Imputing Seasonal Data in an Advanced Indicator with Forecasts from X-13ARIMA-SEATS

Nicole Czaplicki and Yarissa González

U.S. Census Bureau, Washington, DC 20233

Abstract

The Advanced Monthly Retail Trade Survey (MARTS) publishes early sales estimates of retail and food service companies approximately nine working days after the reference month. Retail trade data have strong seasonal patterns and known calendar effects, and the tabulated industry-level estimates are seasonally adjusted. However, the current missing and erroneous data treatment procedures do not fully account for this seasonality, instead relying entirely on respondent-based ratio estimates constructed from current month to prior month data. Such imputation procedures can yield biased estimates, especially when response rates are low and the response mechanism is unlikely to be ignorable. Furthermore, these methods do not utilize additional historic information, at both the individual unit level and at the industry level. In this paper, we demonstrate a novel imputation method for MARTS that utilizes industry level seasonal ARIMA models with calendar effects estimated with X-13ARIMA-SEATS to develop one-step ahead unit level forecast imputed values.

Key Words: Imputation, ARIMA modeling, Regression, RegARIMA, Time Series

Disclaimer: Any views expressed are those of the authors and not necessarily those of the Census Bureau. The Census Bureau has reviewed this data product for unauthorized disclosure of confidential information and has approved the disclosure avoidance applied. (Approval ID: CBDRB-FY19-ESMD-B00011.)

1. Introduction

The Advanced Monthly Retail Trade Survey (MARTS), an economic indicator of total monthly sales in the retail trade and food service industries, publishes early sales estimates approximately nine working days after the reference month. Approximately one month later, these estimates are superseded by preliminary estimates from the Monthly Retail Trade Survey (MRTS). Both MARTS and MRTS are published by the U.S. Census Bureau. These estimates are inputs into quarterly Gross Domestic Product (GDP) published by the Bureau of Economic Analysis. Revisions between MARTS and MRTS estimates are inevitable due to differences in the two survey designs, among other factors. However, large revisions, particularly those that reverse the direction of the seasonally adjusted month-to-month change, receive a high level of scrutiny. The U.S. Census Bureau is researching methodological alternatives for MARTS that may reduce revisions between the MARTS and MRTS estimates. Czaplicki, González, and Bechtel (2018) explored alternative estimators for MARTS and Thompson, Bechtel, and Czaplicki (2018) investigated donor imputation methods using propensity score matching. This paper is a continuation of that effort and looks to improve the current MARTS imputation procedures.

Retail trade data have strong seasonal patterns and known calendar effects, and the tabulated industry-level estimates are seasonally adjusted. The current missing data treatment procedure adjusts an industry-level MARTS ratio estimate of current-month-to-prior-month sales values to the prior month MRTS level. The units included in the MARTS ratio estimate either (1) reported valid values in both consecutive periods or (2) received a subjective imputed value in either the current period, the prior period, or both periods. In the second case, subject matter analysts use historic data from the same unit to derive a current period value (hereafter called analyst imputation); this is discussed further in Section 2. The analyst imputation methods do not utilize additional historic information, at both the individual unit and industry levels, which can lead to biased estimates, especially when the unit response rates are low. In this paper, we demonstrate a novel imputation method for MARTS that utilizes industry level seasonal autoregressive integrated moving average (ARIMA) models (Box and Jenkins 1970) with calendar effects estimated with X-13ARIMA-SEATS to develop one-step ahead unit level forecast imputed values.

MARTS consists of 30 disjoint industries, each with unique response patterns and economic characteristics. These industries are based on the North American Industry Classification System (NAICS) and span the entire retail trade and food services sector of the economy. The diversity of industries and the lack of auxiliary data for predictors makes finding a one-size-fits-all imputation procedure for the survey a challenge. While the proposed method reduces errors and offers better predictions in many industries, it is not superior to the current missing data treatment for all industries.

2. MARTS Background

The MARTS samples approximately 5,500 retail and food service businesses using a stratified probability proportional to size without replacement design, selected from the MRTS sample. The MRTS utilizes a stratified simple random sample without replacement design, selected from the Annual Retail Trade Survey (ARTS) sample. In other words, the MARTS sample is a subsample of MRTS, which is itself a subsample of ARTS. A new MRTS sample is selected approximately every five years. For every MRTS sample, two MARTS samples are selected. The first is introduced at the same time as the new MRTS sample and the second is introduced approximately two and a half years later, halfway through the MRTS sample lifecycle. For both surveys, units that exceed a predetermined industry size cutoff are selected into the sample with certainty, although the size cutoffs for MARTS are often higher than those for MRTS. Thus, a certainty unit in MARTS will also be a certainty unit in MRTS but a certainty unit in MRTS is not necessarily a certainty unit in MARTS. More details about the MARTS design, can be found at: https://www.census.gov/retail/marts/how_surveys_are_collected.html; more details about the MRTS design, can be found at: https://www.census.gov/retail/mrts/how_surveys_are_collected.html.

As the MRTS sample ages, the amount of monthly historic data available for each unit grows. At the start of a new MRTS sample, there is no monthly historic data available for units that are new to the sample. Only units that were included in the previous MRTS sample (mostly large units included in the sample with certainty) would have historic monthly data available at the start of a new sample. Since MARTS is a subset of MRTS, the historic unit level data from MRTS can be used as covariates in imputation for those units also in MARTS. MARTS only collects a single item, current month sales. Thus, there are no current period auxiliary variables from the survey that can aid in MARTS imputation, nor is there any available administrative (monthly tax) data. However, we can use the available MARTS and MRTS historic data to develop imputation methods for MARTS.

Totals for MARTS are estimated using the link relative estimator (Madow and Madow 1978), a synthetic estimator that multiplies a benchmark total from the prior month by an estimate of the current month to prior month change, also called the link relative ratio. Units that provide valid responses to MARTS in the current and prior month provide the data for the link relative ratio. The benchmark total is the preliminary MRTS estimate which uses the Horvitz-Thompson estimator (Horvitz and Thompson 1952) and includes all in-scope tabulation units; ratio imputation is used for nonresponding units. The link relative estimator essentially uses the benchmark value as an estimate of the level of the series and uses the link relative ratio to adjust the prior month level using the current month-to-month change. Equation (1) below defines the link relative estimator used to estimate MARTS industry totals, $\hat{Y}_{t,MARTS}$, where C is the set of MARTS units with data in the current and prior period, $y_{i,t}$ is the value of sales for unit *i* in time *t*, and w_i is the MARTS sample weight. If a unit is a nonrespondent, or fails an edit test, it is not included in the link relative ratio. The link relative estimator serves two purposes: (1) it ensures that the level of the MARTS estimates remains approximately consistent with the MRTS estimate level and (2) it accounts for unit nonresponse, with the assumption that nonrespondents have the same expected current month to prior month change as the estimated link relative ratio.

$$\hat{Y}_{t,MARTS} = \hat{Y}_{t-1,MRTS} \frac{\sum_C w_i y_{i,t}}{\sum_C w_i y_{i,t-1}} \quad (1)$$

