A statistical diagnosis of customer risk-ratings in anti-money laundering surveillance

Bhojnarine R. Rambharat  
Office of the Comptroller of the Currency  
U.S. Department of the Treasury  
Washington, DC 20024  
e-mail: ricky.rambharat@occ.treas.gov

Andrew J. Tschirhart  
Office of the Comptroller of the Currency  
U.S. Department of the Treasury  
Washington, DC 20024  
e-mail: andrew.tschirhart@occ.treas.gov

May 13, 2014

Abstract

A statistical framework is presented to assess customer risk-ratings used in anti-money laundering (AML) surveillance. We analyze data on a sample of 494 customers from a U.S. national bank where the customers are rated from Low to High over 13 time periods. We model these ratings using an ordinal panel data regression framework with random effects, utilizing a set of covariates provided by the bank. We derive the log-likelihood of the model and provide the maximum likelihood estimates (MLEs) of the model parameters. Our findings unveil key policy-related insights, based on the statistical model, about AML surveillance. We provide statistical evidence to support more granular monitoring of highly suspicious customers, which could optimize finite resources in bank operations. Furthermore, we provide two applications using these data, one concerning predictive inference and the other about log-linear modeling. Our analysis provides an approach to diagnose potential limitations with real-time AML surveillance systems. We argue that statistical diagnosis in AML surveillance has invaluable benefits within the micro-sphere of a single financial institution, and, more importantly, that these benefits extend to important public policy issues confronting the global community.

Key words: Ordinal panel data, Regression, Investigation event, Predictive inference, Log-linear modeling, Proportional hazards, Financial crime policy.

1 Introduction

A major policy-related matter in the global community concerns efforts to foil criminal activity in financial markets. Given the developed capacities of modern technologies, channeling criminal funds through financial institutions can transpire seamlessly. Financial proceeds from criminal
transactions can facilitate, among other things, the narcotics trade, human-trafficking, and even terrorism. Moreover, financial crime does not recognize geographic boundaries, hence its effects are widespread and its consequences are detrimental. Timely detection of financial crime helps to maintain the principles of safety and soundness that sustain financial institutions. Therefore, to be effective, surveillance of financial transactions must stymie criminal initiatives before communities are adversely impacted.

Money laundering occupies a central position within the sphere of financial crime. Indeed, key regulations enacted within the past 50 years are designed to detect and prosecute money laundering activities in financial institutions. In the United States, the Bank Secrecy Act (BSA) of 1970 and the Money Laundering Control Act of 1986 are two notable examples of important statutes meant to combat money laundering crimes. A collection of U.S. regulators published official guidance on how to assess BSA-related compliance activities at a given financial institution (Federal Financial Institutions Examination Council, 2010). Following the terrorist attacks on September 11th, 2001, the USA PATRIOT Act provided for additional legislation to combat money laundering, and it gave substantial authority to the U.S. Financial Crimes Enforcement Network (FinCEN) to prosecute instances of money laundering. Similar legislative actions were taken by additional governments, one example being the United Kingdom’s passage of the Anti-terrorism, Crime and Security Act of 2001. These legislative acts serve as blueprints on a “macro-level” to combat money laundering; however, equally significant is the “micro-level” anti-money laundering (AML) surveillance of customers and transactions.

The organizational structure of an AML compliance unit is outlined by the USA PATRIOT Act. On a micro-level, the consequences of a sub-standard AML compliance program are severe as evidenced by the repercussions faced by Riggs Bank in 2004, and Standard Chartered Bank (2012) and HSBC Bank in 2012. Apart from sizable fines that could be levied for failing to comply with AML regulations, the damage to the reputation of a financial institution could be irreparable. Furthermore, as suggested in a report of the U.S. Senate Permanent Subcommittee on Investigations (2012) regarding inadequate AML controls at HSBC, deficiencies in AML compliance programs
could also have severe, negative consequences for national and, indeed, global security.

The “nuts and bolts” of AML surveillance rests within a bank’s AML compliance unit. Financial transactions can be reviewed using manual or automated methods, or a combination of both. This article will address quantitative aspects of monitoring and surveillance at the automated level. Specifically, we will focus on the role of statistical methodology in assessing the quality of an automated monitoring system used for AML surveillance. During his Presidential address to The Royal Statistical Society in 2008, Professor David Hand discussed the “forensic role of statistics” as an instrument to help uncover criminal behavior in money laundering, among other areas, and he argued that these types of statistical applications could benefit the discipline (Hand, 2009). Indeed, it could be argued that correct application of statistical tools in the AML setting is an important aspect of compliance initiatives designed to benefit society.

The role of statistics in fraud, and particularly in AML surveillance, can be measured through evidence in the recent academic literature. Bolton and Hand (2002) present a through review of “statistical fraud” and the value of statistical tools to uncover general types of fraud events. Regarding money laundering, Bolton and Hand (2002) characterize it in terms of three steps: i) placement, which involves the introduction of criminal funds into the financial system, ii) layering, which concerns efforts to conceal criminal transactions beneath ostensibly legitimate ones, and iii) integration, which entails combining funds from criminal and legal transactions. The authors describe the use of rigorous statistical tools to detect suspicious patterns in financial transactions. Sudjianto et al. (2010) outline key challenges for the role of statistics in financial crime data analysis. Additionally, the authors describe the role of sophisticated classification algorithms, including the use of decision trees and support vector machines, to detect criminal activity in financial data. One important point raised in Sudjianto et al. (2010) is that financial crime is a “rare event,” and this could be true for money laundering cases. Therefore, developers of statistical algorithms for analyzing AML data should recognize if rare event effects prevail and leverage robust analytical strategies.

Deng et al. (2009) presents a specific example of statistical methods for AML data analysis. The
authors propose an “active learning through sequential design” algorithm to prioritize and investigate indicia of criminal activity. The authors demonstrate that their approach works favorably relative to other classification schemes. Another challenge in AML data analysis concerns how to uncover potential networks of criminals. The concealment of illicit funds often depends on the cooperation of a group of criminals since monetary proceeds can be easily channeled across national borders. Furthermore, the movement of illegitimate monies is easily facilitated given the technological instruments of modern society. Heard et al. (2010) address the use of Bayesian methods to uncover anomalous behavior in social networks, and their methodology could find application in the effort to expose networks of criminals in the global financial system.

A related application to prioritizing transactions in AML work streams concerns risk-ratings. While there are several varieties of risk-ratings, arguably risk-ratings of customers (or entities) within a financial institution are the most common. A recent report by The Financial Services Authority (FSA) of the U.K. discusses how to address risk in monitoring high-risk customers (The Financial Services Authority, 2011). Typically, a range of risk-ratings with categorical labels such as “Low,” “Medium,” or “High” are utilized to classify customers. In principle, ratings could be assigned manually or through an automated process. Irrespective of how the ratings are developed, they are designed to help focus attention on potentially higher-risk entities. Moreover, AML analysts who assign ratings might update their beliefs about customers with the arrival of new information, either from internal or external sources. Yet, it may also benefit risk prioritization if the customer risk-ratings are somehow diagnosed in a quantitative sense. One could leverage established statistical methodology as diagnostic tools of customer risk-ratings to better inform the relevant bank personnel who use risk-ratings as part of AML surveillance efforts.

The customer risk-rating spectrum influences the type of statistical methodology applicable in a given context. For example, if the ratings are binary (low or high), then logistic regression or a proportional odds model may be appropriate. Generally, if ratings cover a broad spectrum ranging from low to high, then a more sophisticated type of a generalized linear model would apply. Since ratings are typically ordered in terms of severity of risk, models that accommodate an ordinal
outcome would likely accommodate a meaningful analysis of customer risk-ratings. Furthermore, there are different features of risk-rating data to consider. For example, risk-ratings could evolve for a single customer over time. Additionally, ratings might involve a sample of customer ratings at one time point. The more general situation combines the two aforementioned cases where a collection of customers are analyzed over time, resulting in a panel data set. The work in Hamerle and Ronning (1995) is an excellent review of statistical methods used in panel data analysis where qualitative variables (e.g., customer risk-ratings) are studied. Hamerle and Ronning (1995) present several modeling approaches for different types of qualitative data.

