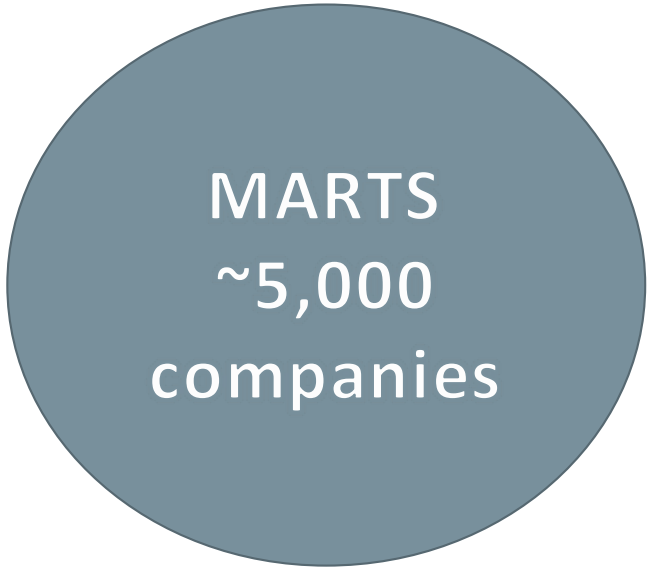
Imputing seasonal data in an advanced indicator with forecasts from X-13ARIMA-SEATS

Nicole Czaplicki and Yarissa González
U.S. Census Bureau

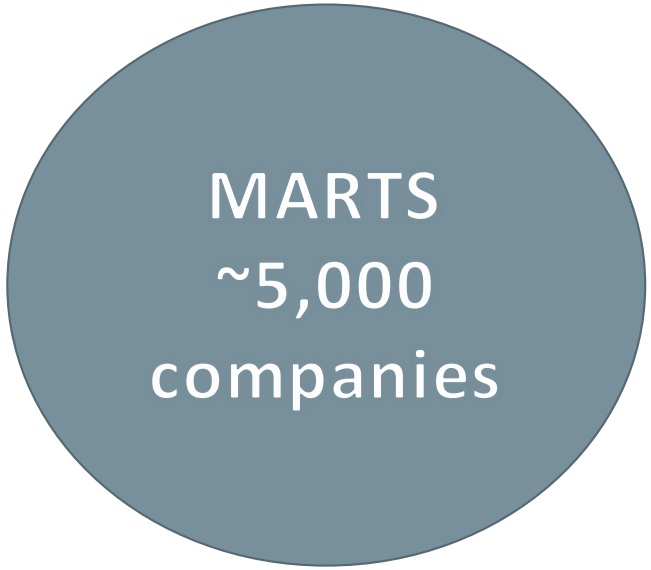
United States™ Census Bureau
U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Monthly Retail Trade



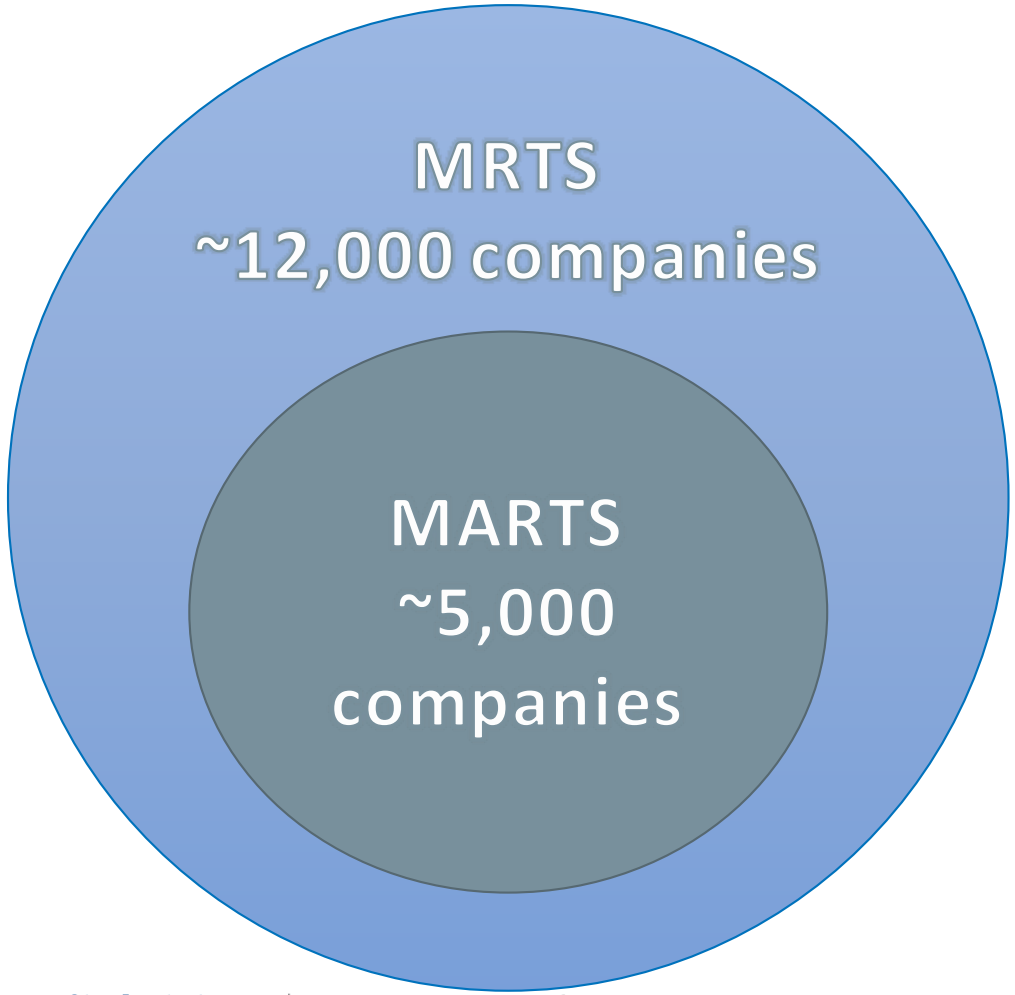
- MARTS - Advance Monthly Retail Trade Survey
  - Economic indicator
  - Input into GDP estimates
  - Collects sales
  - Published approximately 9 business days after the reference month
  - Approximately 7 days to respond
    - Low response rates < 50%

