

## Контроль качества данных в рейтинговых системах

---

---

Russia Risk Conference 2014, 13.11.2014

## Ключевые вопросы

---

- Что контролировать?
- Как контролировать?
- Как использовать?

## Задачи и цели контроля качества данных

---

- ❑ Задачей контроля данных является идентификация и сокращение/устранение ошибок, а также проведение корректирующих и превентивных мер для достижения и поддержания качества данных на высоком уровне.
  
- ❑ Управление качеством данных, используемых для расчета параметров кредитного риска, осуществляется посредством специальных IT инструментов и источников данных, связанных с расчетом риск параметров.

Следующие группы данных, которые используются для расчета первой компоненты Базель II, подлежат контролю:

а) параметры кредитного риска, используемые для расчета требований к капиталу:

- Вероятность дефолта (Probability of Default);
- Убыток в случае дефолта (Loss Given Default);
- Величина кредитного требования подверженного риску дефолта (Exposure at Default);

б) данные, используемые для расчета стоимости обеспечения для применения механизма снижения кредитного риска.

# Что контролировать

---

## ■ Центральный Банк РФ:

Приложение 3. Минимальные требования к качеству данных, используемых кредитными организациями для создания и применения моделей количественной оценки кредитных рисков для целей расчета нормативов достаточности капитала

Первичные и преобразованные данные, используемые для создания и при применении моделей:

- Точность и достоверность данных
- Полнота данных
- Своевременность данных
- Согласованность данных
- Доступность данных
- Контролируемость данных
- Восстанавливаемость данных

Предложенные контроли включают в себя контроли первого и второго уровня, а также контроль источников, хранения и передачи данных.

**Контроль первого уровня** удобно организовать в режиме реального времени на стороне систем. Системы при этом могут блокировать ввод некорректных данных (например, различная величина суммарных активов и пассивов в балансе) или предупреждать о возможных несоответствиях в данных с информированием о необходимости дополнительной верификации.

**Контролируемость, доступность и восстанавливаемость данных**, включая ведение спецификации баз данных, политику доступов и сертификацию данных, используемых для разработки, как правило, осуществляется на стороне ИТ.

## Контроль 2-го уровня: определение периметра контроля

---

Цель – определение зон данных, подлежащих контролю, путем распознавания входных данных для каждого этапа рейтингового процесса. Например, могут выделяться такие зоны:

### Информация, управляемая изнутри

- Количественные данные (отчетность, получаемая банком)
- Качественные данные
- Поведенческая информация
- Данные о структурах групп и др.

### Информация, управляемая извне

- Внешние рейтинги
- Данные Кредитных бюро (если используются)
- Количественные данные, получаемые из внешних источников (отчетность, извлекаемая из внешних источников) и др.

Проблемы, идентифицируемые в первом блоке, как правило, устранимы внутри банка. В случае возникновения сомнений в качестве данных зачастую единственным решением будет отказ от источников и перестроение соответствующим образом процесса.

# Основные группы параметров в рамках контроля 2-го уровня

---

**Полнота** – демонстрирует степень доступности (наличия) данных, используемых для расчета рейтинга

Пример: наличие отчетности  
наличие пропусков в данных

---

**Точность/Надежность** – демонстрирует степень точности/надежности данных

Пример: наличие аномальных значений в данных  
миграции между сегментами

---

**Непротиворечивость** – демонстрирует последовательность и упорядоченность информации, поступающей из различных источников данных, или переменных, которые вместе могут представлять отдельное событие

Пример: разница в качественных оценках, присвоенных участниками процесса  
несогласованность в оценках  
разница между количественными/качественными оценками

---

**Своевременность** – демонстрирует возможность получения и администрирования информации в нужное время и в соответствии с требованиями

Пример: надвигающееся устаревание рейтинга  
своевременность предоставления информации для расчета рейтинга

---

**Прочие метрики** – могут характеризовать отдельные особенности процесса

Пример: доля и размер ручных корректировок рейтинга  
число рейтингов, присвоенных клиенту за определенный временной интервал