Our modeling strategy adopts the framework outlined for panel models of ordinal data with random effects in Hamerle and Ronning (1995). We use this framework to model customer risk-ratings produced by an AML surveillance monitoring system at a bank. It is noteworthy to point out that our intent is to choose an illustrative modeling strategy. Alternatives to the ordinal panel data framework may also be relevant to this analysis. There are numerous examples of established research that illustrates the application of ordinal models, some of which may not necessarily have a strict panel component, but are nonetheless useful for understanding of the statistical modeling objective in our present research problem. For instance, Jansen (1990) uses an ordinal modeling framework in an agricultural setting where damage to plants caused by fungus is studied. The author presents a thorough derivation of the approach used to optimize the statistical model, and the interpretation of parameter estimates allows for understanding of the basic problem. A further example arises in the work of Johnson (1996) where the author describes a Bayesian approach to analyze ordinal data in an essay grading example. A solid methodological discussion of random effects as incorporated in ordinal regression is described in Tutz and Hennevogl (1996). Additionally, a sophisticated modeling framework for ordinal panel data concerning mixed Markov models is analyzed in Böckenholt (1999) where an ordinal Markov chain setup for panels with a large number of time points is studied.

We estimate our ordinal panel data model with random effects using a data set consisting of a random sample of customers over a 13-period time horizon from a U.S. national bank. The raw
data provide ratings on three levels: Low, Medium, and High. We further subdivide the “High” rating into “High I” and “High II” in order to differentiate the effects of external and internal sources of information to an AML compliance unit, and we formally test the hypothesis that “internal intelligence” ratings higher in severity relative to “external intelligence.” Furthermore, we model our data using a set of key covariates that includes information on initial risk-rating profiles, the line-of-business (LOB) subdivisions, law enforcement categories, and internal risk indicators. Additionally, since our panel is somewhat “narrow” as we have a relatively small number of time points, we incorporate time-dummy indicator variables to study temporal effects on the customer risk-rating methodology. We obtain the maximum likelihood estimates of our model parameters, and we analyze the results of various hypothesis tests about specific parameters to enhance our understanding of the variables used in the proposed model.

Our paper is organized as follows. Section 2 describes the public policy context of our problem. In section 3, we provide an overview of our data set. Note that we mask all variables in our effort to conceal any sensitive information. Our objective is to reveal the potential benefits of a careful statistical analysis of these data for the monitoring of AML risk at a financial institution. Section 4 describes the ordinal panel data model that we use for the purposes of statistical analysis. Section 5 presents the estimation results of our model along with our interpretation of key parameters. We also examine the role of some predictive analytics that compare the risk-ratings from our statistical model to the observed risk-ratings. Section 6 describes an application where we also compare model output to observed data; however, we make this comparison with respect to a signal of criminal activity. Section 7 discusses some public policy implications of our work, particularly in the global effort to combat money laundering. Finally, section 8 concludes with a discussion of the key ramifications of AML surveillance for global communities and discusses potential areas of future research in quantitative AML topics. (Section A collects all supplementary material to our analysis.)
2 The public policy context

The laundering of illicit funds has negative consequences such as drug trafficking or terrorist financing, and these could affect the quality of life for ordinary citizens. This context creates an opportunity for public policy officials to charter new initiatives designed to curb money laundering activities. Moreover, effective AML policies might have an impact on areas such as financial inclusion, which involves the provision of financial services to low-income communities, which are particularly vulnerable to money laundering schemes.

On a supranational level, organizations such as the United Nations (UN), the International Monetary Fund (IMF), and the World Bank collect vital information from various member nations on AML surveillance issues. Additionally, regional agencies play an important part in shaping AML public policy initiatives. The Inter-American Development Bank (IADB) in Latin America and the Caribbean, the African Development Bank in Africa, and the Asian Development Bank in Asia are all actively involved in important AML policy-making and implementation for the respective countries under their purview. Indeed, some of these countries are viewed as havens for money laundering, so these institutions must ensure that effective policies are executed to mitigate the impact of criminal activities.

The Financial Action Task Force (FATF) is another example of a key multilateral body that seeks to shape AML public policy problems. The FATF published “The FATF Recommendations” (cf. Financial Action Task Force (2012)) to help national governments enhance AML compliance efforts. In particular, FATF recommends that countries set up a national financial intelligence unit (such as FinCEN in the United States) to collect financial intelligence from relevant financial institutions and disseminate information to law enforcement authorities. Additionally, some national financial intelligence units write “AML regulations” meant to ensure the safety of their citizenry. A cursory scan of media outlets will reveal that some national governments need to make significant strides to implement certain aspects of the FATF guidelines, and severe measures can be levied against a government for non-compliance, including restrictions on access to global financial systems. In fact, the FATF website states that it is “a ‘policy-
making body’ which works to generate the necessary political will to bring about national legis-
slative and regulatory reforms” in order to combat money laundering and terrorist financing (cf.
http://www.fatf-gafi.org/pages/aboutus/).

Arguably, statistical science can add value to the work of the entities mentioned above since
actual numbers (or metrics) can help attest to the value of a given policy directive. There is an
important role for statisticians as they can interpret the data in the context of policies that may be
derived from guidelines issued by such entities as FATF. Additionally, AML public policy issues
also entail a significant amount of subjectivity, and the blend of subject-matter expertise and quan-
titative analysis can provide stronger arguments for or against an initiative. One aspect of the FATF
guidelines is the notion of risk assessment or a risk-based approach. The risk assessment process
is a core part of AML monitoring and surveillance, particularly where quantitative methods might
apply. The ratings that banks assign entities is one example of such an assessment, and we now
turn to a discussion of the customer risk-rating data we use for this analysis.

3 Customer risk-rating panel data

The following sub-sections describe the customer risk-ratings data that we use in this analysis.
As these are sensitive data, we are not able to fully disclose all characteristics of the data set in
detail. We manipulate certain data elements to protect the identity of the bank, the customers, and
the bank’s risk-rating mechanism. Note that parameter estimates and conclusions are based on the
model we design and is not meant to reflect conclusions about the actual system developed by the
bank.

3.1 Overview

Our data consist of periodic information on risk-ratings and associated covariates for a random
sample of 494 customers from a U.S. national bank over 13 time periods. The original sample
contained 500 customers, but we dropped 6 customers due to data integrity issues. As part of our
effort to mask certain facets of the data, we do not disclose the period. Thus, a period could refer to a variety of time units (e.g., hour, day, week, bi-week, month, quarter, etc.), but each period is the same length. The underlying population consists of bank customers who were rated either Low, Medium, or High. The customers come from four of the business departments within the bank. We will define a suspicious customer as one who is tagged with a Medium or High risk-rating (i.e., a non-Low rating) for at least one period during the sample time frame. All of the 494 customers in our sample are “suspicious,” and note that there are no new entrants or exits in the sample data over the 13 periods. Thus we have a balanced panel data set for our analysis. We implement our core data management procedures using the SAS® platform.

Active customers are monitored by the bank’s AML surveillance unit on a periodic basis for a collection of traits and behaviors indicative of AML risk. We define these AML risk indicators, which include information from government and law enforcement agencies and from other business departments within the bank, as “risk conditions.” The AML surveillance unit maps this information to a series of binary and categorical variables that are used as inputs to the bank’s customer risk-rating mechanism. The risk-rating mechanism can be conceptualized as a tool that uses algorithmic logic, among other functionalities, to assign a rating to a given customer or entity. The data collection and risk-rating tool are run periodically, refreshing the risk-rating of the bank’s customers. Hence, the AML surveillance unit monitors its customers on a regular basis for any notable changes in its customers’ risk profiles.

