similar: higher level of frequency of aggregation is more efficient when forecasting aggregated prices than using the respective aggregated level.

Nevertheless, the literature has not explored (i) the importance of the sample size in forecasting efficiency, (ii) longer univariate (ARMA) models, and, (iii) use a more realistic assumption of the error terms in the application of commodity price forecasting. These shortcomings can be attributed to the complexity of theoretical evaluation of long time series models and data limitations. Thus, in order to provide a generalization of the forecasting efficiency in favor of the disaggregated models, we evaluate both: univariate time series varying sample sizes to evaluate the conclusions of the empirical and theoretical literature.

This can be done using Monte Carlo simulation for commodity price predictions. From this perspective, the study aims to improve our understanding on forecasting

efficiency and its interaction with disaggregation levels and sample sizes, and thus corroborate both, previous theoretical and empirical works.

Literature review

Theoretical 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 (1972) 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.

Tiao (1972) 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. Lütkepohl (1987) analyzed theoretically different aggregation levels using short vector ARMA models assuming that the true parameters were unknown. He compared the Mean Squared Error (MSE) of the predictions from disaggregated [short VARMA with bivariate MA(2)] versus the aggregated models [short VARMA with bivariate MA(1)]. His results suggested that aggregated forecast from a disaggregated VARMA models were more accurate than the aggregated forecast from their correspondent aggregated VARMA models.

More recently, in the previous decade, Koreisha and Fang (2004) 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 concluded 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.

Empirical studies

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 MSE of the out-of-sample forecasts 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.

Pena-Levano and Ramirez (2012), explored the effects of the level of aggregation on US agricultural commodities (i.e. cotton, corn and livestock) using also long-order ARMA models. Their conclusions were similar to Ramirez (2012): disaggregated models provided better forecasting than their corresponding aggregated model.

Our study takes into consideration the methodology followed by the theoretical studies and the particular characteristics of the commodity pricing such as cyclical behavior, distribution of the error term, and trend. Thus, we provide with a more comprehensive study that overcomes the limitations of their predecessors by building a Monte Carlo simulation with varying sample sizes and order of the ARMA and VARMA models.

Methodology

In this study we use Monte Carlo simulation methods for the data generating process to compare the disaggregated versus the aggregated models under various conditions. The main interest is the forecasting efficiency difference between both frequency levels.

Here, we first explore the univariate time series case. We vary the order of the ARMA models and the sample sizes. We assume that the prices were homoscedastic and followed a log normal distribution. A total of 8 scenarios for the weekly model were a result of:

- 4 orders of stationary ARMA models: AR(4), AR(12), ARMA(4,1) and ARMA(8,2)

- 2 sample sizes (960 [small] and 9600 [large] weekly observations),

We worked under the assumptions of having 3 levels of aggregations: weekly, monthly and annual. We built the simulation based on the weekly observations and then we aggregated them into monthly and annual observations and estimated their respective optimal monthly and annual models. For the ease of the aggregation procedure, we considered 4 weeks were equivalent to 1 month, and 12 months were equal to 1 year.

The selection of the monthly and annual models were based on the Akaike Information Criterion [AIC] (Akaike 1998) and Parsimony criterion (e.g. models with the least number of parameters). We also evaluated the independence of the error terms 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.

The comparison of forecasting efficiencies between two ARMA models was based on the minimum squared errors (MSE) of the models (Amemiya and Wu 1972, Nijman and Palm 1990). This criterion was the based for the out-of-sample forecasting test, in which we take 20% of the observations for the test.

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. For example, we compared the monthly forecast efficiency of the monthly model versus the monthly forecast efficiency from weekly model (i.e. four forecast weeks were averaged to obtain the monthly forecast). Thus, we used a similar structure as Ramirez (2012), and Pena-Levano and Ramirez (2012) to choose the preferred level of aggregation: the lower the MSE, the more efficient the model.

Results and discussion

In this section, we provided the results from the estimations explained in the previous section. Here we show the comparison in accuracy between the models selected for each of level of aggregations and sample size.

