Finding an Estimator that Minimizes Revisions in a Monthly Indicator Survey

Nicole Czaplicki, Yarissa Gonzalez, Laura Bechtel
U.S. Census Bureau

Disclaimer: Any views expressed are those of the author(s) and not necessarily those of the U.S. Census Bureau.

Outline

- Motivation
- Research purpose
- Background on MARTS and MRTS
- Current estimator
- Estimators considered
- Results
- Conclusions

Monthly Retail Trade at Census

- Two surveys: three releases
  - Advance Monthly Retail Trade and Food Services Survey (MARTS)
    - Advance release of sales
  - Monthly Retail Trade and Food Services Survey (MRTS)
    - Preliminary release of sales
    - “Final” release of sales
- MRTS data are also revised annually with changes from benchmarking to the annual survey.
    - Includes remaining late reported data
    - Data corrections

Retail Trade Indicator Release Example for January

February
Advanced January Estimates from MARTS

March
Preliminary January Estimates from MRTS

April
Final January Estimates from MRTS

Goal: minimize this revision

Motivation

- MARTS is an economic indicators
- Total retail trade and food services sales estimates are inputs to GDP
- **Revisions to the estimates, particularly the seasonally adjusted month-to-month change, are highly scrutinized.**

Research Purpose

Identify an estimator that minimizes revision error between the MARTS estimate and MRTS preliminary estimate

- Using only respondent data
- Including analyst imputes, which are a subset of influential nonrespondents imputed based on historical company performance (current method)

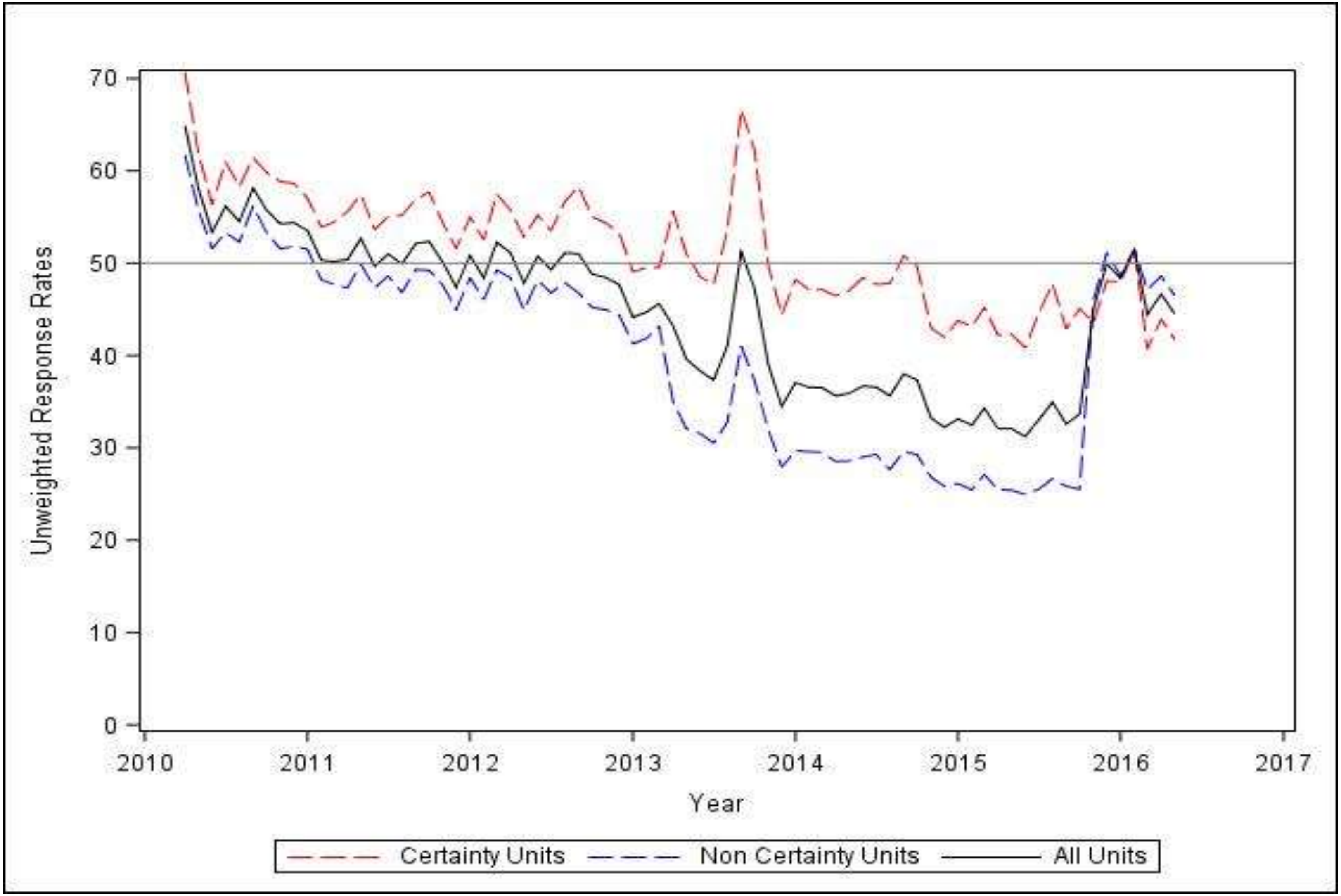
Research Questions

1. Is there an alternative estimator that we can use that would reduce revision error?
2. Are the analyst imputed values reducing revision error?

