Calculating Boundary Limits for the Hidiroglou-Berthelot (HB) Edit used to Flag Administrative Receipts

Melvin McCullough, Michael Kornbau
U.S. Census Bureau, Economic Statistical Methods Division, Washington DC

Abstract

The Census Bureau maintains a comprehensive database of U.S. businesses known as the Business Register for statistical program use. The BR is updated continuously with information available from Census Bureau and other Federal statistical and administrative records programs. The Census Bureau obtains administrative receipts from the Internal Revenue Service (IRS) on a weekly basis for corporations, partnerships, and sole proprietorships. Administrative receipts are used in a variety of business surveys, as a substitute for survey response for non-respondents or to supplement survey-collected receipts. Unedited BR administrative receipts are, at times, a poor substitute for survey-collected receipts. This paper describes the methodology devised to edit administrative receipts for single-location (single-unit) businesses. We used the Hidiroglou-Berthelot (HB) edit to flag units with inconsistent receipts to payroll ratios based on a standardized HB “score” statistic. In order to define boundaries for HB score cutoffs, we devised a logistic regression model that divided the data into three categories: 1. Use for tabulation and imputation 2. Use for tabulation but not for imputation 3. Do not use for tabulation or for imputation.Key Words: Imputation, single-unit, multi-unit

Disclaimer: Any views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

1.0 Background

The Census Bureau receives administrative records for statistical program use from a number of external government agencies including the Internal Revenue Service (IRS), the Social Security Administration (SSA), and the Bureau of Labor Statistics (BLS).

The IRS provides the Census Bureau with administrative payroll, employment and receipts data for businesses. Businesses with paid employees report their payroll and employment either quarterly on IRS Form 941 or annually on IRS Form 944. The data are reported to the IRS based on Employer Identification Number (EIN). A business may have multiple EINs and/or operate establishments in multiple physical locations. The focus of the research for this paper is single-establishment businesses with positive payroll and one EIN, which are less complex to handle for editing of administrative receipts.

The IRS also provides the Census Bureau with business income and receipts data from various business income tax forms including Form 1120 (corporations), Form 1065 (partnerships), and Form 1040, Schedule C (sole proprietorships). All except Form 1040, Schedule C are filed by EIN, though businesses may file consolidated income tax returns, and may use different EINs for payroll tax filing and income tax filing.   Form 1040, Schedule C is filed by Social Security Number (SSN). The Census Bureau maintains SSN-EIN linkages to link the SSN-filed income to the EIN-filed payroll for sole proprietorship employer companies.

Based on the payroll and income tax filings, the IRS provides the Census Bureau with basic information including the following:

- Legal and trade name of business
- Mailing and physical location address
- Legal form of organization
- Principal business activity

1.**1 Introduction**

The IRS provides the Census Bureau with approximately 5.5 million administrative receipt records for single unit businesses annually. These receipts data are posted to the Business Register and can be used as a business measure of size for sampling or for editing and imputing survey-collected data. Before starting our research, the Census Bureau was not editing the administrative receipts data received from the IRS on the Business Register itself, but rather editing these data in a less coordinated effort through and for various survey programs. In addition, the coverage of units on the Business Register in the past was focused on payroll data, rather than receipts data. In order to focus more on receipts data too and to ensure the quality of incoming receipts data, the Census Bureau has begun efforts to develop a more corporate approach to ensuring the quality of receipts data – in particular for single-unit businesses – to begin to explore a strategy to edit administrative receipts. Initially, we explored a few methods including linear regression and the Hidiroglou-Berthelot (HB) edit.

1.2 **General Approach to Outlier Detection**

Our main objective was to identify a test statistic to use in the detection of outliers of current year administrative (ADMIN) receipts, while minimizing errors in the process. Informally, we define a "type1" error as erroneous identification of an establishment's value as a non-outlier. In this case, the test statistic for the establishment is not extreme, but a scatter plot of receipts shows the establishment's value is far removed from the trend of non-outlier points. Similarly, a type II error is erroneous identification of an establishment's value as an outlier; the corresponding test statistic is extreme, but graphically, the receipts value is not. Initially, we investigated the use of linear regression. We regressed current-year ADMIN receipts against current-year payroll and used a pair of regression diagnostics as criteria to flag outliers. Secondly, we tried an HB edit approach in which we used a ratio of current-year to prior-year ADMIN receipts in order to detect ADMIN receipts outliers. Third, we used an HB edit in which we used a ratio of current-year ADMIN receipts to payroll to detect ADMIN receipts outliers.

