Using Local Information to Improve Short-run Corn Cash Price Forecasts

by

Xiaojie Xu and Walter N. Thurman

Using Local Information to Improve Short-run Corn Cash Price Forecasts

Xiaojie Xu

and

Walter N. Thurman*

Paper presented at the NCCC-134 Conference on Applied Commodity Price Analysis, Forecasting, and Market Risk Management

St. Louis, MO, April 20-21, 2015

* Xiaojie Xu is a Ph.D. candidate (xxu6@ncsu.edu), Department of Economics, and Walter N. Thurman is William Neal Reynolds Distinguished Professor (wally_thurman@ncsu.edu), Department of Agricultural and Resource Economics, North Carolina State University. The authors acknowledge Kevin McNew and Geograin, Inc of Bozeman, Montana for generously providing the data used in the analysis in this paper, and participants at the NCCC-134 Conference in St. Louis for helpful comments on an earlier draft.

Using Local Information to Improve Short-run Corn Cash Price Forecasts

Abstract

Using daily prices from 134 corn cash markets from seven Midwestern states, this study examines the increase in short-run cash price forecasting accuracy provided by augmenting futures prices with recent observations from other cash markets. We utilize a Granger-causality-based criterion to determine the structure of the augmented models, i.e. how far to look for potentially relevant forecast information. For about 65% of the markets, the model consisting of prices of the futures market, the specific cash market, and its nearby cash markets ($M2$) forecasts better than the one only incorporating prices of the futures market and the specific cash market ($M1$) for five-, ten-, and thirty-day ahead forecasts based on root mean squared error (RMSE). For short-run forecasts, RMSEs tend not to be significantly different for most of the cash markets investigated, suggesting that the forecast accuracy improvement from including nearby cash markets is only moderate. However, the expanded model ($M2$) tends to significantly outperform the bivariate model ($M1$) more often as the forecast horizon increases.

Keywords: Corn, Cash Price, Futures Price, Forecast, VAR

1 Introduction

The forecast performance of models that incorporate futures prices to forecast cash grain prices have been extensively investigated. Most of the previous literature has focused on forecast accuracy comparisons among different models such as no-change models, econometric models, commercial services, and expert predictions using low frequency data. This study examines forecast accuracy as well, but from the perspective of detailed spatial bid price data. In a data set with hundreds of locations reporting on a daily basis, an issue of modeling parsimony arises: how much spatially explicit information is valuable for the purpose of short-run price forecasting? We propose using a Granger causality criterion to focus on the minimal number of nearby cash markets for a specific cash market to be included in a VAR model.

The motivation behind this idea is as follows. We start by posing the forecasting problem for a specific grain buying location. A nearby farmer, say, is interested in forecasting the bid price at that location over the next 30 days. Information available to the farmer includes the current day's nearby futures quote as well as that day's bid price at a large number of other buying locations. Recent histories of futures prices and other locations' prices are also available. First, while the inclusion of nearby cash markets in a forecast model may lead to low forecast accuracy due to bias from estimation of additional parameters, it also provides useful local information that can improve forecast accuracy by reducing forecast variance. Second, the inclusion of only nearby cash prices in a VAR model is likely to introduce redundant local information; the inclusion of futures prices, with financial information embedded, provides a criterion to limit the number of these cash markets. The remaining specification question is when to stop including nearby cash markets based on the criterion. We consider the relevance of Granger causality.

Several studies (e.g., Bekiros and Diks, 2008) indicated that unidirectional Granger causality can be useful in the prediction of prices. Linking this idea to the problem at hand, we use unidirectional Granger causality to specify models augmented with nearby market data[1] and assess the chosen models in out-of-sample forecast evaluations. This leads to a comparison of two models for each market: one that forecasts a cash price with its own history and that of futures; the other model augmenting the first with price data from nearby markets. The augmented model not only has an appealing statistical property in the sense that histories of futures prices are significantly helpful for cash price forecasts due to unidirectional Granger causality established from futures to cash prices in question, but also incorporates an economic natural way in which producers, physical traders, and hedgers look at the market condition, i.e., they forecast prices of a specific cash market by considering futures prices that convey market-wide changes as well as prices of nearby cash markets that convey local basis momentum.

The current study adds to previous work in several ways. Daily corn cash prices of 182 markets from seven states – Iowa, Illinois, Indiana, Ohio, Minnesota, Nebraska, and Kansas – are analyzed for the periods January 3rd, 2006 – March 24th, 2011. No previous published research has investigated the forecast performance of a model for corn cash prices using such detailed and disaggregated data. While cash prices at delivery markets are usually used in previous literature conducting quarterly or monthly forecasts, we fill the gap of short-run daily forecasts using the rich data that include many non-delivery markets. This study thus addresses forecast users' short-run information needs for making decisions such as daily pricing and storage adjustments, and promotes understandings of price forecasts at non-delivery cash markets, which are not perfectly correlated with prices at delivery points.

[1] Another potential criterion to specify the augmented models is bidirectional Granger causality. Under this circumstance, we face a tradeoff: adding more nearby cash price series to establish unidirectional Granger causality from prices of the futures market to those of the cash market to be forecasted, and thus including more local information and emphasizing the information leadership of the futures market, which can both be helpful to produce better forecasts; but introducing more biases through estimation of additional parameters associated with those added nearby cash markets. For this study, we focus on the application of unidirectional Granger causality as a model specification criterion.

2 Data

The results here are based on data obtained from GeoGrain Inc. They comprise an unbalanced panel of daily corn cash prices. The raw data include over 4,000 markets and cover a 7-year period from September 2005 to March 2011, totaling more than 3.5 million observations. To select markets with the largest numbers of observations, Figure 1 illustrates the 182 markets used in this study. Other markets are eliminated due to high data missing ratios and/or data missing patterns for which cubic spline interpolation does not produce reasonable approximations. On days such as holidays where prices are missing in each market, we omit the observations and assume a smooth continuity of prices (see Goodwin and Piggott, 2001). Other missing prices are approximated by cubic spline interpolation. The percentage of missing observations ranges from 0.3% to 5.2% across markets.

