Solving the mystery of credit scoring models

White paper

As the speed of business continues to accelerate, more and more companies are challenged to keep pace with the demands of their markets, their customers and competitors. The marketplace is shrinking, with distinctions among global, national and regional marketplaces blurring. To increase the efficiency of processing and to meet the demands of time-sensitive customers, businesses must make decisions in real time or near-real time. Larger organizations have been utilizing credit scoring to quickly and accurately assess the risk level of their prospects, applicants and existing customers. Increasingly, midsize and smaller organizations are appreciating the benefits of credit scoring as well.

The credit score is reflected in a number or letter(s) that summarizes the overall risk utilizing available information on the customer. The credit score empowers users to make quick decisions or even to automate decisions. Whereas a decision based on manual review may take hours, a decision based on a credit score will take seconds. However, not all credit score models are created equally. That's why organizations must understand and match their needs to an appropriate credit score model. To select the appropriate model, one must understand not only the similarities and differences of the credit score models available, but also understand and be comfortable with using credit scoring.

To feel confident about using risk scoring in your environment, you need to have a good conceptual and practical understanding of what a risk score is and how it is developed. A risk score is generated when information related to an entity (a business or a consumer) is fed into a risk model. The risk model examines the information and assigns the relative importance of each piece of information, aggregating the individual contributions of each piece of information into one risk score that summarizes the entity's risk level. That's it in a nutshell. Now let's explore in detail each step in the process of generating a risk score.

The risk model

You undoubtedly have heard the word "model" used many different ways. We have forecasting models, revenue models, analytical models, physics models and models of trains, planes and automobiles to name a few. Like a model plane, train or automobile, all models can be thought of as simplistic representations of the real world. As such, models are imperfect facsimiles, never 100 percent accurate. No one expects a model car to replicate the intricacies of its real-life counterpart, and no one should expect other models to be completely accurate either. Besides physical models, there are predictive models, which predict an event based on empirical data — information observable in the real world.

A weather forecast is actually a prediction made by a system that takes empirical data, such as current temperature, humidity, wind conditions and other factors, to determine what the most likely weather conditions will be for the next several days. Weather forecast systems have certainly become more sophisticated over the years with advancements in technology. Whereas earlier weather forecast systems used a few observable factors to make predictions, current systems use hundreds of factors to make weather predictions almost an exact science. As a general rule, the more information that is available, the better the model will perform and the more accurate the prediction will be. However, no one should expect predictive models to be crystal balls. A weather forecast model would be 100 percent accurate only if it were to have perfect information of everything that affects the weather, which is impossible. How often have you dressed for a sunny day, as predicted, only to have it rain or vice versa? Certainly, there have been plenty of times when the forecast was inaccurate. Yet, we still trust in the weather forecast. Why? Well, for the most part, the weather forecast is accurate. It is valid in its prediction and has been consistently reliable over the years.

Credit bureaus offer models that predict credit risk. Similar to the weather forecast system, credit models generate a credit risk prediction based on readily available information. Credit risk models follow the general rule that more information leads to a better prediction. However, they also exhibit the weakness of all models: a small margin of error. Despite this limitation, credit risk models have been used for nearly two decades and repeatedly have proved to be both valid and reliable. Validity and reliability are two key measures that determine a predictive model's usefulness. Validity refers to whether a model does what it is made to do, and reliability refers to a model's validity over time.

Now that we understand what models are, their strengths and limitations, and how their usefulness is measured, let's explore how a model is actually developed.

Model development

To facilitate understanding, we will provide a conceptual explanation of the process of model development, followed by a "real-life" example of each model building step.

1. **Model development is driven by an objective**, which answers the question, "What is the model supposed to predict?"

   Let's say Martians do exist and they are of the scientific sort. One Martian scientist is interested in human beings — specifically, this Martian scientist has a burning desire to predict a human's weight!

2. Once the objective has been defined, the next step is to **gather as much data as possible related to the objective**. A common rule of thumb in model development is "The more data gathered, the better."

   Since the Martian doesn't know much about humans and only wants to predict weight from afar (Mars), he will need to collect all the data he can observe from a distance, including hair color, eye color, skin color, height, waist size, shoe size and so forth.