The link relative estimation procedure is mathematically equivalent to a ratio imputation procedure that uses the link relative ratio to impute missing current values, then adjusts the "imputed" total to the benchmark estimate. Besides this "implied" ratio estimation, MARTS does not perform a generalized imputation procedure for all nonrespondents. However, certain "influential" nonresponding units are imputed by analysts using a combination of past company data, calendar effects, and subject matter knowledge (hereafter referred to as analyst imputes). The procedures for obtaining a replacement value may differ by company, industry, and analyst, but do have common elements. Historic company-level month-to-month changes are reviewed for the current calendar month (e.g. several years of December to January changes would be reviewed when imputing for January). Analysts typically select the unit-level historic month-to-month change from the most recent year with a matching trading day pattern to the current year (same number of Sundays, Mondays, etc.). Applying this historic month-to-month change to the unit's prior period value (reported or imputed) results in an imputed value that accounts for both seasonality and trading day effects. However, finding a matching trading day pattern may require going several years into the past, making it questionable whether that historic month-to-month change is relevant to current economic conditions. For March 2016-February 2017, the matching year is 2012-2013; for March 2017-February 2018 the matching year is 2006-2007.

The rate of analyst imputation differs by industry, with some industries receiving little or no analyst imputation. An analyst imputed value from MARTS is retained in MRTS unless (1) the unit provides a reported value (late reporter), (2) the analyst-imputed value is visibly different from the estimated industry trend, or (3) or the analyst-imputed value fails an edit test. Consequently, the imputed values in MARTS not only affect the MARTS estimates, but the MRTS estimates as well, although to a much lesser degree.

Analyst-imputed values in MARTS are reviewed by more than one subject matter expert to promote consistency across the survey. However, the procedure is subjective, as is the determination of influential units within an industry. Replacing these subjective procedures by a more objective and repeatable procedure would be an enhancement to the survey procedures. But, due to the subjective nature and lack of repeatability of the analyst imputation procedure, we cannot compare the statistical properties of the current imputation procedure to a replacement procedure. Ideally, the replacement procedure will be easily automated and not overly computer resource intensive. Our proposed imputation method incorporates many of the characteristics of the analyst imputation procedure, utilizing past company data and calendar effects, into a fast, repeatable, automated process.

3. Forecast Imputation Method

Retail sales in the United States follow very predictable seasonal patterns. Consequently, the MARTS estimates are seasonally adjusted prior to publication, although unadjusted estimates are also published. The U.S. Census Bureau uses X-13ARIMA-SEATS (X-13A-S) to seasonally adjust the MARTS time series (U.S. Census Bureau 2017). MARTS data are also adjusted for known calendar effects, such as trading day effects, which arise from the changing weekday composition of the month, and moving holiday effects, which arise from certain holidays without a fixed date (i.e. Easter).

In the X-13A-S program, prior to running the seasonal adjustment algorithm, the time series is typically modeled as a regression model with ARIMA errors (a regARIMA model). The regression component of the model may contain pre-defined regressors, such as trading day effects, moving holiday effects (Bell and Hillmer 1983, Findley and Soukup 2000), fixed seasonal effects and outliers, or user-defined regressors. The errors follow an ARIMA process (Box and Jenkins 1970). X-13A-S uses the regARIMA model to extend the time series with forecasts which are later used by the seasonal adjustment algorithm. We investigate whether similar models can also be used to estimate one-step ahead forecasts at the unit level for imputation.

Our goal is to estimate a regARIMA model using the industry time series of total sales, extract the model parameters, and use the parameters on the unit level data to produce a one-step ahead unit level forecast to serve as the imputed value for the unit. This approach combines years of industry-level time series data with unique unit-specific variations to produce a viable imputed value.

We incorporated several elements of the MARTS production procedures for annually selecting regression models for use in the seasonal adjustment process into our regression model building procedure. The industry time series used for model selection and estimation ranged from September 2006 to March 2017. As done for the MARTS seasonal adjustment process, each series was log transformed and included either the six-coefficient trading day regressor (Bell and Hillmer 1983) or the one-coefficient trading day regressor (Gómez and Maravall 1996), as determined by minimum AICC (Akaike information criterion corrected for small sample sizes, Hurvich and Tsai 1989). We tested for the three moving holidays regressors that are considered for MARTS: Easter, Labor Day, and Thanksgiving. The length of the holiday effects corresponded to those used for MARTS in production, an eight day effect preceding Easter (Easter[8]), a nine day effect preceding Labor Day (Labor[9]), and an effect beginning the day after Thanksgiving and lasting until December 24 (Thank[-1]). Moving holiday regressors were selected into the model if the corresponding t-statistic of the model parameter was greater than 1.96 in absolute value (U.S. Census Bureau 2017, p. 71). We did not explicitly identify outliers for the model beyond the automatically identified level shift and additive outliers from X-13A-S.

The ARIMA processes used in X-13A-S typically include both a non-seasonal component and a seasonal component, denoted (p d q) (P D Q) where the lower case letters refer to the non-seasonal orders of the autoregressive component (p), differencing (d), and moving average component (q) and the uppercase letters refer to corresponding seasonal components. Since the parameters estimated from the model will be used on ***unit level data*** to produce a forecast, we only considered autoregressive models, which utilize past values of the time series, discounting moving average models. This allows us to easily apply the estimated model parameters to past values for the unit. This represents a major departure from the models used for MARTS seasonal adjustment procedures. The models used for the production of MARTS seasonally adjusted estimates do not have these restrictions and often have seasonal moving average components, if not non-seasonal moving average components as well.

The ARIMA model selection process was a balancing act between model fit and its usefulness as a generalized imputation procedure. For example, higher order ARIMA models take advantage of correlations at later lags, using data from further in the past compared to more parsimonious models. However, a forecast imputed value for a unit can only be calculated if all of the necessary historic inputs for that unit are available. Figure 1 depicts the amount of historic data available throughout the sample lifecycle for units in the MRTS sample. When a new MRTS sample is introduced, units that were not in the previous MRTS sample (mostly noncertainty units), would be ineligible for imputation with this method until a sufficient period of time passed and all model inputs were available. For example, a (1 1 0) (1 1 0) model includes terms twenty-six months in the past. In practice, this precludes imputation of the majority of the noncertainty units from imputation until they have been in the MRTS sample for more than two years, nearly half of the MRTS sample lifecycle. Therefore, minimizing the time to collect all of the forecast inputs would increase the length of time that the forecast imputation method could be used for all units in the sample. On the other hand, the MARTS data are highly seasonal so including a seasonal component in the model was a priority. Ultimately, we decided to use an order one seasonal autoregressive component for each model. We did consider higher order non-seasonal autoregressive components, but found only modest gains compared to the order one model at the cost of added months of historic data as forecast inputs (that is, greatly restricted availability for imputation). Many economic time series, including retail series, require differencing in order to have stationarity – constant mean, variance, and autocorrelations over time. The assumption that these properties remain constant over time is what makes ARIMA model forecasts viable. Therefore, a first difference was included for all series. The final ARIMA model selected for every series was (1 1 0)(1 0 0). This model uses data from the previous fourteen months to compute the one-step ahead forecast.