Challenges to minimizing revisions

1. Low response rates for MARTS
2. Differences in methodology between MARTS and MRTS

MARTS Unit Response Rates



	MARTS	MRTS
Sample size	5,000 companies	12,000 companies
Sample frame	MRTS sample	Annual Retail Trade Survey sample
Sample design	Stratified PPS -WOR (subsample of MRTS)	Stratified SRS-WOR
Sample redesign cycle	Approximately every 2.5 years	Approximately every 5 years
Time to respond	Approximately 7 business days	Approximately 5 weeks
Imputation	Analyst impute for selected companies	Analyst imputes retained, ratio impute for remaining nonrespondents and edit-failing items
Estimation	Link relative estimator	Horvitz-Thompson estimator
Tabulation industries	30	83

	MARTS	MRTS
Sample size	5,000 companies	12,000 companies
Sample frame	MRTS sample	Annual Retail Trade Survey sample
Sample design	Stratified PPS -WOR (subsample of MRTS)	Stratified SRS-WOR
Sample redesign cycle	Approximately every 2.5 years	Approximately every 5 years
Time to respond	Approximately 7 business days	Approximately 5 weeks
Imputation	**Analyst impute for selected companies**	**Analyst imputes retained, ratio impute for remaining nonrespondents and edit-failing items**
Estimation	**Link relative estimator**	**Horvitz-Thompson estimator**
Tabulation industries	30	83

Link Relative Estimator

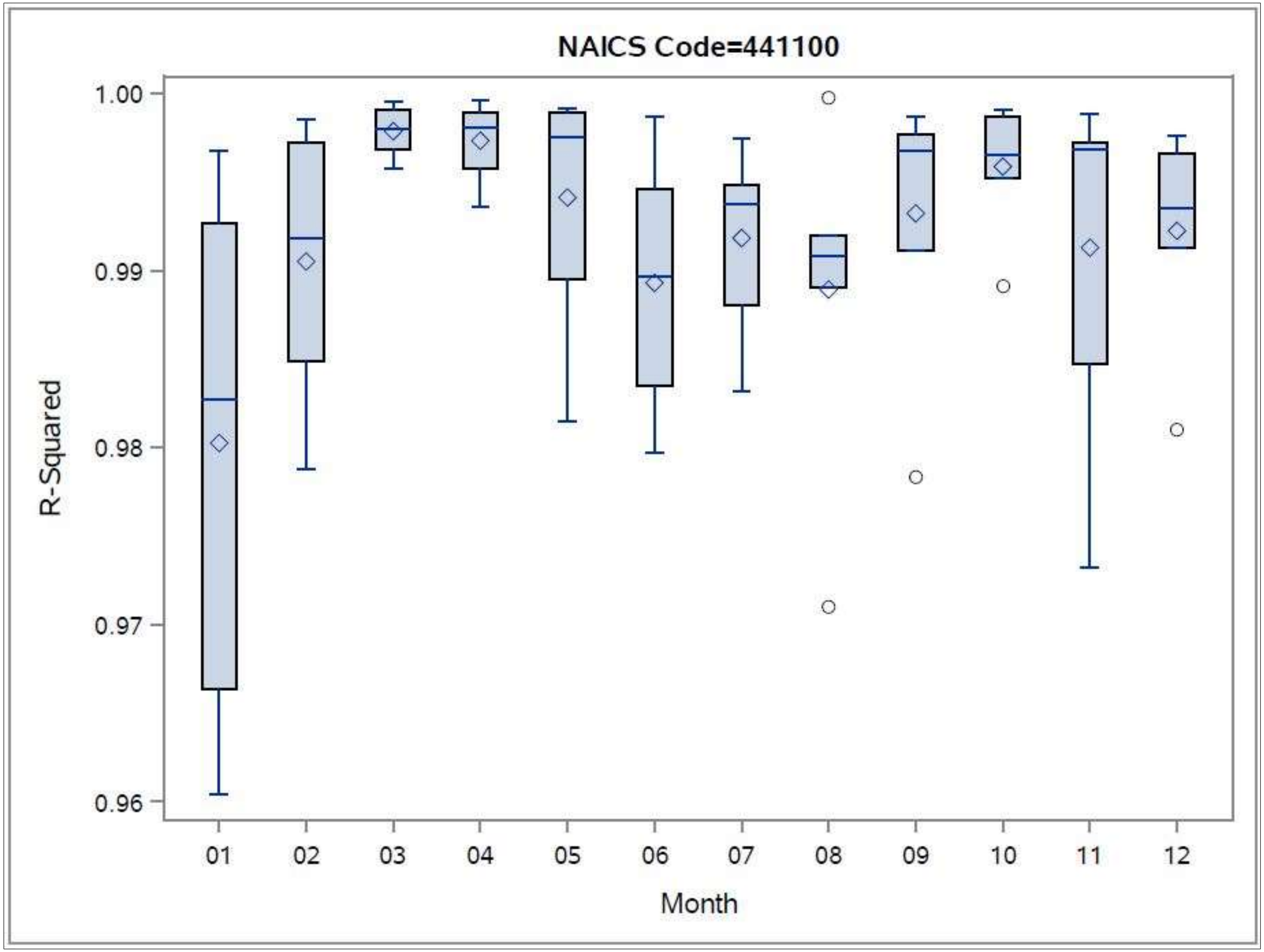
$$\hat{Y}_{LR,t} = B_{t-1}\left[\frac{\sum_{i\in C} w_i y_{t,i}}{\sum_{i\in C} w_i y_{t-1,i}}\right]$$

