Better Identification of Potential Employment Tax Noncompliance Using Credit Bureau Data

Saurabh Datta, Patrick Langetieg, and Brenda Schafer (IRS Research, Applied Analytics and Statistics) [1]

The U.S. income tax system generally requires employers to deposit payroll taxes using either a monthly or semiweekly schedule, depending on the total payroll tax liability the employer reported during the relevant look back period.[2] Any employer that accumulates $100,000 or more in payroll taxes on any day during its monthly or semiweekly deposit period must deposit the tax by the next business day. A failure to make timely deposits results in a failure-to-deposit penalty up to 15 percent and the employer may be subject to subsequent enforcement actions.

Because the IRS does not know the date on which a business makes payroll, the date of the businesses employment tax return filing sets the timeline for the expected next filing. If a return is not subsequently secured, the business is considered potentially noncompliant. This method of identification is not timely, and it is certainly not proactive. In the coming years, the Electronic Federal Tax Payment System (EFTPS) will be used to analyze the past deposits to identify deviations in the deposit patterns to determine potential noncompliance. Until then, the IRS should explore other early detection methods.

If the IRS can more quickly identify employers that have missed—or may miss—one or more scheduled payroll deposits before these employers encounter further financial difficulties, file for bankruptcy, or close down the business, then there is a higher likelihood of collecting the unpaid taxes. This early intervention may even provide useful guidance to the employer regarding the importance of payroll tax compliance and the availability of payment plans that can help the employer meet their payroll deposit responsibilities.

Previous studies have attempted to identify and examine factors that drive noncompliance in the form of tax delinquency and untimely tax deposits in the payroll tax compliance program. These efforts have been limited to IRS internal administrative data and a few surveys (SB/SE (2011); Hopkins and Su (2017)). Hopkins and Su (2017) concludes that including credit bureau data with the IRS administrative data does not contribute to the predictive accuracy of future noncompliance.

This paper discusses preliminary findings that show that matching a homogenous sample of employers with third-party short- and long-term credit bureau credit scores may be useful to proactively identify potential noncompliant employers.

The following sections highlight the sample design and data description for this study, and then the interaction of short- and long-term credit scores with employer payroll tax compliance. The final section summarizes the preliminary findings and discusses opportunities for further research and extensions of this study.

Sample Design

In a previous study undertaken by Hopkins and Su (2017), a sample of 300,000 employers was drawn to analyze the link between detection of potential payroll tax noncompliance using credit bureau data. The sampled employers varied widely in terms of their business capitalization and state of noncompliance.

[1] The views and opinions presented in this paper reflect those of the authors. They do not necessarily reflect the views or the official position of the Internal Revenue Service.

[2] IRS Publication 15, Employer's Tax Guide (Circular E), p. 25, available at https://www.irs.gov/pub/irs-prior/p15--2017.pdf. For Form 943 filers, see IRS Publication 51, Circular A, Agricultural Employer's Tax Guide, p. 15, available at https://www.irs.gov/pub/irs-pdf/p51.pdf.

A stratified sample was drawn and sent to the credit bureau. The sample segmented businesses based on the number of employees (0 to 10, 11 to 25, 26 to 50, and over 50). They were further segmented by total tax liability (less than $2,000, $2,001 to $6,000, $6,001 to $20,000, and over $20,000). Noncompliance was defined based on two criteria: 1) assessment of a failure to deposit penalty; and 2) an unpaid payroll tax liability after receipt of the first IRS notice. The sampling criteria were further restricted to ensure that the sample was statistically representative of all the required variables over 20 quarters starting from the third quarter of Calendar Year 2010 (2010Q3). Unfortunately, the two noncompliance measures intended for this study turned out to be based on rare events in the drawn sample. Even after adequately representing businesses by number of employees and tax liabilities, assessment of a failure to deposit penalty was detected in just four percent of the sample, and an unpaid tax liability after receipt of first IRS notice was realized in only two percent of the sample. The limited occurrence of predicted noncompliance severely restricted the analysis of underlying issues driving noncompliance.

The limitations of the previous sample provided motivation to draw another random sample of 250,000 employers. The reference point for this study is the fourth quarter of Calendar Year 2014 (2014Q4). The study period consists of the eight quarters immediately before and the eight quarters immediately following 2014Q4.