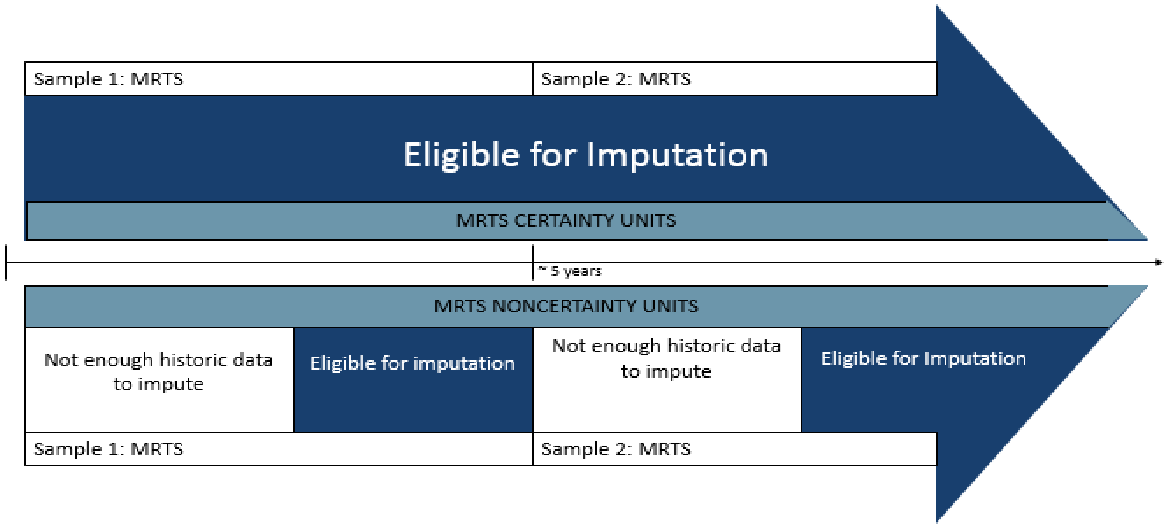

Figure 1. Illustration of available historic data throughout MRTS sample lifecycle.

The model parameters were estimated in SAS using PROC X13 (SAS/ETS 14.1 User's Guide 2015). The input to the procedure was the industry time series beginning September 2006 and ending the month prior to the month being imputed. The log transformation, (1 1 0)(1 0 0) ARIMA model, and the trading day and holiday regressors identified during the model selection stage were all specified in PROC X13. As in the model selection stage, level shifts and additive outliers were automatically identified. However, identified outliers were *only* used for model estimation, and were not used in the calculation of the unit level forecasts. The output saved from PROC X13 included the trading day and holiday components, including their one-step ahead forecasts (corresponding to the current month to be imputed), and the autoregressive model parameter estimates. The product of the trading day and holiday components is the calendar adjustment factor. The past values for the unit are divided by the corresponding monthly calendar adjustment factor when computing the forecast, accounting for the trading day and moving holiday effects appropriate for that month. The calendar adjusted values are then log transformed. The ARIMA model parameter estimates are the coefficients used to combine the log transformed, calendar adjusted, past values for the unit together to estimate the one-step ahead forecast.

The calculation of the imputed value for unit i in time t, $\tilde{y}_{i,t}$, begins with the calendar adjustment and transformation of the unit's historic values. Let $x_{i,t-1} = \ln(y_{i,t-1}/p_{t-1})$ be the calendar adjusted, transformed value for unit i in time t-1, where p_{t-1} is the calendar adjustment factor in month t-1 and let other months be similarly defined. The calendar adjusted, transformed forecast for unit i in time t, $\tilde{x}_{i,t}$ is given by (2), where φ_1 is the parameter estimate for the lag 1 autoregressive component and φ_{12} is the parameter estimate for the lag 12 autoregressive component. See the appendix for a derivation of this expression. The imputed value, $\tilde{y}_{i,t}$, is obtained in (3) by simply exponentiating the transformed forecast and multiplying by the current month calendar adjustment factor.

$$\tilde{x}_{i,t} = (1+\varphi_1)x_{i,t-1} - \varphi_1 x_{i,t-2} + \varphi_{12}x_{i,t-12} - (\varphi_1\varphi_{12} + \varphi_{12})x_{i,t-13} + \varphi_1\varphi_{12}x_{i,t-14} \quad (2)$$

$$\tilde{y}_{i,t} = p_t e^{\tilde{x}_{i,t}} \quad (3)$$

As shown in the appendix, the autoregressive model is applied to the first *difference* of the time series. Thus, the forecast imputation procedure relies on a strong association between a unit's *past* first differences and the current first difference; that is a strong association between past month-to-month changes and the current month-to-month change being imputed. Unusually large or small month-to-month changes that are unlikely to be repeated could result in unrealistic forecast imputed values. These unusual changes may be a result of regular market activities, an especially good or bad month, or a change in company structure, like an acquisition or a divestiture. Whatever the cause, such large swings are unlikely to be a part of a recurring seasonal pattern that will be repeated in the future. To guard against including unusual changes in the regARIMA forecasts, we discarded units identified as outliers via the resistant fences method (Hoaglin, Iglewicz, and Tukey 1986) applied to three sets of month-to-month change distributions: $\frac{y_{t-1}}{y_{t-2}}, \frac{y_{t-12}}{y_{t-13}}$, and $\frac{y_{t-13}}{y_{t-14}}$ within each industry. Each set of distributions was obtained from units with reported values in both the numerator and denominator statistical period, and ratio values that fell outside of the corresponding fences tolerances (three interquartile ranges above the third quartile and below the first quartile) were identified as outliers.

After obtaining forecast imputations, we also applied resistant fences to the ratio of the current month to the prior month$\left(\frac{y_t}{y_{t-1}}\right)$ within each industry, again using only units that had reported values in both statistical periods to create the distributions. We compared the ratio of a unit's forecast imputed value for the current month to the unit's prior month value to this set of tolerances to identify outlying ratios. If the ratio fell outside of the resistant fences bounds, then that unit was not imputed. As discussed in Section 2, the link relative estimator accounts for nonresponse, so it is not necessary to impute for every nonresponding unit.

4. Study Design

To conduct an impartial evaluation, we created sets of MRTS current month respondents for each of the study months from January to December 2016. However, the historic data for these units was a combination of reported and imputed data. We further subset the data to include only those units that were also in the current MARTS sample. Thus, for every unit in the study, we have a "true" value of their monthly sales to compare with the forecast imputed value and can likewise compute a "true" link relative ratio. For industry-level comparisons, we compare the estimated link relative ratio with forecast imputation to the corresponding link relative ratio estimate without imputation.