- $\hat{Y}_{LR,t}$ = link – relative estimator of a 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
- $\left[\frac{\sum_{i\in C} w_i y_{t,i}}{\sum_{i\in C} w_i y_{t-1,i}}\right] = \frac{\hat{Y}_t^C}{\hat{Y}_{t-1}^C}$ = the screened sum ratio (link relative)
- Introduced by Madow and Madow (1978)

Link Relative Assumption Validation

- Key assumption $E\left[y_{t,i} \middle| t-1\right] = \beta_t y_{t-1,i}$
- Tested by estimating:

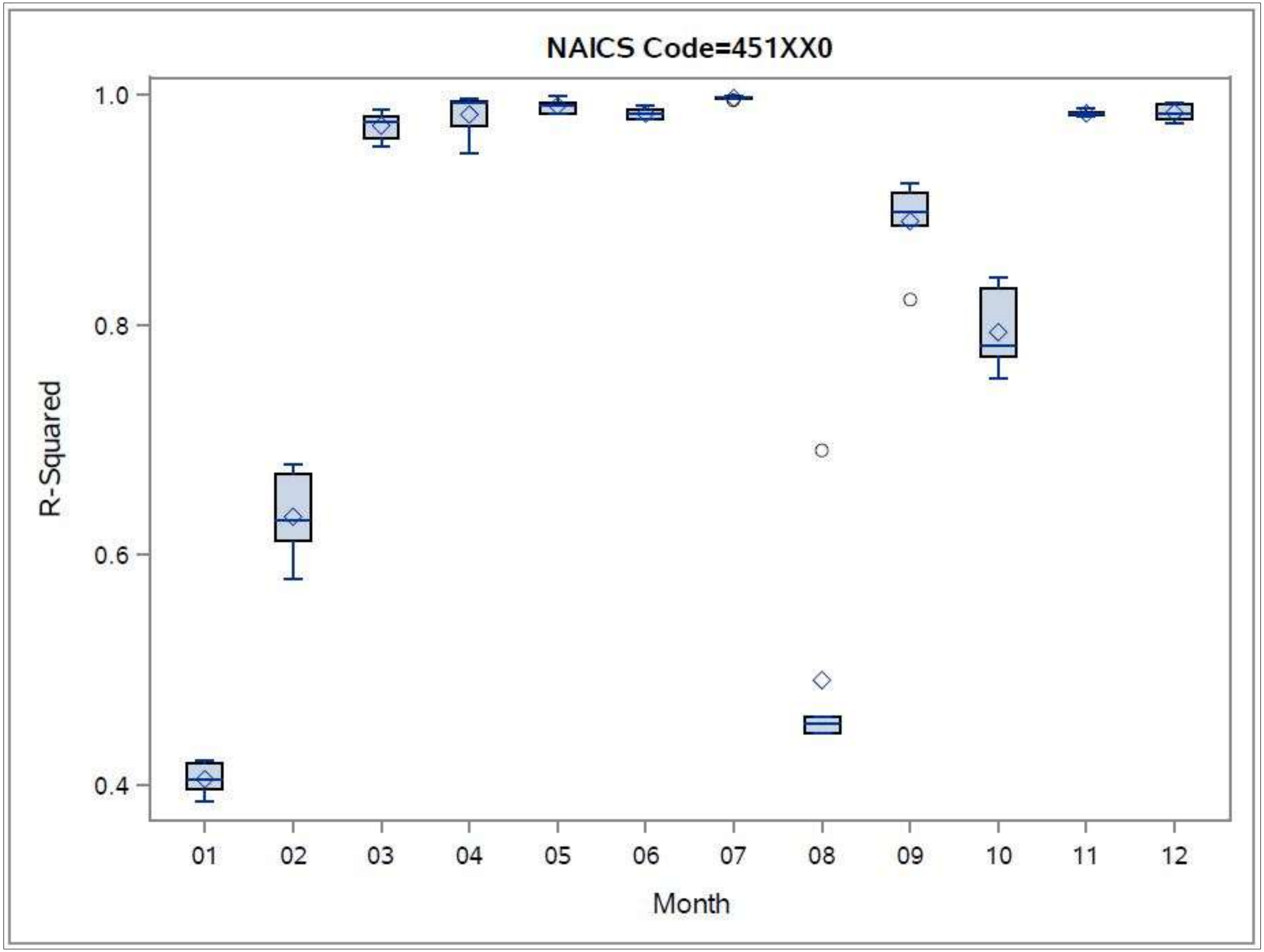
$$y_{i,t} = \beta_t y_{i\ t-1} + e_{i,t}, \qquad e_{i,t}\ \sim(0, \sigma_t^2)$$



Box plots of R^2 statistics by month for the model $y_{i,t} = \beta_t y_{i\ t-1} + e_{i,t}$ for tabulation NAICS 441100, new and used car dealers. Source MARTS and MRTS: May 2010 – May 2016.