3.2 Risk-rating outcomes

The bank’s risk-rating mechanism is a deterministic set of logical tests/rules that map individual risk conditions to a customer-level risk-rating. At a specific point in time, the actual risk-rating that the tool outputs is a categorical variable with three possible values: Low(L), Medium(M), or High(H). A time-series plot of the risk-ratings for four illustrative customers appears in Figure 1. Each risk condition is mapped to a risk score, which indicates the severity of the risk condition. Risk scores for individual risk conditions are then mapped and aggregated to a customer risk-
rating. If a customer has multiple ratings, the highest rating is typically assigned. Customers who are not flagged for any risk conditions are assigned a Low rating. In short, roughly speaking, the risk conditions serve as inputs, the risk scores are weights assigned to the risk conditions, and the risk-ratings are the output.

Figure 1: Customer-based time-series.

3.3 Proposed statistical model

We differentiate between internal versus external sources of information as it relates to the customer risk-ratings. The reason for this differentiation is to enhance the statistical modeling strategy we employ to analyze these data. We separate the High category as defined by the bank into a “high rating derived from external risk conditions” (High I) and a “high rating derived from internal risk conditions” (High II). We remind the reader that, as we discussed in section 3.1, our population from which the sample was drawn only includes customers who had a risk-rating of Medium (M) or greater for at least one period during the sample time horizon, and we label these customers as “suspicious.” Table 1 presents the distribution of customer risk-ratings, by period, taking into
account our adjustment for internal versus external effects in the High category.

Table 1: Frequency distribution of expanded customer risk-ratings (by period).

| Period | Low (L) | Medium (M) | High I | High II |
|--------|---------|------------|--------|---------|
| 1      | 104     | 5          | 382    | 3       |
| 2      | 297     | 19         | 173    | 5       |
| 3      | 291     | 29         | 166    | 8       |
| 4      | 267     | 33         | 175    | 19      |
| 5      | 75      | 27         | 327    | 65      |
| 6      | 65      | 28         | 330    | 71      |
| 7      | 39      | 45         | 307    | 103     |
| 8      | 28      | 12         | 271    | 183     |
| 9      | 20      | 9          | 261    | 204     |
| 10     | 10      | 11         | 251    | 222     |
| 11     | 1       | 7          | 235    | 251     |
| 12     | 3       | 0          | 214    | 277     |
| 13     | 4       | 0          | 193    | 297     |

In addition to the risk conditions and risk-ratings, the panel data set contains additional, supplementary data that could be used as covariates to establish a statistical model of the risk-ratings. First, we have information on the business department(s) within the bank with which each customer transacts. Customer accounts across business departments are linked together. Therefore, if an individual customer interacts with more than one business department, this will be known to the AML surveillance unit. Additionally, business department information is time-variant – i.e., customers can open or close accounts with different business departments during the sample time horizon. Second, we define an “on-board indicator” as the baseline risk-rating assigned to a customer immediately before the sample period begins. This initial rating reflects external intelligence only and does not include any information from the AML surveillance unit. In some sense it is an expression, albeit a crude one, of a prior belief about the nature of AML risk that a given customer poses to the bank. Third, we categorize all risk conditions into four major groups that are mapped to indicator variables in our data, and we refer to these four groups as “law enforcement actions.” This allows the use of information from the risk conditions as an input to our ordinal regression model without using the risk conditions directly. Thus, we are able to mask the precise logic be-
hind the risk conditions, which, arguably, is sensitive information that ought not to be revealed. Finally, we control for latent time effects using a series of time indicators (“dummies”). Since this is a “narrow panel,” we are able to make some inference using the time indicators. However, a panel with more extensive data in the time dimension would require more sophisticated time-series modeling to rigorously address time effects.

3.4 Data structure

The final panel data set contains variables on risk conditions, investigation events, and risk-ratings for each customer \(i = 1, \ldots, N\) in each period \(t = 1, \ldots, T\). Risk conditions are aggregated into four law enforcement actions and are mapped to four indicator variables \(Action I, \ldots, Action IV\). Note that we aggregate the risk conditions in order to protect the confidentiality of the bank’s methodology. Furthermore, we include a full series of time indicators (“time dummies”) for each period, \(Period 1, \ldots, Period 13\), and four indicators \(Dept. A, \ldots, Dept. D\) for whether the customer has an account with a given business department. Table 2 gives the breakdown of risk-ratings by business department. We also include an indicator variable for the initial on-board indicator at Period 1 assigned by the external business departments outside of the AML surveillance unit. This binary risk-rating variable, which is either “Low” or “High I” and serves as a quasi-measure of prior beliefs, allows us to see how AML internal intelligence updates the original, externally determined risk-rating. A full list of covariates is provided in Table 3.

Finally, we remark on some data integrity concerns that we believe merit disclosure. For 191 customers, certain risk conditions erroneously dropped out of the data due to issues with the underlying database platform and/or reporting errors, but we are able to repair the error and interpolate the correct information using the information supplied by the bank. Furthermore, 403 customers have at least one period of data history where business department information was included without any accompanying risk conditions. Since we have data on 494 customers over 13 time periods, this results in \(494 \times 13 = 6,422\) customer-time pairs. Since this information does not appear to be supplied systematically, and since customers can and do interact with different lines of busi-
ness over time, we are concerned that these observations might bias the marginal contribution of business department information in our ordinal regression model. Therefore, the final dataset includes business department information for only those periods where at least one risk condition was concurrently reported.

Table 2: Risk-ratings by business department.

|                  | Low | Medium | High I | High II |
|------------------|-----|--------|--------|---------|
| No Department    | 1204| 0      | 0      | 0       |
| Dept. A          | 0   | 16     | 1186   | 610     |
| Dept. B          | 0   | 7      | 375    | 124     |
| Dept. C          | 0   | 176    | 121    | 484     |
| Dept. D          | 0   | 10     | 1190   | 185     |
| Multiple Departments | 0 | 16     | 413    | 305     |

4 A statistical model of customer risk-ratings

We now discuss the formal statistical modeling framework, the derivation of the log-likelihood, and the optimization methods that we use to analyze the customer risk-rating data set. We use a maximum likelihood estimation approach to optimize the log-likelihood. In addition to making inference from the parameter estimates, we also extend the optimization results to study some applications of core statistical methodologies applied to these AML customer risk-rating data.

4.1 Ordinal panel data regression framework

We use an ordinal panel data regression framework with random effects to model customer risk-ratings. Let

\[ y_{it}^* = x_{it} \beta + \alpha_i + \epsilon_{it} \quad (1) \]

where, \( y_{it}^* \) is the customer risk-rating for customer \( i \) at time \( t \) on a continuous scale, \( x_{it} \) is the \( i \)\(^{th} \) row vector of the covariate matrix at time \( t \), \( \beta \) is a coefficient vector of length \( q \), where \( q \) is the number of covariates, \( \alpha_i \) is a \( N(0, \sigma^2_{\alpha}) \) random variable that represents the random effect, \( \epsilon_{it} \) is an
independent and identically distributed (iid) sequence of \( N(0, \sigma^2) \) random variables that represents
the error disturbance for the data, which is assumed to be independent of \( \alpha_i \), and \( i = 1, \ldots, N \),
\( t = 1, \ldots, T \) are the customer and time indices, respectively. Typically, for identifiability reasons,
\( \sigma^2 = 1 \) and the intercept term is set equal to 0. Since we place a distribution on \( \alpha_i \), we can think
of the foregoing setup in terms of a two-level hierarchal model where the \( \alpha_i \) is the first level of the
hierarchy, and \( y^*_it \) conditional on \( \alpha_i \) is the second level. Note that it is important to bear in mind
that \( y^*_it \) is unobserved.