The study sample was comprised of 70 percent "detected noncompliant employers." A detected noncompliant employer is defined as an employer who received a first notice regarding potentially unpaid payroll taxes at some point during the eight quarters prior to 2014Q4 and whose case ultimately resolved in an assessment of unpaid payroll taxes. The detected noncompliant cases were oversampled to allow us to study this population in greater detail. Note that the remaining 30 percent of the sample is not necessarily payroll tax compliant. They simply were not subject to enforcement action during the eight quarters prior to 2014Q4. The sample was restricted to small businesses with assets below $10 million that filed Form 941. In addition, businesses in the sample must have existed prior to January 1, 2013, to ensure that they had some credit history prior to the study reference date.

The credit bureau matched 160,627 of the 250,000 sampled employers with credit data. Thus, the final sample is comprised of 67 percent detected noncompliant employers and 33 percent other employers.

Data Description

The credit bureau has extensive data, including short- and long-term credit scores, and other credit-risk variables, such as total outstanding balance, Federal and State liens, number of outstanding legal issues, information about credit accounts in collection status, etc. Additionally, the database provides a large collection of firmographic variables, such as industry, business size, location, etc. This study concentrates specifically on the short- and long-term credit scores. The additional variables will be used as an extension to this research.

The short-term credit score predicts the likelihood of defaulting in the next 12 months on a credit obligation that has been past due for more than 91 days. This score makes use of business and consumer variables such as payment history, frequency of payments, and short-term delinquent balances. The score is computed on a scale of 1 to 100, where a higher score is associated with lower risk.

The long-term credit score predicts the probability of bankruptcy or the prospect of defaulting on 75 percent of the credit obligations that are more than 91 days past due. The score is computed using trade, public records, and firmographic data. The primary factors affecting this score are high utilization of credit lines, tax liens and judgments, and bankruptcy filings. As with the short-term credit score, the scale is 1 to 100 with a higher score meaning lower risk.

Industry lenders use both the short- and long-term credit scores when making lending decisions, determining interest rates and risk policies for the businesses (Experian (2016a and 2016b)). The collective use of both the scores provides lenders with important details about the current status of a business and the risk involved in its operations, both in the short- and long-run. The following matrix summarizes the risk categories based on the application of the two credit scores:

TABLE 1. Risk Classification Matrix

Short-Term Risk	Long-Term Risk	
	Low	High
Low	Stable Segment	Medium Risk
High	Slow Recovery	High Risk

SOURCE: RAAS Taxpayer Behavior Lab, May 2017; Experian (2016b).

In Table 1, each category has distinct significance in assessing risk and its potential future implications. Businesses that fall into the Low Short-Term Risk/Low Long-Term Risk category are stable businesses. The businesses in the Low Short-Term Risk/High Long-Term category are able to fulfill their short-term credit obligations but are falling behind in meeting their long-term credit payments. These businesses are considered medium risk. High Short-Term Risk/Low Long-Term Risk businesses are experiencing difficulties in keeping up with their short-term credit obligations but have not been defaulting on long-term credit responsibilities. These businesses are using payment plans to pay off short-term debt, thus they are considered to be in the slow recovery category (Experian (2016b)). Businesses in the fourth category, the High Short-Term Risk/High Long-Term Risk group, are in peril of financial catastrophe.

Following financial industry standards, this paper considers both the short- and long-term credit scores in its analysis. We first explore the separate relationship of the short- and long-term credit scores on the two study segments (detected noncompliant employers and other employers). We then focus the analysis on the interaction of the two credit scores with detected noncompliant employers. In the next section, the paper reports preliminary results from this analysis.

Exploratory Analysis

The credit scores from the credit bureau are classified into five risk categories (low, low-medium, medium, medium-high, and high). For simplification and comparison purposes, the paper clusters the risk categories into two broad categories (low and high) and combines the low and low-medium categories into a single low/low category (see Table 1).

In Table 2, we profile short-term credit score with detected noncompliant employers (referred to as the Detected group) and other employers for three periods: observation period (2014Q4), one year prior to the observation period (2013Q4), and two years prior to the observation period (2012Q4). The table suggests that there is no direct association between a short-term high-risk credit score and payroll tax noncompliance. The probability of being in high versus low doesn't appear to be correlated with Detected cases.