With each approach, we followed the steps below:

- We noted the proportion of establishments to the total classified as outliers.
- We adjusted critical values of the test statistics and re-ran the approach and made comparisons
  based on multiple sets of critical values.
- We created scatter plots to investigate type I and type II errors.
- We investigated the symmetry of outliers about the trend line formed in the scatter plots.

Ultimately, we sought to choose the best critical values of the test statistics that would lead to detecting as many true outliers as possible. At the same time, we sought to minimize the number of plausible values of ADMIN receipts classified as outliers.

In order to make comparisons between the linear regression and the HB edits, we produced scatter plots of each approach by NAICS industry[1]. The HB edits utilize score statistics to differentiate between outliers and non-outliers of ADMIN receipts. Plots denoted HB "outliers" with an asterisk and non-outliers as a dot. Type I errors were noted when the score value was non-significant but the corresponding plotted point was far outside the linear scatter of points formed by ADMIN receipts and payroll. A type II error was noted when HB score was highly significant but the plotted observation was on or near the linear point scatter.

With the linear regression approach, the basis used for outlier identification was a combination of leverage (Hat) and studentized residual (Rstudent). In a similar manner with this approach, we plotted and denoted regression outliers with an asterisk and "non-outlier" with a dot. Again, we noted type I errors and type II errors using the regression approach. After a thorough review of the two approaches, we eventually chose a version of the HB edit method to edit administrative receipts.

Initially, with the HB edit, we set a single set of fixed limits for positive and negative values of HB score. We studied how varying the fixed limits affected the proportion of ADMIN receipts flagged as outliers. We also used the plots to investigate differences between outliers and non-outliers.

1.3 **The Initial HB Edit Study**

For this study, we created a data set from the Business Register of 5,524,948 single-unit establishments with payroll in 2012 and that showed an indication of filing a 2012 income tax return. We divided this data set into a training data set (60%), a test data set (20%), and a cross validation data set (20%). We used the training data set to develop our methods and to compute parameters that we tested on our test data set and validated through our cross-validation set. We deleted small companies where both current-year administrative receipts and payroll were less than $100,000 from the data set prior to the analysis. Small companies can naturally have large swings in the ratio of receipts to payroll, or in the ratio of current-year to prior-year receipts. They also contribute less to overall economic activity.

We experimented with two distinct ratios to test ADMIN receipts using the HB edit. One ratio was formed by the ratio of current- to prior- year administrative receipts. The other ratio was current-year administrative receipts to current-year annual payroll. We prepared scatter plots to evaluate HB edit results for each ratio. One set of plots graphed current-year administrative receipts against prior-year administrative receipts broken out by NAICS industry. The other set of plots graphed current-year ADMIN receipts against current-year payroll. Observations identified as "outliers" based on the HB score statistics were labeled in the plots with an asterisk. We labeled non-outliers with a dot. For each approach, we noted the number of type I and type II errors as previously defined. Based on a thorough review of plots, we decided to utilize the ratio of current year administrative receipts (ADMIN receipts) to payroll as a basis for editing current year ADMIN receipts.

[1] North American Industry Classification System.

1.4 **Calculating HB Parameters**

The NAICS structure uses 6-digit codes to classify a business by what it does. The first two digits define the sector, with subsequent digits providing additional detail. Examples of NAICS sectors include '23' for the construction sector and '42' for the wholesale trade sector. The first four digits define the NAICS industry group.

To calculate HB edit parameters, the first step was to define the ratio for each establishment as:
$R=(X_c/Y_c)$ where

X_c =Current-year best administrative receipts
Y_c=Current-year administrative annual payroll

To apply administrative receipts editing, at least one of the values (X_c,Y_c) had to be \$100,000 or greater. Following this criterion limited the range of the ratio values, X_c and Y_c, which in turn helped us determine "true" outliers with more confidence. We defined the median of these ratios as R_M within NAICS class. The NAICS class was primarily defined as NAICS sector, but several NAICS industry groups were separated out as NAICS classes after a review of receipts- to payroll- ratio distributions revealed significant differences from the remainder of the NAICS sector. Four classes included the more detailed industry groups of 4245, 4247, 4471, and 5251. This implied that the NAICS class for Sector 42 would include all single unit establishments coded in Wholesale Trade, except for those classified as 4245 or 4247. The same structure applied to sectors 44-45 and 52. Based on an analysis of receipts to payroll ratios by industry group, the industry groups 4245, 4247, 4471, and 5251 stood out as different from other industry groups within the same sector.