Next, following Hamerle and Ronning (1995), define the observed customer risk-ratings on a
discrete, ordinal scale such that,

\[
y_{it} = m \Leftrightarrow \theta_{m-1} < y^*_it \leq \theta_m,
\]

where \( m = 1, \ldots, K \). Recall that the \( K = 4 \) customer risk-ratings are as described in section 3
(Low, Medium, High I, and High II). The parameters indexed by \( \theta_m \) represent thresholds (or cut-
offs) that map the latent continuous ratings, as described in equation (1), to observed ratings on a
discrete scale. Note that in our specification, \( \theta_0 = -\infty \) and \( \theta_4 = \infty \), which implies that we only
need to estimate \( \theta_1, \theta_2, \) and \( \theta_3 \). Thus, the full parameter vector is denoted by

\[
\Lambda = (\beta_1, \ldots, \beta_q, \theta_1, \theta_2, \theta_3, \sigma_\alpha)'.
\]

Now using equation (2), set

\[
\Pr (y_{it} = m | x_{it}, \alpha_i, \Lambda) = F (\theta_m - x_{it}/\beta - \alpha_i) - F (\theta_{m-1} - x_{it}/\beta - \alpha_i),
\]

where \( F(\cdot) \) is a cumulative distribution function (cdf). In our modeling of \( \epsilon_{it} \), \( F(\cdot) \) is the standard
Normal cdf denoted by \( \Phi(\cdot) \). Alternative specifications for \( F \) would result in alternative modeling
frameworks, many of which could be adopted for analysis of our customer risk-ratings data. Our
primary concern here, though, is not a specific modeling paradigm, but, rather, a framework that
would allow us to analyze these data to reveal important policy insights useful for AML surveillance.

It is instructive at this point to discuss the covariate set in our proposed ordinal panel data regression model. We have a total of $N = 494$ customers in our data set, and we estimate our model using $T = 12$ periods worth of data. An additional 13th period of customer risk-rating data will be used for an out-of-sample prediction exercise. Recall from section 3 that we have data on (i) initial, binary risk-ratings, (ii) four business department identifiers, (iii) 4 law enforcement action categories, and (iv) 12 time indicators over the data period. Therefore, the dimension of the coefficient vector, $\beta$, is $q = 18$ after adjusting for baseline indicators in the regression. We also need to estimate the variance of the random effects, $\sigma_\alpha^2$ as well as the cutoff parameters $\theta_1$, $\theta_2$, and $\theta_3$. Hence, we have a total of 22 elements in our parameter vector, $\Lambda$. We now discuss how to derive the log-likelihood of the model, and also how to compute maximum likelihood estimates.

### 4.2 Log-likelihood derivation

The likelihood function for our proposed ordinal panel data model, as a function of the parameter vector $\Lambda$, is given by

$$L(\Lambda) = f(y_{11}, \ldots, y_{it}, \ldots, y_{NT}|x_{11}, \ldots, x_{it}, \ldots, x_{NT}, \Lambda),$$

where $1 < i < N$ and $1 < t < T$. Using first principles, we derive an expression for the likelihood function. We make two simplifying assumptions: i) the risk-ratings are independent across customers, and ii) for a given customer, conditional on $\alpha_i$, the risk-ratings are independent across time. The former assumption is simplifying because it is conceivable that customers could be related, and the nature of those relationships could be non-trivial. However, no particulars of such relationships are disclosed in our original data set, most likely because of the sensitivity of the data elements. Nonetheless, should an understanding of potential relationships between customers become obvious, it ought to be taken into account as part of the modeling framework.
Now, based on equation (3), we set
\[
\Pr (y_{it} = m | x_{it}, \alpha_i, \Lambda) = \Phi (\theta_m - x_{it} \beta - \sigma_\alpha \tilde{\alpha}_i) - \Phi (\theta_{m-1} - x_{it} \beta - \sigma_\alpha \tilde{\alpha}_i),
\]
where
\[
\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du
\]
is the cdf of a \( N(0, 1) \) random variable, which is used to model \( \epsilon_{it} \) in equation (1). The quantity \( \tilde{\alpha}_i \) is a standardized \( N(0, 1) \) random variable where \( \tilde{\alpha}_i = \alpha_i / \sigma_\alpha \). In order to derive the expression for \( L(\Lambda) \), we need to integrate out the random effect, \( \tilde{\alpha}_i \), using its probability density function (pdf), denoted by \( f(\tilde{\alpha}_i) \), which is the pdf of a \( N(0, 1) \) random variable. Under the assumption of independence of the customer risk-ratings across customers as well as across time for each customer, we have
\[
L(\Lambda) = \prod_{i=1}^{N} \int_{R} \prod_{t=1}^{T} \prod_{m=1}^{K} \left[ \Pr (y_{it} = m | x_{it}, \alpha_i, \Lambda) \right]^{1 \{y_{it} = m\}} f (\tilde{\alpha}_i) d\tilde{\alpha}_i,
\]
where we refer to the expression in equation (4) for \( \Pr (y_{it} = m | x_{it}, \alpha_i, \Lambda) \). Note that for each customer, we first take the product across the \( K \) categories using the expression in equation (4), which is raised to the power \( 1 \{y_{it} = m\} \) since we want to capture the probability of being in a specific category. The cases where the indicator variable is equal to 0 do not contribute to the product across the \( K \) categories. Next, we take the product across time using the independence of the ratings across time. Subsequently, these two products inside the integral are integrated with respect to the distribution of the standardized random effect, \( f(\tilde{\alpha}_i) \). Finally, the outside product is formed using the independence of the risk-ratings across customers. Therefore, we can express the full log-likelihood as
\[
l(\Lambda) = \log L(\Lambda) = \sum_{i=1}^{N} \log \left( \int_{R} \prod_{t=1}^{T} \prod_{m=1}^{K} \left[ \Pr (y_{it} = m | x_{it}, \alpha_i, \Lambda) \right]^{1 \{y_{it} = m\}} f (\tilde{\alpha}_i) d\tilde{\alpha}_i \right),
\]
where, again, we reference equation (4) for \( \Pr (y_{it} = m | x_{it}, \alpha_i, \Lambda) \) in the above expression. We
next describe the optimization methods that result in maximum likelihood estimates of the model parameters.

4.3 Model optimization methods

An important objective of this data-analytic exercise is to estimate the parameters of our proposed model given in equations (1, 2, 4) using the log-likelihood function derived in equation (6). We implement an MLE-based optimization routine to estimate model parameters. The complication with estimation arises through the random effect, \( \alpha_i \). The random effect introduces the integral in equation (6), which requires numerical techniques for optimization.

We first estimate the log-likelihood using an MLE approach based on a mean-variance adaptive Gauss-Hermite quadrature. The software Stata® is used to implement the optimization procedure. The command `xtoprobit` is used to execute the optimization, which results in MLE-based point estimates of \( \Lambda \). The parameters of the asymptotic Normal distribution for the parameter estimates are also reported as they are central to some key inferences that are made. Note that our model does not include an intercept parameter, as this is incorporated into the cut-off parameters \( \theta_1, \theta_2, \) and \( \theta_3 \). We now turn to a full discussion of our empirical results in the next section.

5 Empirical diagnosis of customer risk-ratings

We report parameter estimation results for the proposed ordinal panel data regression model of customer risk-ratings. While the actual ordinal panel data model of the customer risk-ratings is straightforward, the interpretation of the parameter estimates can be delicate as it has important consequences for AML surveillance, particularly for policy-makers in this sphere. Hence, we are more concerned with the “diagnosis” of the output from our modeling and analysis of these data.
5.1 Estimation of ordinal panel data model

The results in Table 3 include the maximum likelihood estimates, standard errors, z-statistics, p-values, and 95% confidence intervals for all model parameters. The z-statistic and p-value allows testing of the significance of each parameter in the regression model at a level $\alpha$, which we have set equal to 5%. The reported p-value is $\Pr(Z > |z|)$, where $z$ is the z-statistic.

