FS 11-03 August 2011

Conclusive Evidence on the Benefits of Temporal Disaggregation to Improve the Precision of Time Series Model Forecasts

Octavio A. Ramirez

Professor and Head of the Department of Agricultural and Applied Economics, The University of Georgia

---

Dept. of Agricultural & Applied Economics
College of Agricultural & Environmental Sciences
University of Georgia

Conclusive Evidence on the Benefits of Temporal Disaggregation

to Improve the Precision of Time Series Model Forecasts

Octavio A. Ramirez[1]

[1] Corresponding author, Professor and Head of the Department of Agricultural and Applied Economics, The University of Georgia, 301 Conner Hall, Athens, GA 30602, e-mail: oramirez@uga.edu, phone: (706) 542-2481, fax: (706) 542-0739.

Abstract: Simulation methods are used to measure the expected differentials between the Mean Square Errors of the forecasts from models based on temporally disaggregated versus aggregated data. This allows for novel comparisons including long-order ARMA models, such as those expected with weekly data, under realistic conditions where the parameter values have to be estimated. The ambivalence of past empirical evidence on the benefits of disaggregation is addressed by analyzing four different economic time series for which relatively large sample sizes are available. Because of this, a sufficient number of predictions can be considered to obtain conclusive results from out-of-sample forecasting contests. The validity of the conventional method for inferring the order of the aggregated models is revised.

Keywords: Data Aggregation, Efficient Forecasting

The issue of whether prediction accuracy can be improved by building time series models on the basis of disaggregated data and averaging their forecasts up to the desired level of aggregation has been sporadically explored in the literature for nearly four decades. Amemiya and Wu (1972) theoretically investigated a scenario where the disaggregated series followed autoregressive {AR(1) and AR(2)} processes, deriving formulas to compute the prediction efficiency losses due to non-overlapping temporal aggregation and concluding that they could be substantial.

Tiao (1972) similarly analyzed the case of integrated moving average processes and found that significant gains in forecasting accuracy could be achieved by using disaggregated data, especially for short-term predictions, when the process is non-stationary. Wei (1978) expanded those results to seasonal autoregressive integrated moving average (SARIMA) models. He showed that the loss in forecasting efficiency through aggregation could be substantial if the non-seasonal component of the disaggregated series is non-stationary. Lütkepohl (1984) and Wei (1990) extended the previous findings to multivariate cases. Other related contributions include Rose (1977), Wei (1978), Tiao and Guttman (1980), Weiss (1984), and Gonzalez (1992). All of the previously cited works and results, however, are theoretical and asymptotic in nature, and assume that the model parameters are known.

Notably, Lutkepohl (1987) compared the Mean Square Error (MSE) of the forecasts from disaggregated versus aggregated ARMA models when the parameters have to be estimated. He showed that, asymptotically, the MSE matrix for the vector of 1-step to r-step ahead aggregate predictions from a disaggregate ARMA model strictly dominates the MSE matrix for the prediction vector from the corresponding aggregate model. In addition to being asymptotic, however, his analyses were limited to a “very special” type of AR process. Lacking a broader result, he concludes that "it seems that temporal aggregation generally results in efficiency losses in forecasts if the ARMA coefficients are estimated" and "overall, the results suggest that using disaggregate data is preferable." On the feasibility of obtaining broader theoretical results, Koreisha and Fang (2004) opine that it is too complicated to analytically compute the effect of parameter estimation errors on the prediction MSE for models as simple as an ARMA(1,1).

Lutkepohl (1987) also points out that "the fact that in practice temporally aggregated data sometimes do provide better forecasts than the disaggregate data (e.g. Abraham 1982) cannot be explained by asymptotic theory" and provides the following possible reasons for this apparent contradiction: a) the sample sizes involved are too small for the asymptotic theory to be valid, b) the assumptions, such as structural stability over time, underlying asymptotic theory are violated in practice, and c) a specific estimate obtained on the basis of a single realization of the data generating process is not a good approximation to the actual forecast MSE of the underlying process. In short, he questions the applicability of asymptotic results under small sample conditions and whether empirical out-of-sample forecasting contests based on a limited number of predictions provide reliable estimates of the actual MSEs.

Several empirical studies have failed to unequivocally support the previously discussed theoretical results (Koreisha and Fang, 2004). Butter (1976), for example, compared monthly versus quarterly models of the difference between the yield on mortgages and the yield on government loans using data from 1961 to 1974. He generated quarterly forecasts based on a quarterly ARMA(1,1) model and on the average of the monthly predictions from an AR(1), and found the quarterly model to be generally more accurate than the monthly model when the forecast horizon was more than one quarter. Nijman and Palm (1990) compared annual predictions from annual, quarterly and monthly IMA(1) models of the Netherlands GNP estimated using data from 1957 to 1984. Their results suggest that the MSE of the forecasts from the quarterly models is lower than that of the annual models but further disaggregation into monthly data hardly yields extra information to forecast that series on an annual basis.

Lutkepohl (1987, pp 250-257) provides examples involving U.S. personal consumption expenditures and U.S. gross private domestic investment where it is not obvious that the forecasts from the disaggregate (quarterly) models outperform those from the aggregate (semi-annual and annual) models. More recently Silvestrini et al (2008) forecast France's annual budget deficits using both monthly and annual information and conclude that "the one-step-ahead predictions based on the temporally aggregated models generally outperform those delivered by standard monthly ARIMA models." However, it is important to keep in mind that the out-of-sample comparisons in all of those examples are based on a handful of observations.

Georgoutsos, Kouretas and Tserkezos (1998) conducted an empirical evaluation based on monthly and quarterly Structural Vector Autoregressive (SVAR) models of the interrelationships between the rate of growth of the real monetary aggregate ($M1$), the industrial production index, and the 3-month treasury bills rate. A total of 24 years of data (1964-1987) were used to build the models and four years were saved to assess their ability to make *ex-post* quarterly forecasts. They report the standard deviation of the actual and predicted values, Theil's inequality coefficients, and MSE bias and variance/covariance decomposition results and conclude that "all the measures of forecasting accuracy assume larger values when the aggregate data is used instead of the aggregated forecasts from the monthly observations." The magnitudes of the forecast MSE reductions, however, are not provided by these authors.

More recently Koreisha and Fang (2004) examined the forecasting accuracy of models based on aggregated versus disaggregated data where the disaggregated predictions are immediately updated as new information becomes available. These authors assume that the disaggregated ("monthly") data are generated by hypothetical short-order {up to ARMA(2,1) and ARMA(1,2)} models and theoretically show that new monthly observations becoming available during the current quarter could be used to improve the performance of quarterly forecasts in the case of short-term predictions. A recent synopsis of past literature on this issue is presented in Silvestrini and Veredas (2008).

In summary, previous studies have shed light on the potential gains in forecasting efficiency that could be achieved by using more disaggregated time series models. However, the principle of building models and making predictions based on the lowest possible level of data disaggregation even when only more aggregate level forecasts are desired has not been adopted in practice, and there does not seem to be a concerted effort to collect more disaggregated data for the purpose of increasing forecasting efficiency.

This might be due to the fact that, as previously discussed, the stronger results in favor of disaggregated models are theoretical in nature and there is still doubt as to whether those results actually hold in small sample applications. Related concerns include the relatively short orders of the ARMA models that have been theoretically and empirically investigated to date and the lack of generally applicable measurements of the efficiency gains that could be expected under conditions where the model parameters have to be estimated. Further, the potential benefits of disaggregation down to weekly data have not been investigated. A final issue is that the few empirical evaluations available to date have yielded mixed results, perhaps because they have been based on out-of-sample forecasting contests where the limited number of predictions available for the comparisons might have obscured the outcomes.

This research attempts to address the previous issues by measuring the exact MSE differentials between the disaggregated and the aggregated model forecasts through simulation methods. This allows for comparisons including long-order ARMA models, such as those expected with weekly data, under realistic conditions where the parameter values have to be estimated. The weakness of past empirical evidence is addressed by analyzing four different time series for which a large number of observations are available. Because of this, a sufficient number of predictions can be considered to obtain conclusive results from the out-of-sample forecasting contests. Comprehensive comparisons involving annual forecasts from annual, quarterly, monthly and weekly models; quarterly forecasts from quarterly, monthly and weekly models; and monthly forecasts from monthly and weekly models are presented.

A final issue explored in this research relates to Brewer's (1973) formula, which has been used in numerous past studies (up to Silvestrini et al 2008) to ascertain the ARMA order of aggregate models based on that of the "originating" disaggregated model. It has been noted that this formula implies that the ARMA orders (p and q) do not decrease with aggregation, which contradicts empirical observation (Rossana and Seater, 1995). This is another instance where a theoretical result does not appear to match what is being experienced in practice. This research assesses whether such high ARMA orders are really necessary to properly model the aggregated processes corresponding to particular disaggregated time series.

The Data Series

Four common, readily available data series are used in the research: the US oil spot price (1326 weekly observations beginning 2/22/1985), the 10-year US bond yield (2332 weekly observations beginning 11/13/1965), the US/Japan exchange rate (1942 weekly observations beginning 6/8/1973), and the US Federal funds rate (2120 weekly observations beginning

1/7/1970). All data was downloaded from the IHS Global Insight data services website (http://www.ihs.com). The four series were subject to augmented Dickey-Fuller and Phillips-Perron tests with an intercept and a linear time trend. The null hypothesis of a unit root could not be rejected at an α of 0.10 in any of the cases. In addition, the more general test suggested by Ouliaris, Park and Phillips (1989) was conducted assuming a third-degree polynomial trend. The unit root hypothesis again held in all four series (α=0.10). Alternatively, through the same battery of tests (α=0.001), it is concluded that the first-differenced data is stationary.