Next, we produced the following transformation of the ratio for each establishment:

$$SR = (R/R_M)-1 \text{ if } R \geqslant R_M$$
$$= 1-(R_M/R) \text{ if } 0<R \leqslant R_M$$

Another transformation using SR above was a size-effect transformation: [2]$ESR=SR*\{\max(X_c*R_m,Y_c\}^{0.5}$. ESR takes into account the magnitude of the numerator and denominator, and consequently assigns more outliers to large businesses that produce more economic activity. The exponent U=0.5 can be adjusted to any value from 0 to 1. We determined that the size parameter U=0.5 was the most effective at identifying outliers after experimenting with several values of the exponent. In section 1.5 below, we describe our approach for deciding on a value of the size parameter U.

We then determined the first and third quartiles and medians for the transformed ratios of SR and ESR within NAICS class. Then we computed the following four measures of variation for each NAICS class:
$D_{SR,\,Q1}=\max\ \{(SR_M-SR_{Q1}),|0.05*SR_M|\}$
$D_{SR,\,Q3}=\max\ \{(SR_{Q3}-SR_M),|0.05*SR_M|\}$

$D_{ESR,\,Q1}=\max\ \{(ESR_M-ESR_{Q1}),|0.05*ESR_M|\}$

[2] Sigman Richard S (2005) Statistical Methods used to Detect Cell-level and Respondent-Level Outliers in the 2002 Economic Census of the Services Sector

$D_{ESR,\,Q3}$=max {(ESR_{Q3}-ESR_M),|0.05*ESR_M|}

The seven values R_M, SR_M, ESR_M, $D_{SR,Q1}$, $D_{SR,Q3,}$ $D_{ESR,Q1,}$ and $D_{ESR,Q3}$ as defined were utilized to create the following normalized HB test statistics:

QSR= ((SR-SR_M)/$D_{SR,\,Q3}$) if $SR \geqslant SR_M$
QSR= ((SR-SR_M)/$D_{SR,\,Q1}$) if $SR \leqslant SR_M$

QESR= ((ESR-ESR_M)/$D_{ESR\,,Q3}$) if $ESR \geqslant ESR_M$
QESR= ((ESR-ESR_M)/$D_{ESR,\,Q1}$) if $ESR \leqslant ESR_M$

Then the normalized test statistics QSR and QESR were utilized to calculate the H-B score statistic as follows:

Score= max{QSR,QESR} if $R \geqslant R_M$
Score= min{QSR,QESR} if $R \leqslant R_M$

1.5 **Setting the Size Parameter U**

We shall describe our approach we followed to determine the optimum value of U to utilize in the equation that calculates ESR. In order to accomplish this, we varied the value of U from 0.3 to 0.7 by 0.1. With each value of U={0.3,0.4,0.5,0.6,0.7}, we re-ran the HB edit and evaluated the results. Note that in these runs, we used fixed limits to test for statistical significance of the HB score statistics. As we varied the U-parameter and fixed HB score limits in these initial runs, we created scatter plots. In the plots, we looked at the proportion of establishments by industry labeled as “outliers” based on each set of fixed limits and each value of U. We also looked at the symmetry or lack of symmetry of the outliers about the trend line that associates current ADMIN receipts to payroll. After making many runs and investigating many plots, we chose U=0.5. U=0.5 showed a balance between flagging ADMIN receipts outliers and failing to flag plausible ADMIN receipts values. The value U=0.5 produced outliers that were symmetrical about the trend line that relates ADMIN receipts to payroll.

1.6 **Determining HB Score Limits**

To determine if receipts for a single-unit establishment were an outlier with the HB edit, we assigned score cutoffs. Those with an HB score outside of the cutoffs were designated as an outlier. If the score was within the cutoffs, the receipts value was accepted.

To determine score cutoffs, we looked at the relationship between the HB score and whether or not 2012 administrative receipts agreed closely to receipts reported in the 2012 Economic Census. We expected that true outliers were more likely to disagree with survey-collected data.