We conducted an empirical simulation study for each month in 2016 using the observed monthly industry response rates to randomly select nonresponding units. Observed response rates were calculated for both MARTS certainty and noncertainty units for each month. While examining the distribution of sales by MARTS certainty unit respondents and nonrespondents by industry, we observed differences in the distribution of sales, indicating a relationship between response propensity and unit size in many industries. Although all of the units included in the certainty strata exceed some pre-determined size cutoff at the time of sampling, there can still be great disparity of sizes within the certainty strata. Many retail industries are highly concentrated in the upper tail of the distribution, with the top three units sometimes accounting for 30-60% of the total industry sales. Understanding the frequency with which these very large units do not respond to MARTS, and thereby how often they need to be imputed, is crucial to simulating nonresponse in the data in a way that mimics a real production setting. In Figure 2 we use fictitious data to illustrate a pattern we observed in the distribution of nonrespondents and respondents in the certainty stratum in some industries. In this example, the largest units in the industry are nonrespondents and the distribution of sales for respondents differ from that of the nonrespondents. While not every industry demonstrated such a stark contrast between certainty unit respondents and nonrespondents, this pattern did emerge frequently and prompted us to consider ways to incorporate this pattern into our nonresponse simulation.

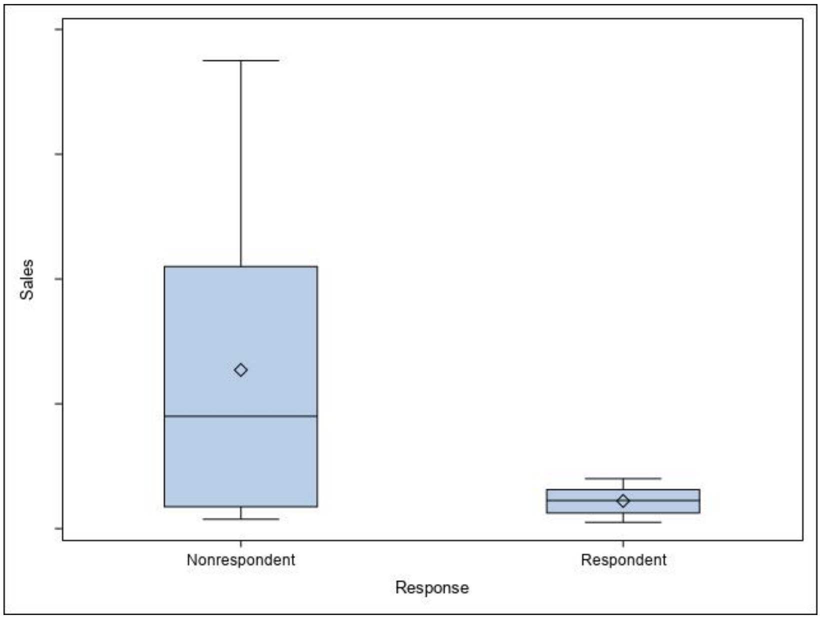

Figure 2. Example of the observed distribution of sales for respondents and nonrespondents in the certainty stratum using fictitious data

We decided to split the certainty strata in each industry into two groups for each month based on that month's sales using the cumulative square root of the frequency method (Kish 1965 p. 105). Observed response rates were calculated from the full MARTS sample for the noncertainty stratum and for the two certainty subgroups within industry. These observed rates were used as the response propensities in the corresponding groups within industry for the nonresponse simulation. However, to avoid automatically including or excluding units from our simulation, observed response rates of 0.0 were replaced with a response propensity of 0.05 and observed response rates of 1.0 were replaced with a response propensity of 0.95. We induced nonresponse 500 times for each month using a missing at random response mechanism within size category and the response propensities described above. Each set of monthly simulations is independent of the other months, imputed values from one month are not carried over to the following month.

For a given month, the benchmark total for the prior period is constant throughout the simulation: only the estimate of the link relative ratio changes from simulation to simulation. Therefore, our evaluation focuses on the estimates of the link relative ratios.

Let S_t be the set of tabulation units in a given industry in the respondent test deck at time t and let $K_{t,r}$ and $N_{t,r}$ be the sets of respondents and nonrespondents of size $k_{t,r}$ and $n_{t,r}$, respectively, in simulation r. Then the estimated link relative ratio for simulation r, using the forecast imputation method, $\hat{L}_{r,f}$, is defined as:

$$\hat{L}_{r,f} = \frac{\sum_{K_{t,r}} w_i y_{i,t} + \sum_{N_{t,r}} I_{i,r} w_i \tilde{y}_{i,t,r}}{\sum_{K_{t,r}} w_i y_{i,t-1} + \sum_{N_{t,r}} I_{i,r} w_i y_{i,t-1}}$$

Where w_i is the MARTS weight, $y_{i,t}$ is the reported sales for *i* in time *t*, $\tilde{y}_{i,t}$ is the imputed sales value for unit *i* in time *t*, and $I_{i,r} = 1$ if the nonresponding unit *i* was imputed in simulation *r* for time *t*, and 0 otherwise.

Likewise, the estimated link relative ratio for simulation *r* without imputation, $\hat{L}_{r,n}$, is:

$$\hat{L}_{r,n} = \frac{\sum_{K_{t,r}} w_i y_{i,t}}{\sum_{K_{t,r}} w_i y_{i,t-1}}$$

We computed the mean absolute error of the link relative ratio and the mean squared prediction error within each industry, for each month. The mean absolute error provides an estimate of the average deviation of the estimated link relative ratio from the true link relative ratio. Large errors between the estimated link relative ratio and the true link relative ratio are likely to result in large revisions between MARTS and MRTS estimates. The mean absolute error of the link relative ratio with (MAE_f) and without (MAE_n) imputation are defined as:

$$MAE_f = \frac{1}{500}\sum_{r=1}^{500}\left|\hat{L}_{r,f} - L_{true}\right| \qquad MAE_n = \frac{1}{500}\sum_{r=1}^{500}\left|\hat{L}_{r,n} - L_{true}\right|$$

The mean absolute prediction error (MSPE) measures accuracy of unit level predictions. For units for which a forecast imputed value was not obtained, we use the ratio impute implied by the estimated link relative ratio with imputation. With this measure, we can compare how well the forecast imputation procedure predicts the individual unit sales compared to the ratio imputation implied by the link relative estimator.