The differenced series, however, seem to experience periods of high and low volatility over time (Figures 1 to 4). Specifically, the standard deviation (STD) of the last 380 or so observations of Series 1 (the US oil spot price) is over three times that of the previous 940[1]. In the case of Series 2 (the 10-year US bond yield), the STD of observations 720 to 1150 is about 2.5 times higher than in the rest of the data[1]. Series 3 (the US/Japan exchange rate) seems to be almost twice as volatile during the first 840 periods[1]. Finally, the STD of Series 4 (the US Federal funds rate) is approximately three times higher during the first 890 observations[1]. Therefore, in all four series, the first-differenced data are re-scaled using the appropriate standard deviation ratios in order to achieve a more homogeneous level of volatility over time (for the purposes of peer-review, the re-scaled data are shown in Figures 1S-4S). It is important to note that without re-scaling the ARMA orders needed to attain independently distributed errors are excessive (p>24 and q>24 in the case of Series 2 and 4).

The next step is to determine the appropriate ARMA order for each of the first-differenced re-scaled series. For this purpose, a total of 625 ARMA(p,q) models encompassing all possible p-q combinations (p=0,…,24 and q=0,…,24) are estimated. In the case of Series 1, an ARMA(16,11) exhibits the lowest Akaike Information Criterion (AIC=4,030.03). However, a more parsimonious ARMA(8,7) still appears to have independently distributed errors[2] and shows a nearly identical AIC of 4,030.11. The latter is thus selected as the appropriate order. For Series 2, the minimum AIC (-4,754.61) corresponds to an ARMA(14,13), but an ARMA(6,3) still exhibits independent errors[2] and has the eight lowest AIC (4,750.22). The more parsimonious specification is also chosen in this case. The ARMA order with the smallest AIC (p=11, q=11) is selected for Series 3 because there is no substantially more parsimonious model with seemingly independent errors[2]. Finally, an ARMA(13,19) with an AIC of 2,308.90 is chosen for Series 4 since this is the least parameterized model with independent errors[2].

Simulation Analyses

In the following analyses, it is only assumed that the previously selected ARIMA data-generating processes are generally representative of the dynamic behavior of economic variables which are observed on a weekly basis. While it is not implied that they are the exact models underlying the observed oil prices, bond yields, and exchange and federal funds rates, from here on, they will be referred to as the “true” models.

The next step in the analysis is to compute the Mean Square Errors of the monthly, quarterly and annual out-of-sample forecasts from those “true” models, i.e. assuming that their parameter values are known. The forecast errors are defined as the differences between the average of the first four (for “monthly”), twelve (for “quarterly”) and forty eight (for “annual”) periods ahead predictions from these weekly models and the average of the values “observed” during those periods[3]. The observed values and predictions, of course, are for the original (non-differenced) series. The MSE of the forecasts from the corresponding aggregate (monthly, quarterly and annual) models are needed as well. In the case of the monthly models, for example, the forecast errors are defined as the differences between the averages of the one (for monthly), three (for quarterly) and twelve (for annual) period-ahead predictions and the averages of the four, twelve and forty eight weekly values observed during those periods.

As previously mentioned, Koreisha and Fang (2004) derived formulas to compute the MSE of the forecasts from monthly AR(1), AR(2), MA(1), MA(2), ARMA(1,1), ARMA(1,2) and ARMA(2,1) models with known parameter values, and from their corresponding aggregate (quarterly and annual) models. These authors assumed that the aggregate models conform to the result of Brewer (1973) which states that if the disaggregate model is ARMA(p,q) the aggregated time series follows and ARMA(p, [p+1+(q-p-1)/m]) where [k] denotes the integer part of k and m is the number of periods being aggregated. For future reference, it is important to point out that the orders prescribed for the aggregate series (p and [p+1+(q-p-1)/m]) are only maximum (Stram and Wei, 1986) in the sense that there could be more parsimonious ARMA specifications that also exhibit independently distributed errors. Also note that the procedures utilized by Koreisha and Fang (2004) require derivation of the parameter values implied for the aggregate models based on those assumed for the disaggregate model. This derivation becomes exponentially complex for larger values of p and q, which might explain why they limit their examples to aggregations of monthly models with $p \leq 2$, $q \leq 2$ and $p+q \leq 3$.

Given that the objectives of this research include exploring the potential forecasting efficiency gains from utilizing more disaggregate (weekly) models, which realistically exhibit much larger p and q orders, and to obtain results pertaining to small sample applications where the parameter values are unknown and have to be estimated, and for other key reasons that will become apparent later in the paper, an alternative simulation-based procedure to compute estimates for the MSE of the forecasts from the disaggregate and aggregate models is developed. This procedure (A) consists of the following steps:

1) A sample of T weekly observations is simulated from an ARMA(p,q) process using the order and parameter values corresponding to each of the four “true” models discussed in the previous section. Since the simulated observations are actually first-differences, the corresponding level values are recovered assuming zero as the starting level.

2) The first 1,000 observations from both the first-difference and the level samples are discarded to eliminate the impact of the starting values for the ARMA simulates, which were also set to zero.

3) The last 192 observations are also excluded and saved for the out-of-sample forecast evaluation.

4) The remaining T-1,192 observations from the first-difference sample are utilized to estimate an ARMA(p,q) model using the Gauss 9.0 ARIMA module. The same ARIMA module is used to produce 1- to 192-period-ahead forecasts from the estimated model. Those predictions are then translated into level forecasts by adding the appropriate number of predicted differences to the last observed level value.

5) The resulting forecasts and the last 192 observations excluded and saved from the simulated levels sample are used to compute the errors of the model’s monthly, quarterly and annual one-, two-, three- and four-period-ahead predictions. In the annual case[3], for example, the level forecasts for weeks 1 to 48, 49 to 96, 97 to 144 and 145 to192 are averaged and subtracted from the average of the observed values during those four periods to obtain the corresponding prediction errors.

6) The same T-1,192 first-difference observations are then aggregated into monthly, quarterly and annual data, i.e. averaged over every 4, 12 and 48 time periods respectively.

7) Monthly, quarterly and annual ARMA(p*,q*) models are then estimated on the basis of that data. The order of those models is determined utilizing the previously discussed result of Brewer (1973).

8) One-, two-, three- and four-period-ahead difference forecasts from those models are also obtained using the Gauss 9.0 ARIMA module, and prediction errors are computed by subtracting their corresponding levels forecasts from the appropriate observed averages of the weekly levels data.

9) Quarterly and annual predictions and prediction errors are analogously computed for the monthly model, and annual predictions and prediction errors for the quarterly model, in order to allow for forecasting precision comparisons across monthly versus quarterly and annual, and quarterly versus annual levels of aggregation as well.

10) Steps 1-9 above are repeated 10,000 times to obtain 10,000 of prediction errors corresponding to each of the previously discussed models and forecasting timeframes. Those errors are squared and averaged to obtain the desired MSE estimates.

A secondary issue of interest related to step 7) above is that while the orders for the aggregate models determined by applying Brewer's (1973) formula are guaranteed to make their error terms independently distributed, it is possible that more parsimonious models can be found which: 1) also exhibit independent errors, and 2) produce more efficient forecasts under finite sample conditions. The following procedure (B) is utilized in order to investigate the first part of this hypothesis:

1) 1,000 samples of T=5,000 weekly observations are simulated from ARMA(p,q) models with the orders and parameter estimates of each of the four "true" first-difference models discussed in the previous section.
2) The first 1,000 observations from each sample are discarded to eliminate the impact of the starting values for the ARMA simulates, which were set to zero.
3) The remaining T-1,000=4,000 observations in each sample are aggregated into monthly data.
4) 1,000 ARMA models of increasing order (p=1, q=0; p=0, q=1; p=1, q=1; p=2, q=0; p=0, q=2; p=2, q=1; p=1, q=2; p=2, q=2; …) are estimated on the basis of the 1,000 monthly samples.
5) Box-Pierce tests with 120 lagged residuals are conducted and the lowest ARMA(p,q) order under which less than 100 out of 1,000 estimated models fail this test at α=10% and less than 10 at α=1% is selected as the most parsimonious monthly specification that is likely to exhibit independent errors.
6) The above steps are repeated with T=13,000 and T=49,000 to obtain 12,000 and 48,000 weekly observations which are then aggregated into 1,000 quarterly and 1,000 annual observations and used to determine the order of the most parsimonious quarterly and annual models likely to exhibit independent errors.

Results

As previously indicated, procedure (A) is applied assuming that the true models are the ARMA(p,q) data-generating processes implied by the estimated US oil spot price, 10-year US bond yield, US/Japan exchange rate, and US federal funds rate first-difference models discussed in a previous section. Since the presumed true models are bona fide estimates of actual economic time-series processes, the results should be generally representative of what could be expected in practice. The first set of results corresponds to a scenario where T=49,192 weekly observations are repeatedly simulated and 48,000 of them are made available for model estimation. This means that the monthly, quarterly and annual models are estimated with 12,000, 4,000 and 1,000 observations respectively. At these sample sizes the parameter estimates tend to be close but are still not identical to their "true" values. Therefore, this scenario only approximates the "known parameter value" assumption underlying previous theoretical analyses.