The Census Bureau conducts the Economic Census every five years, including 2012. For an Economic Census year, we considered administrative receipts to be “close” to Census receipts if:

$0.9*X_E \leqslant X_A \leqslant 1.1*X_E$ where
X_E: Economic Census receipts (EC receipts)

X_A: Administrative receipts

For single unit establishments with reported Economic Census receipts and administrative receipts, we computed the following recode variable Z:

Z=0 if $0.9*X_E \leqslant X_A \leqslant 1.1*X_E$
Z=1 otherwise

The variable Z was set to one for a significant difference (greater than a 10% threshold) between Economic Census and administrative receipts. We formulated two logistic regression models with Z as the response, and NAICS class and HB edit score as the predictors. One model included only observations with negative HB scores. The other model included observations with non-negative HB scores. We found that positive score values were more widely dispersed than negative score values by NAICS class.

Note that HB edit score is a continuous variable, so it is represented in the model by one predictor. NAICS class is discrete, so it is represented by M-1 predictors where M is the number of discrete NAICS Class values. The form of the logistic regression model is:

$$ln\left(\frac{p_i}{1-\ p_i}\right) = \ \beta_0 + \ \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \ \beta_m X_{mi}$$
(1)

Where p_i is the probability that Z_i=1 for observation i.
X_{1i} is the HB edit score for observation i and
X_{2i} through X_{Mi} are the M-1 predictor variables for the M NAICS classes.
The fitted logistic response function can be expressed as:
[3] $\pi_i=[1+\exp(-b`X)]^{-1}$ (2)

The probability that an administrative receipts value disagrees (differs by more than 10%) with Economic Census receipts is:

$P(Z_i=1)=p_i=\frac{\exp(b_0+\sum_{j=1}^{m} b_j X_{ij})}{1+\exp(b_0+\sum_{j=1}^{m} b_j X_{ij})}$

The probability of agreement is $1-p_i$, which reaches a maximum when the HB score is close to zero and decreases as the HB score increases. We defined two sets of HB score limits as follows:

L_{j2}=Low score limit for NAICS class j, outer ($L_{j2}<0$)
L_{j1}=Low score limit for NAICS class j, inner ($L_{j1}<0$)
U_{j1}=High score limit for NAICS class j, inner ($U_{j1}>0$)
U_{j2}=High score limit for NAICS class j, outer ($U_{j2}>0$)

The limits above are based on a percentage of the maximum posterior probability of agreement, which is the max$\{1-p_i\}$. Let P_{50} be 50% of the maximum posterior probability of agreement. Let P_{90} be 90% of the posterior probability of agreement. For negative HB

[3] Neter J., Kutner M., Wasserman W., Applied Linear Statistical Models page 573-574

score values, solving for score in (2), we obtained the inner and outer lower HB score limits as:

L_{j1}=-(1/b_1) [ln(1/P_{90}-1)+b_0+b_j] for NAICS class j
L_{J2}=-(1/b_1)[ln(1/P_{50}-1)+b_0+b_j] for NAICS class j

For positive score values, we obtained the inner and outer upper HB score limits as:

U_{j1}=-(1/b_1)[ln(1/P_{90}-1)+b_0+b_j] for NAICS class j
U_{j2}=-(1/b_1)[ln(1/P_{50}-1)+b_0+b_j] for NAICS class j

If the HB score statistic was outside of the outer limits (score<L_{j2} or score>U_{j2}) then administrative receipts was considered as "non-tab, non-impute" for Census Bureau usage. If the HB score statistic is inside of the outer limits, but outside of the inner limits, administrative receipts was flagged as "use for tabulation, but not for imputation". If the HB score statistic was inside of the inner limits, administrative receipts was flagged as "may be used for tabulation and imputation". The decision to use the 90th percentile and the 50th percentile to define the limits is somewhat arbitrary, but seemed reasonable to Census staff.

Note that with each new year, updated administrative receipts and payroll are loaded into the Business Register. As a result, the seven HB parameters are re-calculated so that flags can be set for the updated data. The HB score limits based on the recode variable Z are updated only once every 5 years at the time of the Economic Census.

1.7 **Production Results**

The Census Bureau implemented the new methodology for flagging administrative receipts in production on its Business Register in 2016, covering data from 2014 and 2015. Table A below shows results for these two years.

The HB score statistics were utilized to flag ADMIN receipts with the following flag types:

- A:Use for tabulation and imputation[4]
- T:Use for tabulation but not for imputation
- U:Do not use for tabulation or for imputation

Table A: **Distribution of Flags**

YEAR	Flag Value	Frequency	Percent
2014	A	3,220,421	82.61
2014	T	553,323	14.19
2014	U	124,695	3.20
2015	A	3,020,699	82.51
2015	T	535,611	14.63
2015	U	104,550	2.86

[4] "Use for tabulation" means the data can be acceptably included with other aggregated data that also includes responses from Census Bureau surveys for the purposes of tabulation. Likewise, "Use for Imputation" means the data can be acceptably included in a commingled universe with Census Bureau survey response data and other administrative data to be used for survey imputation.