We computed the MSPE separately by **MRTS** certainty units and noncertainty units. Recall from Section 3 that the historic data needed to calculate unit level imputations should be available for most MRTS certainty units throughout the entire MRTS sample lifecycle. However, the universe of noncertainty (small) MRTS units is large, as are the sampling intervals. Consequently, it is unlikely that a MRTS noncertainty business will be selected in adjacent samples, and the majority of noncertainty MRTS units will not have historical data until 14 months after the sample is introduced. Our study period, the year 2016, occurred late in the MRTS sample lifecycle. Thus, the necessary unit level historic data for the MRTS noncertainty units should be available. By separating the MSPE by certainty status when comparing to the imputed values implied by the link relative estimator, we can see if the performance of forecast imputation is consistent across unit size categories.

$$MSPE_f = \frac{1}{500}\sum_{r=1}^{500}\frac{\sum_{N_{t,r}}\left[I_{i,r}\left(\tilde{y}_{i,t,r} - y_{i,t}\right)^2 + (1 - I_{i,r})\left(\hat{L}_{r,f}y_{i,t-1} - y_{i,t}\right)^2\right]}{n_{t,r}}$$

$$MSPE_n = \frac{1}{500}\sum_{r=1}^{500}\frac{\sum_{N_{t,r}}\left[\left(\hat{L}_{r,n}y_{i,t-1} - y_{i,t}\right)^2\right]}{n_{t,r}}$$

5. Results

For each month, we conducted Wilcoxon signed rank tests (α=0.05) to determine whether there was a difference in the link relative ratio estimates with and without the forecast imputation. While the vast majority of the link relative ratio estimates with forecast imputation were statistically different from their no imputation counterparts, 38 out of the 360 monthly estimates across the 30 MARTS industries were not statistically different. Where there were differences in the link relative estimates, we compared the MAE and computed the frequency at which the MAE was smaller with and without forecast imputation. We also computed the ratios of the MAE without forecast imputation to the corresponding measure with imputation. Ratios greater than one indicate the estimates with forecast imputation outperformed the corresponding estimates without imputation. Table 1 summarizes these results. The medians reported in Table 1 are computed over all months in the study, whether there was a significant difference in the link relative ratios or not.

NAICS	*No diff in ratio*	*Smaller with Imputation*	*Smaller without Imputation*	*Median* $\left(\frac{MAE_n}{MAE_f}\right)$
441100	0	3	9	0.58
441200	1	6	5	1.59
441300	3	5	4	1.35
442000	0	6	6	1.18
443120	1	6	5	1.21
443X00	2	8	2	1.84
444100	0	7	5	1.26
444200	1	7	4	1.62
445100	0	1	11	0.74
445X00	0	9	3	1.51
446000	1	9	2	1.75
447000	0	3	9	0.30
448110	3	9	0	2.79
448120	0	9	3	1.46
4481L0	1	10	1	1.88
448200	1	11	0	2.02
448310	3	5	4	1.73
4511X0	1	9	2	1.92
451XX0	5	7	0	3.39
452110	2	6	4	1.96
452120	5	4	3	2.12
452910	0	11	1	7.34
452990	2	5	5	1.17
453210	0	12	0	2.78
453930	0	9	3	1.67
453XX0	0	9	3	1.44
454100	3	9	0	2.53
454310	0	3	9	0.63
454X00	1	9	2	1.54
722000	2	5	5	1.44

Table 1. Summary of mean absolute error by industry. Source: MARTS and MRTS January 2016-December 2016

However, the proposed imputation method does appear to be a viable alternative for reducing error in the estimate of the link relative ratio for many of the MARTS industries. For the remaining 26 industries, the median MAE ratio is greater than one, signifying smaller MAE with forecast imputation in a majority of the months in the study. Figure 4 shows the relative MAE for the four industries with the largest median MAEs, indicating a preference for imputation. The relative MAEs without imputation for these industries are larger than those in Figure 3, offering opportunities for improvement. For these industries, the forecast imputation procedure produces link relative ratio estimates with smaller relative MAE than the corresponding estimates without imputation for most of the months in the study, demonstrating consistent performance of forecast imputation throughout the study year.

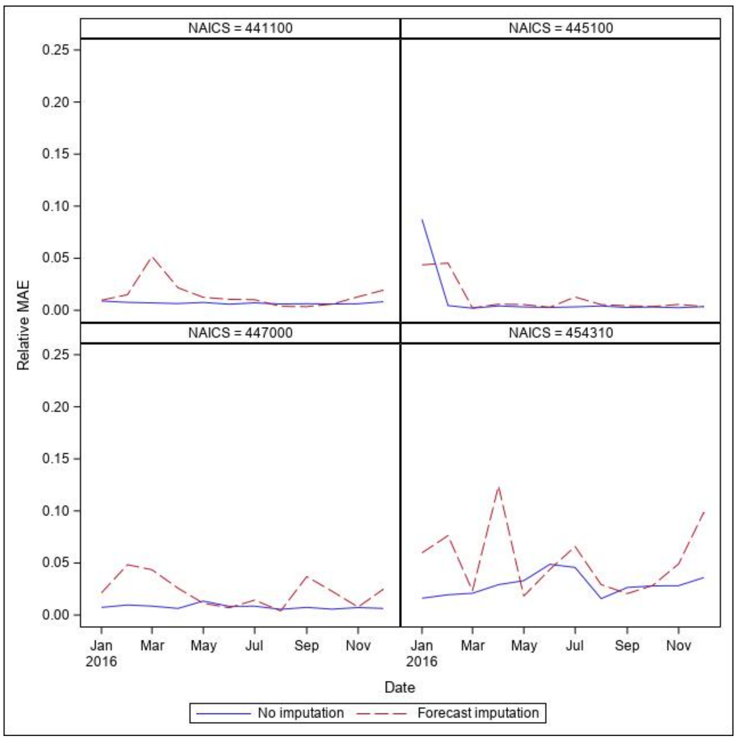

Figure 3. Relative mean absolute error of the link relative ratio estimates for selected industries Source: MARTS and MRTS January 2016-December 2016

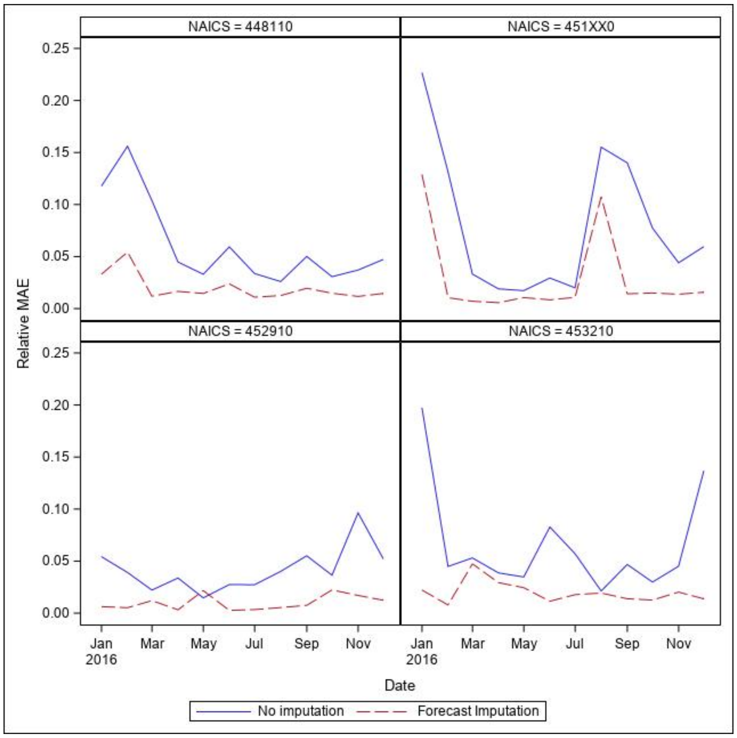

Figure 4. Relative mean absolute error of the link relative ratio estimates for selected industries. Source: MARTS and MRTS January 2016-December 2016

The results for the mean square prediction error are presented in Table 2. Note that three industries have only certainty units. As in Table 1, median MSPE ratios greater than one indicate the MSPE was smaller with imputation. For many industries, the MSPE was reduced by the addition of forecast imputation. The reductions in MSPE were often larger and more frequent for the MRTS certainty units compared with the noncertainty units in the same industry. The MSPE results closely align with those observed for the MAE, especially for certainty units. The four industries that had smaller mean absolute errors without forecast imputation also demonstrate smaller MSPEs without imputation. Likewise, the industries that showed a reduction in mean absolute error with forecast imputation also produce smaller MSPEs with forecast imputation, particularly for certainty units.