The results, presented in the left half of Table 1 {Model 1(L) to Model 4(L)}, are generally consistent across the four data series in regard to the forecasting efficiency gains that can be expected from using models with lower levels of data aggregation. In the case of one-period-ahead forecasts, disaggregation from monthly to weekly (M to W) data reduces the MSEs by an average of 32.41% (range of 25.38% to 39.41%). Interestingly, the gains from quarterly to weekly (Q to W) and annual to weekly (A to W) disaggregation are even higher (average of 45.98% and 44.23% respectively). Disaggregation from quarterly to monthly (Q to M), annual to monthly (A to M) and annual to quarterly (A to Q) data yields marked MSE reductions (average of 35.25%, 41.38% and 32.51 %) as well. Improvements in the precision of the two-, three- and four-period-ahead predictions are lower but remain notable at all levels of disaggregation, averaging 14.85% for two periods ahead, 8.90% for three periods ahead and 6.72% for four periods ahead.

The next set of results relate to the hypothesis that lower orders than those prescribed by Brewer's (1973) formula for the aggregate ARMA models can be found which also exhibit independent errors. Table 2 presents the orders of the monthly, quarterly and annual models implied by that formula for each of the four weekly processes under consideration, and the orders obtained on the basis of the second procedure (B) described in the previous section. Note that there is always a model order which clearly exhibits independent errors and is substantially more parsimonious than Brewer's. In fact, aggregate quarterly and annual models with independent errors can be obtained with very low ($p \leq 2$, $q \leq 1$) ARMA orders in three of the four cases, even when the Brewer's formula suggests that much higher orders are needed. In addition, note that the average AIC values are always lower in the more parsimonious models at all levels of aggregation. These results support the general observation that, in practice, the optimal order of ARMA models rapidly declines with aggregation.

The results in Table 2 also illustrate the loss of information about the dynamic behavior of the time series that can occur with aggregation. In the case of Series 4, for example, the weekly model is an ARMA(13,19) with highly significant parameters, which theoretically involves several overlapping cycles of various lengths. And the most parsimonious monthly model with independent errors is still an ARMA(6,5). In contrast, according to Procedure B, the quarterly and annual models are both ARMA(0,1) with barely significant MA parameters.

The next question is whether the more parsimonious models identified through Procedure B exhibit smaller forecasting errors than those specified using Brewer's formula. Under an infinite sample size, it stands to reason that the forecasts from all models with independently distributed errors should have the same MSE regardless of their ARMA order. In other words, adding statistically unnecessary parameters should not affect the asymptotic efficiency of the forecasts. Alternatively, in finite-sample applications, it seems that more parsimonious models, so long as they have independent errors and a lower AIC, should yield more accurate predictions.

In order to explore the validity of this claim, Procedure A is applied with a sample size of SS = 3840 weekly = 960 monthly = 320 quarterly = 80 annual simulated observations, which is a high upper-bound for what might be available in practice, and under two sets of aggregate model orders: the ones prescribed by Brewer's (1973) formula and the parsimonious specifications identified through Procedure (B). Table 3 shows the percentage increase in the MSE of the forecasts resulting from the use of the more heavily parameterized models. In the case of the forecasts from the monthly models (M-M, M-Q and M-A) small but consistent MSE increases are observed even at this relatively large sample size of 960 monthly observations. The differences in this case are not striking because there is still an ample number of observations available to estimate the excessively parameterized monthly models.

As expected, since there are only 320 observations available for model estimation, much larger differences are generally found when comparing the MSE of the forecasts from the quarterly models (Q-Q and Q-A). In the case of the annual predictions (A-A), the MSE increases range between 11 and 23 percent. Finally note that the magnitudes of the differences seem to hold for the longer-term (up to four-period-ahead) forecasts. In short, it appears that the formula developed by Brewer (1973) is a very conservative upper bound for the ARMA order required for aggregated models to exhibit independently distributed errors. As a result, in small sample applications, it seems best to seek and utilize more parsimonious ARMA orders.

The previous finding is also pertinent for the next step in this research where MSE differentials for increasing levels of data aggregation are evaluated under small sample conditions because, in such cases, it becomes difficult to estimate the more aggregated models with the long ARMA orders implied by Brewer's formula. Specifically, for the selected small sample size of 1920 weekly observations, there are only 160 and 40 data points available to estimate the quarterly and annual models. Even with samples of 160 observations, convergence issues are frequently encountered when attempting to repeatedly estimate heavily parameterized

ARMA models (which is required to implement procedure A). In addition, the previous results confirm that the more lightly parameterized models exhibit a superior forecasting performance. Thus, for the small sample evaluations, Procedure A is again applied using the most parsimonious specifications reported in Table 2 for the aggregated models.

The results from these small sample evaluations are presented in the right half of Table 1 {Model 1(S) to Model 4(S)} to allow for an easy comparison with the large sample results {Model 1(L) to Model 4(L)}. Overall, it appears that the forecasting efficiency gains that can be expected from using models with lower levels of data aggregation under small sample size conditions are not much different from those observed in the large sample scenario. However, depending on the circumstances, they can be somewhat larger or smaller.

In order to explore the reason for this ambiguity, the Model 1 and Model 3 forecast MSEs which underlie the results in Table 1 are presented in Table 4. As expected, the predictions from the models estimated on the basis of small samples (SS=1,920) always exhibit larger MSEs. The relative MSE differences, however, vary widely. In the case of Model 1, for example, the MSE of the one-period-ahead annual forecasts from the weekly model (W-A) is 12.88% larger when the model is estimated with 1,920 instead of 48,000 observations. In contrast, the MSE of the annual forecasts from the annual model (A-A) is only 4.33% bigger. As a result, the forecasting precision gains from annual-to-weekly disaggregation (A to W in Table 1) are 35.28%-30.00%=5.28% smaller when SS=1,920. Alternatively, in the case of Model 3, the MSE of the one-period-ahead quarterly forecasts from the monthly model (M-Q) is only 1.26% larger at SS=1,920 while the MSE of the quarterly forecasts from the quarterly model (Q-Q) is 9.29% bigger. As a result, the forecasting precision gains from quarterly-to-monthly disaggregation (Q to M in Table 1) are 38.91%-34.09%=4.88% higher when SS=1,920.

To summarize the small-sample results (Table 1), disaggregation from monthly to weekly (M to W) data reduces the one-period-ahead forecast MSEs by an average of 32.21 % (range of 25.37 to 38.98%). The gains from quarterly to weekly (Q to W) and annual to weekly (A to W) disaggregation are even higher (average of 46.06% and 43.51% respectively). Disaggregation from quarterly to monthly (Q to M), annual to monthly (A to M) and annual to quarterly (A to Q) data yields marked MSE reductions (average of 35.88%, 41.58% and 32.23 %) as well. As in the large sample scenario, the improvements in the precision of the two-, three- and four-period-ahead predictions are lower but remain substantial at all levels of disaggregation (average of 14.28% for two periods ahead, 8.31% for three periods ahead and 6.34% four periods ahead).

Precision gains of these magnitudes should be very appealing to time series forecasters and justify a strong recommendation to collect more disaggregated data and build models at the lowest possible level of disaggregation. As previously explained, these efficiency gain estimates relate to four time-series data-generating processes that are generally representative of the dynamic behavior of economic variables which are observed on a weekly basis (oil prices, bond yields, exchange and federal funds rates). However, they are still theoretical measures of expected gains based on simulation analyses.

Given the mixed empirical evidence discussed in the introduction, a valid question is whether such gains will manifest themselves in the typical out-of-sample forecast (OSF) contest where NS sub-samples of T-NS-K+1, T-NS-K+2,…, T-K observations are extracted from an actual sample of size T and one- to K-period-ahead predictions are obtained from models based on each of those sub-samples in order to compute the OSF MSEs. Table 5 presents the results of such analyses for the four datasets considered in this study and K=4. The statistics under the All/480 heading are obtained when the maximum number of complete annual observations available (27 for series 1, 48 for series 2, 40 for series 3 and 44 for series 4) are included in the analyses and the last 480 weeks are used in the OSF evaluations. Thus the annual, quarterly and monthly OSF MSE comparisons are based on 10, 40 and 120 observations respectively.

First note that in the case of the one-period-ahead forecasts (1 P-A), on average across the four series, the actual MSE differentials are somewhat close to those obtained in the theoretical simulation-based analyses. There is, however, quite a bit of variability in the results across series. In the case of Series 3, for example, the MSE of the annual forecasts from the annual model is 25.73% and 51.07% lower than the MSEs of the annual forecasts from the monthly and the quarterly models. That is, disaggregation from annual to quarterly and monthly data substantially increases the annual forecast MSE. Additional contradictions with the theoretical findings appear in the two-period-ahead (2 P-A) predictions, with the annual forecasts from the weekly model (A to W) in Series 1, the annual forecasts from the monthly model (A to M) in Series 2, as well as the annual forecasts from the monthly and quarterly models (A to M and A to Q) in Series 3 exhibiting noticeably larger MSEs than those from the corresponding annual models.

More pronounced and frequent contradictions are observed in the three- and four-period-ahead predictions to the point where, even on average, in more than half of the cases, the aggregate models show lower MSEs than their disaggregated counterparts. Note, however, that all of these “reversals” correspond to annual or quarterly forecast comparisons, which are only based on 10 and 40 observations, respectively. In short, although the number of observations included in the OSF contest is larger than in past studies, frequent and substantial inconsistencies with the theoretical results are still observed, even in the case of one-period-ahead forecasts.

Fortunately, in three of the four series, there are sufficient observations available to significantly increase the number used in the OSF evaluations while still having a suitable amount left for model estimation. These more robust comparisons are also presented in Table 5 under the heading “All/Min=960”, which means that the smallest sub-sample used for model estimation included the first 960 weekly (20 annual, 80 quarterly, 240 monthly) observations and the remaining were used in the OSF contest. The numbers included in the contests at the various levels of aggregation are shown under the #OSF column.