First, note that the overall proportion of establishments flagged with 'U' is relatively small as compared to the overall total. Secondly, note that there is reasonable consistency between proportions by flag type ('A', 'T', 'U') over the two years. Table B below presents flag values, frequencies, and percent for a selected number of NAICS classes based on the most recent production run. A review of Table B shows that the percent of cases flagged as 'U' over 2014 and 2015 are reasonably consistent by NAICS class. Note that with the exception of NAICS class 5251, the percent of establishments flagged with 'U' ranges between 1.7 to 4.1 percent. The percent of establishments flagged with 'T' ranges between 12 to 22 percent. The percent of establishments flagged with 'A' ranges from 74 to 85 percent. The HB methodology, as implemented, was successful at flagging a relatively small number of establishments as outliers flagged as 'U'. Moreover, the proportion of establishments flagged as 'A' is a large enough portion of the total number, that there remains a large imputation base. Note that the percent of flagged establishments by flag type for NAICS class 5251 are unique as compared to the proportion by flag type for most other groups. We offer two reasons for this. First, the HB parameters for the four-digit NAICS class (medians and quartiles) are much different. Secondly, the cell size for the group 5251 is quite small relative to other NAICS class groups as displayed in column 3 of Table B.

Table B: **Distribution of Flags by NAICS Class**

NAICS Class	Flag Value	2014 FREQ	2014 Percent	2015 FREQ	2015 Percent
23	A	385,674	80.29	364,864	79.61
23	T	82,285	17.13	81,621	17.81
23	U	12,395	2.58	11,806	2.58
31	A	158,706	85.47	148,578	85.02
31	T	22.599	12.17	21,971	12.57
31	U	4,382	2.36	4,217	2.41
42	A	170,180	74.87	158,065	74.38
42	T	47,949	21.09	45,760	21.53
42	U	9,182	4.04	8,687	4.09
44	A	353,572	84.77	329,279	84.11
44	T	56,430	13.53	55,201	14.10
44	U	7,072	1.70	6,999	1.79
5251	A	469	29.52	454	36.35
5521	T	277	17.43	204	16.33
5251	U	843	53.05	591	47.32
56	A	165,534	81.24	155,903	81.24
56	T	29,867	14.66	28,288	14.66
56	U	8,346	4.10	7,486	4.10

1.8 **Evaluation of HB Score Limits**

We found that the distribution of positive HB score values from the 90th to 99th percentile range from roughly 5 to 40 for most NAICS class groups. The distribution of negative score values from the 90th to the 99th percentile ranged from roughly -2.5 to -14.5. Note that columns 4-6 in Table C and Table D show the respective estimated coefficients based on the logistic regression runs for positive and negative HB scores. A review of these two sets of estimated coefficients in Table C and Table D reinforces our opinion that separate logistic regression models were required for negative and positive HB scores. Note that the estimated coefficient for HB score from column 7 of table C is negative (-0.0515). In contrast, the estimated coefficient for negative values of HB score found in column 7 of Table D is positive (0.1686).

Note that Table C and Table D below display the upper and lower HB score limits as defined in section 1.6. Columns 7 and 8 of Table C and Table D show the respective upper and lower HB score limits based on the ratio of ADMIN receipts to payroll for processing year 2012. A review of these limits for positive (Table C) and negative (Table D) HB score values show that the respective score limits are not symmetric about zero. A review of Table C below shows that the upper outer HB score limits for most NAICS class groups range from 22 to 36. The upper inner score limits range from 3 to 12 for most groups. For negative HB score values, column 8 of table D shows that the lower outer limits vary between -7 to -12. Column 7 of Table D shows that the lower inner limit HB score varies between -1.3 to -3.3.