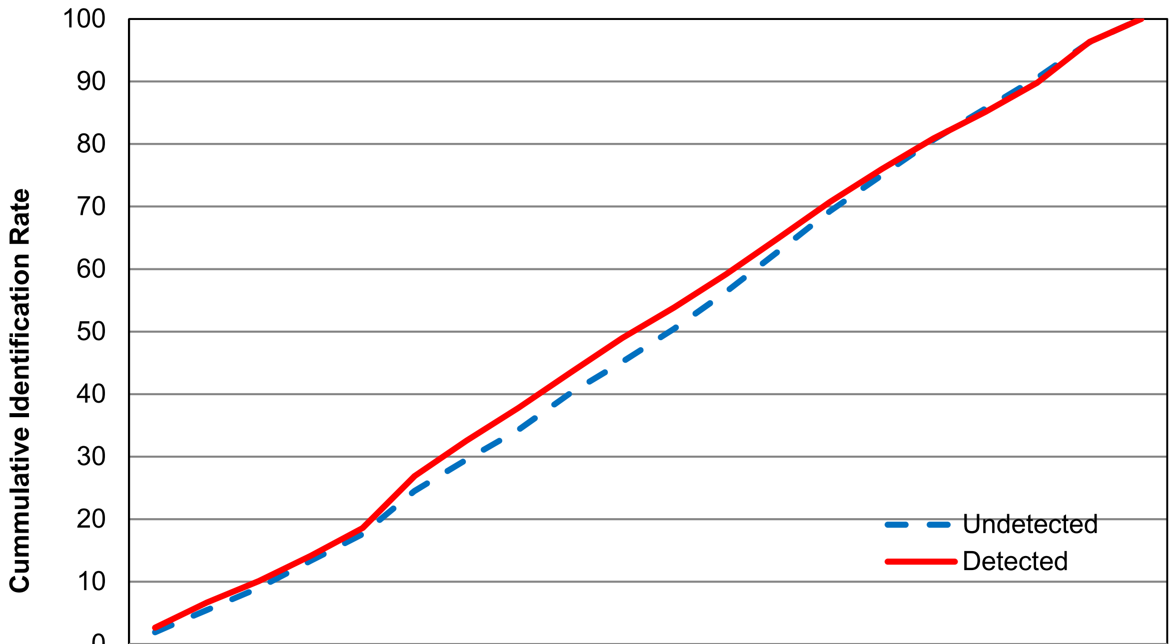
TABLE 2. Distribution of Low and High Short-Term Credit Risk Over Time

Item	2014Q4 (Observation Period)		2013Q4 (One Year Prior)		2012Q4 (Two Years Prior)	
	Detected	Other	Detected	Other	Detected	Other
Low	81.41%	82.58%	81.86%	82.58%	81.56%	81.68%
High	18.59%	17.42%	18.14%	17.42%	18.44%	18.32%
Total	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

SOURCE: RAAS Taxpayer Behavior Lab, May 2017.

Figure 1 ranks the Detected and Other cases in the descending order of short-term risk. The lower ventiles on the X-axis are associated with higher risk; as one moves left to right on the X-axis, the level of risk goes down. The scale of the Y-axis is the cumulative identification rate. The identification rate is defined as the number of Detected or Undetected cases in each ventile. We compute the identification rates separately for both detected and undetected cases. Based on this graph, it appears that the identification rate of Detected cases is slightly better than the Other cases, but the difference does not appear to be material. For instance, 14.2 percent of the Detected cases are within the top two ventiles of highest risk. Similarly, 13.4 percent of the Other cases are within the same range. Therefore, the short-term credit score doesn't provide substantial insight in distinguishing noncompliant employers.

FIGURE 1. Identification Rate of Detected and Other Cases in 2014Q4 Based on Short-Term Credit Score



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Undetected	1.9	5.5	9.0	13.4	17.6	24.5	29.5	34.3	40.2	45.2	50.5	56.4	62.8	69.3	75.0	80.7	85.7	90.6	96.3	100.
Detected	2.6	6.7	10.1	14.2	18.6	26.9	32.6	37.8	43.4	49.0	53.9	59.2	64.9	70.7	76.0	80.9	85.2	89.8	96.3	100.

Ventiles of the Population (Ordered from High Risk to Low Risk)

SOURCE: RAAS Taxpayer Behavior Lab, May 2017.