Monthly Retail Trade



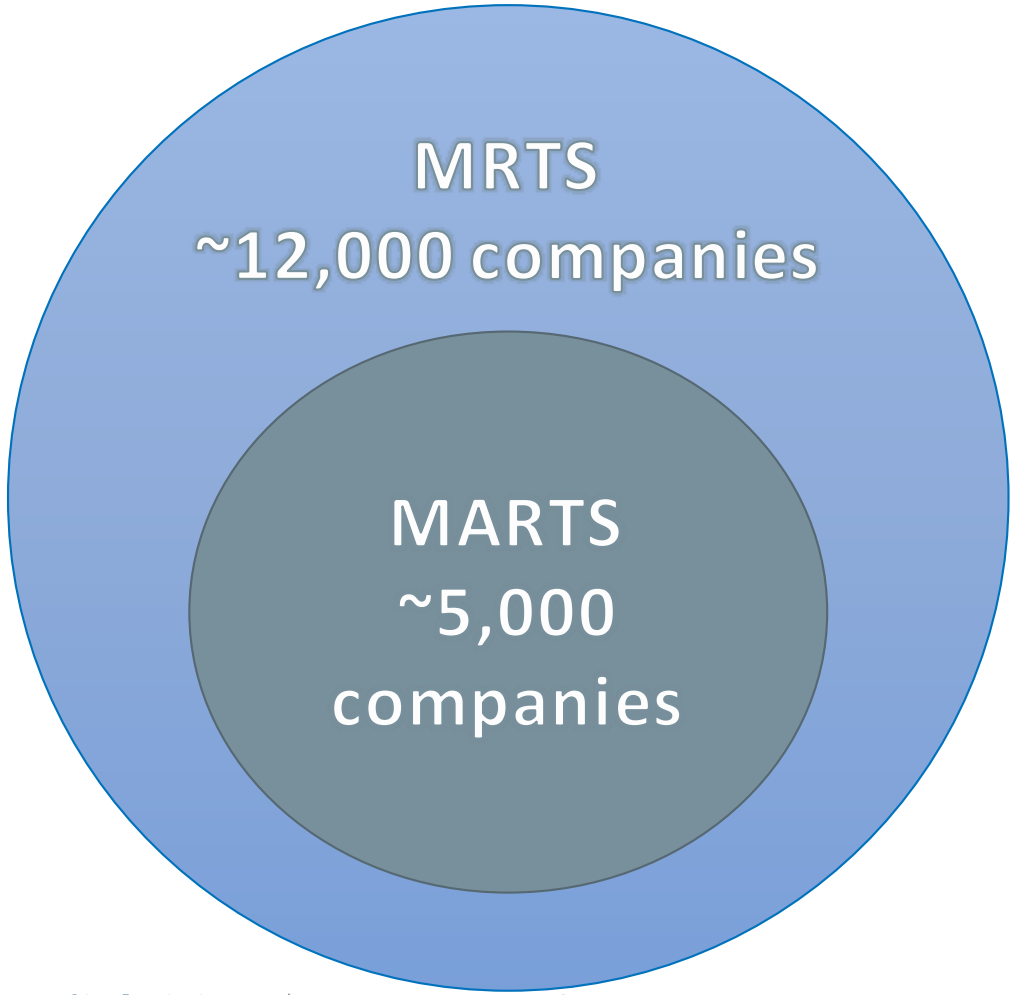
- MARTS - Advance Monthly Retail Trade Survey
  - Selected (large) nonresponding companies imputed by analysts
  - Link relative estimator