Recall that the distribution of flags are based on the score limits that are set by two logistic regression models, one for negative values of HB score, and the other for non-negative values of score. Again, the maximum posterior probability that the EC to administrative receipts differ by not more than 10% should attain a maximum at or very near zero (as shown by the plot in Appendix A). In particular, for the NAICS class group 3100, the maximum is 0.8297 as shown in column 3 of Table C. We arbitrarily chose 90% of the maximum to set the inner limits (negative and positive) for HB score. We chose 50% of the maximum to set the outer limits (negative and positive) of the HB score statistic. Reviewing the plot in Appendix A, shows that HB score attains a maximum of about 0.825 at score=0. To obtain an approximation to the HB score upper inner limit from the plot, multiply 0.825*0.90=0.7425. Find the point 0.7425 on the vertical axis and project a horizontal line across to intersect the logistic regression curve. At this point of intersection, if you drop down a perpendicular line to the horizontal axis, you get an estimate of approximately 10 for the upper inner HB score limit. The actual score limit based on a review of column 7 of Table C is 8.58. We see that the plot is consistent with our computational limit. Similarly, using 50% of the maximum for the upper outer limit (0.825*0.50=0.4125) and projecting we obtain an estimate of about 38. The actual upper limit based on column 8 of Table C is 36.24 for NAICS class 3100. The main point here is that the HB score limits based on our computations are consistent with what we see graphically.

Based on the plot in appendix A, as the HB score values increase, the posterior probability that EC and ADMIN receipts agree significantly decreases. For example, at score=60, the probability of agreement between EC and ADMIN receipts is approximately 0.20. For HB score=120, the probability of agreement, decreases to near zero. This reinforces the notion that the HB score reliably predicts outliers.

1.9 **Conclusion**

We were able to successfully use an HB edit for administrative receipts on the Business Register for single-unit establishments. This will provide the Census Bureau with a standardized and centralized source for assessing the quality of these data and for using them in a more consistent manner where their quality is deemed acceptable for use in tabulation and imputation. We did not have success in extending the use of this type of edit to multi-unit establishments. Multi-unit establishments have different data collection units between the IRS and the Census Bureau, and more complex relationships between administrative receipts and payroll.

We devised a set of limits for the HB score statistics that would be consistent over time while flagging only the ADMIN receipts outliers. We exploited reported EC receipts to check the reliability of current- year ADMIN receipts data for the Economic Census year 2012. We know that current-year ADMIN receipts should be close in value to EC receipts for most industries as represented by NAICS class. Based on prior investigations, we found that ADMIN receipts and payroll are well correlated within NAICS class. Based on reviews of the HB score statistics we noted that the distribution of score$\geqslant$0 are much different from the distribution of score<0. Based on all this prior information, we formulated a pair of logistic regression models with a change variable Z as the response and HB score and NAICS class as the predictors.

The Census Bureau is interested in identifying when administrative receipts may be acceptable for a tabulation of receipts, or as an imputation for survey-collected data. In order to accomplish this, we created a pair of limits for both positive and negative values of the HB score statistics. We based those limits on an arbitrarily selected percentage of a maximum probability of agreement between EC and ADMIN receipts for HB score at or near zero. For positive and negative score values and for most NAICS classes, this maximum probability of agreement varies from 0.65 to 0.85. We utilized ninety percent of the maximum probability of agreement (P_{90}) to set the inner limits for HB score statistics. We utilized fifty percent of the maximum probability of agreement to set the outer limits of HB score. As the HB Score Statistic increases in value, the probability of agreement decreases. As the HB score gets increasingly negative, the probability of agreement decreases. We found that the magnitudes of the HB score statistics were reliable in flagging ADMIN receipts outliers.

We found that the methodology met the following overall objectives. First, the overall proportion of establishments by flag type was reasonable. Initially, we expected to flag about one to four percent of establishments as outliers. Secondly, the proportion of establishments by flag type is reasonably consistent for most NAICS classes. The production results showed that the methodology flagged about the same proportion by flag type over time. Moreover, production results showed that the methodology flagged about the same proportion by flag type by NAICS class over time. Finally, the models showed that the magnitudes of the HB score statistics were reliable test statistics to differentiate between good and outlier values of current-year administrative receipts.