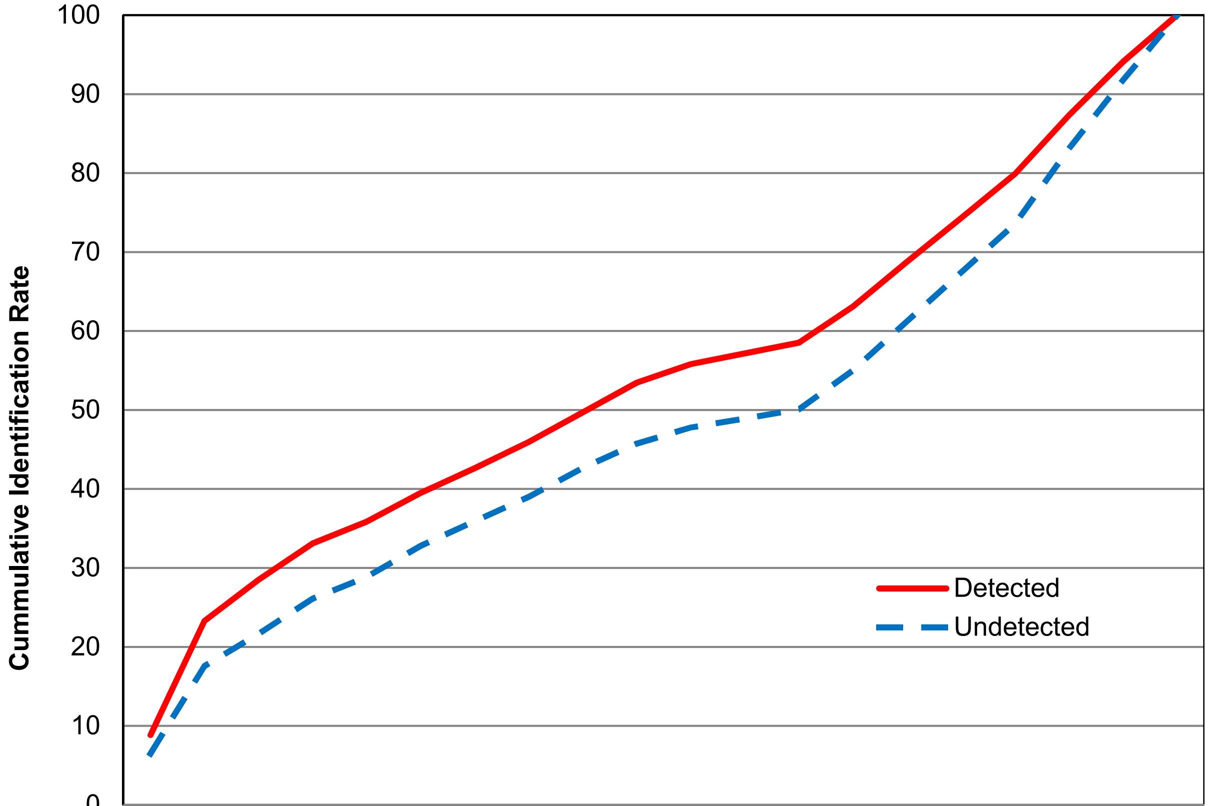
Similarly, Table 3 profiles the Detected and Other cases against the long-term credit score. The identification rate of the Detected cases is higher than that of the Other cases, but there is no clear correspondence between high risk and Detected cases per se.

TABLE 3. Distribution of Low and High Long-Term Credit Score Over Time

Item	2014Q4 (Observation Period)		2013Q4 (One Year Prior)		2012Q4 (Two Years Prior)	
	Detected	Other	Detected	Other	Detected	Other
Low	76.61%	82.58%	75.71%	81.68%	73.76%	79.28%
High	23.39%	17.42%	24.29%	18.32%	26.24%	20.72%
Total	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

SOURCE: RAAS Taxpayer Behavior Lab, May 2017.

FIGURE 2. Identification Rate of Detected and Other Cases in 2014Q4 Based on Long-Term Credit Score



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Detected	8.8	23.3	28.5	33.1	35.9	39.5	42.6	45.9	49.7	53.5	55.8	57.2	58.5	63.1	68.8	74.3	79.9	87.3	94.1	100.
Undetected	6.5	17.6	21.7	26.1	28.9	32.8	35.9	39.0	42.7	45.7	47.8	48.9	50.1	55.0	61.3	67.4	73.6	83.1	91.8	100.

Ventiles of the Population (Ordered from High Risk to Low Risk)

SOURCE: RAAS Taxpayer Behavior Lab, May 2017.