Table 3: MLE results: point estimates and 95% confidence intervals for all model parameters.

| Parameter               | Estimate | Std. error | z-statistic | p-value | 95% lower | 95% upper |
|-------------------------|----------|------------|-------------|---------|-----------|-----------|
| Random effect ($\sigma^2_\alpha$) | 1.164    | 0.117      | 9.955       | 0.000   | 0.935     | 1.393     |
| Low cutoff ($\theta_1$)  | 1.749    | 0.157      | 11.10       | 0.000   | 1.440     | 2.057     |
| Mid cutoff ($\theta_2$) | 2.276    | 0.162      | 14.04       | 0.000   | 1.958     | 2.593     |
| High cutoff ($\theta_3$) | 5.657    | 0.178      | 31.83       | 0.000   | 5.309     | 6.005     |
| Initial rating          | 2.423    | 0.154      | 15.72       | 0.000   | 2.121     | 2.725     |
| Dept. A                 | 1.897    | 0.102      | 18.59       | 0.000   | 1.697     | 2.097     |
| Dept. B                 | 1.440    | 0.130      | 11.07       | 0.000   | 1.185     | 1.695     |
| Dept. C                 | 1.188    | 0.109      | 10.88       | 0.000   | 0.974     | 1.402     |
| Action I                | 3.705    | 0.128      | 28.85       | 0.000   | 3.453     | 3.957     |
| Action II               | 3.051    | 0.279      | 10.95       | 0.000   | 2.505     | 3.597     |
| Action III              | 11.16    | 3.428      | 0.003       | 0.997   | -6.708    | 6.730     |
| Period 2                | -1.159   | 0.111      | -10.46      | 0.000   | -1.376    | -0.942    |
| Period 3                | -1.229   | 0.111      | -11.09      | 0.000   | -1.446    | -1.012    |
| Period 4                | -1.111   | 0.110      | -10.06      | 0.000   | -1.327    | -0.894    |
| Period 5                | 0.082    | 0.108      | 0.756       | 0.450   | -0.130    | 0.293     |
| Period 6                | -0.374   | 0.107      | -3.488      | 0.000   | -0.584    | -0.164    |
| Period 7                | -0.292   | 0.105      | -2.787      | 0.005   | -0.497    | -0.087    |
| Period 8                | 0.522    | 0.110      | 4.729       | 0.000   | 0.306     | 0.738     |
| Period 9                | 0.666    | 0.112      | 5.966       | 0.000   | 0.447     | 0.884     |
| Period 10               | 0.721    | 0.112      | 6.420       | 0.000   | 0.501     | 0.941     |
| Period 11               | 0.932    | 0.114      | 8.165       | 0.000   | 0.709     | 1.156     |
| Period 12               | 1.103    | 0.116      | 9.483       | 0.000   | 0.875     | 1.331     |

The random effect is measured by the parameter $\sigma^2_\alpha$. The z-statistic and resulting p-value show that this parameter is significant; hence, the additional layer in the modeling hierarchy is relevant to our model. The cutoff parameters ($\theta_1$, $\theta_2$, $\theta_3$) are also statistically significant. Recall that on the continuous scale, ratings below $\theta_1$ translate to the Low category, ratings between $\theta_1$ and $\theta_2$ translate to the Medium category, ratings between $\theta_2$ and $\theta_3$ translate to the High I category,
and ratings greater than $\theta_3$ translate to the High II category. It is worth pointing out that the 95% confidence intervals for $\theta_1$ and $\theta_2$ overlap. Therefore, the upper boundary for the Medium rating is not significantly different from its lower boundary – i.e., the boundary that demarcates the Low rating. Therefore, this shows some evidence that there is not a clear distinction between the Low and Medium ratings in these data.

The situation is strikingly different when statistically testing the difference between $\theta_2$ and $\theta_3$, the lower bound for the High I rating and the upper bound for the High I rating, respectively. (Note that $\theta_2$ is also the lower bound for the High II rating.) The 95% upper endpoint for $\theta_2$ is about 2.6 while the 95% lower endpoint for $\theta_3$ is about 5.3, which is a notable difference of about 2.7. Since the thresholds defined by $\theta_2$ and $\theta_3$ convey information about the high-risk customers, the inference about these parameters could provide valuable information to AML surveillance units. These results justify the need for more granular monitoring of high-risk customers as there is a clear distinction of customers among the high-risk population.

Additionally, the initial on-board indicator, which can be construed as “real-time prior” information on a customer, is expected to contain useful information for AML surveillance. Unlike the ordinal outcome risk-rating, this initial indicator can be used to essentially reaffirm that the financial institution collects useful data on its customers, and that these data are reliable for monitoring and surveillance purposes. The initial rating on a customer is binary (i.e., “High I” vs. “Low”), and since it is positively associated with the risk-ratings over time, we can infer that the bank’s initial rating is a strong signal of future customer risk-ratings.

The coefficients of the business department identifiers are positive and significant, which implies that relative to the baseline department, which we arbitrarily designate as Dept. D, these three departments signal more risky customer ratings. Regarding the law enforcement action identifiers, we choose Action IV as the baseline. The coefficients of Action I and Action II are positive and significant meaning that these actions are likely associated with higher customer risk-ratings relative to Action IV. Action III, however, has a positive but statistically insignificant coefficient, which suggests that any inference about its effect on the customer risk-rating outcomes is inconclusive.
In fact, the standard error of the estimated coefficient for Action III is too high to reach any meaningful conclusion about the effect of the Action III category. (We also compute robust standard errors for these parameter estimates, and we present the results in the Appendix – cf. section A.)

The coefficients of the time indicators, denoted by period 2 through period 12, provide intriguing insights about the temporal aspects of the customer risk-rating process. We set period 1 as the baseline time point. We observe for periods 2 through 4 as well as periods 6 and 7, the estimates are negative and statistically significant. Period 5 has a positive coefficient, but it is not statistically significant. Thus, prior to (and including) period 7, customers were rated “less risky,” on average, relative to their period 1 rating. This seems to corroborate the findings on the estimate of the coefficient of the initial on-board indicator discussed above. The initial indicator may influence the period 1 rating such that a customer is labeled higher than might otherwise be the case.

An interesting feature of the temporal effects surface with the coefficients of the indicators for period 8 through period 12 where the estimates are positive and statistically significant. Thus it appears that customers are rated “more risky,” on average, relative to the period 1 rating for the later time periods in the data set. One might conjecture that customers are monitored with increased scrutiny during these periods, which results in higher ratings. The key observation in this context is the switch that occurs in the sign of the estimated coefficients of the time dummies. The presence of this switch could validate the assertion that learning takes place within the AML surveillance machinery. It ought not to be surprising that with the accumulation of evidence over time, signals of customer risk will surface and hence suspiciousness metrics, such as the customer risk-ratings, will respond accordingly to this enhanced intelligence.

5.2 Benchmarking the model-predicted risk-ratings

We explore how to possibly “validate” the proposed ordinal panel data regression model of the customer risk-rating data by measuring to what degree the observed ratings and model-predicted ratings agree with each other. First, the results from periods 1 to 12 in Table 4 show a strong degree of association between observed and predicted ratings. Initially, there is a strong degree of positive
association, as shown by the sample correlation, and this association slightly deteriorates over the ensuing periods. The trend, however, is not monotone as we see in period 5 there is an increase in the association measure. Throughout periods $t = 1$ through $t = 12$, the correlation between the model-predicted ratings and the observed ratings is strong and positive.
Table 4: Predictive contingency tables for in-sample periods $t = 1$ to $t = 12$.

| Predicted ($t=1$) | Current ($t=1$) | Predicted ($t=2$) | Current ($t=2$) | Predicted ($t=3$) | Current ($t=3$) |
|-------------------|----------------|------------------|----------------|------------------|----------------|
|                   |                |                  |                |                  |                |
| **1**             | 104            | **1**            | 297            | **1**            | 291            |
| **2**             | 0              | **2**            | 0              | **2**            | 0              |
| **3**             | 0              | **3**            | 5              | **3**            | 106            |
| **4**             | 0              | **4**            | 3              | **4**            | 0              |
| **Correlation:**  | 0.9787         | **Correlation:** | 0.4943         | **Correlation:** | 0.4956         |
| **Predicted ($t=4$)** | **1**  | **2**            | **3**          | **Correlation:** | **Correlation:** |
| **1**             | 267            | **2**            | 0              | **1**            | 63             |
| **2**             | 0              | **2**            | 1              | **2**            | 0              |
| **3**             | 116            | **3**            | 5              | **3**            | 14             |
| **4**             | 0              | **4**            | 0              | **4**            | 0              |
| **Correlation:**  | 0.5140         | **Correlation:** | 0.9094         | **Correlation:** | 0.7724         |
| **Predicted ($t=5$)** | **1**  | **2**            | **3**          | **Correlation:** | **Correlation:** |
| **1**             | 73             | **2**            | 0              | **1**            | 20             |
| **2**             | 0              | **2**            | 1              | **2**            | 0              |
| **3**             | 5              | **3**            | 3              | **3**            | 5              |
| **4**             | 0              | **4**            | 0              | **4**            | 0              |
| **Correlation:**  | 0.7084         | **Correlation:** | 0.8154         | **Correlation:** | 0.8002         |
| **Predicted ($t=6$)** | **1**  | **2**            | **3**          | **Correlation:** | **Correlation:** |
| **1**             | 10             | **2**            | 1              | **1**            | 20             |
| **2**             | 0              | **2**            | 1              | **2**            | 0              |
| **3**             | 5              | **3**            | 5              | **3**            | 5              |
| **4**             | 0              | **4**            | 0              | **4**            | 0              |
| **Correlation:**  | 0.7431         | **Correlation:** | 0.8198         | **Correlation:** | 0.9224         |
A key objective of any type of statistical modeling rests with how the results are used to leverage “predictive benefits.” For example, an AML analyst can use the model-predicted ratings with out-of-sample data to predict how to categorize customers in the period(s) ahead. Table 5 presents the association measure between the model-predicted ratings and observed ratings for data outside of the time period used to estimate the model. This type of predictive analytics could help an institution’s AML surveillance unit better manage resources and focus their attention on the riskiest customers. Indeed, AML surveillance units could use this type of predictive analytics to optimize their finite resources for future monitoring cycles.