Models selected

Table 1 presents the best model orders selected for each of the series according to the AIC and parsimony criteria for the weekly, monthly, and annual aggregation levels. Specifically, we show the models with the orders that achieved very low AICs while being reasonably parsimonious and exhibiting independently distributed residuals.

Table 1 Best ARMA models order selected for scenario

True Model	n	Estimated models						Independence test
		Weekly	AIC	Monthly	AIC	Annual	AIC	P-value
AR(4)	960	AR(4)	429	AR(1)	-232	AR(1)	-70	0.26
	9600	AR(4)	4085	AR(1)	-2880	AR(1)	-753	0.99
AR(12)	960	AR(12)	-8454	AR(3)	-1984	AR(1)	-125	0.80
	9600	AR(12)	-84019	AR(3)	-19512	AR(1)	-1246	0.93
ARMA(4,1)	960	ARMA(4,1)	-4041	ARMA(1,1)	-1107	AR(1)	-100	0.56
	9600	ARMA(4,1)	-40201	ARMA(1,1)	-11011	AR(1)	-1126	0.82
ARMA(8,2)	960	AR(5,1)	-4007	AR(3)	-520	ARMA(1,1)	-73	0.79
	9600	AR(8,2)	-39863	AR(3)	-4926	ARMA(1,1)	-597	0.99

AIC refers to the AIC criterion. For the independence test, we tested the Box-Ljung test for the residuals of the weekly model. N = number of weekly observations

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) and Pena-Levano and Ramirez (2012). Likewise, the estimated weekly models provided similar coefficients and orders compared to the true model, except in the case of ARMA(8,2) under 960 observations. This is because the coefficients of the last terms were

close to zero, and under a low number of observations, the model estimated them as zero values.

As expected, larger the number of observations drives higher the multiplied effect of the AIC. Additionally, all of the selected models presented independency among their error terms according to the Box-Ljung test. Interestingly, higher the number of observations led to a higher level of p-value for independence. Thus, 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.

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

Tables 2 display the results of the one-period-ahead forecasting contests to predict monthly time series. Here we compared the disaggregated weekly model versus the aggregated monthly models.

Table 2. Results of comparison between weekly and monthly models

Monte Carlo Model	n	Comparison between models (MSE)		
		W - M	M - M	Δ%MSE (W-M)
AR(4)	960	0.196	0.122	-60%
	9600	0.185	0.138	-34%
AR(12)	960	0.002	0.004	60%
	9600	0.002	0.004	65%
ARMA(4,1)	960	0.012	0.020	38%
	9600	0.015	0.025	40%
ARMA(8,2)	960	0.056	0.070	20%
	9600	0.014	0.202	93%

W-M: Estimated weekly model used to predict monthly price
M-M: Monthly model used to forecast monthly price
This table shows the mean square error (MSE) of each model under the out-of-sample forecasting. The number of observations used for this test is 20% of the total observations (n).
Δ%MSE is the difference between the disaggregated vs. aggregated model to predict aggregated prices. A positive number reflects the benefit of using a disaggregated model.

The comparisons show that there are substantial efficiency gain when using the AR(12), ARMA(4,1) and ARMA(8,2) weekly models to forecast complex monthly prices relative to the monthly AR(3), ARMA(1,1) and AR(3), respectively, under large sample sizes (n=9600 weeks or 20 years of data). Likewise, there is a significant gain in using the weekly models over their respective monthly models for monthly quarterly prices (between 20-60% gain), which corroborates the results of Ramirez (2012) and Pena-Levano and Ramirez (2012). Nevertheless, when the time series is relative not complex [i.e. the true model is AR(4)], the conclusion is reversed. This suggests that under short ARMA disaggregated orders, there is no gain from forecasting using the disaggregated model.