Clearly, involving a higher number of observations in the OSF comparisons makes the resulting statistics more consistent with the outcomes of the simulation analyses. Although there are still a few instances where the MSEs of the forecasts from the aggregate models are somewhat lower than those from their disaggregated counterparts, the overall pattern as well as the four-series averages now strikingly favor the disaggregated models at all forecasting horizons. Depending on the level of disaggregation, average efficiency gains range from 31.64% (Q to M) to 54.82% (A to W) for the one-period-ahead, from 9.3% (Q to M) to 21.36% (Q to W) for the two-period-ahead, from -0.62% (Q to M) to 13.53% (A to M) for the three-period-ahead, and from -0.70% (Q to M) to 10.09% (M to W) for the four-period-ahead forecasts.

In short, the OSF contests provide convincing empirical evidence to the substantial benefits of disaggregation at various levels ranging from annual to weekly data. They also shed light on the reason why the evidence from previous studies has been mixed, by showing that OSF comparisons based on less than 10 annual and 40 quarterly observations are subject to substantial sampling error and can therefore yield unreliable results.

Conclusions and Recommendations:

The first scenario explored in the paper is based on very large samples, which is approximately equivalent to assuming that the true model parameters are known. Under such scenario, consistent improvements in forecasting accuracy resulting from higher levels of data/model disaggregation are found for four different data-generating processes that are broadly representative of the dynamic behavior of economic time series variables. Efficiency gains are high across all possible levels of disaggregation, i.e. from annual to quarterly, monthly and weekly; from quarterly to monthly and weekly; and from monthly to weekly. Asymptotic gains have been theoretically measured in previous studies, but only under small-order ARMA models with hypothetical parameter values, i.e. not necessarily realistic data-generating processes.

In the second scenario, the true parameter values are unknown and have to be estimated on the basis of a realistic sample size of 1920 weekly observations. This is still a simulation-based analysis which allows for a relatively precise measurement of the theoretically expected gains. Under this scenario, the efficiency gains from increased disaggregation are similar to those obtained when using large samples. This comparison had not been previously accomplished due to the difficulties in analytically computing the expected forecasting errors under finite sample sizes. That is, the simulation-based procedures advanced in this paper appear to be the only feasible way to obtain the necessary measurements. So, it is established that the forecasting efficiency gains expected from temporal disaggregation are not hindered when the model parameters have to be estimated on the basis of small samples.

Thirdly, on average, similarly high levels of accuracy improvements are observed when the actual data series are modeled and subjected to a traditional out-of-sample forecasting contest. This finding is aided by the fact that the time series being evaluated are substantially longer than those used previous studies, and that several series are considered in order to obtain this average result. In fact, when only ten years of data are used in the contests, the average gains are not as conclusive and the improvements from disaggregation are far from obvious in some of the individual series, which explains the mixed findings reported in past papers.

It is hoped that the previously discussed results, which include theoretical measurements under small sample conditions and more realistic ARMA orders and parameter values, and a convincing empirical validation of those measurements, will increase awareness of the substantial accuracy gains that can be achieved by using more disaggregated data and models for time series forecasting. Another important finding of this study is that Brewer's formula only provides a very conservative upper bound for the ARMA order required for the aggregate models corresponding to a particular disaggregate process. It seems likely that, in most cases, much more parsimonious specifications for the aggregate models can be found that exhibit independent errors and provide more accurate forecasts in small sample applications.

References

Abraham B. 1982. Temporal aggregation and time series. *International Statistical Review* **50**: 285-291.

Amemiya T, Wu R. 1972. The effect of aggregation on prediction in the autoregressive model. *Journal of the American Statistical Association* **67**: 628–632.

Brewer K. 1973. Some consequences of temporal aggregation and systematic sampling for ARMA and ARMAX models. *Journal of Econometrics* **1**: 133–154.

Butter F. 1976. The use of monthly and quarterly data in an ARMA model. *Journal of Econometrics* **4**: 311–324.

Georgoutsos D, Kouretas G, Tserkezos D. 1998. Temporal aggregation in structural VAR models. *Applied Stochastic Models and Data Analysis*. **14**:19-34.

Gonzalez P. 1992. Temporal aggregation and systematic sampling in structural time series models. *Journal of Forecasting* **11**: 271-281.

Koreisha S and Fang Y. 2004. Updating ARMA predictions for temporal aggregates. *Journal of Forecasting*. **23**: 275–296

Lütkepohl H. 1984. Forecasting contemporaneously aggregated vector ARMA processes. *Journal of Business and Economic Statistics* **2**: 201–214.

Lütkepohl H. 1987. *Forecasting Aggregated Vector ARMA Processes*. Springer-Verlag: Berlin.

Man K, Tiao G. 2006. Aggregation effect and forecasting temporal aggregates of long memory processes. *International Journal of Forecasting* **22**: 267– 281.

Nijman T., Palm F. 1990. Predictive accuracy gain from disaggregate sampling in ARIMA models. *Journal of Business & Economic Statistics* **8**: 405–415.

Ouliaris S, Park J, Phillips P. 1989. Testing for a unit root in the presence of a maintained trend. In *Advances in Econometrics and Modeling*, Raj B. (eds). Kluwer Academic Publishers: Dordrecht/Boston/London.

Rossana R, Seater J. 1995. Temporal Aggregation and Economic Time Series. *Journal of Business & Economic Statistics*, **13**(4): 441-451

Rose E. 1977. Forecasting aggregates of independent ARIMA processes. *Journal of Econometrics* **5**: 323-345.

Silvestrini A, Veredas D. 2008. Temporal aggregation of univariate and multivariate time series models: a survey. *Journal of Economic Surveys* **22**: 458–497.

Silvestrini A, Salto M, Moulin L, Veredas D. 2008. Monitoring and forecasting annual public deficit every month: the case of France. *Empirical Economics* **34**: 493–524.

Stram D, Wei W. 1986. Temporal aggregation in the ARIMA process. *Journal of Time Series Analysis* **7**: 279–292.

Tiao G. 1972. Asymptotic behavior of temporal aggregates of time series. *Biometrika* **59**: 525–531.

Tiao G, Guttman A. 1980. Forecasting contemporal aggregates of multiple time series. *Journal of Econometrics* **12**: 219-230.

Wei W. 1978. Some consequences of temporal aggregation in seasonal time series models. In *Seasonal Analysis of Economic Time Series Analysis*, Zellner A (eds). US Department of Commerce, Bureau of the Census: Washington, DC.

Wei W. 1978. The effect of temporal aggregation on parameter estimation in distributed lag models. Journal of Econometrics **8**: 237-246.

Wei W. 1990. *Time Series Analysis*. Addison-Wesley: New York.

Weiss A. 1984. Systematic sampling and temporal aggregation in time series models. *Journal of Econometrics* **26**: 271-281.

Footnotes

[1] With t-test(s) rejecting the null hypotheses of equal variance across periods (α<0.001).

[2] The errors are considered independent if Ho can't be rejected at an α=0.25 by the Box-Pierce test with 120 lagged residuals.

[3] To simplify the computations required later in the analyses, a month, quarter and year are assumed to consist of four, twelve and forty-eight weeks respectively. Thus, for the remainder of the paper, month, quarter and year will used to denote simulated data-periods of these weekly lengths. In addition, "observed" and "observation" will be used in reference to simulated data.

Table 1: Percentage forecast MSE reductions resulting from different levels of disaggregation under large (L) and small (S) sample sizes (SS=48,000 and SS=1,920 observations).

Model 1(L)	**1 P-A**	**2 P-A**	**3 P-A**	**4 P-A**	**Model 1(S)**	**1 P-A**	**2 P-A**	**3 P-A**	**4 P-A**
M to W	26.39	12.37	8.08	5.74	**M to W**	25.61	10.73	5.68	2.51
Q to W	43.54	15.46	6.98	4.34	**Q to W**	42.40	11.30	0.22	-0.39
A to W	35.28	8.48	6.21	5.35	**A to W**	30.00	0.32	0.97	1.35
Q to M	34.02	11.34	4.79	2.88	**Q to M**	34.19	10.64	2.78	1.58
A to M	32.22	7.59	5.41	4.68	**A to M**	30.49	3.89	3.19	3.57
A to Q	24.09	5.64	4.27	3.68	**A to Q**	23.84	4.69	4.56	4.26
Model 2(L)	**1 P-A**	**2 P-A**	**3 P-A**	**4 P-A**	**Model 2(S)**	**1 P-A**	**2 P-A**	**3 P-A**	**4 P-A**
M to W	38.44	15.17	8.68	5.74	**M to W**	38.89	15.56	9.02	6.24
Q to W	43.74	14.04	8.53	6.86	**Q to W**	44.20	14.03	8.72	7.01
A to W	42.23	16.18	10.13	7.63	**A to W**	43.26	16.42	10.06	7.65
Q to M	31.03	9.67	5.92	4.91	**Q to M**	31.24	9.22	5.60	4.59
A to M	39.07	14.98	9.26	6.91	**A to M**	39.79	14.66	8.62	6.35
A to Q	31.57	11.99	7.35	5.57	**A to Q**	32.62	12.18	7.15	5.34
Model 3(L)	**1 P-A**	**2 P-A**	**3 P-A**	**4 P-A**	**Model 3(S)**	**1 P-A**	**2 P-A**	**3 P-A**	**4 P-A**
M to W	39.41	15.85	9.03	6.56	**M to W**	38.98	15.38	8.82	6.51
Q to W	46.48	19.45	12.45	9.24	**Q to W**	50.17	26.45	19.13	15.67
A to W	47.36	18.29	11.45	8.37	**A to W**	48.59	18.73	11.59	8.53
Q to M	34.09	14.97	9.82	7.52	**Q to M**	38.91	22.53	16.91	14.29
A to M	44.51	17.51	11.07	8.01	**A to M**	45.96	18.07	11.32	8.26
A to Q	34.74	12.93	7.72	5.19	**A to Q**	31.15	7.92	3.06	0.69
Model 4(L)	**1 P-A**	**2 P-A**	**3 P-A**	**4 P-A**	**Model 4(S)**	**1 P-A**	**2 P-A**	**3 P-A**	**4 P-A**
M to W	25.38	13.34	8.46	7.08	**M to W**	25.37	12.68	7.42	6.57
Q to W	50.17	24.77	12.81	10.39	**Q to W**	47.48	22.55	11.00	8.67
A to W	52.02	20.79	13.00	9.82	**A to W**	52.18	20.40	12.41	8.92
Q to M	41.85	20.32	10.10	8.48	**Q to M**	39.19	17.97	8.52	7.29
A to M	49.70	19.82	12.34	9.30	**A to M**	50.09	19.94	12.40	9.03
A to Q	39.63	15.44	9.65	7.09	**A to Q**	41.32	16.43	10.34	7.56