Table 5: Predictive contingency table for out-of-sample period \( t = 13 \).

| Current \((t=13)\) | 1 | 2 | 3 | 4 |
|---------------------|---|---|---|---|
| 1                   | 3 | 0 | 1 | 0 |
| 2                   | 0 | 0 | 0 | 0 |
| 3                   | 5 | 5 | 180 | 3 |
| 4                   | 0 | 0 | 98 | 199 |

Correlation: 0.6667

If we analyze the contingency tables for all periods \((t = 1 \text{ through } t = 13)\), we see that most of the agreement between the model-predicted ratings and observed ratings are in categories 3 and 4, which are the riskiest categories! Note that the overall correlation is moderately strong given its value of about 0.667. As we described in section 3, categories 3 and 4 separate out two levels of high-risk customers where category 3 consists of “High I” customers and category 4 consists of “High II” customers. The strong agreement between the model and data for the higher ratings would likely be of interest to AML surveillance units and law enforcement personnel. Therefore, real-life benefits could be realized through statistical modeling and analysis of these data.
6  Application: AML investigation events

We analyze the relationship between the model-predicted risk-ratings and the observed risk-ratings relative to a signal of criminal activity. Moreover, we study this relationship over time to see if there are any useful insights to gather from its underlying dynamics. The indicia of criminal activity we analyze concerns a specific internal investigation.

6.1  Investigation events

Risk conditions can be conceptualized as a series of proxies for a latent measure of the likelihood that a customer will engage in future suspicious activity. Specifically, these risk conditions are leading indicators of suspicious behavior. However, if the bank believes that a suspicious event has already transpired, it flags the customer with an “investigation event.” We are not able to disclose the precise nature of an investigation event since we wish to protect the confidentiality of our data set. However, one can think of an investigation event as some event, internal within the bank, that causes a customer (or entity) to be scrutinized more intensely.

Investigation events are not determined by the same algorithm used to calculate risk-ratings, but are assigned separately, using both quantitative and qualitative inputs. Investigation events are lagging indicators of risk as they indicate previous suspicious behavior, but they are also leading indicators of further suspicious activity, as previous suspicious activity is a predictor of future risk. As such, investigation events are also one of the internal risk conditions included in the bank’s risk-rating process. Table 6 presents the number of investigation events by period.

Additionally, because investigation events are assigned independently of the risk-rating process, they are an exogenous input to the risk-rating algorithm. Ideally, the bank could assess its risk-rating logic by testing whether high-ratinged customers are more likely to engage in future criminal activity than low-ratinged customers. However, because the bank receives limited information from law enforcement, it does not have enough information to perform this test. Investigation events, which are initiated by the bank when it believes that suspicious activity has already taken
place, can be used as a proxy for determining whether a customer’s activity is tied to criminal
behavior, and so we can use the investigation events as one way to assess the customer risk-rating
model.

Table 6: Number of investigation events (by period).

| Period | Number of Investigation Occurrences |
|--------|-------------------------------------|
| 0      | 8                                   |
| 1      | 11                                  |
| 2      | 11                                  |
| 3      | 10                                  |
| 4      | 39                                  |
| 5      | 9                                   |
| 6      | 54                                  |
| 7      | 128                                 |
| 8      | 43                                  |
| 9      | 33                                  |
| 10     | 39                                  |
| 11     | 23                                  |
| 12     | 26                                  |
| 13     | 0                                   |

6.2 Log-linear modeling of investigation events

The relationship between the current and the model-predicted ratings can be further investigated
using information on an investigation event (IE), and this study can be tracked over time. Specif-
ically, we can compare the model-predicted customer risk-ratings to the current customer risk-
ratings accounting for whether or not an investigation event transpired, and we can do this over the
time frame in the data set. Tables 7 through 12 present this information over the 12 time periods in
the in-sample data set.
### Table 7: Predicted ratings vs. Current ratings, Period = 1 and 2

| Current rating | Predicted rating No investigation event | Predicted rating Investigation event |
|---------------|--------------------------------------|----------------------------------|
|               | 1 2 3 4                              | 1 2 3 4                          |
| 1             | 99 5  | 5                                  |
| 2             | 369 7 | 4 2                                |
| 3             | 1 2   |                                    |
| 4             | 1     |                                    |

### Table 8: Predicted ratings vs. Current ratings, Period = 3 and 4

| Current rating | Predicted rating No investigation event | Predicted rating Investigation event |
|---------------|--------------------------------------|----------------------------------|
|               | 1 2 3 4                              | 1 2 3 4                          |
| 1             | 284 7                               | 7                                 |
| 2             | 1 15 13                             | 2 1                               |
| 3             | 104 59                              | 1                                  |
| 4             | 2 6                                 |                                    |

### Table 9: Predicted ratings vs. Current ratings, Period = 5 and 6

| Current rating | Predicted rating No investigation event | Predicted rating Investigation event |
|---------------|--------------------------------------|----------------------------------|
|               | 1 2 3 4                              | 1 2 3 4                          |
| 1             | 69 2 4                              |                                    |
| 2             | 24 1 2                              |                                    |
| 3             | 5 3 316 3                             | 1 2                              |
| 4             | 5 60                                 |                                    |
### Table 10: Predicted ratings vs. Current ratings, Period = 7 and 8

| Current rating | Predicted rating | Predicted rating |
|----------------|------------------|------------------|
|                | No investigation event | Investigation event |
| 1              | 28 1              | 10               |
| 2              | 3                | 42               |
| 3              | 15 86 171        | 10 25            |
| 4              | 7 55             | 8 33             |

### Table 11: Predicted ratings vs. Current ratings, Period = 9 and 10

| Current rating | Predicted rating | Predicted rating |
|----------------|------------------|------------------|
|                | No investigation event | Investigation event |
| 1              | 10               | 10               |
| 2              | 1                | 8                |
| 3              | 5 5 237 5        | 8 1              |
| 4              | 63 135           | 1 5              |

### Table 12: Predicted ratings vs. Current ratings, Period = 11 and 12

| Current rating | Predicted rating | Predicted rating |
|----------------|------------------|------------------|
|                | No investigation event | Investigation event |
| 1              | 1                | 7                |
| 2              | 5 216 4          | 10               |
| 3              | 3 242            | 6                |
| 4              | 3 268            | 6                |
Tables 7 through 12 present the joint empirical distribution of the predicted ratings and the current ratings over time, but conditional on an investigation event. A log-linear modeling framework would potentially shed light on the relationship between and investigation event and the current rating, model-predicted rating, and time. There are different types of log-linear models that could be estimated including the independence model, conditional model(s), two-way models, and a full, saturated model. We opt to estimate the saturated model as it sheds some more insights relative to the other ones. The results of the log-linear estimation of the saturated model are given in Table 13.