Monthly Retail Trade



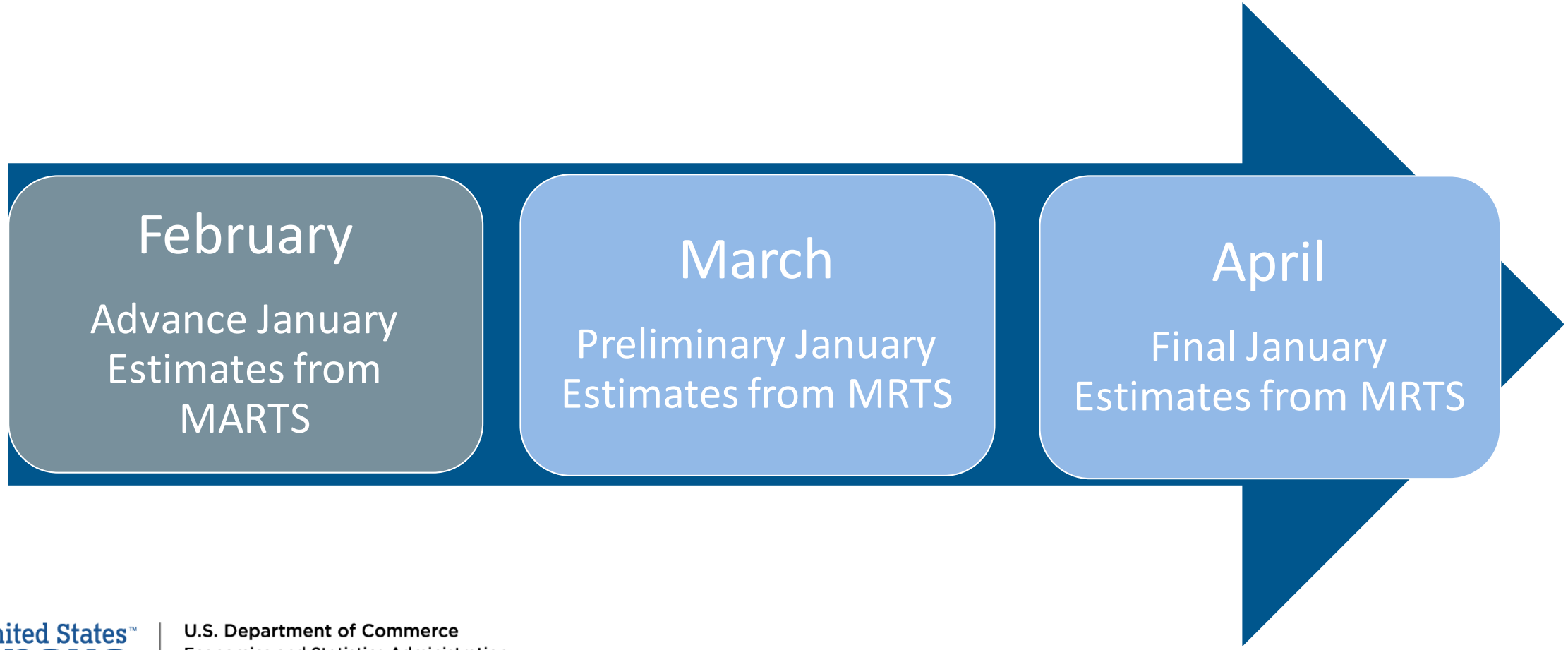
- MRTS – Monthly Retail Trade Survey
  - Input into GDP estimates
  - Collects sales and inventories
  - Published 6 weeks after reference month
  - Approximately 5 weeks to respond

Monthly Retail Trade



- MRTS – Monthly Retail Trade Survey
  - Analyst imputes retained, ratio imputation for remaining nonrespondents
  - Horvitz-Thompson estimator

Retail Trade Indicator Release Example for January



Link Relative Estimator

$$\hat{Y}_{LR,t} = B_{t-1}\left[\frac{\sum_{i\in C} w_i y_{t,i}}{\sum_{i\in C} w_i y_{t-1,i}}\right]$$

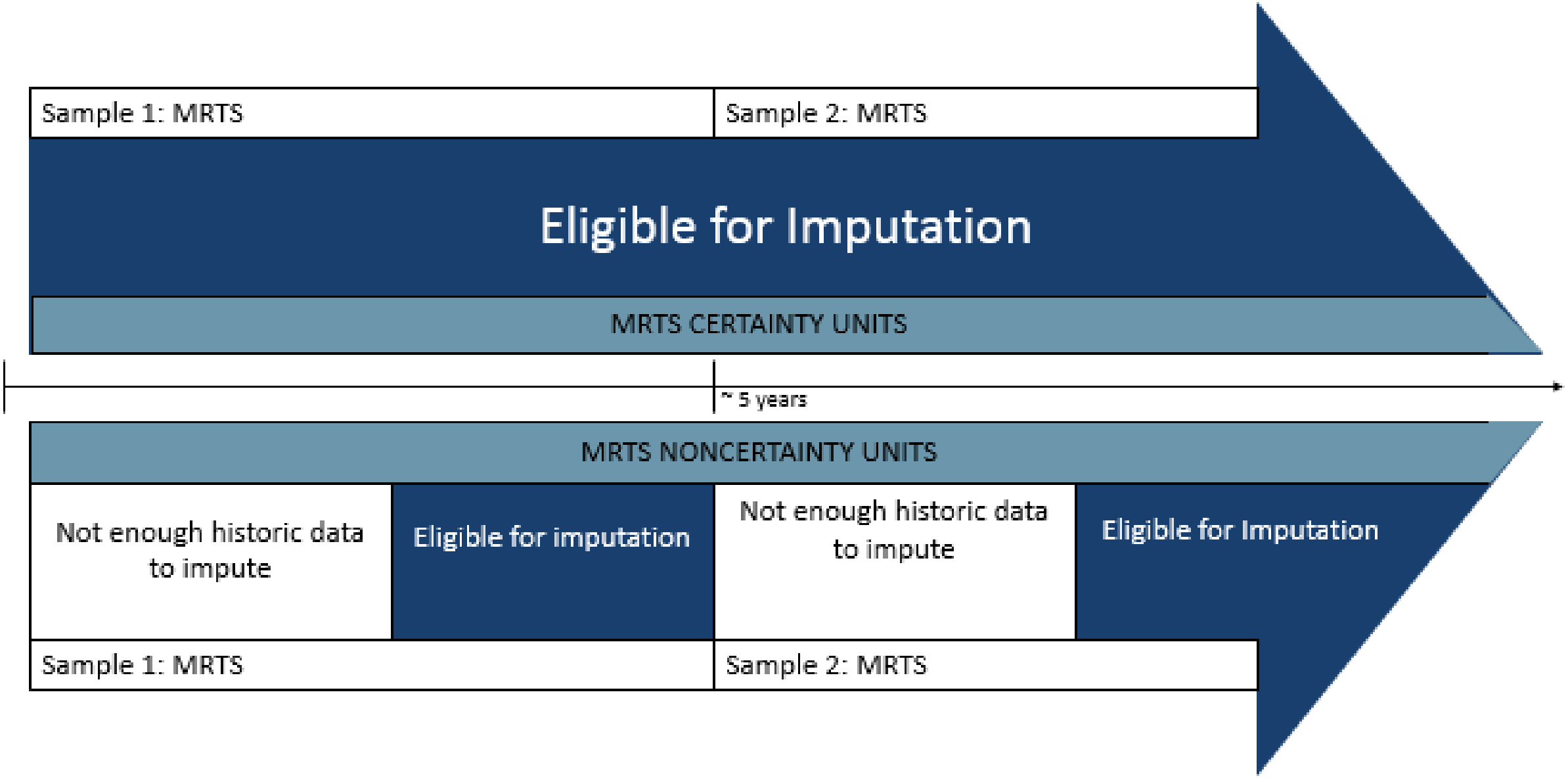
- $\left[\frac{\sum_{i\in C} w_i y_{t,i}}{\sum_{i\in C} w_i y_{t-1,i}}\right]$ = the link relative ratio
- $\hat{Y}_{LR,t}$ = link relative estimate of the total for period t
- B_{t-1} = benchmark value associated with period $t-1$
  - MRTS Preliminary total
- C = set of units with usable data for both periods t and $t-1$
- w_i = sample weight for unit i
- Introduced by Madow and Madow (1978)