Tabulation NAICS 451XX0

- General Bookstores
- Specialty, Religious, and Used Bookstores
- School and College Bookstores
- Musical Instrument and Supplies Stores
- News Dealers and Newsstands
- Prerecorded Tape, Compact Disc, and Record Stores



Box plots of R^2 statistics by month for the model $y_{i,t} = \beta_t y_{i\ t-1} + e_{i,t}$ for tabulation NAICS 451XX0. Source MARTS and MRTS: May 2010 – May 2016.

Modified Link Relative Estimator

$$\hat{Y}_{MLR,t} = \left[B_{t-1} - \sum_{i \in S} \widetilde{w}_i y_{t-1,i}\right]\left[\frac{\sum_{i \in \acute{C}} w_i y_{t,i}}{\sum_{i \in \acute{C}} w_i y_{t-1,i}}\right] + \sum_{i \in S} \widetilde{w_i} y_{t,i}$$

- $\acute{C}$ = set of units with usable data for both periods t and $t-1$ that are **included** in the calculation of the screened-sum ratio

- S = set of units with usable data for both periods t and $t-1$ that are **excluded** in the calculation of the screened-sum ratio
  - Referred to as “estimate separately” units
  - Group 1: “high impact” units currently excluded from MARTS estimation
  - Group 2: Group 1 units plus the analyst imputed units
- $\widetilde{w_i}$=final weight from benchmark survey
- Developed at Census for the Manufacturers’ Shipments, Inventories, and Orders Survey (M3)

Weighting Class Estimator

$$\hat{Y}_{WC,t} = \sum_{p=1}^{P} \sum_{i \in r_p} w_i \, y_{t,i} \frac{\sum_{i \epsilon s_p} w_i}{\sum_{i \epsilon r_p} w_i}$$

- $i \in r_p$ indicates unit i is in the respondent set for weighting cell p
- $i \in s_p$ indicates unit i is in the sample for weighting cell p
- Estimate inverse-probability of selection as $\frac{\hat{N}_{s_p}}{\hat{N}_{r_p}} = \frac{\sum_{i \in s_p} w_i}{\sum_{i \in r_p} w_i}$
- Weighting cells will be defined by certainty status
- Sarndal and Lundstrom (2005, p. 70), Oh and Scheuren (1983), Kalton and Kasprzyk (1986), Little (1986)

Weighting Class Estimator

$$\hat{Y}_{WC,t} = \sum_{p=1}^{P} \sum_{i \in r_p} w_i \, y_{t,i} \frac{n_p}{m_p}$$

- n_p = sample size in cell p
- m_p = number of respondents in cell p

- Vartivarian and Little (2002) recommend using unweighted response rate.

Ratio Estimator

$$\hat{Y}_{R,NAICS_j,t} = \sum_{p=1}^{P} \sum_{i \in r_p} w_i y_i \frac{\sum_{i \epsilon s_p} w_i x_i}{\sum_{i \epsilon r_p} w_i x_i}$$

- x_i = value of auxiliary variable for unit i
  - MRTS prior month's sales
- For the ratio estimator, we assume
  - $y_i = \beta x_i + \varepsilon_i$ where $\varepsilon_i \sim (0, x_i \sigma^2)$