Table 13: Log-linear modeling of investigation events as a function of current ratings, model-predicted ratings, and time.

| Covariate           | Estimate | Std. error | z-statistic | p-value   |
|---------------------|----------|------------|-------------|-----------|
| Intercept           | -2.672   | 0.696      | -3.838      | 1.240e-04 |
| Current rating      | -0.559   | 0.453      | -1.232      | 0.218     |
| Predicted rating    | -1.667   | 0.470      | -3.550      | 3.850e-04 |
| Time                | 0.510    | 0.110      | 4.634       | 3.580e-06 |
| Current × Predicted | 0.674    | 0.154      | 4.367       | 1.260e-05 |
| Current × Time      | -0.295   | 0.0678     | -4.357      | 1.320e-05 |
| Predicted × Time    | 0.308    | 0.0673     | 4.571       | 4.850e-06 |
| Current × Predicted × Time | -0.0498 | 0.0198 | -2.513 | 0.0120 |

Supplement to Table 13

Null deviance: 840.58 on 92 degrees of freedom
Residual deviance: 461.54 on 85 degrees of freedom
AIC: 633.99
Number of Fisher Scoring iterations: 5

The results in Table 13 reveal some interesting observations. Marginally, the current rating and predicted rating are negatively associated with an investigation event, with the latter being statistically significant. However, when evaluated in isolation, these estimated coefficients do not provide deep insights. The estimated coefficient on the time covariate is positive and significant, which signifies that investigation events are more likely as time progresses. The interaction terms provide some more information about how these covariates relate to investigation events. The
interaction between current and predicted ratings is positive, which follows from the interpretation attributed to their marginal impact on investigation events. As time progresses, the current rating is *negatively* associated with investigation events (as measured by the interaction between current rating and time), and the predicted rating is *positively* associated with investigation events (as measured by the interaction between current rating and time). The three-way interaction between current rating, predicted rating, and time says that over time, the interaction between current rating and predicted rating is negatively associated with an investigation event.

The model that we estimate is meant to provide important, high-level insights to policy-makers. It is limited, however, in terms of modeling, for example, time, but it nonetheless provides information on the impact of time in this problem. The negative association between the current and predicted ratings with investigation events across time is apparent when examining the count data in Tables 7 through 12. Consider the counts highlighted in **bold font**. These are cases where the model predicts a rating of 3 (High I) or 4 (High II), but the actual, current rating is 1 (Low) or 2 (Medium), *and* an investigation event transpired. There are also instances where the model predicts a rating of 1 (Low) or 2 (Medium), but the actual, current rating is 3 (High I) or 4 (High II), *and* an investigation event did *not* transpire. The former situation could entail more “risk,” which highlights the importance of enhancing AML surveillance with useful statistical analysis. It could be argued that these cases ought to warrant scrutiny as well even though the bank’s risk-rating is “Low” or “Medium.” Additionally, we have only studied one type of outcome (i.e., investigation events), but AML analysts who are privy to more sensitive data can analyze different outcome measures and gauge what impact a variable might have through statistical tools like log-linear models.

## 7 Implications for public policy

The results of statistical analyses, such as our study of customer risk-rating data, could be used to inform policy-makers in the anti-money laundering sphere. Recall, in section 5 above, we
develop a conceptual approach for a “statistical diagnosis” of AML customer risk-ratings. Here we consider ways that a statistical framework can be extended to additional components of a financial institution’s AML program. Additionally, we investigate how statistical tools could be used by multilateral institutions to help identify sources or trends in money laundering from a supranational or macro-level analysis. In particular, statistical analysis can inform financial institutions how to implement a risk-based approach to AML and use predictive analytics to identify emerging sources of AML risk. The implications of statistical analyses can supplement information to policy-makers, thereby helping them to make better decisions that benefit global communities.

7.1 The Risk-Based Approach

The Financial Action Task Force (FATF), a supranational body dedicated to advancing policy recommendations to combat money laundering and the financing of terrorism, released The FATF Recommendations in 2012 (Financial Action Task Force, 2012). The very first recommendation is that both countries and individual financial institutions should adopt a “risk-based approach” in assessing, analyzing, and mitigating AML risk. The concept of “risk” has certain implications for the role of statistical analysis within an AML framework. Chief among them is that estimates of AML risk are necessarily uncertain. At the level of an individual financial institution, uncertainty arises because feedback from law enforcement on the true nature of a suspicious customer or transaction is limited, so the outcome of a suspicious activity is often uncertain. A statistical framework may allow management to quantify this uncertainty, which may not be readily feasible when using non-statistical methods for flagging AML risk. At a country level, policy-makers using statistical tools could use resulting measures of uncertainty to inform future regulations and guidance to reduce uncertainty and combat money laundering initiatives.

An empirical diagnosis of AML data such as those used to risk-rating an institution’s customer base can not only advise a financial institution on which customers may require more enhanced monitoring, but can also help institutions to evaluate the power of existing models and determine what improvements, if any, could be made to the monitoring framework to reduce uncertainty
around risk estimates. For example, an institution might test alternative modeling specifications, or it may determine that current data collection efforts are insufficient to assess AML risk with the degree of confidence desired by management. Moreover, statistical comparisons could be made across different customer ratings, or even within a customer rating as we demonstrated with the “High” category (recall High I and High II). This type of analysis could better inform decision-makers about how to grapple with different types of customers, or how to understand potential granularities among a specific class of customers. In the case of high risk customers, our analysis might help management to manage the risk associated with this important category.

The FATF Recommendations also apply a risk-based approach within the context of ongoing monitoring of country, customer, and product-level risk assessments. In the case of customers, risk profiles can change over time, which necessitates some level of ongoing customer monitoring, and in this respect, statistical diagnosis may provide useful policy-related insights. At a customer level, if a customer’s behavior significantly changes, this may indicate the need to refresh information on this customer. At an aggregate level, if goodness-of-fit tests or other statistical diagnoses indicate that model performance has deteriorated, this may encourage analysts to respecify the model. Analysts could perform model specification tests to reassess which variables to include in the model and which type of functional form transformations may be most appropriate. These types of statistical exercises can help institutions to pro-actively address a significant complication of any monitoring framework: that AML risk constantly changes as criminals seek new ways to penetrate the global financial system.

Furthermore, a risk-based approach can help banks to manage ongoing customer risk, once it has been appropriately assessed and a framework has been developed for ongoing monitoring. For example, a hierarchical framework could be considered, with high-risk customers modeled separately at one level, and incorporated into a larger framework at another level. Such a hierarchical model would be roughly commensurate to “enhanced monitoring” of high-risk customers consistent with a risk-based approach, while still giving sufficient attention to the entire customer base.
7.2 Predictive Analytics

Policy-makers, both within a financial institution and at a larger level, can use predictive analytics derived from the models which were developed using a risk-based approach to identify nascent sources of risk in the near-to-medium term horizon. Both standard predictive methods, such as the prediction exercise that we did in Section 5.2, and simulation-based or Bayesian methods could be used to leverage historical and current information to identify near-term future risk. At a financial institution level, this could help stakeholders decide where to concentrate finite monitoring resources in the near-term and where to direct future model development efforts. This need not be limited to regression models incorporating an ordinal outcome, or even to regression models at all. More basic measures of association could be used to help policymakers understand which risks are likely to be correlated and how certain predictors may act in tandem with one another.

At a higher level, FIU units at multilateral financial institutions such as the IMF or regional bodies such as the Inter-American Development Bank, the African Development Bank, etc. could use a predictive framework to identify future sources of national or supranational risk. As with financial institutions, multilateral bodies also have finite resources, so a statistical framework could help policymakers steer these resources to the highest areas of concern. Furthermore, predictive or simulation-based modeling could not only help multilateral bodies monitor future AML risks, but could also influence higher-level policies regarding how the institution serves and interacts with the rest of the financial system.