3. Next, the **data is analyzed** through a statistical process, usually call **regression**, to create a model. Common statistical packages used for regression include SAS, SPSS and even Microsoft® Excel! The statistical process looks for any significant relationships between what the model wants to predict — called the ***dependent***, or ***predicted variable*** — and all the possible variables that can influence the predicted variable. The variables that have a relationship with the predicted variable are called ***independent***, or ***predictor variables***. Once the regression process identifies all significant relationships between the predicted variable and the predictor variables, then a regression equation is generated that can be used to calculate the predicted variable when the predictor variable

values are known. The model can be fine-tuned at this point to be more predictive by segmenting the population used to create the model into more homogeneous, or similar groups. Conceptually, this step is the most difficult to understand, so let's continue with the example, which will clarify this part of modeling.

The Martian now needs to look for relationships between the predicted variable, weight, and the predictor variables. Because the Martian is unable to observe the weight of people from afar, he actually will need to abduct a large number of people from Earth! This group of people who have been abducted is called a ***sample population****. The size of the sample population has to be large enough to be representative of the entire population of people on Earth. Now the Martian can weigh all the people to capture the predicted variable values, capture other human features to get potential predictor variable values and then dump all the data into the regression process. At this point, the regression process will look for any significant relationships between weight and all the other potential predictor variables, such as hair color, eye color, skin color, height, waist size, shoe size and so on. For simplicity, let's say that there are only significant relationships between weight and waist size and between weight and height. Then the regression process will generate an* ***equation*** *that best estimates a person's weight based on waist size and height. Let's say the equation is as follows: Weight (in Earth pounds) =* ***1.4 x waist size (in inches) + 1.7 x height (in inches)****.*

If the Martian then observes someone on Earth to have a waist size of 30 inches and a height of 70 inches, the person's predicted weight will equal 1.4 x 30 + 1.7 x 70, or 161 pounds. Based on the sample population, 161 pounds can be considered the likely weight of a person with that waist size and height. Now, is there any way to make weight prediction more accurate? Looking at the sample population, is there a dimension by which the Martian can ***break up the population into smaller, more homogenous groups****? Yes, there is, and that dimension is gender. That makes sense. The average man with the same waist size and height as a woman will typically weigh more because men have a higher percentage of muscle mass than women and muscle mass weighs more than fat. So, the sample population is grouped by gender and run through the regression process again. Now there are two regression equations to predict weight, one for males and one for females. The male weight equation is weight = 1.5 x waist size + 1.8 x height, and the female weight equation is weight = 1.3 x waist size + 1.6 x height.*

Given the waist size of 30 inches and a height of 70 inches again, the predicted weight of a man is 171 pounds and the predicted weight of a woman is 151 pounds. The Martian has a more accurate weight prediction model! The process of dividing the sample population into more homogeneous groups is called ***segmentation****. This weight model now has two segments.*

4. Now that a **model** is created, it **needs to be validated**. When a model is being developed, a small number of subjects from the sample population is randomly selected to validate the model after it is created. This excluded group is called the **holdout sample**. Once the model is developed, the predicted variable value for the holdout sample is compared with the actual value being predicted. The closer the predicted variable is to the actual value, the better the model is preforming.

   The Martian would hold out a group of people to be excluded from those used for model development. Once the model is developed, the predicted weight of the holdout group is compared with their actual weight. The closer the predicted weight is to the actual weight, the better the model is performing.

Those are the basics of model development. **If you understand this, you understand how all predictive models are developed**. For example, let's consider **Experian's Commercial Intelliscore℠ model** and how it was developed.

The development of Commercial Intelliscore followed the exact same steps described above.

1. **Define the objective** — The objective of Commercial Intelliscore is to predict the likelihood of a business becoming greater than 90 days delinquent within a 6-month window.

2. **Collect data** — Experian® analyzed data on 3.4 million businesses in its database, including commercial credit information, business demographic information and public record information.

3. **Analyze data** — The Commercial Intelliscore model was developed through regression. However, unlike in the simplistic Martian example, where there were two segments and two predictor variables, Commercial Intelliscore has six segments (also known as scorecards) and more than 50 predictor variables. The six segments are based on prior trade experience, amount of trade activity, presence of public record information and business size.