The data set analyzed covers a six-year period from January 2006 to March 2011, totaling 1316 observations for each market. Futures prices of CBOT (Chicago Board of Trade) contracts also are collected by GeoGrain Inc. As in Yang *et al.* (2001), futures prices of the nearest maturity contract are used until the first day of the delivery month and then those of the next nearest maturity contract are used. This futures price series is highly liquid and trading is active. Prices (cents per bushel) are converted to their natural logarithms[2].

Figure 2 plots the price series of the futures and all of the 182 cash markets. As one might expect, the suite of prices move together. The correlation coefficient between the price series of the futures market and that of individual cash markets ranges from 0.9879 to 0.9972. Unit root tests (not reported here) – the augmented Dickey-Fuller test (Dickey and Fuller, 1981), the Phillips-Perron test (Phillips and Perron, 1988), and the Kwiatkowski-Phillips-Schmidt-Shin test (Kwiatkowski *et al.,* 1992) – show all of the price series to be non-stationary in levels but stationary in differences at the 5% significance level.

3 Using Granger Causality to Specify a Forecast Model

For each of 182 cash markets, we construct a time series model that predicts the current price change in market c using its own past price changes, past changes in futures prices, and past changes in nearby markets:

$$\Delta x_t = \mu + \sum_{i=1}^{k} \Gamma_i \Delta x_{t-i} + e_t, \tag{1}$$

where $x_t = (f_t \; c_t \; nc_t^0 \cdots nc_t^m)'$, f denotes the futures market, nc^j denotes the $j-th$ most nearby cash market of market c for $j = 0, \ldots, m$ (if $j = 0$, no nearby cash market is included in the model), and m is the smallest

[2] Log prices are used because the log transformation stabilizes the variance of the underlying raw series and could be beneficial for forecasting (Lükepohl and Xu, 2012). Readers are referred to Lükepohl and Xu (2012) for a detailed investigation into conditions under which taking logs is beneficial for forecasting.

number of nearby markets identified for market c such that prices of market c are unidirectionally Granger caused by those of the futures market, i.e., the following conditions are satisfied: $\Gamma_{1,12} = \ldots = \Gamma_{k,12} = 0$ is not rejected, and $\Gamma_{1,21} = \ldots = \Gamma_{k,21} = 0$ is rejected, where $\Gamma_{i,pq}$ denotes the $pq-th$ element of the matrix Γ_i[3]. To be more specific, m is found by testing from 0 up to 15 nearby cash markets for market c until the first time unidirectional Granger causality from prices of the futures market to those of market c is established.

Of the 182 markets examined, in 139 (76%) unidirectional Granger causality from futures to local cash price is eventually found. The process was stopped once the VAR included 15 nearby markets. Table 1 summarizes the distribution of the minimal number of nearby markets. The 134 cash markets with the identified minimal number ranging from 1 to 15 are used in the next step to assess forecast performance. A simple geographical analysis as in Figure 3 shows that the majority of the cash markets with the minimal number greater than 15 concentrate in Iowa. While cash prices incorporate information from the futures market, they are also determined by local supply and demand conditions. Notably, cash markets are dense in Iowa, which is the largest corn-producing state (National Agricultural Statistics Service, 1997, 2010).

4 The Role of Market Density

Market density varies spatially: nearby markets are farther away for some cash markets than for others. This could have two effects. Considering the possibility of information from nearby markets being more redundant in areas dense in markets than in those where nearby markets are farther away, the inclusion of nearby markets may be less useful in providing local information where density is huge. Thus, more local markets are needed to establish unidirectional causality from futures to cash price. On the other hand, in regions sparse in markets, including nearby markets necessarily includes farther away markets that might be less relevant to the cash market in question, and similarly, more of them are required for the establishment of the unidirectional causal relationship. Thus we construct a measure of market density and investigate its relationship to the minimal number of nearby markets for each cash market.

The measure md for a specific cash market $c = 1, 2, ..., 182$ is formulated as $md^c = \sum_{i \neq c}^{182} \frac{1}{d_{i,c}} = \sum_{i \neq c} \frac{1}{\sqrt{(long_i - long_c)^2 + (lat_i - lat_c)^2}}$, where $d_{i,c}$ is the distance between markets i and c, and $long$ and lat represent longitude and latitude, respectively. Vector $md = (md^1 \cdots md^{182})$ thus contains market density for all cash markets. Without loss of generality, md is normalized by its largest element. The 2nd and 4th rows of Table 2 list the average market density and average minimal number by state. Numerical results of the

[3] The $I(1)$ price data is incorporated in the first difference form into the VAR model whose optimal lag is determined based on the chi-squared distributed test statistics $LR = (T - c)(\log|\Sigma_r| - \log|\Sigma_u|)$, where T is the number of observations, c is a degrees of freedom correction factor (Sims, 1980), and $|\Sigma_u|$ and $|\Sigma_r|$ are determinants of the error covariance matrices from the unrestricted and restricted models, respectively, with the maximum lag fixed at 10 and the significance level at 5%. This significance level is used for all tests performed in this study.

average market density turn out to be consistent with what is visually evident in Figure 1: it is high in Iowa, low in Nebraska, and intermediate in Minnesota, Kansas, Ohio, Illinois, and Indiana. As discussed before, the average minimal number is large for areas dense (Iowa) and sparse (Nebraska) in markets, but small for those whose density is intermediate (Minnesota, Kansas, Ohio, Illinois, and Indiana).

We also calculate the normalized average market density for each state using only the 15 nearby markets of a specific cash market. The numerical results are listed in the 3rd row of Table 2. The conclusion for the average market density by state generally still holds.

5 Forecast Performance Comparisons

Two VAR models mentioned in Section 1 are considered for out-of-sample forecast performance comparisons for each of the 134 markets (see Section 3): the one consisting of prices of the futures market and a specific cash market ($M1$), and the one further incorporating nearby cash markets identified for the specific cash market ($M2$). We conduct recursive forecasts for each model and compare their results to determine if there are significant differences in forecast performance over a post-estimation sample: October 1st, 2010 – March 24th, 2011[4].