NAICS	*Certainty* *Smaller with Imp.*	*Certainty* *Smaller without Imp.*	*Certainty* *Med* $\left(\frac{MSPE_n}{MSPE_f}\right)$	*Noncert.* *Smaller with Imp.*	*Noncert.* *Smaller without Imp.*	*Noncert.* *Med* $\left(\frac{MSPE_n}{MSPE_f}\right)$
441100	5	7	0.93	1	11	0.67
441200	10	2	1.27	7	5	1.48
441300	7	5	1.40	6	6	0.99
442000	8	4	2.52	5	7	0.98
443120	5	7	0.75	8	4	1.30
443X00	9	3	3.82	6	6	0.95
444100	6	6	1.06	2	10	0.63
444200	11	1	5.79	7	5	1.81
445100	2	10	0.29	5	7	0.77
445X00	8	4	1.83	10	2	1.07
446000	11	1	5.50	6	6	1.29
447000	1	11	0.28	3	9	0.82
448110	11	1	2.88	11	1	7.11
448120	9	3	1.42	10	2	2.22
4481L0	8	4	1.06	8	4	1.80
448200	12	0	4.46	9	3	2.88
448310	10	2	3.04	4	8	0.74
4511X0	9	3	3.31	11	1	2.41
451XX0	11	1	21.73	11	1	1.86
452110	9	3	2.97	0	0	n/a
452120	8	4	4.65	0	0	n/a
452910	11	1	27.99	0	0	n/a
452990	7	5	1.20	7	5	1.43
453210	10	2	6.02	8	4	1.36
453930	10	2	9.64	7	5	1.06
453XX0	7	5	1.58	6	6	1.01
454100	9	3	1.73	5	7	0.92
454310	3	9	0.54	2	10	0.60
454X00	6	6	0.96	8	4	1.19
722000	8	4	1.53	3	9	0.95

Table 2. Summary of mean squared prediction error of MRTS certainty and noncertainty units by industry. Source: MARTS and MRTS January 2016-December 2016

With only modest improvements in MSPE for the noncertainty units in many industries, we wondered if imputing the noncertainty units actually improved the estimate of the link relative ratio or whether most of the error reduction was a result of imputing the MRTS certainty units. Therefore, we recalculated estimates of the link relative ratio, this time only imputing for the MRTS certainty units. The mean absolute error results imputing only for MRTS certainty units are summarized in Table 3. Many industries showed an increase in the frequency in which the estimates with imputation had smaller errors than the estimates without imputation when we impute only the MRTS certainty units. This suggests that we may only need to impute the MRTS certainty units to see improvements in the mean absolute error of the link relative estimates compared to the no imputation scenario. This approach has the advantage of keeping the set of units that are imputed constant throughout the sample lifecycle since many of the noncertainty units cannot be imputed early in a new sample due to the unavailability of unit level historic data.

NAICS	*No diff in ratio*	*Smaller with Imputation*	*Smaller without Imputation*	*Median* $\left(\frac{MAE_n}{MAE_f}\right)$
441100	0	6	6	1.00
441200	0	12	0	1.28
441300	1	8	3	1.31
442000	1	6	5	1.68
443120	0	6	6	1.02
443X00	1	9	2	1.44
444100	0	7	5	1.22
444200	0	12	0	1.60
445100	0	2	10	0.66
445X00	0	8	2	1.13
446000	2	8	2	1.76
447000	0	3	9	0.39
448110	0	11	1	1.79
448120	0	7	5	1.05
4481L0	0	10	2	1.50
448200	1	10	1	2.10
448310	0	9	3	1.29
4511X0	1	10	1	1.97
451XX0	2	10	0	2.47
452110	2	6	4	1.96
452120	5	4	3	2.12
452910	0	11	1	7.34
452990	3	2	7	1.28
453210	0	11	1	2.95
453930	0	11	1	1.82
453XX0	1	11	0	1.53
454100	3	9	0	2.69
454310	0	3	9	0.74
454X00	0	11	1	1.67
722000	2	8	2	1.23

Table 3. Summary of mean absolute error results imputing only for nonrespondents who are also MRTS certainty units. Source: MARTS and MRTS January 2016-December 2016

There were some industries, however, where imputing both the MRTS certainty and noncertainty units resulted in a greater reduction in error compared with imputing only the MRTS certainty units. Three clothing related industries, NAICS 448110 (Men's Clothing Stores), 448120 (Women's Clothing Stores), and 448200 (Shoe Stores), demonstrated smaller MAE and MSPE for the noncertainty units when the noncertainty units were imputed with the forecast imputation method. We include results for Men's Clothing Stores but observed similar patterns in the other two industries. Figure 5 compares the ratio of the MSPE with imputation to the MSPE without imputation separately for MRTS certainty and noncertainty units under the two imputation scenarios (imputing all units or imputing only the MRTS certainty units). As with Figures 3 and 4, values greater than one indicate a reduction in error with forecast imputation. For Men's Clothing Stores, the forecast imputation procedure yields lower MSPE for the MRTS noncertainty units compared to the link relative ratio imputation. A similar pattern is observed at the total level with the relative MAE shown in Figure 6. For these industries, the forecast imputation procedure is effective for both large and small units.