4. **Validate the model** — The Commercial Intelliscore model was validated when it was created and continues to be validated regularly by Experian and its clients that use the model to score their portfolio of accounts.

If you understood the above conceptual, "real-life" and Commercial Intelliscore examples of how a model is developed, then you will understand all the other predictive models out there. This fundamental understanding will help you assess which predictive model is best for your business needs.

Credit risk model selection

As we have said, not all credit risk models are created equally. So, how do you choose the right model for you? Besides understanding how a model is developed and what it predicts, knowing how well it predicts is essential. A risk table, or performance chart, is created for each model and is used to gauge a model's performance. The risk table displays basic score distribution information. Along with the table, the

Kolmogorov-Smirnov (KS) statistic may be provided as well. The KS statistic, which is named after its two creators, is a widely accepted measure of model performance. Let's go over the risk table below to understand the information provided.

	A	B	C	D	E	F	G	H	I	J
	Score interval	% bads in interval	# of goods	# of bads	Cum. goods	Cum. bads	Cum. good %	Cum. bad %	% cum. diff.	Odds of goods to one bad
1	82.93–100.00	0.92%	991	9	991	9	5.27%	0.75%	4.52%	107.7
2	79.67–82.92	1.29%	987	13	1,978	22	10.52%	1.83%	8.69%	76.5
3	77.17–79.66	1.51%	985	15	2,963	37	15.76%	3.08%	12.68%	65.2
4	74.99–77.16	1.66%	983	17	3,946	54	20.99%	4.50%	16.49%	59.2
5	72.89–74.98	1.79%	982	18	4,928	72	26.21%	6.00%	20.21%	54.9
6	70.82–72.88	2.11%	979	21	5,907	93	31.42%	7.75%	23.67%	46.4
7	68.76–70.81	2.30%	977	23	6,884	116	36.62%	9.67%	26.95%	42.5
8	66.74–68.75	2.47%	975	25	7,859	141	41.80%	11.75%	30.05%	39.5
9	64.71–66.73	2.67%	973	27	8,832	168	46.98%	14.00%	32.98%	36.5
10	62.67–64.70	2.96%	970	30	9,802	198	52.14%	16.50%	35.64%	32.8
11	60.49–62.66	3.46%	965	35	10,767	233	57.27%	19.42%	37.85%	27.9
12	58.05–60.48	3.82%	962	38	11,729	271	62.39%	22.58%	39.80%	25.2
13	55.03–58.04	4.32%	957	43	12,686	314	67.48%	26.17%	41.31%	22.1
14	51.42–55.02	4.64%	954	46	13,640	360	72.55%	30.00%	42.55%	20.6
15	46.70–51.41	5.65%	943	57	14,583	417	77.57%	34.75%	42.82%	16.5
16	40.41–46.69	6.53%	935	65	15,518	482	82.54%	40.17%	42.38%	14.3
17	31.67–40.40	8.14%	919	81	16,437	563	87.43%	46.92%	40.51%	11.3
18	19.10–31.66	10.85%	892	108	17,329	671	92.18%	55.92%	36.26%	8.3
19	5.31–19.09	15.98%	840	160	18,169	831	96.64%	69.25%	27.39%	5.3
20	0.00–5.30	36.91%	631	369	18,800	1,200	100.00%	100.00%	0.00%	1.7

First, all the businesses used to develop the model are scored, then rank ordered from lowest risk to highest risk. Then, the businesses are divided into twentiles — equal 5 percent segments. Column A shows the **score interval** for each of the 20 rows. The score range is 0 to 100, with 0 indicating highest risk and 100 indicating lowest risk.

So, the lowest-risk 5 percent of the businesses scored between 82.93 and 100 (cell A1). The second-lowest-risk 5 percent scored between 79.67 and 82.92 (A2), and so on. The worst-scoring 5 percent scored between 0.00 and 5.30 (A20).

Column B shows the **percentage of bad businesses** within each score interval. A bad account for this model is defined as a business that becomes more than 90 days delinquent within one year. The percentage of bads increases as we move down the column, which is to be expected since moving down the column means we are looking at higher-risk businesses. If we did not see this pattern of increasing percentages of bads, then the model would not be valid.