Considered Estimators

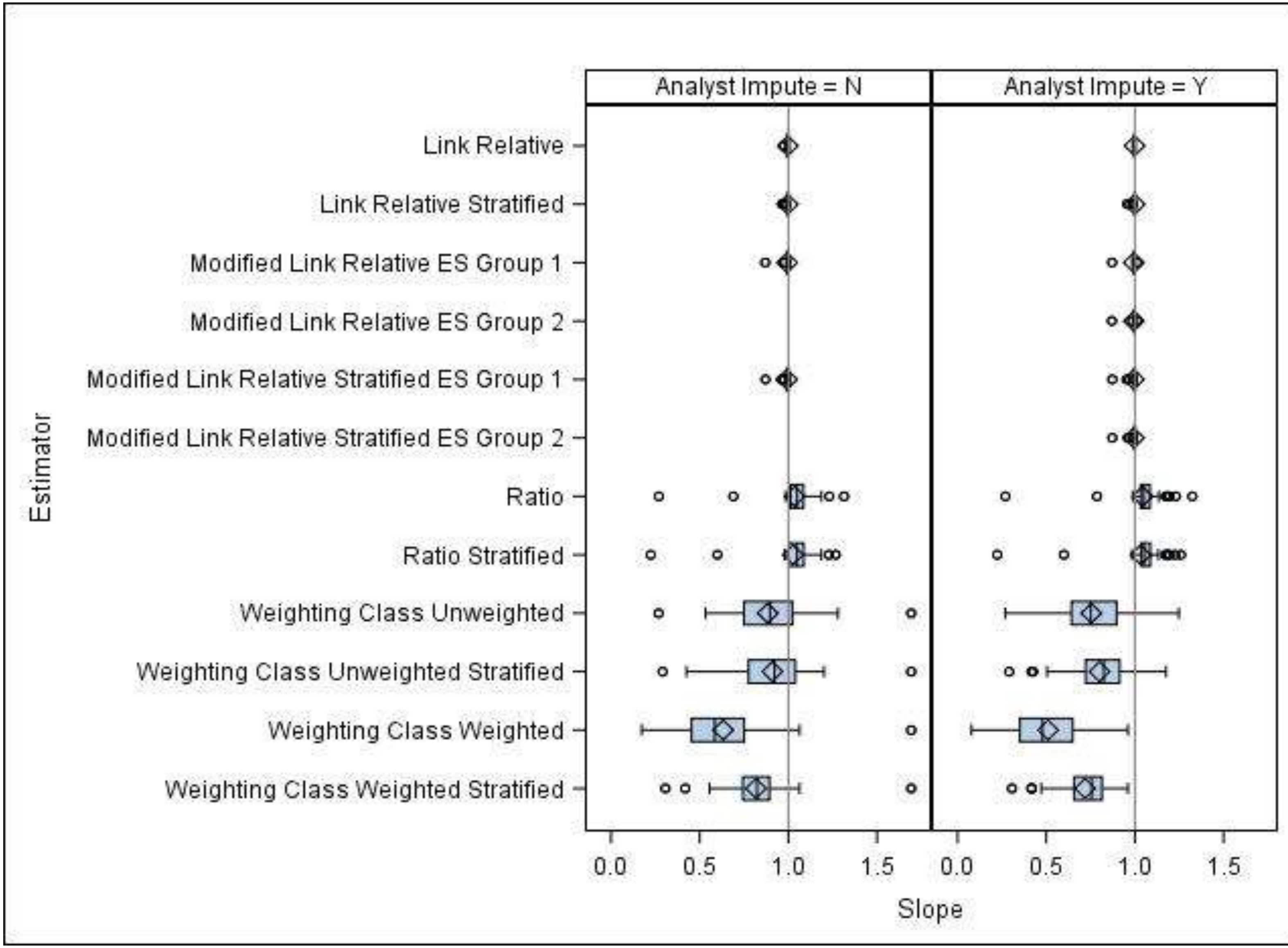
Estimator	Stratified or non-stratified	Include or exclude weights in adjustment factor
Link Relative	Both	n/a
Modified Link Relative group 1	Both	n/a
Modified Link Relative group 2	Both	n/a
Weighting Class	Both	Both
Ratio	Both	Include weights

Evaluation Criteria

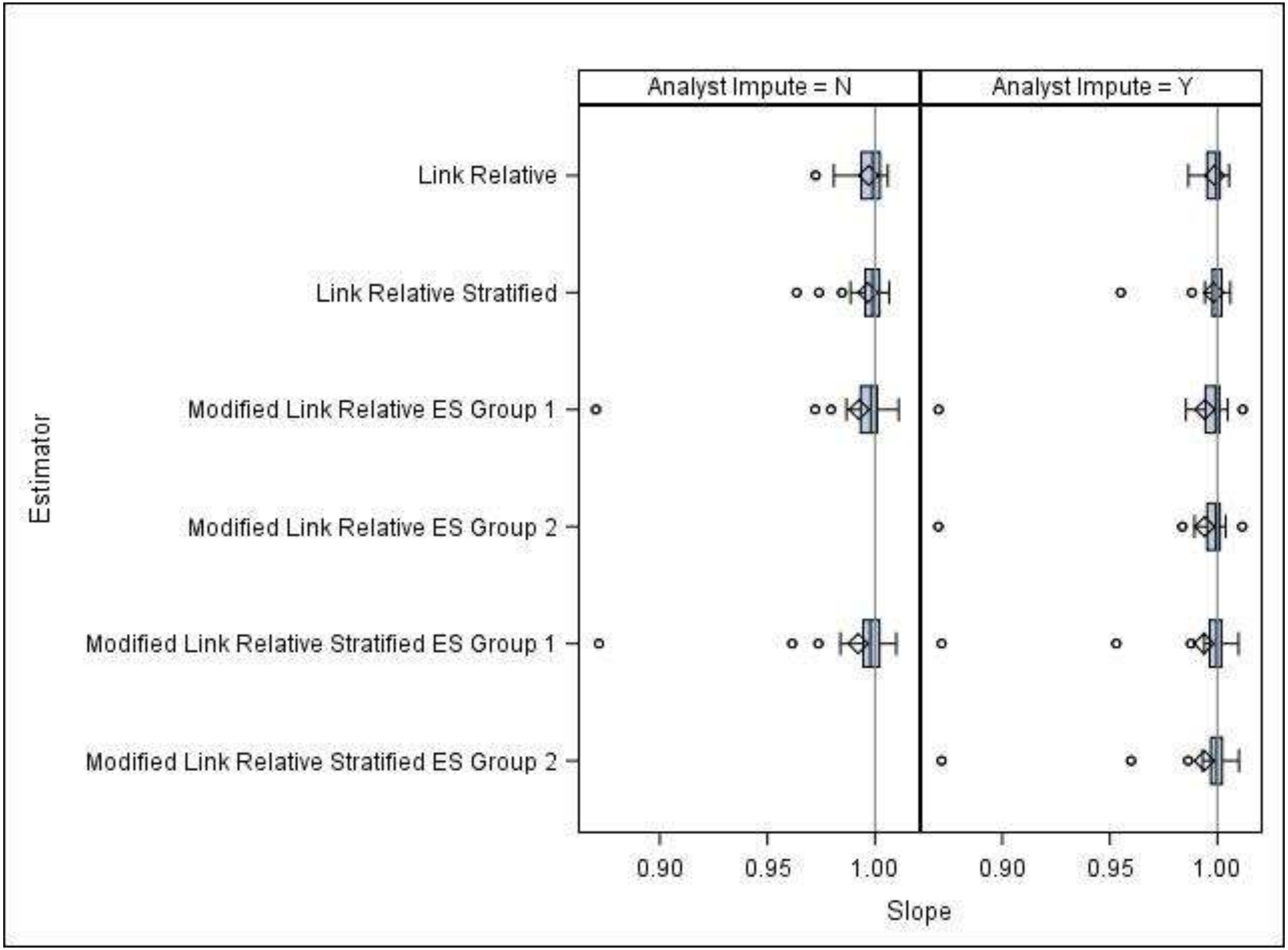
- Regression: Ideally MARTS perfectly predicts MRTS so β_t should be close to 1.0