Table 3. Results of comparison of forecasting annual models

True Model	n	Comparison between models (%MSE)					
		W - A	M - A	A - A	Δ%MSE (W-A)	Δ%MSE (W-M)	Δ%MSE (M-A)
AR(4)	960	0.121	0.010	0.010	-1108%	-1163%	4%
	9600	0.115	0.039	0.033	-247%	-198%	-16%
AR(12)	960	0.000	0.001	0.008	94%	66%	83%
	9600	0.002	0.001	0.013	87%	-170%	95%
ARMA(4,1)	960	0.002	0.002	0.008	79%	22%	73%
	9600	0.005	0.008	0.016	70%	37%	52%
ARMA(8,2)	960	0.006	0.019	0.040	85%	68%	52%
	9600	0.007	0.026	0.055	87%	72%	53%

W-A: Estimated weekly model used to predict annual prices
M-A: Monthly model used to forecast annual prices
A-A: Annual model used to forecast annual prices
This table shows the mean square error (MSE) of each model under the out-of-sample forecasting. The number of observations used for this test is 20% of the total observations (n).
Δ%MSE is the difference between the disaggregated vs. aggregated model to predict aggregated prices. A positive number reflects the benefit of using a disaggregated model.

In the case of annual forecasts, the results show a similar trend. For the short order time series [AR(4)], the results show that the monthly model performed better than the monthly model to forecast the annual prices. Interestingly, the results for the comparison

between monthly and annual prices is not that conclusive. When looking at small sample sizes, the monthly model did a better job, but the situation was reversed under large sample sizes. 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 versus 240 for the monthly comparisons under low sample size, whereas ten times more observations were used in both cases when there is large number of observations.

In contrast, overall, for the long order AR and ARMA models, the most disaggregated models performed better than their corresponding aggregated models, and additionally higher the level of disaggregation led to higher accuracy in the forecasting. This can be attributed to the fact that under a complex time series, a disaggregated model can capture better the variation of the previous prices and thus be more precise.

In summary, the results for time series 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 when the price series is complex. This is consistent with Tiao (1972), Amemiya and Wu (1972), Koreisha and Fang (2004) and Ramirez (2012) studies. In contrast, under simpler time series, an aggregated model can perform better. In practice, the first case is a more realistic scenario, because price modeling is complex in nature and, as used in Pena-Levano and Ramirez (2012), prices are generally modeled under long ARMA models.

Conclusions and direction of future research

Data simulated through Monte Carlo simulation were evaluated in this study under two different factors: sample size and order of the time series. The price commodities behaved under log normal distribution which is the common criterion for pricing commodities and stocks which is common in agricultural applications. The time series were selected to be stationary and homoscedastic, thus not transformations or GARCH modeling was required.

The commodity prices were aggregated into three different levels of aggregation: weekly, monthly and annual, with two sizes of samples (20 and 200 years of observations) under long and short ARMA models. The simulated series were used to estimate the models according to AIC and parsimonious criteria in which it was verified that the residuals were independent and identically distributed.

Under the different scenarios, disaggregation levels effectively provided an efficiency gain in forecasting under long ARMA models and large sample sizes, and the best models for this, in the three commodities were the weekly models. The same behavior was consistent for small sample sizes and long ARMA models.

Interestingly, the conclusion is reversed for short AR models. One explanation of is that for this case, due to the simplicity of the time series, it is not necessary to use a disaggregated version. Thus, an aggregated version is sufficient to capture more optimally the variation of the time series.

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, and Pena-Levano and Ramirez (2012) for commodity price forecasting.

For agriculture, the most appropriate modeling involves generally high order time series modeling due to the complexity of the commodity pricing. Therefore, our study suggests that it would be more appropriate to use a more disaggregated model in order to achieve a higher performance in forecasting aggregated series.

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

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.

ETF. 2012. "After Roller-Coaster 2010 & 2011, Cotton's Newfound Surplus Weighs On Prices." Last Modified April 2012 Accessed May 2012. http://www.etf.com/sections/features-and-news/3665-after-roller-coaster-2010-a-2011-cottons-newfound-surplus-weighs-on-prices?nopaging=1.

FAO. 2004. "Falling commodity prices and industry responses: Some lessons from the international coffee crisis." http://www.fao.org/docrep/006/y5117e/y5117e03.htm.

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. 2012. "Analysis of Efficiency Gains in Commodity Price Forecasting Using ARMA Models with Varying Levels of Data Aggregation." University of Georgia.

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.

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.