Notes: M to W, Q to W, and A to W compare the MSEs of the monthly, quarterly and annual forecasts from the respective aggregate models versus those from a weekly model. Q to M and A to M compare the MSEs of the quarterly and annual forecasts from the respective aggregate models versus those from a monthly model. A to Q compares the MSE of the annual forecast from the annual model versus the MSE of the annual forecast from the quarterly model. 1 P-A, 2 P-A, 3 P-A, and 4 P-A refer to the one-, two- three- and four-period-ahead forecast comparisons.

Table 2: ARMA model orders prescribed by Brewer's (1973) formula versus those of the most parsimonious models with independently distributed errors identified through Procedure (B).

	Brewer's			**Most Parsimonious**		
Model 1	**Order**	**AIC**	**P-Value**	**Order**	**AIC**	**P-Value**
Monthly	(8,8)	3733.53	0.960	(2,1)	3731.14	0.618
Quarterly	(8,8)	4983.66	0.936	(2,1)	4981.49	0.608
Annual	(8,9)	5937.57	0.964	(1,1)	5933.35	0.661
	Brewer's			**Most Parsimonious**		
Model 2	**Order**	**AIC**	**P-Value**	**Order**	**AIC**	**P-Value**
Monthly	(6,6)	4344.05	0.874	(3,2)	4341.98	0.7095
Quarterly	(6,6)	5460.10	0.904	(1,1)	5457.56	0.6660
Annual	(6,6)	6657.37	0.904	(0,1)	6651.39	0.6646
	Brewer's			**Most Parsimonious**		
Model 3	**Order**	**AIC**	**P-Value**	**Order**	**AIC**	**P-Value**
Monthly	(11,11)	4387.51	0.9207	(3,3)	4372.39	0.6702
Quarterly	(11,11)	5552.32	0.9398	(5,5)	5545.55	0.7548
Annual	(11,12)	6940.78	0.9753	(0,1)	6931.24	0.6638
	Brewer's			**Most Parsimonious**		
Model 4	**Order**	**AIC**	**P-Value**	**Order**	**AIC**	**P-Value**
Monthly	(13,15)	3691.55	0.9826	(6,5)	3688.94	0.7752
Quarterly	(13,14)	5218.38	0.9824	(0,1)	5209.31	0.5784
Annual	(13,14)	7003.40	0.9836	(0,1)	6984.81	0.6749

Notes: P-value is for testing the null hypothesis of independent errors using the Box-Pierce statistic with 120 lagged residuals. The AIC and p-values are averages for 1000 models estimated on the basis of 1000 simulated observations each (see Procedure B).

Table 3: Percentage increase in the MSE of the forecasts from the models with the ARMA orders prescribed by Brewer's (1973) formula versus the most parsimonious orders with independently distributed errors identified through Procedure B, under a sample size of 3,840 weekly observations.

	Model 1				**Model 2**			
	1 P-A	**2 P-A**	**3 P-A**	**4 P-A**	**1 P-A**	**2 P-A**	**3 P-A**	**4 P-A**
M-M	2.238	1.892	1.525	1.135	1.181	1.065	1.269	1.078
M-Q	1.717	0.924	0.924	0.750	1.119	0.741	0.431	0.392
M-A	1.012	0.529	0.482	0.671	0.529	0.343	0.347	0.505
Q-Q	4.622	4.420	4.341	3.698	4.625	4.529	4.291	3.772
Q-A	4.159	2.993	3.493	3.537	4.185	2.990	2.749	2.568
A-A	14.512	12.966	11.614	10.401	13.194	13.557	12.188	11.680
	Model 3				**Model 4**			
	1 P-A	**2 P-A**	**3 P-A**	**4 P-A**	**1 P-A**	**2 P-A**	**3 P-A**	**4 P-A**
M-M	1.671	1.753	1.879	1.829	3.017	2.769	3.089	3.457
M-Q	1.861	1.660	1.032	1.451	3.085	3.565	3.373	2.850
M-A	1.367	1.391	1.151	0.934	3.317	2.514	2.455	2.497
Q-Q	1.832	1.564	1.900	1.975	5.951	7.134	7.040	6.581
Q-A	1.663	2.354	1.642	1.334	6.925	5.863	5.261	4.507
A-A	19.589	21.801	21.491	20.603	21.791	22.836	21.886	21.472

Notes: M-M, M-Q and M-A refer to the monthly, quarterly and annual forecasts from the monthly models, Q-Q and Q-A refer to the quarterly and annual forecasts from the quarterly models, and A-A refers to the annual forecast from the annual model. 1 P-A, 2 P-A, 3 P-A, and 4 P-A refer to the one-, two- three- and four-period-ahead forecast comparisons.

Table 4: Forecast MSE from Models 1 and 3 for under large (L) and small (S) sample sizes (48,000 and 1,920 observations).

Model 1(L)	**1 P-A**	**2 P-A**	**3 P-A**	**4 P-A**	**Model 3(S)**	**1 P-A**	**2 P-A**	**3 P-A**	**4 P-A**
W-M	1.87	6.27	11.21	16.13	**W-M**	2.95	11.14	19.22	27.31
W-Q	5.00	19.1	29.86	36.95	**W-Q**	8.50	32.37	55.25	78.36
W-A	16.21	39.49	53.84	65.63	**W-A**	31.89	128.7	227.3	328.5
M-M	2.51	7.03	11.89	16.54	**M-M**	4.84	13.17	21.08	29.21
M-Q	5.72	19.24	29.09	36.22	**M-Q**	10.42	34.10	56.77	79.64
M-A	16.09	38.07	52.64	64.15	**M-A**	33.52	129.7	228.0	329.5
Q-Q	8.69	21.53	29.93	36.8	**Q-Q**	17.06	44.02	68.32	92.92
Q-A	17.63	37.76	51.90	63.69	**Q-A**	42.71	145.8	249.2	356.6
A-A	23.15	39.62	54.37	66.53	**A-A**	62.03	158.3	257.1	359.1
Model 1(L)	**1 P-A**	**2 P-A**	**3 P-A**	**4 P-A**	**Model 3(L)**	**1 P-A**	**2 P-A**	**3 P-A**	**4 P-A**
W-M	1.83	6.03	10.62	15.08	**W-M**	2.88	10.88	18.99	27.17
W-Q	4.77	17.54	26.69	33.30	**W-Q**	8.35	32.22	54.97	77.85
W-A	14.36	34.97	49.12	60.9	**W-A**	31.66	128.2	226.5	327.5
M-M	2.48	6.88	11.56	16.00	**M-M**	4.75	12.93	20.87	29.08
M-Q	5.58	18.39	27.32	33.81	**M-Q**	10.29	34.01	56.62	79.34
M-A	15.04	35.31	49.54	61.33	**M-A**	33.38	129.4	227.5	328.8
Q-Q	8.45	20.74	28.69	34.81	**Q-Q**	15.61	40.00	62.78	85.78
Q-A	16.84	36.05	50.13	61.97	**Q-A**	39.30	136.6	236.0	338.9
A-A	22.19	38.21	52.37	64.34	**A-A**	60.15	156.9	255.8	357.4
% Diff 1	**1 P-A**	**2 P-A**	**3 P-A**	**4 P-A**	**% Diff 3**	**1 P-A**	**2 P-A**	**3 P-A**	**4 P-A**
W-M	2.19	3.98	5.56	6.96	**W-M**	2.43	2.39	1.21	0.52
W-Q	4.82	8.89	11.88	10.96	**W-Q**	1.80	0.47	0.51	0.66
W-A	12.88	12.93	9.61	7.77	**W-A**	0.73	0.39	0.35	0.31
M-M	1.21	2.18	2.85	3.37	**M-M**	1.89	1.86	1.01	0.45
M-Q	2.51	4.62	6.48	7.13	**M-Q**	1.26	0.26	0.26	0.38
M-A	6.98	7.82	6.26	4.60	**M-A**	0.42	0.23	0.22	0.21
Q-Q	2.84	3.81	4.32	5.72	**Q-Q**	9.29	10.05	8.82	8.32
Q-A	4.69	4.74	3.53	2.78	**Q-A**	8.68	6.73	5.59	5.22
A-A	4.33	3.69	3.82	3.40	**A-A**	3.13	0.89	0.51	0.48

Note: W-M, W-Q and W-A refer to the monthly, quarterly and annual forecasts from the weekly model; M-M, M-Q and M-A refer to the monthly, quarterly and annual forecasts from the monthly model; Q-Q and Q-A refer to the quarterly and annual forecasts from the quarterly model; and A-A refers to the annual forecasts from the annual model. 1 P-A, 2 P-A, 3 P-A, and 4 P-A refer to the one-, two- three- and four-period-ahead forecast comparisons.