In the first row, the lowest-risk 5 percent of businesses, the percentage of bad businesses that fall into this row is 0.92 percent (B1). In row 20, the worst-scoring 5 percent of businesses, the percentage of bad businesses is 36.91 percent (B20). Column C displays the **number of good accounts** for each row, and column D displays the **number of bads** in each row. Again, one can see that the number of bads dramatically increases as we move down the column, from low risk to high risk.

Column E displays the **cumulative number of goods**, which is summing the total number of goods at and above a given row. For example, the cumulative goods for row 2, 1,978 (E2), is calculated by adding the number of goods for row 2, 987 (C2), and the number of goods for row 1, 991 (C1). This means that there are 1,978 businesses that scored at or above 79.67.

Column F, cumulative bads, is calculated the same way using the number of bads. Column G is the **cumulative good percentage**, which is the percentage of all the good businesses at and above a given row. The total number of good businesses is 18,800 (E20). Therefore, to calculate the cumulative good percentage for row 5, 26.21 percent (G5), the cumulative goods for row 5, 4,928 (E5), is divided by the total number of goods, 18,800 (E20).

Column H, the **cumulative bad percentage**, is calculated the same way using bads. Column I is the **cumulative percentage difference** between the cumulative percentage good (G) and cumulative percentage bad (H). For example, for row 10, the cumulative percentage difference, 35.64 percent (I10), is calculated by subtracting 16.50 percent (H10) from 52.14 percent (G10). If there is no difference between the cumulative good percentage and cumulative bad percentage, then we can conclude that the model cannot differentiate between good and bad businesses. Conversely, we can conclude that the greater the cumulative percentage difference, the better the model is at accurately assessing risk. **The greatest cumulative percentage difference value is the model's KS statistic**. Column J displays the **odds of goods to one bad**. For row 1, the value is 107.7, which means there are 107.7 good businesses in this row for every one bad business.

Credit risk model score application

Using the credit risk model score enables quick, accurate and consistent decisions. The score is unbiased and objective over time and across reviewers. The examples below show how the score can be applied for enhanced decisioning.

<table>
<tr><th>Risk level</th><th>Risk averse goal: reduce risk</th><th>Standard</th><th>Aggressive goal: increase sales</th></tr>
<tr><td rowspan="5">100
Low risk
↕
0
High risk</td><td>Automatically accept</td><td rowspan="2">Automatically accept</td><td rowspan="3">Automatically accept</td></tr>
<tr><td>Review</td></tr>
<tr><td rowspan="3">Automatically reject</td><td>Review</td></tr>
<tr><td rowspan="2">Automatically reject</td><td>Review</td></tr>
<tr><td>Automatically reject</td></tr>
</table>

Automated decisioning is applicable for large-volume, low-dollar transactions. This chart shows the flexibility of scores based on automated decisioning. A risk-averse business can use the score to automatically accept applicants that scored very low risk, manually review those that scored in the low-risk range and automatically reject those that scored mid to high risk. A business with standard risk tolerance would raise the level of risk tolerance, and the level would be raised even more for aggressive businesses.

Current review	Low risk score (poor risk)	Medium risk score (marginal risk)	High risk score (good risk)
Poor risk	Reject automatically	Reject automatically	Marginally approve
Marginal risk	Reject automatically	Marginally approve	Approve with favorable terms
Good risk	Marginally approve	Approve with favorable terms	Approve with best terms

For those businesses that review all transactions, the risk score can be integrated with the current review process to enhance accurate risk assessment and decisioning. This chart is an example of integrating the risk score with the existing review method to decision applications.

The risk score has diverse utility. As described above, it can be used for application processing in determining approval or decline as well as setting prices, premiums, down payments, credit limits, terms and conditions. The score also can be used for effective portfolio management. Scoring the portfolio enables portfolio risk evaluation and risk distribution assessment. Low-risk accounts can be identified for cross-sell and up-sell campaigns. Collections efforts can be prioritized based on risk score. Renewals can be automated for low-risk accounts, and high-risk accounts can be terminated. The risk score also can be used for effective marketing by targeting low-risk prospects instead of mass-marketing. Targeting low-risk prospects will result in lower-risk applicants, which ultimately will lead to a healthier portfolio.

600 City Parkway West
Orange, CA 92868
800 520 1221

www.experian.com