Figure 2 is similar to Figure 1, as it represents the identification rates of detected noncompliant employers and other employers with long-term credit risk. The graph suggests that the long-term credit score provides substantial insight in distinguishing the Detected cases from the Other cases because there is a clear separation of the two lines in Figure 2. For instance, 33.1 percent of the Detected cases are within the top two ventiles of highest risk, but just 26.1 percent of the Other cases lie within the same range.

Next, we study whether using both the short- and long-term credit scores results in better identification of different risk categories than studying them separately. Tables 4A and 4B report the Detected and Other cases, respectively, in the same risk category matrix for the observation period 2014Q4.

Tables 4A and 4B show that a higher percentage of Detected cases are in the High Risk and Medium Risk segments compared to the Other cases. Conversely, a higher percentage of the Other cases fall in the Low Risk and Slow Recovery segments compared to the Detected group. Based on these two tables, it appears that the application of both scores simultaneously can provide better identification of potential payroll noncompliance.

TABLE 4A. Distribution of Detected Cases Across Two Credit Scores Concurrently for 2014Q4

Short-Term Risk	Long-Term Risk	
	Low	High
Low	Stable Segment (67.9%)	Medium Risk* (13.5%)
High	Slow Recovery** (8.8%)	High Risk (9.8%)

* Potential risk of financial instability in the long run.
** Slow payment on credit obligations, but chances of surviving through financial hardship is positive.
SOURCE: RAAS Taxpayer Behavior Lab, May 2017.

TABLE 4B. Distribution of Other Cases Across Two Credit Scores Concurrently for 2014Q4

Short-Term Risk	Long-Term Risk	
	Low	High
Low	Stable Segment (72.4%)	Medium Risk* (10.0%)
High	Slow Recovery** (10.1%)	High Risk (7.5%)

* Potential risk of financial instability in the long run.
** Slow payment on credit obligations, but chances of surviving through financial hardship is positive.
SOURCE: RAAS Taxpayer Behavior Lab, May 2017.

For Table 5, because 66.7 percent of the sample used for this paper is comprised of employers with a detected payroll tax underpayment, a percentage higher than 66.7 percent in any of the cells will imply that the application of credit risk scores results in a better identification of detected cases. Table 5 reports this computation for the 2014Q4 period.

TABLE 5. Detected Noncompliance Rates by Risk Category for 2014Q4

Short-Term Risk	Long-Term Risk	
	Low	High
Low	Stable Segment (65.3%)	Medium Risk* (73.0%)
High	Slow Recovery** (63.7%)	High Risk (72.0%)

* Potential risk of financial instability in the long run.
** Slow payment on credit obligations, but chances of surviving through financial hardship is positive.
SOURCE: RAAS Taxpayer Behavior Lab, May 2017.

The same information from Table 5 is transformed into Table 6A by netting out the overall detection rate from each cell. A positive net percentage suggests improvement in identification due to the use of credit risk scores.

TABLE 6A. Net Improvement of Detected Noncompliance Rates by Risk Category for 2014Q4

Short-Term Risk	Long-Term Risk	
	Low	High
Low	Stable Segment (-1.4%)	Medium Risk* (6.3%)
High	Slow Recovery** (-3.0%)	High Risk (5.3%)

* Potential risk of financial instability in the long run.
** Slow payment on credit obligations, but chances of surviving through financial hardship is positive.
SOURCE: RAAS Taxpayer Behavior Lab, May 2017.

The results on Table 6A suggest that conjoint application of the short- and long-term credit risk scores helps in better identification of Medium- and High-Risk segments. The improvement in identification ranges from 5.3 percent to 6.3 percent. The application of both of the credit risk scores is important to clearly identify businesses that belong to different risk categories and each risk category requires specific treatment since businesses in each category face different challenges (Experian, 2016b). Application of long-term credit risk score only segments the sample into low and high risk categories but fails to fully identify the risk segments namely, stable, slow recovery, medium, and high risk. This limitation is circumvented by applying the short-term risk score in conjunction with the long-term risk score. By only applying the long-term credit risk score the low risk category is under identified by -1.6 percent and the High-Risk category is over identified by 5.9 percent. Whereas, applying both the risk scores, the Medium- and the High-Risk segments are not only identified independently but also overidentified by 6.3 and 5.3 percent, respectively. It is important to segregate the Stable segment from Slow Recovery segment since the latter categories require more urgent attention and potential intervention.