Predictive inference might help policy-makers to draft policies that cater to risks that may not have been recognized without predictive modeling. Note that we are not arguing that predictive models or inferential strategies can predict whether or not a financial crime event will transpire. Rather, we argue that responsible use of predictive metrics can help policy-makers to better cope with possible outcomes that they may not have otherwise contemplated. Thus, predictive exercises could supplement a risk-based strategy to the extent that they can identify and possibly manage risk that are of importance to AML policy architects.
8 Discussion

Should criminals be able to use the global financial system to channel their illicit proceeds, there will likely be detrimental effects on society. While there are many important aspects of AML surveillance, the present study argues that there can be a role for statistical analysis of AML data. While statistical analysis can, in no way, eradicate financial crimes, we argue that it can add efficiencies to the overall monitoring and surveillance process. To adequately combat money laundering requires the expertise of compliance analysts responsible for manual reviews in addition to the sophistication afforded by statistical algorithms designed to detect anomalous behavior.

As this modest analysis illustrates, standard modeling paradigms can be used to diagnose how a bank’s customers are perceived in terms of AML risk. We show, using a generalized linear model, how an institution can potentially understand what variables influence AML risk, which in our study relates to customer risk-ratings. Our analysis also illustrates the importance of the interactions between the covariates used in our analysis, and how these interactions evolve over time. Moreover, we provide evidence to argue for more granular monitoring of high risk customers as a possible means to mitigate AML risk.

The modeling framework we use for customer risk-ratings could be enhanced, particularly if more data are available. We showed the benefits of a prediction exercise given the data we use, but it could be imagined that more accurate predictions might be made with a richer data set. As noted above, high risk customers ought to be subjected to more enhance, granular monitoring. Indeed, a potential refinement of our proposed model would be to use a hierarchical framework to first model high risk customers at one level, and then conditional on the “high risk level,” a second layer of modeling could accommodate all customers. One could use such a modeling approach to better manage AML risk in customer risk-ratings, which is only one component of AML surveillance. As sub-standard AML monitoring entails non-trivial consequences for a financial institution, robust data analysis should be leveraged where possible to add efficiencies to the surveillance strategy.

Disclaimer: The views expressed in this paper are solely those of the authors and neither re-
A Appendix: Robustness checks

In addition to the MLE point estimates reported in table 3, we also re-ran the specification using Huber-White (robust) standard errors, which are included in table 14. All \( p \)-values which are significant under classical assumptions remain significant using robust standard errors. On the other hand, the robust standard error for Action III is smaller than the classical standard error, such that the Action III parameter becomes significant. Because of the curious nature of this result, we used the classical standard errors for our main empirical diagnosis, prediction exercises, and log-linear modeling.

Furthermore, we conducted a series of robustness checks surrounding our random-effects ordinal probit specification. By default, the \texttt{xtoprobit} command uses 12 integration points, but we also ran specifications with alternate numbers of integration points (8 and 16) and obtained similar results. We conducted a likelihood-ratio (LR) test against a standard ordinal probit regression (i.e., without a random-effect) which confirmed that the random-effect parameter is significant. We also replicated our results using the NLMIXED procedure in SAS® 9.3 to confirm that our convergence was not conditional on our choice of statistical software. Results of our robustness checks are available upon request from the authors.
Table 14: MLE results: point estimates and 95% confidence intervals for all model parameters.

| Parameter               | Estimate | Robust Std. error | z-statistic | p-value | 95% lower | 95% upper |
|-------------------------|----------|-------------------|-------------|---------|-----------|-----------|
| Random effect ($\sigma^2_\alpha$) | 1.164    | 0.111             | 10.49       | 0.000   | 0.946     | 1.381     |
| Low cutoff ($\theta_1$)  | 1.749    | 0.154             | 11.37       | 0.000   | 1.447     | 2.050     |
| Mid cutoff ($\theta_2$)  | 2.276    | 0.175             | 13.02       | 0.000   | 1.933     | 2.618     |
| High cutoff ($\theta_3$) | 5.657    | 0.201             | 28.14       | 0.000   | 5.263     | 6.051     |
| Initial rating          | 2.423    | 0.184             | 13.15       | 0.000   | 2.062     | 2.784     |
| Dept. A                 | 1.897    | 0.138             | 13.79       | 0.000   | 1.628     | 2.167     |
| Dept. B                 | 1.440    | 0.199             | 7.231       | 0.000   | 1.050     | 1.830     |
| Dept. C                 | 1.188    | 0.151             | 7.871       | 0.000   | 0.892     | 1.484     |
| Action I                | 3.705    | 0.155             | 23.88       | 0.000   | 3.401     | 4.009     |
| Action II               | 3.051    | 0.267             | 11.43       | 0.000   | 2.528     | 3.574     |
| **Action III**          | **11.16**| **0.266**         | **41.98**   | **0.000**| **10.64** | **11.68** |
| Period 2                | -1.159   | 0.105             | -11.07      | 0.000   | -1.364    | -0.954    |
| Period 3                | -1.229   | 0.113             | -10.92      | 0.000   | -1.450    | -1.008    |
| Period 4                | -1.111   | 0.107             | -10.38      | 0.000   | -1.320    | -0.901    |
| Period 5                | 0.0816   | 0.0845            | 0.966       | 0.334   | -0.0841   | 0.247     |
| Period 6                | -0.374   | 0.0798            | -4.690      | 0.000   | -0.530    | -0.218    |
| Period 7                | -0.292   | 0.0951            | -3.064      | 0.002   | -0.478    | -0.105    |
| Period 8                | 0.522    | 0.0990            | 5.269       | 0.000   | 0.328     | 0.716     |
| Period 9                | 0.666    | 0.0901            | 7.384       | 0.000   | 0.489     | 0.842     |
| Period 10               | 0.721    | 0.0903            | 7.986       | 0.000   | 0.544     | 0.898     |
| Period 11               | 0.932    | 0.0788            | 11.84       | 0.000   | 0.778     | 1.087     |
| Period 12               | 1.103    | 0.0665            | 16.58       | 0.000   | 0.973     | 1.233     |

References

Böckenholt, U. (1999). Measuring change: Mixed Markov models for ordinal panel data. British Journal of Mathematical and Statistical Psychology 52(1), 125–136.

Bolton, R. J. and D. J. Hand (2002). Statistical fraud detection: a review. Statistical Science 17(3), 235–255.

Deng, X., V. R. Joseph, A. Sudjianto, and C. F. J. Wu (2009). Active learning through sequential design, with applications to detection of money laundering. Journal of the American Statistical Association 104(487), 969–981.

Federal Financial Institutions Examination Council (2010). Bank Secrecy Act/Anti-Money Laun-
dering Examination Manual. Washington, DC, USA.

Financial Action Task Force (2012). The FATF Recommendations. Paris, France.

Hamerle, A. and G. Ronning (1995). Panel analysis for qualitative variables. In *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, Chapter 8, pp. 402–451. Plenum Press.

Hand, D. J. (2009). Modern statistics: the myth and the magic. *Journal of the Royal Statistical Society (Series A)* 172(2), 287–306.

Heard, N., D. Weston, K. Platanioti, and D. Hand (2010). Bayesian anomaly detection methods for social networks. *The Annals of Applied Statistics* 4(2), 645–662.

Jansen, J. (1990). On the statistical analysis of ordinal data when extravariation is present. *Journal of the Royal Statistical Society (Series C)* 39(1), 75–84.

Johnson, V. E. (1996). On Bayesian analysis of multirater ordinal data: An application to automated essay grading. *Journal of the American Statistical Association* 91(433), 42–51.

Sudjianto, A., M. Yuan, D. Kern, S. Nair, A. Zhang, and F. Cela-Díaz (2010). Statistical methods for fighting financial crimes. *Technometrics* 52(1), 5–19.

The Financial Services Authority (2011). Banks’ management of high money-laundering risk situations: How banks deal with high-risk customers (including politically exposed persons), correspondent banking relationships and wire transfers. London, United Kingdom.

Tutz, G. and W. Hennevogl (1996). Random effects in ordinal regression models. *Computational Statistics & Data Analysis* 22(5), 537–557.

U.S. Senate Permanent Subcommittee on Investigations (2012). U.S. Vulnerabilities to Money Laundering, Drugs, and Terrorist Financing: HSBC Case History. Washington, DC, USA.