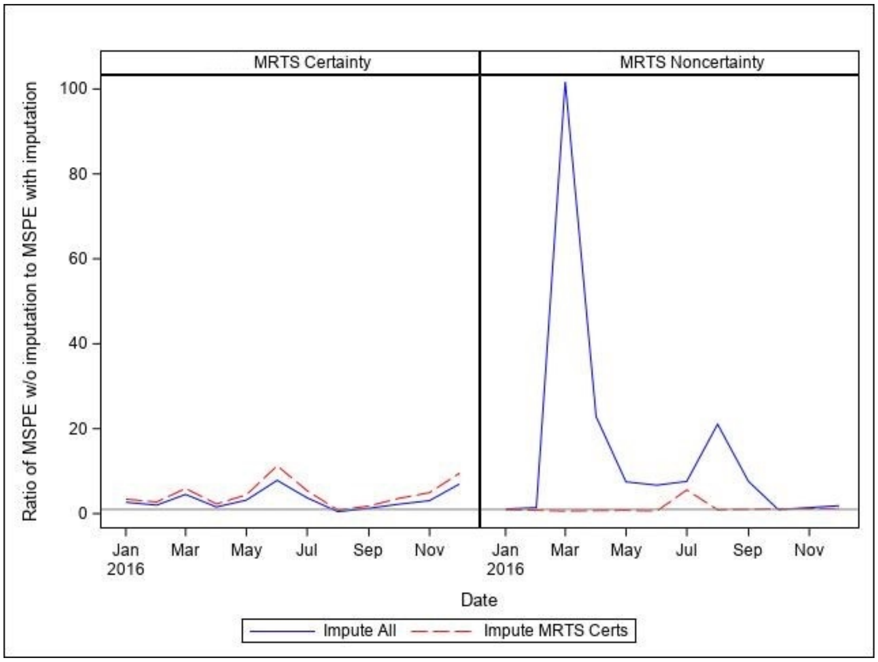

Figure 5. Ratio of mean squared prediction error without imputation to mean squared prediction error with imputation under two imputation scenarios: imputing all eligible nonrespondents, and imputing nonrespondents who are also MRTS certainty units for NAICS 448110, men's clothing stores. Source: MARTS and MRTS January 2016-December 2016

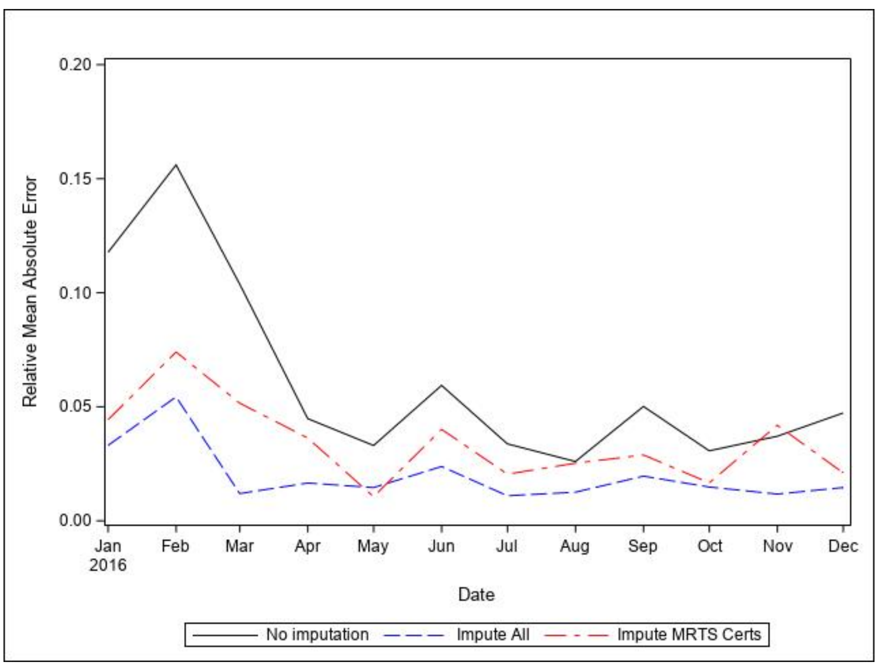

Figure 6. Relative mean absolute error under three scenarios: no imputation, imputing all eligible nonrespondents, and imputing nonrespondents who are also MRTS certainty units for NAICS 448110, men's clothing stores. Source: MARTS and MRTS January 2016-December 2016

Finally, we examined some industry characteristics to see if there was anything common to those industries which performed better with imputation or among those that performed better without. We observed that industries with a small number of very large units often performed better with imputation. Among industries with many similar sized units, none of which were particularly large, no imputation was often preferred. In an effort to quantify this concept, for each industry, each month, we estimated the Herfindahl-Hirschman (HH) Index, a measure of market concentration (Bondarenko, 2019), using the complete MRTS sample. Higher values of the HH Index indicate a higher market concentration.
We compare the median MAE ratio imputing for both certainty and noncertainty units to the median HH Index in Figure 7. Although there is only a weak positive relationship between the median ratio values and the HH Index, all industries with a median HH Index above 275 had a median MAE ratio greater than one, indicating a preference for forecast imputation. This finding makes sense and validates the current analyst imputation practices. Analysts impute few, if any, units in the industries where we estimated a very low market concentration because in these industries individual imputes have very little impact on the estimated link relative ratio. These are the same industries in which our simulation results showed a preference for no imputation. Conversely, in the industries with high estimated market concentration, we find smaller MAE with imputation. For these highly concentrated industries, the largest units often drive the estimated link relative ratio estimates. When these very large units are missing, including forecast imputed values for these large units reduces the error in the link relative ratio estimates.

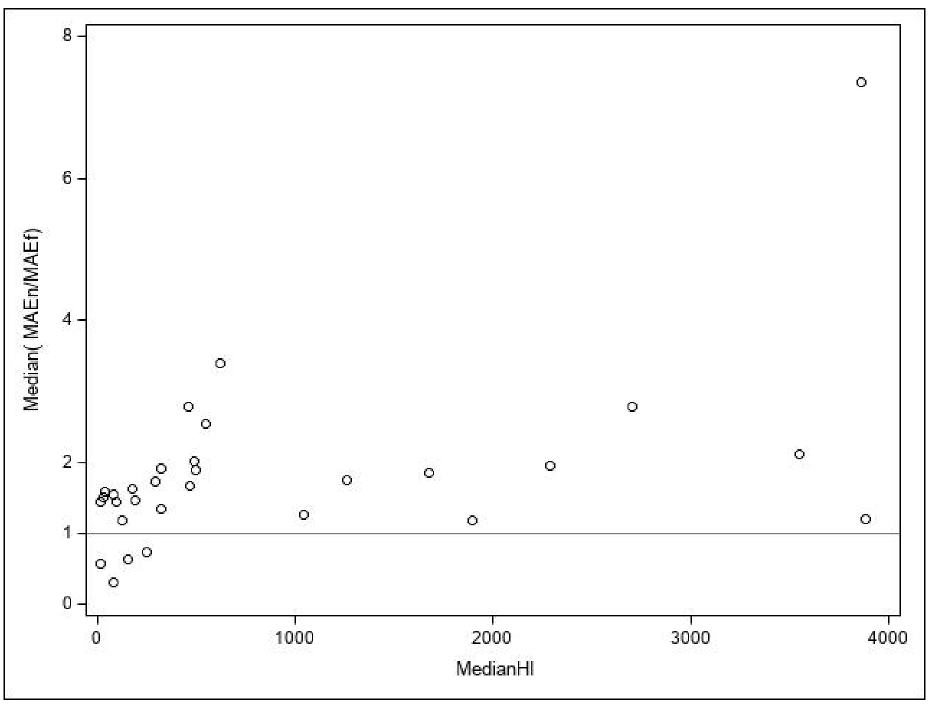

Figure 7. Median HH Index vs Median MAE ratio for each industry in MARTS.
Source: MARTS and MRTS January 2016-December 2016