Table 5: Percentage forecast MSE reductions resulting from different levels of disaggregation in out-of-sample forecasting (OSF) contests based on the actual data.

	Model 1 - All/480					**Model 1 - All/Min=960**				
	1 P-A	**2 P-A**	**3 P-A**	**4 P-A**	**#OSF**	**1 P-A**	**2 P-A**	**3 P-A**	**4 P-A**	**#OSF**
M to W	45.94	20.32	11.25	6.39	120.00	45.94	20.32	11.25	6.39	120.00
Q to W	53.72	18.11	-4.55	-8.02	40.00	53.72	18.11	-4.55	-8.02	40.00
A to W	60.73	-16.29	-27.05	-15.50	10.00	60.73	-16.29	-27.05	-15.50	10.00
Q to M	33.39	11.34	-5.67	-10.49	40.00	33.39	11.34	-5.67	-10.49	40.00
A to M	58.91	33.44	30.07	31.08	10.00	58.91	33.44	30.07	31.08	10.00
A to Q	44.89	8.90	7.71	14.01	10.00	44.89	8.90	7.71	14.01	10.00
	Model 2 - All/480					**Model 2 - All/Min=960**				
	1 P-A	**2 P-A**	**3 P-A**	**4 P-A**	**#OSF**	**1 P-A**	**2 P-A**	**3 P-A**	**4 P-A**	**#OSF**
M to W	50.89	24.59	16.53	15.84	120.00	44.55	19.11	12.63	9.42	288.00
Q to W	54.04	17.31	-5.92	3.07	40.00	51.28	15.71	3.79	6.89	96.00
A to W	43.75	13.15	3.28	-11.16	10.00	47.11	12.92	13.84	5.25	24.00
Q to M	33.74	2.96	-17.98	-6.38	40.00	34.05	7.37	-1.95	1.43	96.00
A to M	25.38	-8.79	-25.02	-30.89	10.00	39.26	5.25	5.77	-2.22	24.00
A to Q	57.27	31.13	19.36	5.95	10.00	45.97	21.99	16.19	6.18	24.00
	Model 3 - All/480					**Model 3 - All/Min=960**				
	1 P-A	**2 P-A**	**3 P-A**	**4 P-A**	**#OSF**	**1 P-A**	**2 P-A**	**3 P-A**	**4 P-A**	**#OSF**
M to W	50.48	25.23	19.20	11.58	120.00	37.83	24.54	14.99	10.49	192.00
Q to W	54.37	26.15	3.96	9.50	40.00	58.58	28.75	6.73	7.31	64.00
A to W	38.77	19.14	0.13	-11.47	10.00	47.53	24.40	18.65	6.02	16.00
Q to M	34.90	14.94	0.22	6.26	40.00	30.26	17.73	7.98	6.97	64.00
A to M	-25.73	-29.23	-39.17	-43.60	10.00	4.09	-6.35	4.19	2.73	16.00
A to Q	-51.07	-46.05	-55.35	-92.26	10.00	-15.73	-15.10	0.54	-5.84	16.00
	Model 4 - All/480					**Model 4 - All/Min=960**				
	1 P-A	**2 P-A**	**3 P-A**	**4 P-A**	**#OSF**	**1 P-A**	**2 P-A**	**3 P-A**	**4 P-A**	**#OSF**
M to W	15.93	11.49	11.23	14.08	120.00	12.93	12.56	12.77	14.08	240.00
Q to W	56.79	26.19	15.99	13.53	40.00	50.18	22.88	13.81	11.46	80.00
A to W	68.96	26.89	9.48	-1.85	10.00	63.91	25.77	11.89	3.62	20.00
Q to M	37.75	6.75	-0.38	0.84	40.00	28.85	0.74	-2.86	-0.71	80.00
A to M	68.28	27.91	12.85	4.38	10.00	63.92	26.61	14.10	7.83	20.00
A to Q	60.92	25.61	14.15	6.60	10.00	58.19	23.81	12.56	6.31	20.00

Table 5 (continued): Percentage forecast MSE reductions resulting from different levels of disaggregation in out-of-sample forecasting (OSF) contests based on the actual data.

	Averages - All/480					**Averages - All/Min=960**				
	1 P-A	**2 P-A**	**3 P-A**	**4 P-A**	**#OSF**	**1 P-A**	**2 P-A**	**3 P-A**	**4 P-A**	**#OSF**
M to W	40.81	20.41	14.55	11.97	120.00	35.31	19.13	12.91	10.09	210.00
Q to W	54.73	21.94	2.37	4.52	40.00	53.44	21.36	4.94	4.41	70.00
A to W	53.05	10.72	-3.54	-9.99	10.00	54.82	11.70	4.33	-0.15	17.50
Q to M	34.95	9.00	-5.95	-2.44	40.00	31.64	9.30	-0.62	-0.70	70.00
A to M	31.71	5.84	-5.32	-9.76	10.00	41.55	14.74	13.53	9.85	17.50
A to Q	28.00	4.90	-3.53	-16.43	10.00	33.33	9.90	9.25	5.16	17.50

Notes: All/480 indicates that the maximum number of complete annual observations available was included in the analyses and the last 480 weeks were used in the OSF evaluations. All/Min =960 indicates that the smallest sub-sample used for model estimation included the first 960 weekly observations and the remaining were used in the OSF contest. #OSF refers to the average number of observations available for the OSF contest. All other notation is as defined in the previous tables.

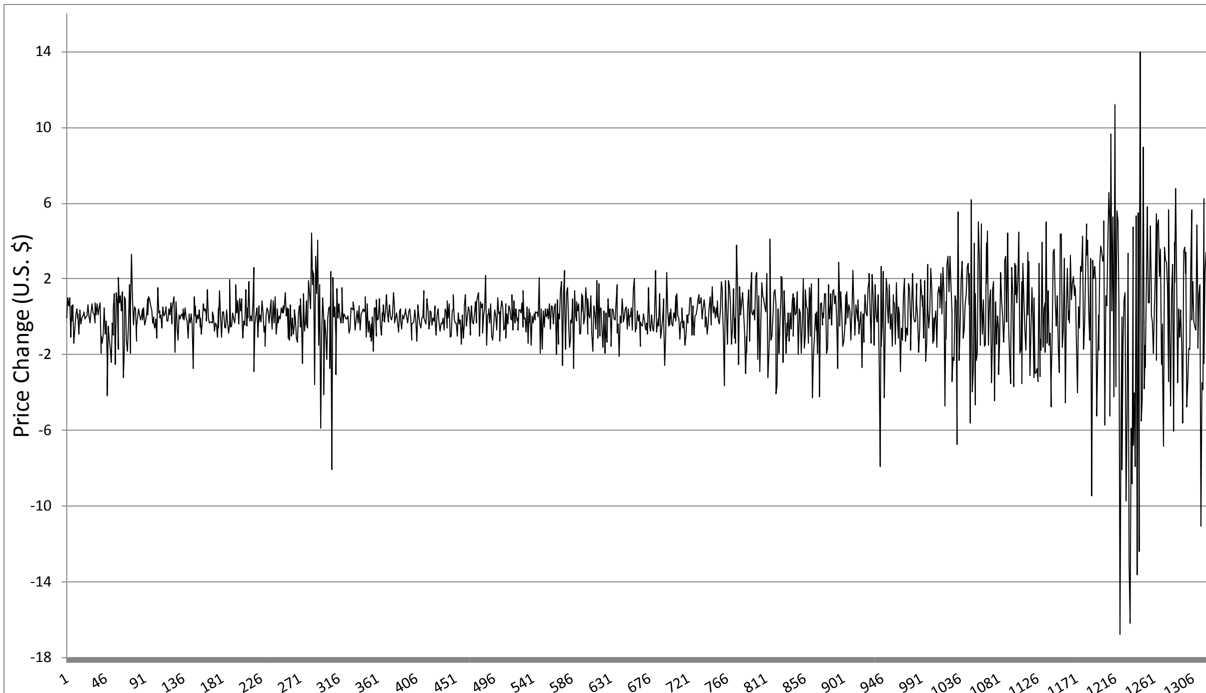

Figure 1: 1985-2010 Weekly US Oil Spot Price (first differences)