Developing an imputation method for MARTS

- What are the characteristics of the microdata?
  - Seasonal
  - Trading day effects
  - Moving holiday effects
  - Skewed

Developing an imputation method for MARTS

- What data are available?
  - Time series of estimated industry totals
  - Unit level monthly historic data from MRTS


Sample 1: MRTS
Sample 2: MRTS
Eligible for Imputation
MRTS CERTAINTY UNITS
~ 5 years
MRTS NONCERTAINTY UNITS
Not enough historic data to impute
Eligible for imputation
Not enough historic data to impute
Eligible for Imputation
Sample 1: MRTS
Sample 2: MRTS

RegARIMA Forecast Approach

- PROC X13
- Fit regARIMA models for each MARTS industry time series
- Extract ARIMA model parameters and calendar adjustment factors
- Compute one-step-ahead unit level forecasts = imputed values
- Quality control on model input and output

RegARIMA Models

- Regression + ARIMA

$$\ln(Y_t) = \widehat{\beta}' \boldsymbol{X}_t + Z_t$$

where Y_t = time series -> MARTS industry total sales

$\boldsymbol{X}_t$ = regression matrix [trading day, automatic outliers

moving holiday effects (Easter, Labor Day, Thanksgiving)]

$\widehat{\beta}'$ = parameter estimates

Z_t = ARIMA process

[(1 1 0)(1 0 0), uses data at t-1, t-2, t-12, t-13, t-14]

Imputing with a regARIMA model

$z_{i,t-1} = \ln\left(\frac{y_{i,t-1}}{a_{t-1}}\right)$ $y_{i,t-1}$ = sales for unit i at $t-1$

a_{t-1} = calendar adjustment factor at $t-1$

(trading day and moving holiday effects)

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

$$\tilde{y}_{i,t} = a_t e^{\tilde{z}_{i,t}}$$

Study design

- January –December 2016
- Calculate observed MARTS response rates for each month
  - 3 subgroups within industry – large certainty, small certainty, noncertainty
- Monthly datasets of MARTS units with reported data
- Induce nonresponse 500x – missing at random using observed response rates as response propensities within industry subgroups
- Impute and calculate link relative ratio with imputed values
  - Impute all units
  - Impute MRTS certainty units only

Evaluation

- Mean square prediction error (MSPE)
  - Evaluation at the unit level
  - Compare regARIMA forecast performance to implied ratio impute from link relative estimator
  - Computed separately for MRTS certainty and MRTS noncertainty units
- Mean absolute error (MAE) of link relative ratio
  - Evaluation at the tabulation level
  - Compare estimates of link relative ratio to “true” link relative ratio from completed data set

Results: Three patterns

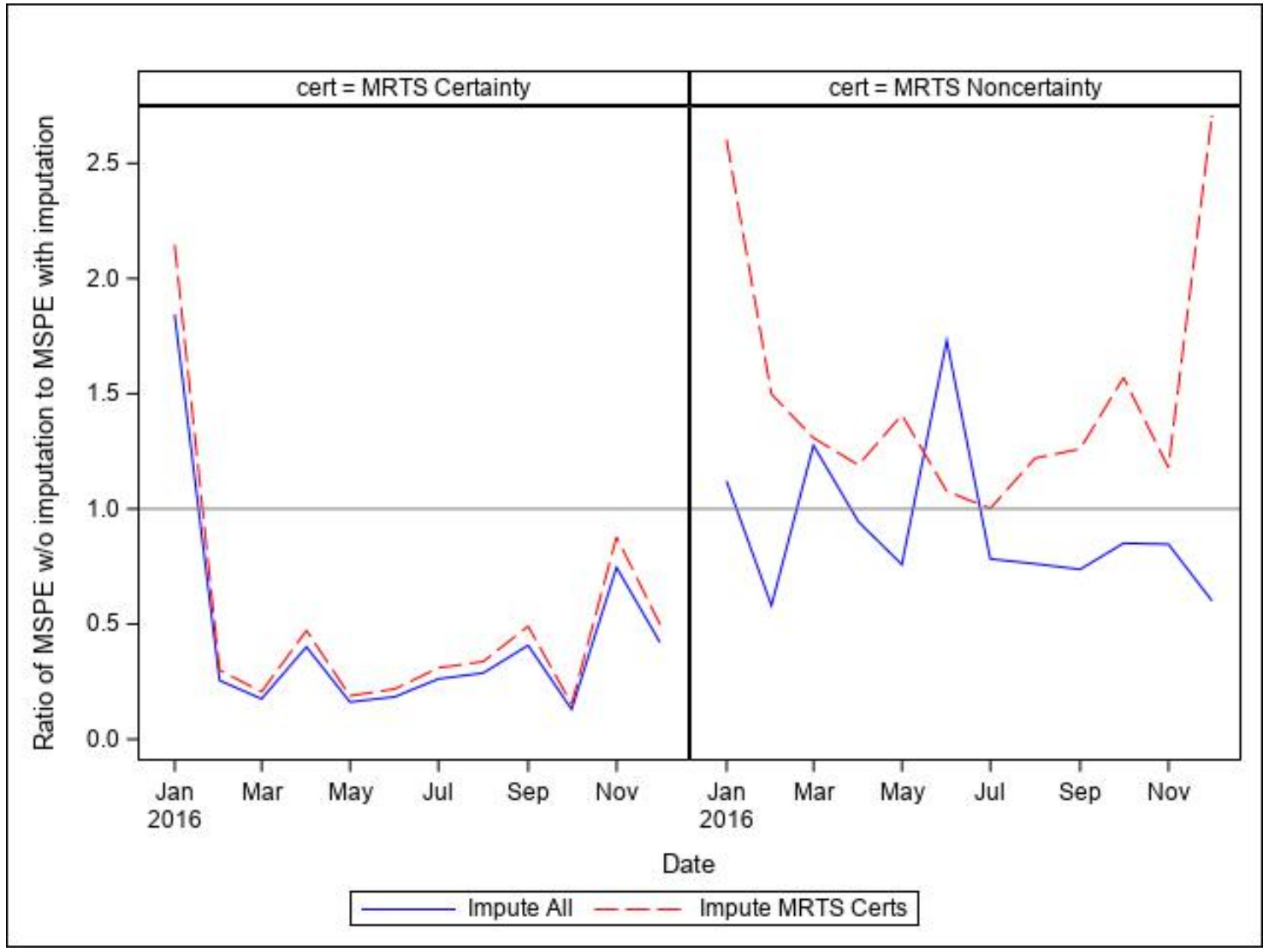
- No improvement with forecast imputation
- Improvement with forecast imputation for all units
- Improvement with forecast imputation for MRTS certainty units only