The VAR models and their optimal lags are reestimated for each new forecast. The comparison is accomplished by using a (modified) Diebold-Mariano (MDM) test of significant differences in mean squared errors (MSEs) (Harvey *et al.*, 1997). The test is based on $d_t = e_{1t}^2 - e_{2t}^2$ at a given forecast horizon h, where e_{1t} and e_{2t} are forecast errors at time t based on two alternative models. The forecast comparison test statistic is given as:

$$MDM = \left[\frac{n+1-2h+n^{-1}h(h-1)}{n}\right]^{1/2} \left[n^{-1}\left(\gamma_0 + 2\sum_{k=1}^{h-1}\gamma_k\right)\right]^{-1/2} \bar{d}, \tag{2}$$

where $\bar{d}$ is the sample mean of d_t, $\gamma_0 = n^{-1}\sum_{t=1}^{n}(d_t - \bar{d})^2$ is the variance of d_t, and $\gamma_k = n^{-1}\sum_{t=k+1}^{n}(d_t -$

[4]While forecast performance comparisons cannot be conducted for 48 cash markets based on the model specification criterion used in the current study, whether local information helps improve forecasts can still be assessed. One simple way is to compare the accuracy of a bivariate model that forecasts a cash price with its own history and that of futures with an augmented model that further incorporates price data from a number of nearby markets. In this exercise, the number of nearby markets to include is determined by a minimum out-of-sample RMSE criterion. Empirical results are presented in Table 3. It is evident that adding local information generally improves forecasts. This indicates that while the unidirectional Granger causality criterion cannot be used to specify price forecast models for the 48 markets, other model specification criteria can be used. For example, if the number of parameters is a concern then the benefit from including nearby markets has to be large to outweigh the penalty from additional parameters, Bayesian information criterion (BIC), known to lean in the direction of model parsimony, is a possible candidate. This study focuses on using unidirectional causality to specify forecast models and thus only concentrates on the 134 cash markets. Future studies comparing different criteria for commodity cash price (or cash-futures basis) forecasts are of interest given the possibility that different market participants have different concerns over model specifications.

$\overline{d})(d_{t-k} - \overline{d})$ is the $k-th$ auto-covariance of d_t for $k = 1, \ldots, h-1$ and $h \geq 2$. Under the null hypothesis that MSEs generated by both models are equal, the MDM test follows a $t-$distribution with $n-1$ degrees of freedom.

Root mean squared errors (RMSEs) and the MDM test results for five-, ten-, and thirty-day (denoted as $h = 5$, 10, and 30) ahead forecasts are calculated, but not reported here to save space[5].

There are three significant implications of these numerical results. First, for around 65% of the 134 cash markets, adding nearby markets improves forecast accuracy from the perspective of RMSEs at all of the three horizons[6]. The average RMSEs for cash markets for which $M2$ forecasts better are 0.05469, 0.06806, and 0.09095 based on $M2$ and 0.05536, 0.06889, and 0.09191 based on $M1$ for horizons five-, ten-, and thirty-day, respectively. A simple geographical analysis in Figure 4 reveals that adding nearby markets does not help for most cash markets in Nebraska no matter which horizon is considered. This suggests that forecast accuracy improvements through the inclusion of nearby markets do not outweigh forecast accuracy reductions due to bias from estimation of additional parameters for those cash markets in Nebraska in the short run. The 5th, 6th, and 7th rows of Table 2 list by state the percentage of cash markets for which $M2$ forecasts better at horizons five-, ten-, and thirty-day, respectively. Comparing the results for Nebraska with those for other states, it is evident that $M2$ does not have advantages over $M1$ in forecasting in an area where nearby markets are far away and market density is low.

Second, the RMSEs based on $M1$ and $M2$ differ slightly. The MDM test finds 12, 14, and 18 significant cases for horizons five-, ten-, and thirty-day, respectively, with corresponding cash markets plotted in Figure 5. It can be seen that $M1$ outperforms $M2$ significantly more often at the horizon five-day, especially for cash markets in Nebraska, but $M2$ tends to outperform $M1$ significantly more often as the forecast horizon increases to ten- and thirty-day. For each of the 134 cash markets, 95% forecast intervals based on $M1$ and $M2$ also are constructed[7]. It is found that 99.15%, 99.12%, and 98.92% of actual values fall in the forecast intervals at horizons of five, ten, and thirty days, respectively, regardless of the market and model in question.

Third, the percentages of the average reductions in RMSEs by switching from $M1$ to $M2$ for cash markets for which $M2$ forecasts better are 1.22%, 1.20%, and 1.04% at horizons five-, ten-, and thirty-day, respectively.

---

[5] Detailed numerical results are available upon request. Five-, ten-, and thirty-day correspond to one-, two, and six-week ahead forecasts, respectively.

[6] It should be noted that whether adding nearby markets improves forecast accuracy is based on the comparison between $M1$ and $M2$ only in the current study. Therefore, even for the 35% of the 134 markets for which adding local information through the unidirectional Granger causality criterion does not improve forecast accuracy, there could exist other nearby market selection criteria that can reveal the benefit from incorporating local information.

It is also of interest to investigate forecast performance of VARMA models. However, compared with VAR models, order selection is more complex, and parameter estimates are much more involved for VARMA models. Finite-order VAR models may be taken as approximations of VARMA models.

[7] Plots of prediction intervals are available upon request.

It is beyond the scope of this study to evaluate the detailed economic value of utilizing $M2$ for forecast purposes, but as a useful reference work, Colino and Irwin (2010) compare accuracy between futures-based and futures-outlook composite forecasts for hog prices using four outlook programs – Illinois/Purdue, Iowa, Missouri, and USDA – and find that the average RMSE reduction by switching from the futures-based to futures-outlook composite forecasts is 1.18% for the horizon of one quarter ahead for the four programs considered[8]. They further indicate that a RMSE reduction of 2.20% translates to $25,300 for a risk-averse hog producer with production of 10,000 head per year. Thus, considering the short horizons in this study, the economic value of RMSE reductions by adopting $M2$ is nontrivial for at least some cash markets because the maximum reduction can reach 3.68%, 3.87%, and 2.58% at horizons of five, ten, and thirty days.