$$\hat{Y}_{t,j,MRTS\ Prelim} = \beta_t \hat{Y}^{k}_{t,j,MARTS} + \varepsilon_t,\ \varepsilon_t \sim (0, \sigma_j^2)$$

- Mean absolute revision error for month-to-month percent change

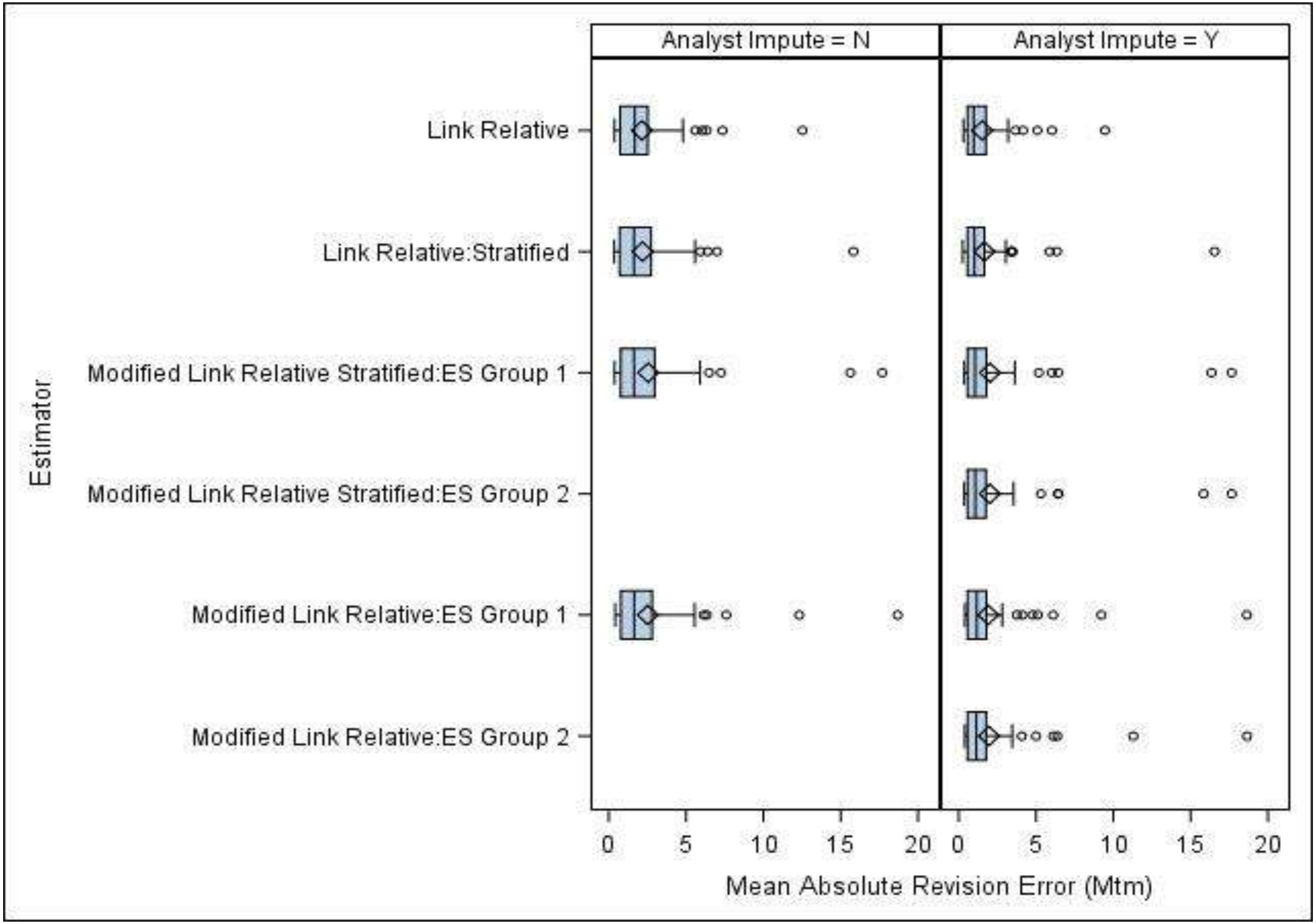


Distribution of slope parameters by estimator from the 30 tabulation NAICS regression models. Source: MARTS May 2010-May2016