Tables 6B and 6C report the same analysis as shown in Table 5, but for the quarters 2013Q4 and 2012Q4, that is, one and two years prior to the observation point, respectively.

TABLE 6B. Net Improvement of Detected Noncompliance Rates by Risk Category for 2013Q4

Short-Term Risk	Long-Term Risk	
	Low	High
Low	Stable Segment (-1.5%)	Medium Risk* (6.3%)
High	Slow Recovery** (-2.8%)	High Risk (5.3%)

* Potential risk of financial instability in the long run.
** Slow payment on credit obligations, but chances of surviving through financial hardship is positive.
SOURCE: RAAS Taxpayer Behavior Lab, May 2017.

TABLE 6C. Net Improvement of Detected Noncompliance Rates by Risk Category for 2012Q4

Short-Term Risk	Long-Term Risk	
	Low	High
Low	Stable Segment (-1.5%)	Medium Risk* (6.3%)
High	Slow Recovery** (-3.3%)	High Risk (4.8%)

* Potential risk of financial instability in the long run.
** Slow payment on credit obligations, but chances of surviving through financial hardship is positive.
SOURCE: RAAS Taxpayer Behavior Lab, May 2017.

The results from Tables 6B and 6C suggest that the conjoint risk scores identify the Medium-Risk and High-Risk cases better than the observed overall rate of 66.7 percent and much earlier than the observation point. Applying only the long-term risk score to the analysis presented in Table 6B, under identifies the Low-Risk category by 1.7 percent and over identifies the High-Risk category by 5.7 percent. Similarly, on Table 6C, application of the long-term risk score over identifies the High-Risk category by only 5 percent but under identifies the Low-Risk segment by 1.6 percent. However, in both the cases, the long-term High-Risk category fails to distinguish between Medium-Risk and High-Risk segments. This demarcation is important since Medium- and High-Risk segments have very different characteristics and they need different types of intervention or treatment. Earlier detection of potentially noncompliant cases may help the IRS recover unpaid payroll taxes, prevent the accumulation of further unpaid payroll taxes, provide guidance to potentially noncompliant employers on the importance of remaining tax compliant, and information on payment options that will help them stay in compliance.

Tables 7, 8, and 9 examine whether the application of both the short- and long-term credit scores help in identifying potential payroll noncompliance in specific segments (legal issues, credit card balance greater than $5,000, and age of business) of the Detected cases.

TABLE 7. Net Improvement of Detected Noncompliance Rates for Cases Having Legal Issues for 2014Q4†

Short-Term Risk	Long-Term Risk	
	Low	High
Low	Stable Segment (-2.8%)	Medium Risk* (-4.1%)
High	Slow Recovery** (8.3%)	High Risk (20.4%)

† The percentages in parentheses represent the net percentage of detected noncompliance cases with legal issues in excess of the overall rate of 24.5 percent.
* Potential risk of financial instability in the long run.
** Slow payment on credit obligations, but chances of surviving through financial hardship is positive.
SOURCE: RAAS Taxpayer Behavior Lab, May 2017.

In Table 7, legal issues are defined as the existence of tax liens at Federal, State and local tax levels, bankruptcies, credit accounts in collections, and Uniform Commercial Code (UCC) filings on a business as of 2014Q3. Based on this analysis, application of both the short- and long-term credit scores can help identify potential payroll noncompliance among employers with legal issues by 8.3 percent and 20.4 percent respectively for the Slow Recovery and High-Risk segments of the sample. In Table 7, applying only the short-term risk score (and ignoring the long-term risk score) would result in under-identification of the short-term low risk segment by 3 percent and over-identification of the short-term High-Risk segment by 14.3 percent. Moreover, this classification results in combining the Stable and Medium-Risk segments into one Low-Risk category and the Slow Recovery and High-Risk segments into one High-Risk category. This combination of dissimilar categories into one category may not result in efficient intervention and treatment since their characteristics are inherently different.

Table 8, examines Detected cases with an average credit card balance of $5,000 across all the credit channels at the reference point 2014Q4. The results indicate that this analysis could identify potential payroll noncompliance among employers in the Slow Recovery group. This is an important segment because businesses in this category are working to improve their credit rating and may be very receptive to outreach and education about the importance of compliance and payment options.