6 Conclusion

By examining 182 corn cash markets for the periods January 3rd, 2006 – March 24th, 2011, this paper determines that Granger causality can usefully be used to parsimoniously specify VAR price forecasting models. VAR models with increasing numbers of nearby markets are considered until unidirectional Granger causality from futures to the local cash market price is established. Then the forecasting efficiency of the arrived-at VAR model is contrasted with a VAR that uses only the histories of the local cash market and the nearby futures price. For about 65% of the 134 cash markets whose identified minimal numbers of cash markets range from 1 to 15, we find that $M2$ (the more expansive model) forecasts better than $M1$ (the bivariate VAR using only local cash price and futures) at all horizons investigated. For most of these 134 cash markets, the difference in forecast accuracy of the two models is not statistically significant based on the MDM test (Harvey *et al.*, 1997). When it is, $M2$ tends to outperform $M1$ more often as the forecast horizon increases. The general rule of thumb supported by this study is: when a VAR model is used for short-run corn cash price forecasts, the identified nearby cash markets included in the model can improve the forecast accuracy moderately. Specifically, the average reductions in RMSEs by using $M2$ instead of $M1$ for cash markets for which $M2$ forecasts better are 1.22%, 1.20%, and 1.04% at horizons of five, ten, and thirty days, respectively.

The improved forecast accuracy by adopting M2 benefits from a statistical property that unidirectional Granger causality is helpful for forecasting and an economic natural way in which people who participate in

[8] Colino and Irwin (2010)'s work is selected as a reference because futures-outlook composite forecasts can be treated as incorporating additional local information into futures-based forecasts through outlook programs. This corresponds to the idea of using local information to improve forecasts in this study to some extent. Separate from one-quarter ahead forecasts, Colino and Irwin (2010) also provide results for two- and three-quarter ahead forecasts. One-quarter ahead forecasts are selected to illustrate the average RMSE reduction because they represent a horizon that is closest to those considered in the current study for which the longest is thirty-day (six-week) ahead.

agricultural commodity markets look at the market condition, i.e., they use the futures market to observe market-wide changes and local cash markets to observe local basis momentum. These benefits could outweigh forecast accuracy reductions due to estimation of additional parameters through including nearby cash markets in a VAR model.

Using an economic and statistical appropriate approach to select nearby markets, future research could be extended to examine whether local information helps improve cash-futures basis forecasting, which is important to agribusiness because it is the key to successful hedging.

References

Bekiros, S.D., and C.G.H. Diks. 2008. "The Relationship between Crude Oil Spot and Futures Prices: Cointegration, Linear and Nonlinear Causality." *Energy Economics* 30:2673-2685. DOI: 10.1016/j.eneco.2008.03.006

Colino, E.V., and S.H. Irwin. 2010. "Outlook vs. Futures: Three Decades of Evidence in Hog and Cattle Markets." *American Journal of Agricultural Economics* 92:1-15. DOI: 10.1093/ajae/aap013

Dickey, D.A., and W.A. Fuller. 1981. "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root." *Econometrica* 49:1057-1072. DOI: 10.2307/1912517

Harvey, D., S. Leybourne, and P. Newbold. 1997. "Testing the Equality of Prediction Mean Squared Errors." *International Journal of Forecasting* 13:281-291. DOI: 10.1016/S0169-2070(96)00719-4

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt, and Y. Shin. 1992. "Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root." *Journal of Econometrics* 54:159-178. DOI: 10.1111/1467-9892.00213

Lükepohl, H., and F. Xu. 2012. "The Role of the Log Transformation in Forecasting Economic Variables." *Empirical Economics* 42:619-638. DOI: 10.1007/s00181-010-0440-1

National Agricultural Statistics Service. 1997. "Usual Planting and Harvesting Dates for U.S. Field Crops." http://usda.mannlib.cornell.edu/usda/nass/planting//1990s/1997/planting-12-05-1997.pdf.

National Agricultural Statistics Service. 2010. "Field Crops Usual Planting and Harvesting Dates." http://usda.mannlib.cornell.edu/usda/current/planting/planting-10-29-2010.pdf.

Phillips, P.C.B., and P. Perron. 1988. "Testing for a Unit Root in Time Series Regression." *Biometrica* 75:335-346. DOI: 10.2307/2336182

Sims, C.A. (1980). Macroeconomics and Reality. Econometrica, 48, 1-48. DOI: 10.2307/1912017

Yang, J., D.A. Bessler, and D.J. Leatham. 2001. "Asset Storability and Price Discovery in Commodity Futures Markets: A New Look." *Journal of Futures Markets* 21:279-300. DOI: 10.1002/1096-9934(200103)21:3<279::AID-FUT5>3.3.CO;2-C