Unit level prediction errors: No imputation better

$$\frac{MSPE_{noimp}}{MSPE_{impute}}$$

NAICS 447000

Gasoline Stations

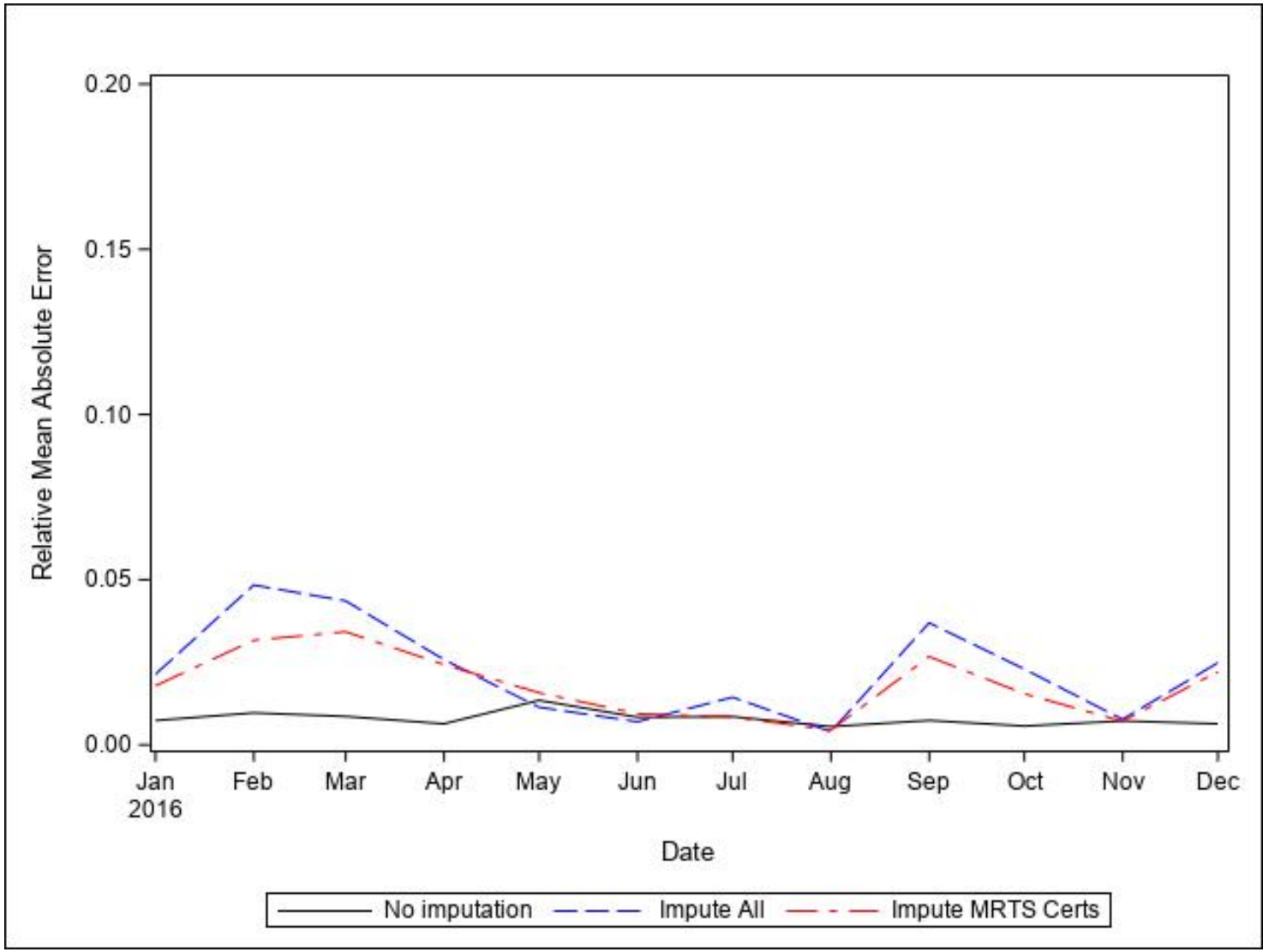


Effects on Total: No imputation better

Relative MAE

NAICS 447000

Gasoline Stations

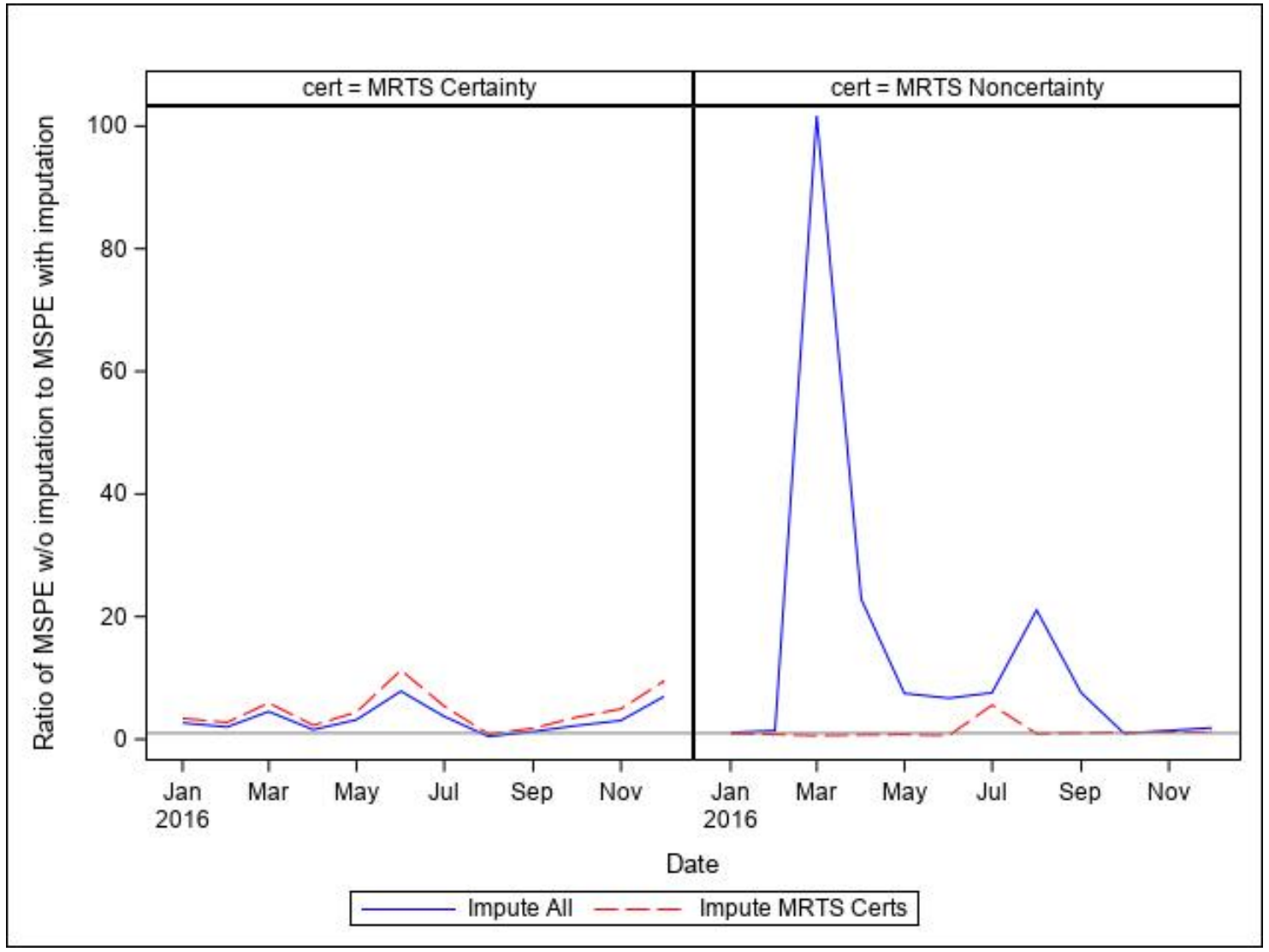


Unit level prediction errors: Impute all units

$$\frac{MSPE_{noimp}}{MSPE_{impute}}$$

NAICS 448110

Men's Clothing Stores