Distribution of slope parameters for the link relative and modified link relative estimators from the 30 tabulation NAICS regression models Source: MARTS May 2010-May2016

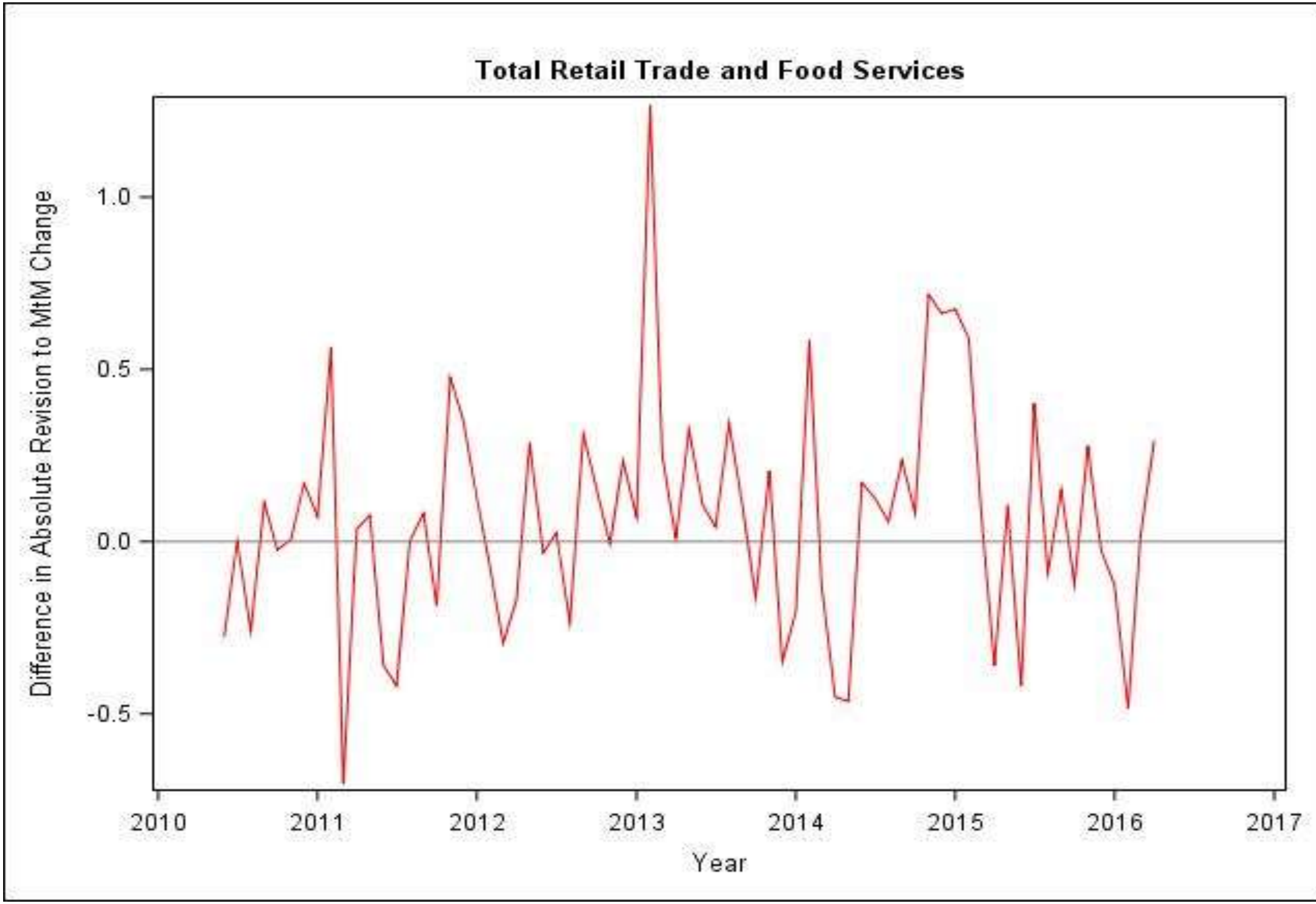
Mean Absolute Revision Error Month-to-Month Changes Total Retail Trade and Food Services		
Estimator	Current analyst imputes	No analyst imputes
Link Relative	0.309	0.374
Link Relative: Stratified	0.277	0.354
Modified Link Relative: ES group 1	0.411	0.462
Modified Link Relative: ES group 2	0.396	n/a
Modified Link Relative Stratified: ES group 1	0.377	0.435
Modified Link Relative Stratified: ES group 2	0.363	n/a
Ratio	6.399	6.146
Ratio: Stratified	66.203	5.995
Weighting class: Weighted	75.854	45.048
Weighting class: Unweighted	24.611	5.960
Weighting class: Stratified weighted	29.388	12.205
Weighting class: Stratified unweighted	17.825	3.906