Appendix: Illustrations and Tables

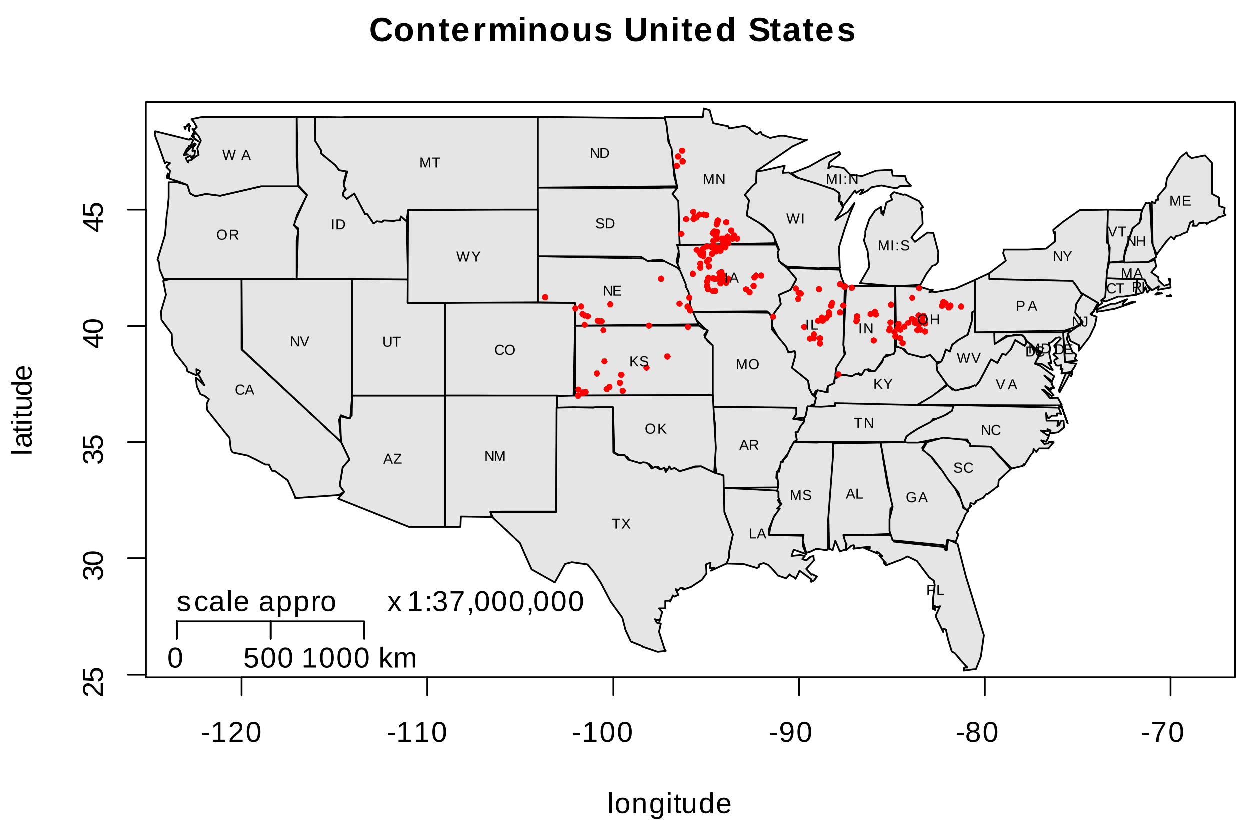

Figure 1: The 182 Cash Markets

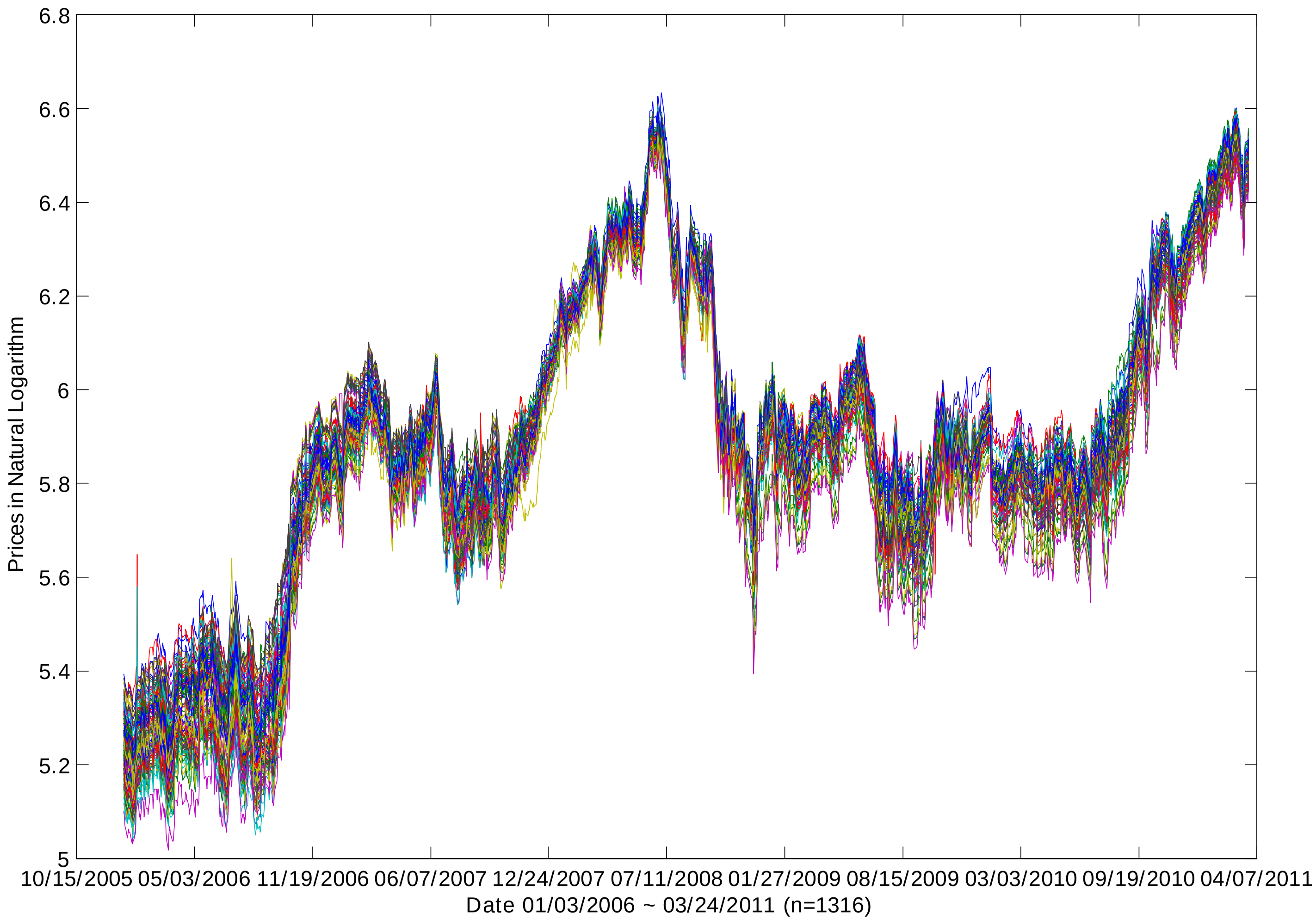

Figure 2: Price Series of the Futures