6. Conclusions

The high degree of seasonality in retail data and the lack of other current period predictors make time series forecasting methods an appealing choice for imputation in the Advance Monthly Retail Trade Survey. By incorporating a simple ARIMA model and known calendar effects, such as trading day effects and moving holiday effects, the forecast imputed value incorporates the seasonality and calendar variation typically experienced by units within the industry. The forecast imputation method described here utilizes the unit's unique historical data to compute a forecast imputed value for the current month. Our simulation study provided evidence that the addition of forecast imputation can improve the accuracy of the link relative estimates for most of the industries in MARTS. This methodology performed particularly well in industries with a high market concentration as measured by the HH Index. Industries that demonstrated better performance without this additional imputation, by contrast, had relatively low market concentration. In industries with low market concentration, the link relative ratio estimate using only respondent data already estimates the true link relative ratio well and the forecast imputation method presented here does not reduce the mean absolute error of the estimate. Furthermore, the forecast imputation procedure presented here is not equipped to handle sudden, unexpected price changes, like those that frequently occur in price dependent industries such as gasoline stations. In such industries, the link relative estimator without imputation may be better suited.

However, given that the link relative estimator will be used to estimate totals, not all nonresponding units may need to be imputed. The simulation study provides convincing evidence of precision improvements in the when restricting the forecast imputation procedure to the MRTS certainty units for many MARTS industries. Imputing only the MRTS certainty units opens up the possibility of using models that extend further into the past since these units often have several years of historic data available extending into the previous MRTS sample; we consider this a topic for future research. That said, it appears that precision improvements in several clothing industries, including men's and women's clothing stores and shoe stores require forecast imputations for both certainty and noncertainty MRTS units.

For future research on the forecast imputation method, we would like to delve deeper into quality of the unit level imputations. First, we want to explore how to improve the quality control method used to exclude unlikely imputations. Another refinement to the proposed imputation method could include a way to "update" the one-step ahead forecast with information from the current month reporting units. Instead of only using the current month reported values to develop the fences for the resistant fences quality control measure, we could perhaps use these data to improve the imputed values by incorporating late-breaking industry trends. We are also planning to explore methods for estimating the variance of the link relative estimates incorporating the uncertainty due to the forecast imputation procedure.

As mentioned in Section 4, MARTS response patterns may differ by unit size. Another area for future research is investigating the performance of the forecast imputation procedure by modeling the response propensity is a function of the unit's size. Such research may offer a more realistic simulation of the response behavior seen in production and allow us to better asses the merits of forecast imputation when used with the link relative estimator.

Acknowledgements

The authors would like to thank Laura Bechtel, Katherine Jenny Thompson, Demetra Lytras, Xijian Liu, Scott Scheleur, and Bonnie Kegan for their thoughtful comments on previous versions of this paper.

References

Bell, W. R. and Hillmer, S. C. (1983), "Modeling Time Series with Calendar Variation" *Journal of the American Statistical Association,* 78, 526-534.

Bondarenko, P. (2019, May 30). *Herfindahl-Hirschman index*. Retrieved from Encyclopedia Britannica: https://www.britannica.com/topic/Herfindahl-Hirschman-index

Box, G. E. P. and Jenkins G. M. (1970), Time Series Analysis, Forecasting and Control, San Francisco, CA: Holden-Day.

Czaplicki, N., González,Y, and Bechtel, L. (2018). "Finding an estimator that minimizes revisions in a monthly indicator survey." *Proceedings of the FCSM Research Conference.*

Findley, D. F. and Soukup, R. J. (2000), "Modeling and Model Selection for Moving Holidays," *Proceeding of the American Statistical Association, Business and Economic Statistics Section*, 102–107, https://www.census.gov/ts/papers/asa00_eas.pdf

Gómez, V. and Maravall, A. (1996), "Programs TRAMO and SEATS, Instructions for the User (Beta Version: September 1996)," Banco de España – Servicio de Estudios, Documento de Trabajo no. 9628 (English version).

Hoaglin, D.C., Iglewicz, B. and Tukey, J.W. (1986) "Performance of Some Resistant Rules for Outlier Labeling," Journal of the American Statistical Association, 81, 396, 991-999

Horvitz, D.G. and Thompson, D.J. (1952). "A generalization of sampling without replacement from a finite universe," *Journal of the American Statistical Association,* 47, 663-685.

Hurvich, C. M. and C. Tsai (1989). Regression and time series model selection in small samples. Biometrika 76, 297-307.

Kish, L. (1965), Survey Sampling, New York, NY: John Wiley & Sons, Inc.

Madow, L.H. & Madow, W.G. (1978). "On Link Relative Estimators," *ASA Proceedings of the Section on Survey Research Methods*, 1978, 534-539.

"SAS/ETS(R) 14.1 User's Guide". SAS/ETS(R) 14.1 User's Guide. N.p., n.d. Web. 13 Jul. 2015.

Thompson, K., Bechtel, L., and Czaplicki, N. (2018). "Evaluating Hot Deck with Propensity Score Matching For the Advance Monthly Retail Trade Survey". *Proceedings of the FCSM Research Conference.*

U. S. Census Bureau (2017), *X-13ARIMA-SEATS Reference Manual*, Version 1.1, Time Series Research Staff, Center for Statistical Research and Methodology, U.S. Census Bureau, U. S. Department of Commerce. www.census.gov/ts/x13as/docX13AS.pdf

Appendix: Derivation of the one-step ahead forecast for a (1 1 0)(1 0 0) ARIMA model

Let $y_1, y_2 \dots y_T$ be a time series.
Let $x_t = \ln(y_t/p_t)$ be the calendar adjusted, log transformed value of the series at time t, where p_t is the calendar adjustment factor described in section 3.

Define the first difference of x_t as:

$$d_t = x_t - x_{t-1}$$

Then the (1 1 0)(1 0 0) ARIMA model for x_t (a (1 0)(1 0) ARMA model for d_t) is:

$$(1 - B\varphi_1)(1 - B^{12}\varphi_{12})d_t = e_t \quad e_t \sim N(0, \sigma^2)$$

Note: B is the backshift operator (i.e. $Bd_t = d_{t-1},\ B^{12}d_t = d_{t-12}$)

$$d_t = \varphi_1 d_{t-1} + \varphi_{12} d_{t-12} - \varphi_1\varphi_{12} d_{t-13} + e_t$$

Substituting $x_t - x_{t-1}$ in for d_t

$$x_t - x_{t-1} = \varphi_1(x_{t-1} - x_{t-2}) + \varphi_{12}(x_{t-12} - x_{t-13}) - \varphi_1\varphi_{12}(x_{t-13} - x_{t-14}) + e_t$$

$$\tilde{x}_t = (1 + \varphi_1)x_{t-1} - \varphi_1 x_{t-2} + \varphi_{12} x_{t-12} - (\varphi_1\varphi_{12} + \varphi_{12})x_{t-13} + \varphi_1\varphi_{12} x_{t-14} + e_t$$