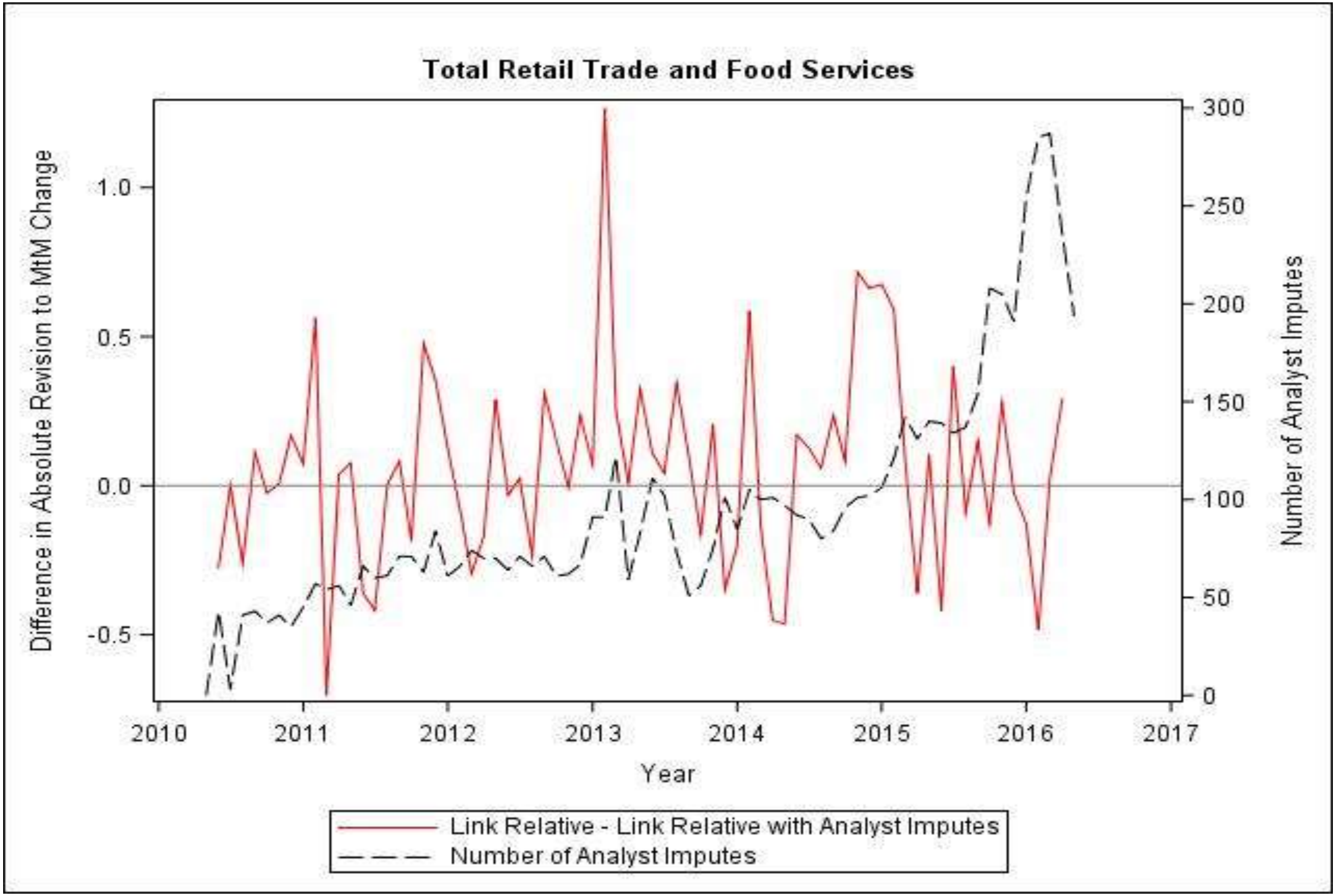
Distribution of mean absolute revision error for month to month changes at the tabulation industry level for the link relative and modified link relative estimators. Source: MARTS May 2010-May2016

Research Questions

1. Is there an alternative estimator that we can use that would reduce revision error?
2. Are the analyst imputed values reducing revision error?



Absolute revisions to month-to-month changes using the link relative estimator without analyst imputes – absolute revisions to month-to-month changes using the link relative estimator with analyst imputes for MARTS Source: MARTS May 2010-May2016



Absolute revisions to month-to-month changes using the link relative estimator without analyst imputes – absolute revisions to month-to-month changes using the link relative estimator with analyst imputes for MARTS. Number of analyst imputes plotted on the second y-axis. Source: MARTS May 2010-May2016

Conclusions

- Provided strong evidence that the currently used version of link relative estimator is effective at reducing revision error
  - Modified link relative estimator and stratified link relative estimator had inconsistent performances
  - Nonresponse adjustment estimators did not reduce revision error
- Found evidence that the current imputation procedure reduces revision error
  - We looked at aggregate results – did not look at impact of individually imputed values
  - It is possible that the improvements may be due to a select few observations

Contact Information

- Nicole.Czaplicki@census.gov
- Yarissa.Gonzalez@census.gov
- Laura.Bechtel@census.gov

Acknowledgements

- Thank you to MARTS estimation research team
  - Adam Smeltz
  - Artaud Hourticolon Retzler
  - Andromache Mason