and All of the 182 Cash Markets

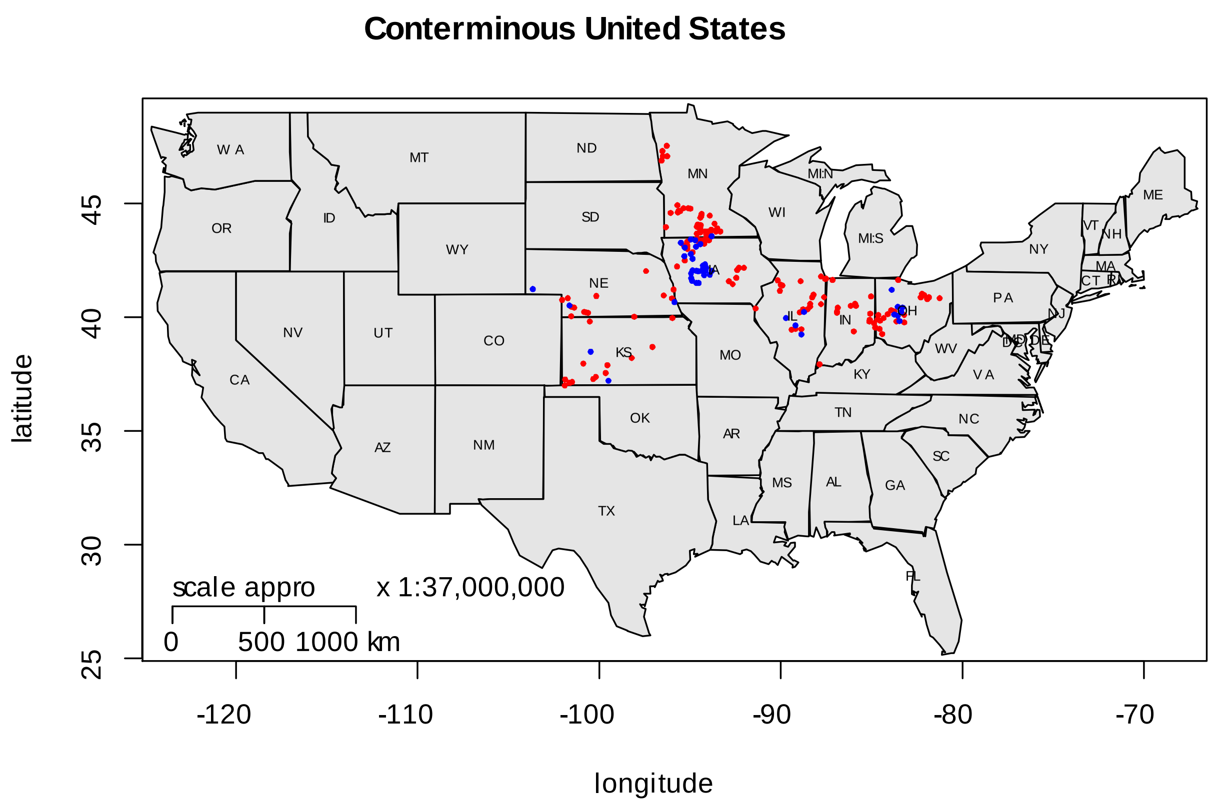

Figure 3: Red (Blue) Points Represent Cash Markets with the Minimal Number Ranging from 0 to 15 (Greater than 15)