Table C: Estimated Coefficients and HB Score Limits for Positive Score Values

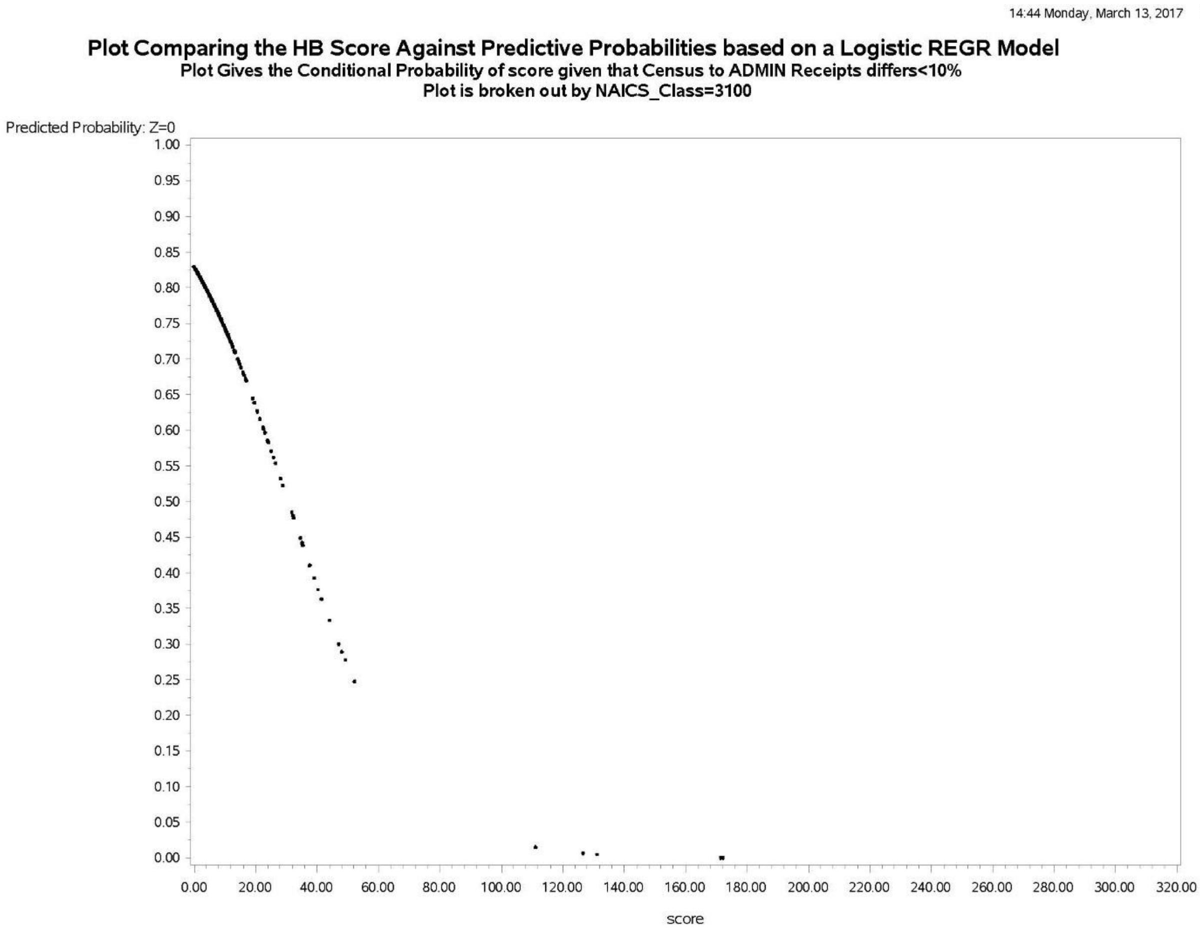
NAICS_CLASS	Predicted PROB. OF No Agreement	Maximum PROB. Of Agreement	Intercept	Estimated Coefficient Of NAICS Class	Estimated Coefficient Of HB SCORE	HB SCORE Upper Inner Limit	HB SCORE Upper Outer Limit
0000	0.2186	0.7814	0.9189	0.2949	-0.0515	6.81	32.18
1100	0.3131	0.6869	0.9189	0.1928	-0.0515	12.22	34.14
2100	0.3801	0.6199	0.9189	-0.4893	-0.0515	3.82	23.86
2200	0.2494	0.7506	0.9189	0.1234	-0.0515	5.99	30.11
2300	0.2919	0.7081	0.9189	-0.0928	-0.0515	5.10	27.70
3100	0.1703	0.8297	0.9189	0.6047	-0.0515	8.58	36.24
4200	0.2318	0.7682	0.9189	0.2190	-0.0515	6.43	31.24
4245	0.2128	0.7872	0.9189	0.3298	-0.0515	7.00	32.62
4247	0.2266	0.7734	0.9189	0.2490	-0.0515	6.59	31.61
4400	0.2202	0.7798	0.9189	0.2856	-0.0515	6.76	32.06
4471	0.2724	0.7276	0.9189	0.00357	-0.0515	5.47	28.74
4800	0.4434	0.5566	0.9189	-0.7513	-0.0515	3.18	21.74
5100	0.2892	0.7108	0.9189	-0.0795	-0.0515	5.15	27.84
5200	0.3600	0.6400	0.9189	-0.4034	-0.0515	4.06	24.63
5251	0.2371	0.7629	0.9189	-0.1900	-0.0515	0.00[5]	23.52
5300	0.3456	0.6544	0.9189	-0.3404	-0.0515	4.25	25.21
5400	0.2793	0.7207	0.9189	-0.0310	-0.0515	5.33	28.36
5500	0.3177	0.6823	0.9189	0.0214	-0.0515	9.24	31.02
5600	0.2902	0.7098	0.9189	-0.0846	-0.0515	5.13	27.79
6100	0.2631	0.7369	0.9189	-0.0510	-0.0515	3.69	27.30
6200	0.2266	0.7734	0.9189	0.2488	-0.0515	6.58	31.61
7100	0.2690	0.7310	0.9189	0.0207	-0.0515	5.54	28.93
7200	0.2214	0.7786	0.9189	0.2787	-0.0515	6.73	31.97
8100	0.2676	0.7324	0.9189	0.0279	-0.0515	5.57	29.01