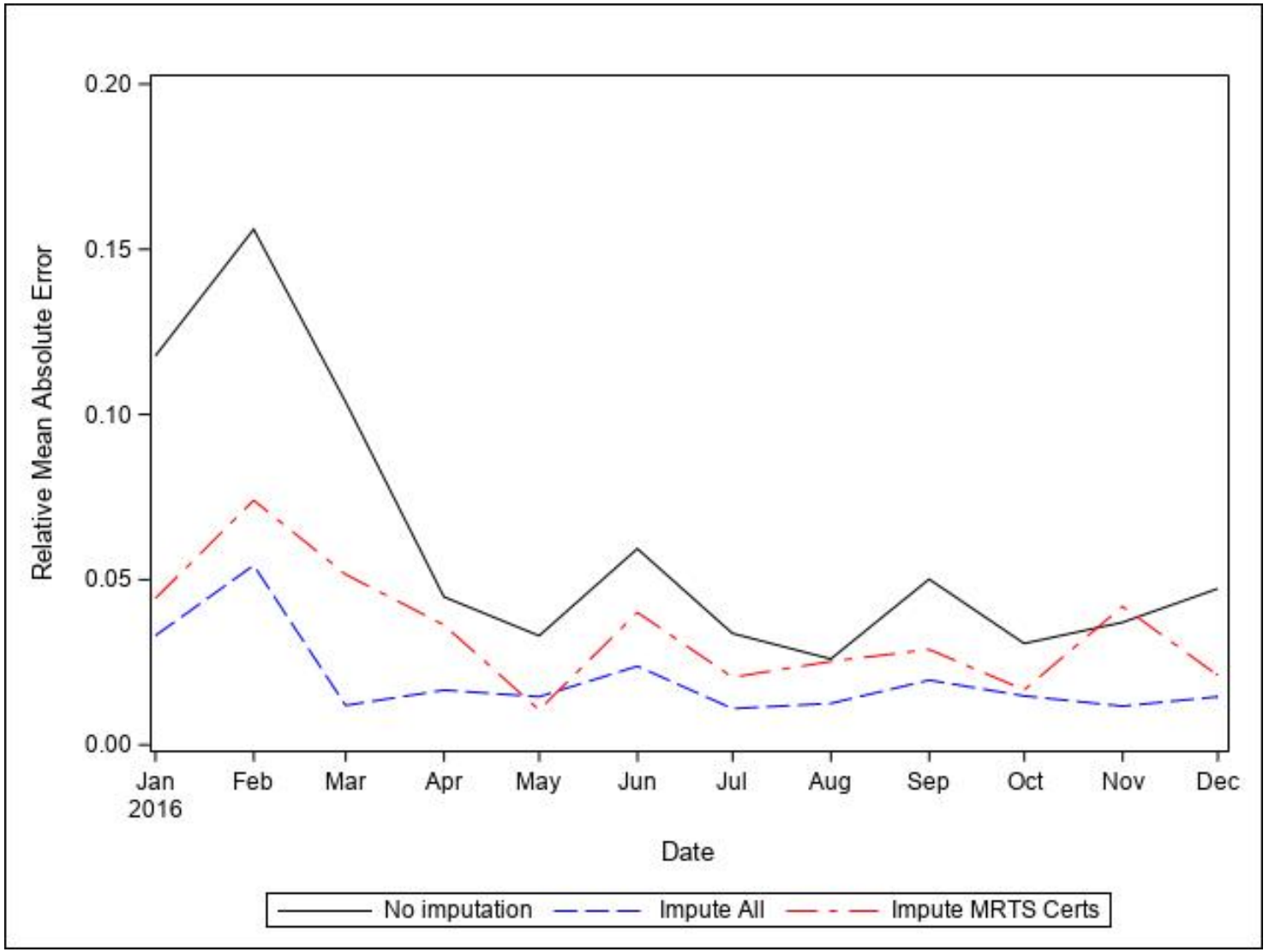


Effects on Total: Impute all units

Relative MAE

NAICS 448110

Men's Clothing Stores

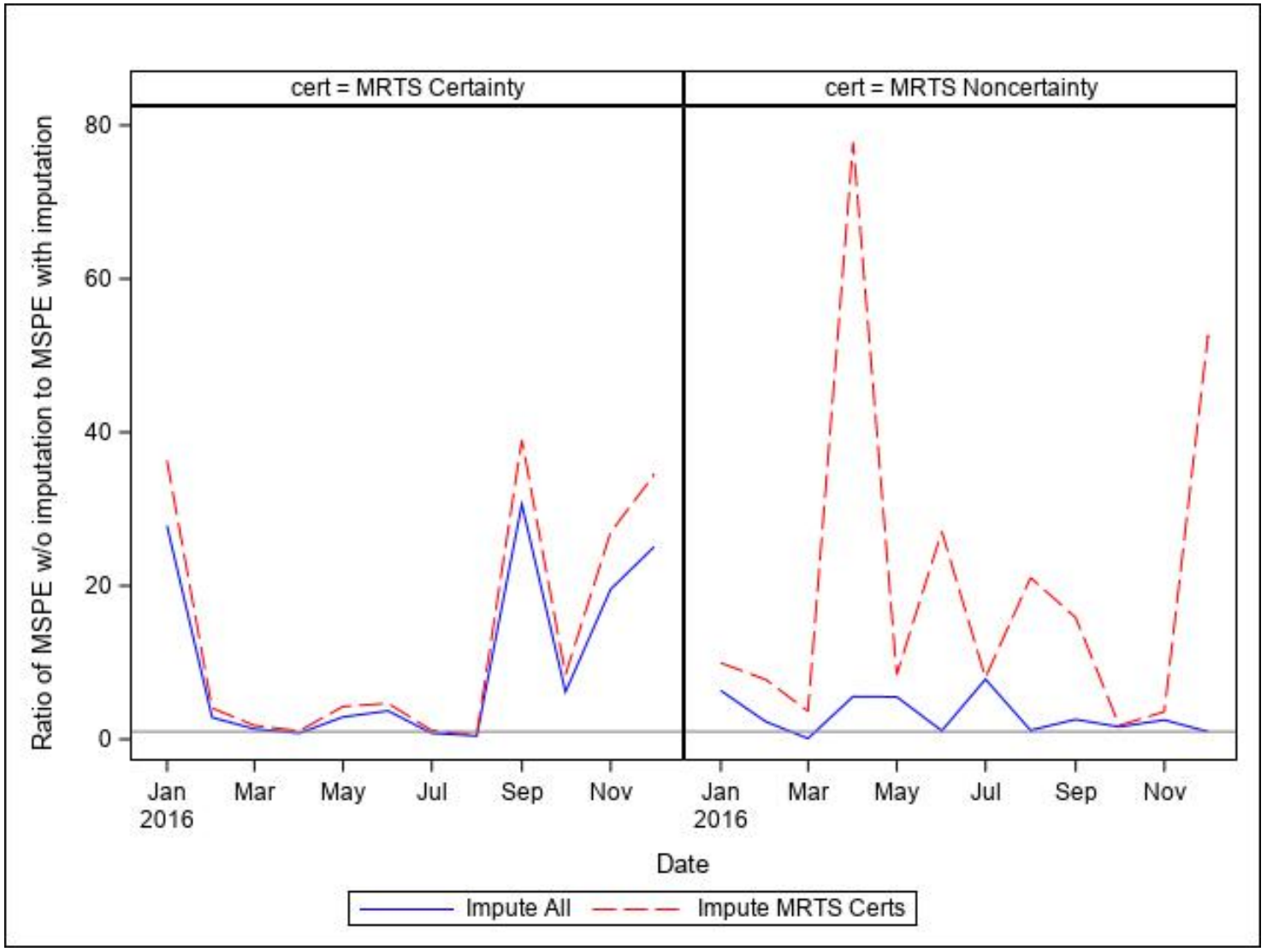


Unit level prediction errors: Impute MRTS certs

$$\frac{MSPE_{noimp}}{MSPE_{impute}}$$

NAICS 4511X0

Sporting Goods and Hobby Stores

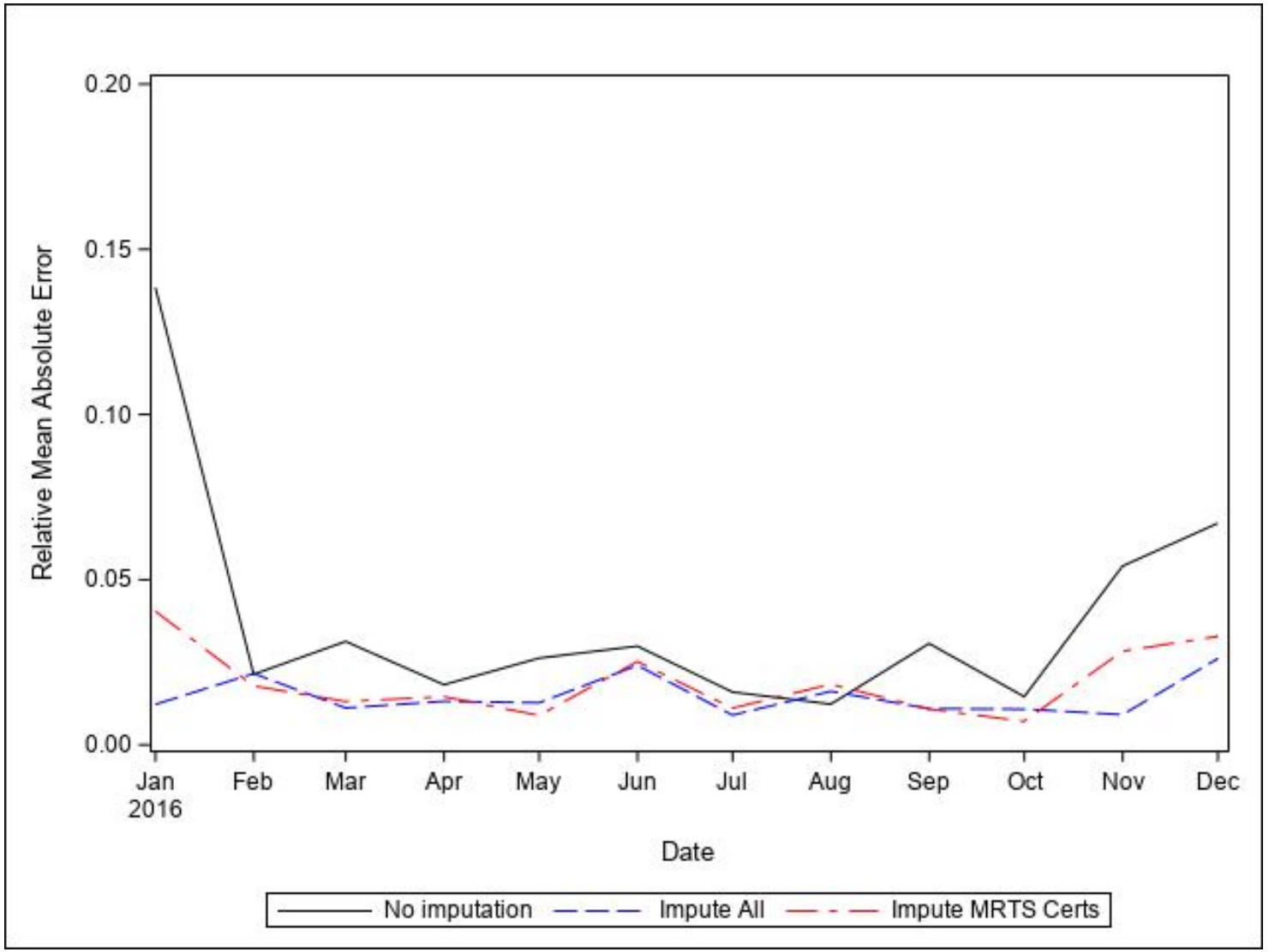


Effects on Total: Impute MRTS certs only

Relative MAE

NAICS 4511X0

Sporting Goods and Hobby Stores



Conclusions

- RegARIMA forecast imputation reduces MSPE and MAE for many MARTS industries
- However, there are some MARTS industries where the link relative estimator without imputation better accounts for nonresponse
- Most effective for large units: limited use for smaller units
- Imputing only the MRTS certainty units may be enough to see the benefits of imputation and offers consistent methodology throughout the sample lifecycle

Future Work

- Repeat simulation with response propensities as a function of size
- Refinements to RegARIMA forecast imputation method
  - "Update" one-step ahead forecast with information from current month reported data
  - If we only impute MRTS certainty units, could look at models that use data further in the past
- Incorporating uncertainty due to imputation into variance estimates

Questions?

Nicole.Czaplicki@census.gov
Yarissa.Gonzalez@census.gov