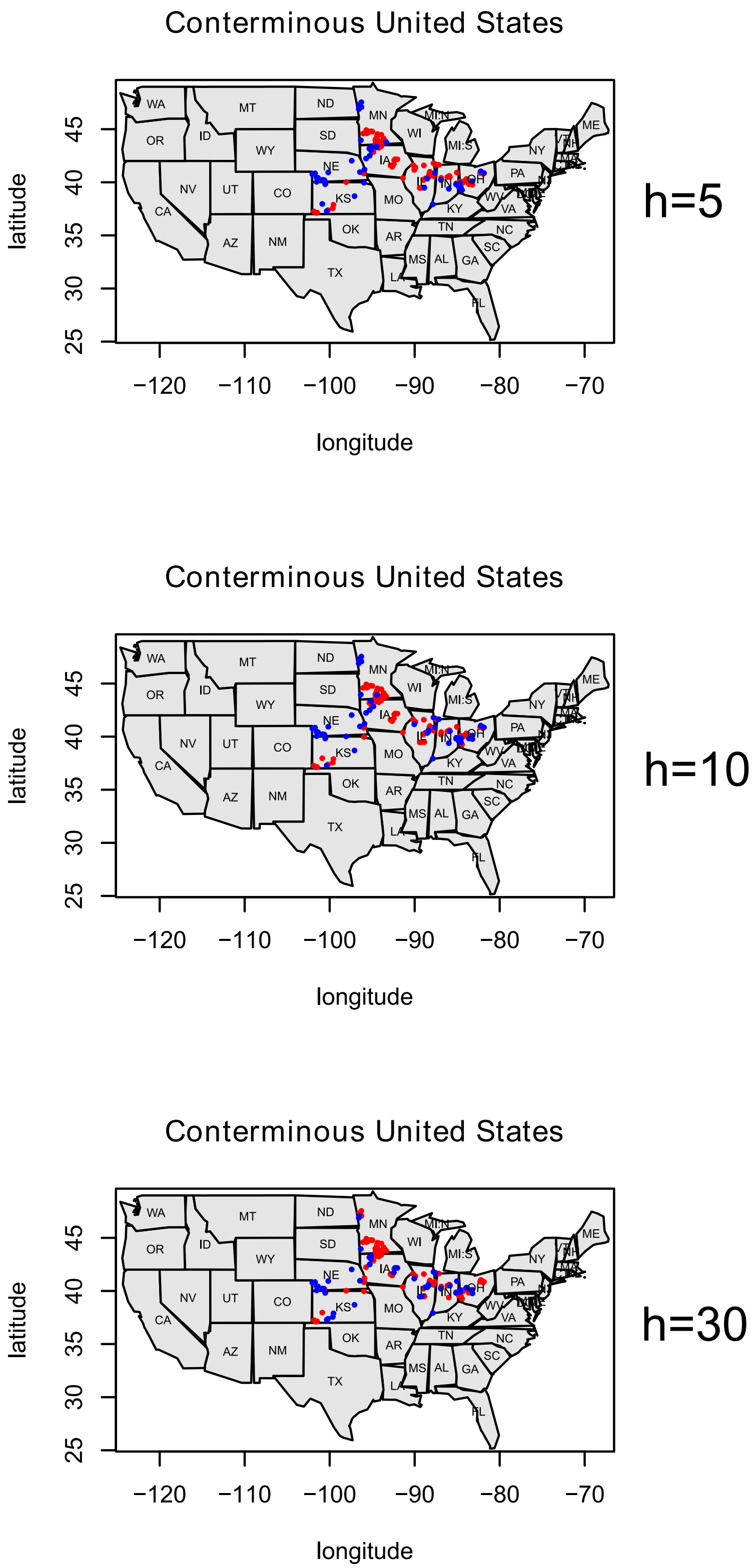

Figure 4: Red (Blue) Points: $M2$ ($M1$) Forecasts Better based on RMSE

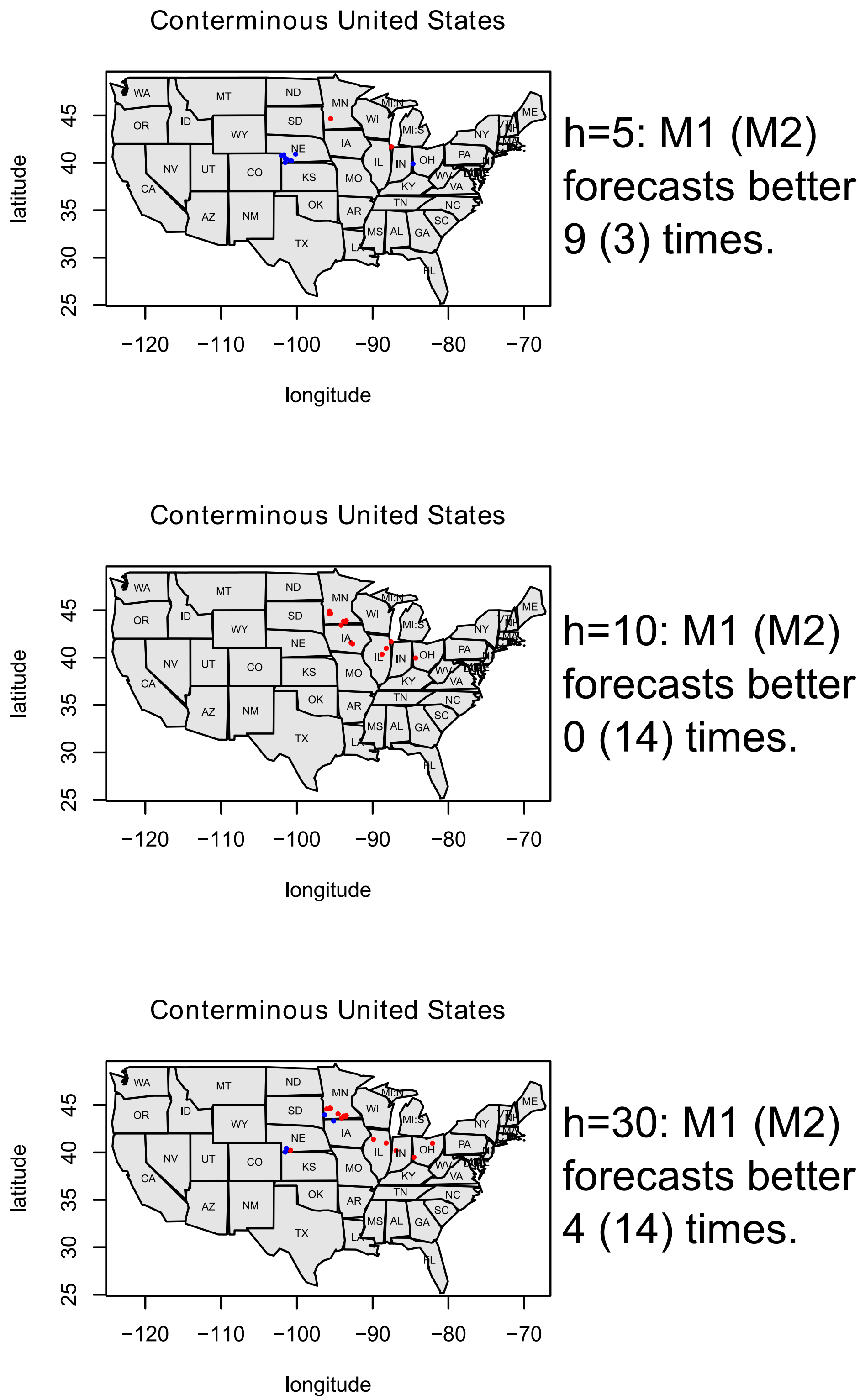

Figure 5: Red (Blue) Points: $M2$ ($M1$) Forecasts Better based on RMSE; the MSEs of $M2$ and $M1$ are significantly different based on the (modified) Diebold-Mariano (MDM) test for each point

Table 1: The Distribution of the Minimal Number of Nearby Markets

Minimal Number	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	>15
Number of Markets	5	28	12	16	13	14	9	4	4	6	9	3	6	3	4	3	43