[5] For this NAICS_CLASS group the estimate was poor with a very large variance

Table D: Estimated Coefficients and HB Score Limits for Negative Score Values

NAICS_CLASS	Predicted PROB. OF No Agreement	Maximum PROB. Of Agreement	Intercept	Estimated Coefficient Of NAICS Class	Estimated Coefficient Of HB SCORE	HB SCORE Upper Inner Limit	HB SCORE Upper Outer Limit
0000	0.1683	0.8317	1.2674	0.3345	0.1686	-3.03	-11.52
1100	0.2582	0.7418	1.2674	-0.2107	0.1686	-2.13	-9.40
2100	0.3715	0.6285	1.2674	-0.7416	0.1686	-1.55	-7.75
2200	0.2108	0.7892	1.2674	0.0558	0.1686	-2.53	-10.39
2300	0.2110	0.7890	1.2674	0.0514	0.1686	-2.51	-10.36
3100	0.1819	0.8181	1.2674	0.2361	0.1686	-2.83	-11.10
4200	0.2803	0.7197	1.2674	-0.3245	0.1686	-1.98	-9.01
4245	0.4635	0.5365	1.2674	-1.1205	0.1686	-1.28	-6.82
4247	0.1755	0.8245	1.2674	0.2808	0.1686	-292	-11.29
4400	0.1793	0.8207	1.2674	0.2537	0.1686	-2.86	-11.17
4471	0.2131	0.7869	1.2674	0.0393	0.1686	-2.49	-10.32
4800	0.2129	0.7871	1.2674	0.0416	0.1686	-2.50	-10.33
5100	0.2165	0.7835	1.2674	-0.0188	0.1686	-2.23	-10.02
5200	0.2403	0.7597	1.2674	-0.1162	0.1686	-2.25	-9.74
5251	0.3100	0.6900	1.2674	-0.4594	0.1686	-1.86	-8.60
5300	0.2739	0.7261	1.2674	-0.2927	0.1686	-2.02	-9.12
5400	0.1729	0.8271	1.2674	0.2977	0.1686	-2.94	-11.35
5500	0.4521	0.5479	1.2674	-1.0718	0.1686	-1.32	-6.94
5600	0.1656	0.8344	1.2674	0.3500	0.1686	-3.05	-11.58
6100	0.1913	0.2087	1.2674	0.1741	0.1686	-2.72	-10.85
6200	0.1475	0.8525	1.2674	0.4873	0.1686	-3.33	-12.17
7100	0.1955	0.8045	1.2674	0.1474	0.1686	-2.67	-10.74
7200	0.1499	0.8501	1.2686	0.4681	0.1686	-3.29	12.09
8100	0.1491	0.8509	1.2686	0.4741	0.1686	-3.30	12.11

Appendix A


14:44 Monday, March 13, 2017 1
Plot Comparing the HB Score Against Predictive Probabilities based on a Logistic REGR Model
Plot Gives the Conditional Probability of score given that Census to ADMIN Receipts differs<10%
Plot is broken out by NAICS_Class=3100
Predicted Probability: Z=0
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0.00
20.00
40.00
60.00
80.00
100.00
120.00
140.00
160.00
180.00
200.00
220.00
240.00
260.00
280.00
300.00
320.00
score