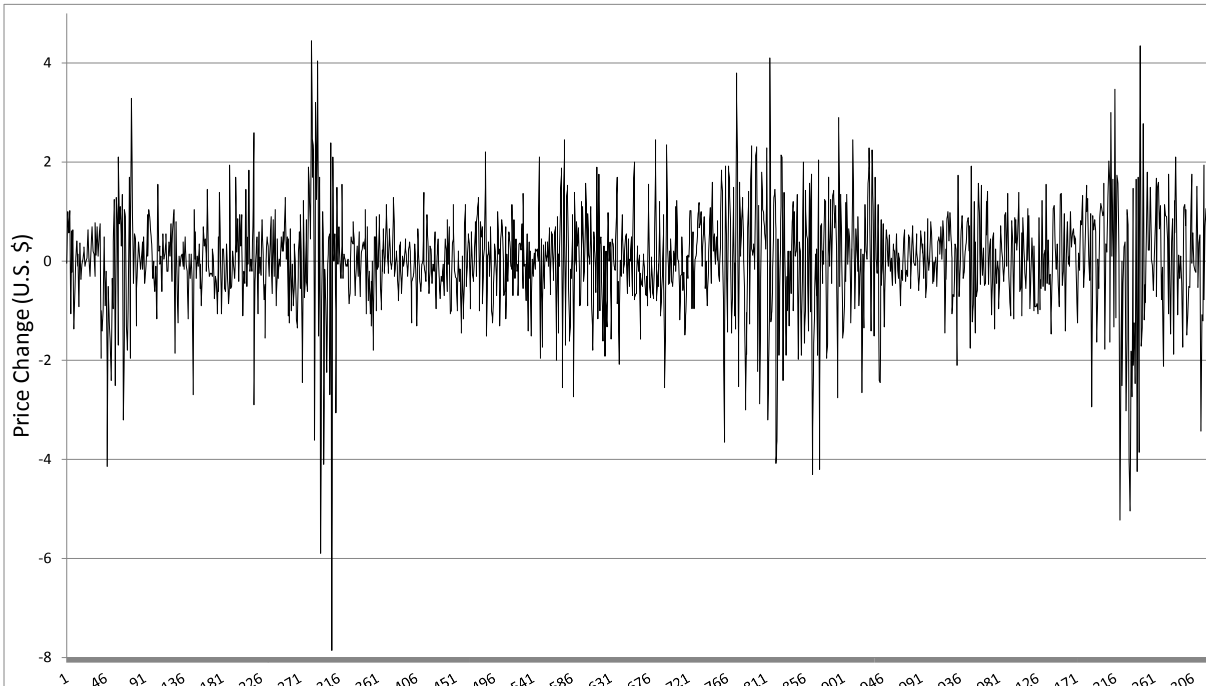

Figure 1S: 1985-2010 Weekly US Oil Spot Price (re-scaled first differences)

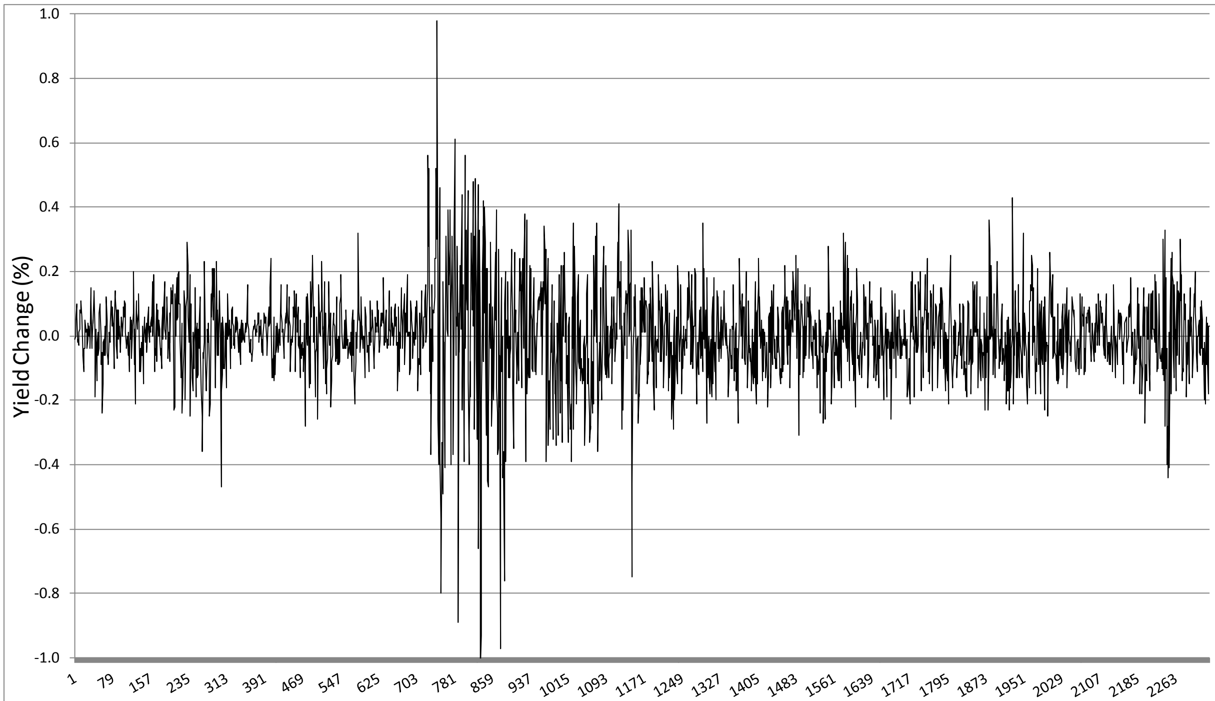

Figure 2: 1966-2010 Weekly 10-Year US Bond Yield (first differences)

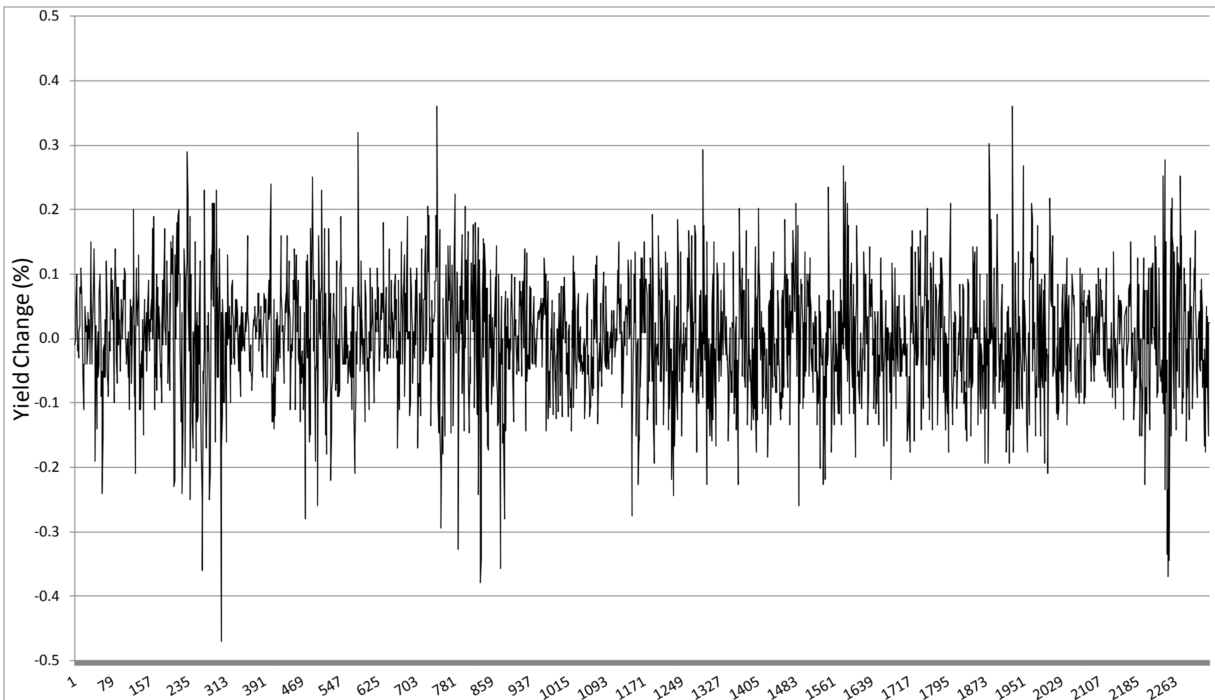

Figure 2S: 1966-2010 Weekly 10-Year US Bond Yield (re-scaled first differences)

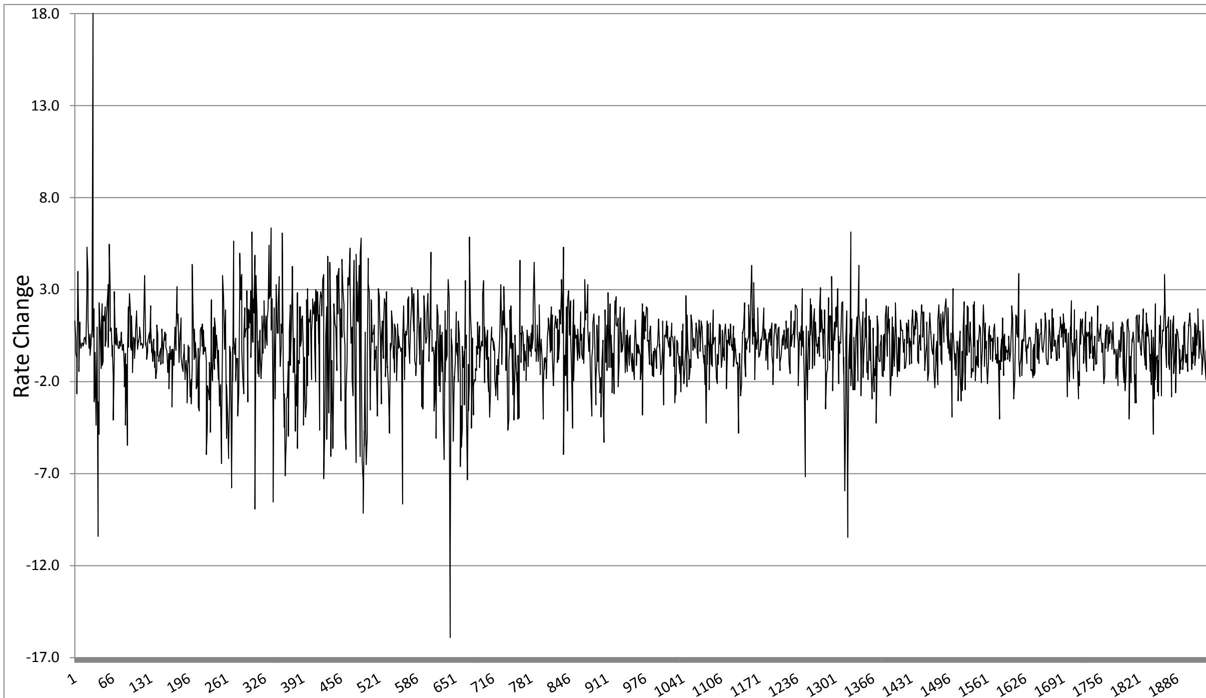

Figure 3: 1974-2010 Weekly U.S./Japan Exchange Rate (first differences)