Table 2: The Average Market Density, Average Minimal Number, and Percentage of Cash Markets for which $M2$ Forecasts Better based on RMSE by State

State	Iowa	Minnesota	Kansas	Ohio	Illinois	Indiana	Nebraska
Average Market Density	0.3231	0.2733	0.1821	0.1738	0.1731	0.1727	0.1316
Average Market Density based on 15 nearby markets	0.1880	0.1477	0.1524	0.1170	0.0968	0.1012	0.0721
Average Minimal Number	10.4000	5.7714	4.8125	8.4074	7.6957	3.6000	9.7500
Percentage of Cash Markets for which $M2$ Forecasts Better based on RMSE at $h = 5$	15/24 62.50%	27/34 79.41%	7/12 58.33%	11/18 61.11%	16/19 84.21%	10/14 71.43%	2/13 15.38%
Percentage of Cash Markets for which $M2$ Forecasts Better based on RMSE at $h = 10$	19/24 79.17%	27/34 79.41%	9/12 75.00%	9/18 50.00%	13/19 68.42%	8/14 57.14%	1/13 7.69%
Percentage of Cash Markets for which $M2$ Forecasts Better based on RMSE at $h = 30$	16/24 66.67%	31/34 91.18%	6/12 50.00%	12/18 66.67%	10/19 52.63%	9/14 64.29%	3/13 23.08%

Notes: For cash markets with the minimal number being greater than 15, 16 is used to calculate the average minimal number for each state.

Table 3: RMSEs for 48 Cash Markets based on the Bivariate Model and Augmented Models with 1 to 15 Nearby Markets

Market Index	Market Density	$h = 5$ the lowest RMSE	$h = 5$ the associated number of nearby markets	$h = 10$ the lowest RMSE	$h = 10$ the associated number of nearby markets	$h = 30$ the lowest RMSE	$h = 30$ the associated number of nearby markets
1 (IA)	0.2520	0.0540	1	0.0674	3	0.0904	8
2 (IA)	0.2650	0.0545	3	0.0678	3	0.0906	5
3 (IA)	0.3140	0.0544	7	0.0677	12	0.0899	10
4 (IA)	0.2491	0.0542	6	0.0671	13	0.0896	10
5 (IA)	0.3211	0.0544	4	0.0680	4	0.0906	10
6 (IA)	0.2492	0.0544	1	0.0681	1	0.0918	6
7 (IA)	0.2794	0.0543	2	0.0675	9	0.0896	9
8 (IA)	0.2530	0.0545	1	0.0679	5	0.0910	9
9 (IA)	0.3345	0.0538	1	0.0671	1	0.0898	10
10 (IA)	0.2848	0.0543	3	0.0676	3	0.0900	5
11 (IA)	0.3177	0.0538	6	0.0671	7	0.0895	7
12 (IA)	0.3180	0.0540	4	0.0673	3	0.0897	10
13 (IA)	0.3430	0.0541	1	0.0675	2	0.0901	10
14 (IA)	0.3154	0.0539	5	0.0671	4	0.0898	11
15 (IA)	0.3212	0.0540	5	0.0672	5	0.0899	5
16 (IA)	0.2753	0.0539	4	0.0671	4	0.0895	7
17 (IA)	0.2533	0.0545	1	0.0677	11	0.0900	10
18 (IA)	0.3944	0.0527	14	0.0651	14	0.0899	14
19 (IL)	0.1739	0.0517	5	0.0639	5	0.0799	1
20 (IN)	0.1758	0.0538	4	0.0691	4	0.0852	9
21 (OH)	0.1301	0.0533	15	0.0672	15	0.0917	5
22 (OH)	0.2172	0.0518	12	0.0660	13	0.0894	1
23 (OH)	0.2083	0.0518	12	0.0660	13	0.0901	1
24 (OH)	0.1841	0.0527	11	0.0672	11	0.0909	2

Table 3 (continued)

Market Index	Market Density	$h = 5$ the lowest RMSE	the associated number of nearby markets	$h = 10$ the lowest RMSE	the associated number of nearby markets	$h = 30$ the lowest RMSE	the associated number of nearby markets
25 (OH)	0.1745	0.0524	11	0.0671	11	0.0906	2
26 (OH)	0.1963	0.0525	12	0.0669	12	0.0910	4
27 (NE)	0.1398	0.0553	0	0.0699	0	0.0928	1
28 (IL)	0.1428	0.0522	0	0.0657	0	0.0845	1
29 (KS)	0.0987	0.0552	13	0.0683	13	0.0899	15
30 (KS)	0.0949	0.0574	0	0.0716	0	0.0945	0
31 (NE)	0.1633	0.0546	2	0.0699	0	0.0903	1
32 (IL)	0.1502	0.0547	2	0.0684	2	0.0841	2
33 (IL)	0.1977	0.0544	2	0.0686	1	0.0855	0
34 (IA)	0.3240	0.0539	14	0.0663	14	0.0895	8
35 (IA)	0.2771	0.0538	0	0.0662	0	0.0891	0
36 (IA)	0.3447	0.0525	1	0.0645	12	0.0880	4
37 (IA)	0.2958	0.0541	0	0.0666	0	0.0899	0
38 (IA)	0.2615	0.0543	0	0.0669	0	0.0898	4
39 (IA)	0.3315	0.0537	15	0.0659	15	0.0894	6
40 (KS)	0.1088	0.0538	11	0.0658	11	0.0887	4
41 (MN)	0.3255	0.0543	10	0.0664	8	0.0924	8
42 (KS)	0.1002	0.0556	2	0.0686	15	0.0912	15
43 (OH)	0.0893	0.0503	13	0.0698	13	0.1130	10
44 (NE)	0.0657	0.0549	0	0.0694	0	0.0915	2
45 (OH)	0.1131	0.0525	2	0.0648	5	0.0863	5
46 (IA)	0.3827	0.0539	3	0.0666	3	0.0909	3
47 (IA)	0.3258	0.0550	8	0.0660	7	0.0906	7
48 (OH)	0.1885	0.0520	12	0.0658	13	0.0913	3