TABLE 8. Net Improvement of Detected Noncompliance Rates for Cases Having Average Balance of $5,000 Across All Credit Lines for 2014Q4[3]

Short-Term Risk	Long-Term Risk	
	Low	High
Low	Stable Segment (1.8%)	Medium Risk* (-11.3%)
High	Slow Recovery** (14.8%)	High Risk (-3.4%)

* Potential risk of financial instability in the long run.
** Slow payment on credit obligations, but chances of surviving through financial hardship is positive.
SOURCE: RAAS Taxpayer Behavior Lab, May 2017.

[3] The percentages in parentheses represent the net percentage of detected noncompliance cases with an average balance of $5000 in excess of the overall rate of 11.3 percent.

Table 9 studies the association of the credit risk scores with the age of the business. We hypothesize that the newest businesses (less than 3 years old) are more vulnerable to lower credit scores and potential tax noncompliance. This table suggests that application of both credit scores helps in identifying potential noncompliance among new businesses that are in Medium-Risk and High-Risk categories. In Table 9, using just the long-term risk score results in the under-identification of the Low-Risk category by 1.7 percent and over-identification of the High-Risk category by 6.2 percent. But using only the long-term risk score fails to distinguish between the Medium-Risk and High-Risk segments, which have very distinct characteristics. Similarly, applying just the short-term risk score results in under-identification of the Low-Risk and over-identification of the High-Risk categories by 0.3 and 1.5, respectively. This result suggests that it is important to use both risk scores since they are complementary to each other, representing different risk segments.

TABLE 9. Net Improvement of Detected Noncompliance Rates for Businesses With Age Less Than 3 Years for 2014Q4[4]

Short-Term Risk	Long-Term Risk	
	Low	High
Low	Stable Segment (-1.7%)	Medium Risk* (7.5%)
High	Slow Recovery** (-1.2%)	High Risk (4.4%)

* Potential risk of financial instability in the long run.
** Slow payment on credit obligations, but chances of surviving through financial hardship is positive.
SOURCE: RAAS Taxpayer Behavior Lab, May 2017.

Conclusion and Future Research

This paper performs a preliminary analysis of short- and long-term credit scores related to detected noncompliant employers vs. other employers. It provides exploratory evidence that concurrent application of both the short- and long-term credit scores may help in identifying potential payroll tax noncompliance. This may help the IRS in understanding the behavior of these businesses, enable the IRS to take proactive steps to recover unpaid payroll taxes, and educate employers about payment plans that can help them meet their payroll tax obligations.

An extension of this paper will include examining the association between changes in credit score and noncompliance by matching credit bureau data with IRS administrative data. Furthermore, we plan to examine whether a change from the Low-Risk category to the High-Risk category is associated with future noncompliance. A comprehensive understanding of the causality between the two credit scores may also help in understanding how credit risk scores may help in identifying noncompliance with suitable lags after controlling for other factors in an econometric model. Another extension will be to use the Markov Transition Matrix from the credit risk modeling literature. A Transition Matrix structure (Jones (2005); Dobrow (2016)) can be employed to study the relationship between transition of credit risk categories and potential future noncompliance. This structure will help in detecting whether the initial credit risk state or movement from different risk categories in the past is associated with future employment tax noncompliance. Appropriate techniques need to be employed to identify the optimum look back period to effectively predict the timing of potential future noncompliance.

[4] The percentages in parentheses represent the net percentage of detected noncompliance cases (with age of the business being less than 10 years) in excess of the overall rate of 11.7 percent.

References

Dobrow, R. (2016). *Introduction to Stochastic Processes with R.* Wiley.

Experian. (2016a). *Experian Intelliscore Plus V2*. Retrieved from Experian.com: https://www.experian.com/assets/business-information/brochures/intelliscore-plus-v2-product-sheet.pdf.

Experian. (2016b). *Experian Stability Risk Score*. Retrieved from Experian.com: https://www.experian.com/assets/business-information/brochures/financial-stability-risk-score-ps.pdf.

Jones, M. (2005). *Estimating Markov Transition Matrices Using Proportions Data: An Application to Credit Risk.* International Monetary Fund, Monetary and Financial Systems Department. IMF.

SB/SE Research New Carrollton/Richmond. (2011). *Benefit and Cost Evaluation of the Federal Tax Deposit Alert (FTD) Program* (Vol. NCH0149).

Hopkins, C. and K. Su (2017). Supplementing IRS Data with External Credit Report Data in Employment Tax Predictive Models. *IRS Research Bulletin* (elsewhere in this volume).