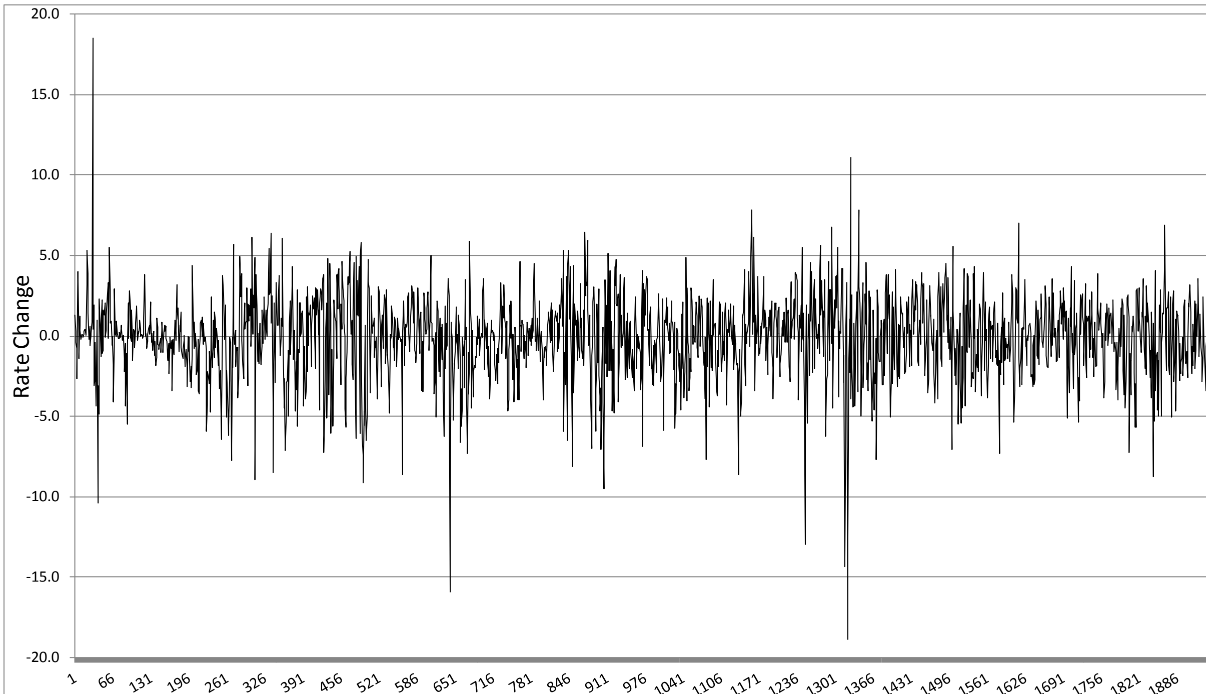

Figure 3S: 1974-2010 Weekly U.S./Japan Exchange Rate (re-scaled first differences)

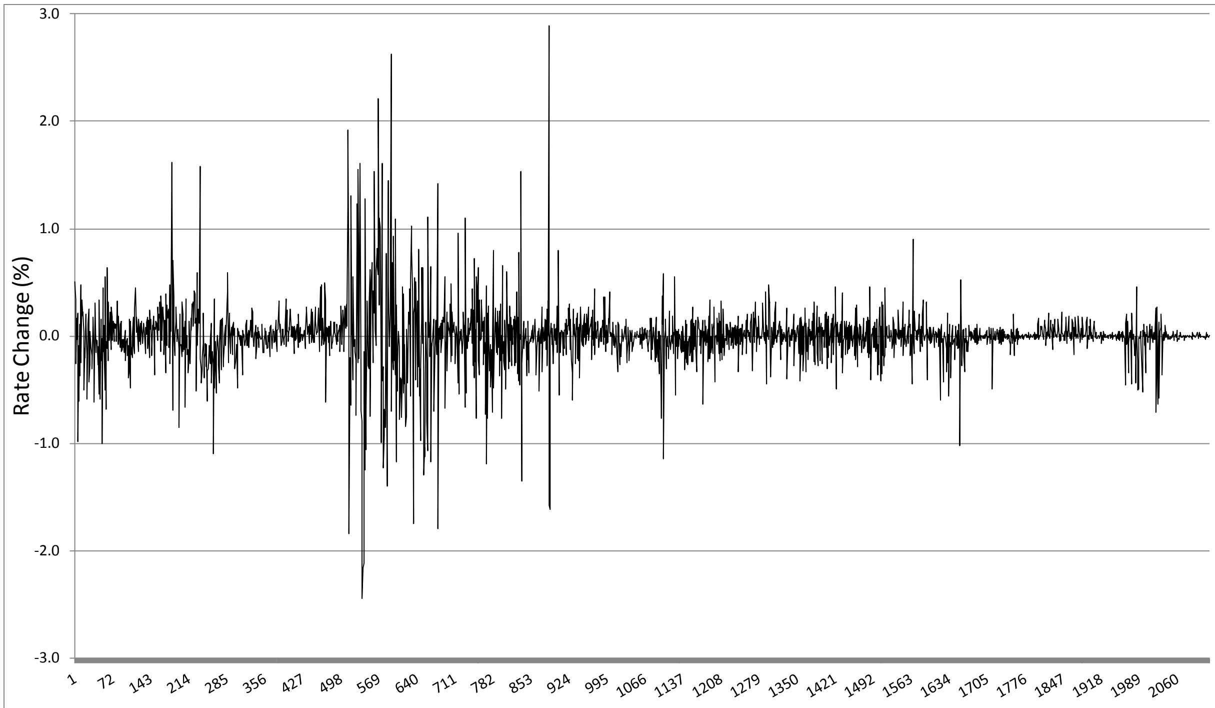

Figure 4: 1971-2010 Weekly U.S. Federal Funds Rate (first differences)

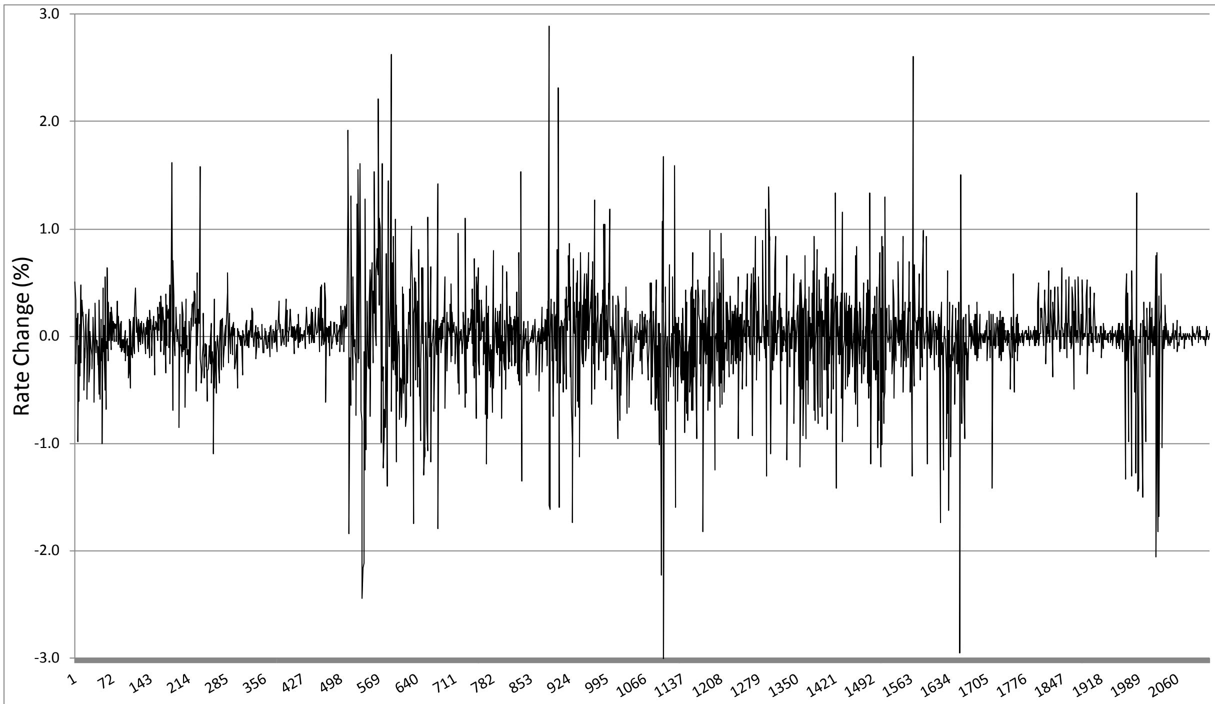

Figure 4S: 1971-2010 Weekly U.S. Federal Funds Rate (re-scaled first differences)