# Как контролировать

---

- Для каждого из показателей необходимо разработать **метрики** качества данных
- Каждое значение метрики качества данных может использоваться на индивидуальной либо консолидированной основе. Наиболее распространенными видами метрик являются:
  - Базовое значение: представляет собой индивидуальное значение, относящееся к каждому конкретному наблюдению. (Пример: сумма задолженности отдельного контрагента в общем кредитном портфеле)
  - Агрегированное значение (Сумма, Количество, Среднее):
    - Сумма значений определенных наблюдений в общем наборе данных;
    - Количество определенных наблюдений в общем наборе данных;
    - Среднее значение определенных наблюдений в общем наборе данных.
  - Отклонение от целевого значения : представляет собой разницу между установленным целевым значением и полученным наблюдением.
  - Процентный индекс: отражает процентную долю определенного значения (или группы значений) в определенной выборке.
  - Отклонение от исторического среднего: отражает разницу между текущей величиной и средним историческим значением, нормализованную на историческое стандартное отклонение. (Пример: отклонение текущей средней величины фактора от среднего значения, используемого при разработке модели)

# Как контролировать

- Примеры метрик:

Отсутствие отчетности			
C1 Клиенты, с рейтингами, присвоенными без отчетности			
пороговые значения			
Числитель : C1	Границы	Желтая	5%
Знаменатель : Все клиенты, с присвоенными рейтингами		Красная	10%

Несогласованность качественных оценок			
CO8 Число присвоенных рейтингов с измененными качественными оценками			
CO8.1 Улучшение оценки			
CO8.2 Ухудшение оценки			
CO8.3 Оценки совпадают			
CO9 Общее число присвоенных рейтингов за период			
CO10 = CO8.1+CO8.2 Число измененных оценок за период			
Пороговые значения			
Числитель: CO10	Границы	Желтая	10%
Знаменатель : CO9		Красная	15%

- Для каждой метрики необходимо определить пороговые значения



# Пороговые значения

---

**Пороговое значение** определяет предел контрольной величины, выход за который означает смену статуса для уровня качества данных.

**Пороговую величину характеризуют:**

- Значение
- Направленность
- Число уровней

В обязательном порядке должен быть внедрен по меньшей мере один уровень. Однако наиболее распространенным методом при назначении числа является «метод светофора»:

- **«желтый»** пороговый уровень призван оповещать о возможных надвигающихся проблемах. В этом случае полезным будет понять причины и принять меры по выводу метрики в «зеленую» зону.
- **«красный»** пороговый уровень сигнализирует о возникших проблемах и призывает принять срочные меры по выявлению причин и повышению статуса до по крайней мере «желтого» уровня

## Назначение и управление пороговыми значениями

Не существует единого подхода к назначению пороговых значений. Значения выбираются и обосновываются с учетом особенностей разработанных моделей и рейтингового процесса, используя аналитические и статистические (если применимо) методы.

Пороговые значения требуют периодического пересмотра (не реже одного раза в год). Однако пересмотр может проводиться и чаще, при возникновении структурных изменений.

# Агрегирование пороговых значений

Помимо метрик, определенных на нижнем уровне, полезно также иметь механизм агрегирования оценок с целью понимания ситуации на уровне групп метрик или даже всего портфеля.

В этом случае каждой из метрик присваивается вес в общей сумме в зависимости от значимости/критичности метрики. В случае применения принципа «светофора» механизм может выглядеть следующим образом:

- Если пороговое значение = «зеленый», то  $Q(i) = W(i) * 1$
- Если пороговое значение = «желтый», то  $Q(i) = W(i) * 0.5$
- Если пороговое значение = «красный», то  $Q(i) = W(i) * 0$

Агрегирование:  $Q = \sum q(i)$

Метрика	Вес метрики, W(i)	Пороговое значение	Итоговый вес, Q(i)
разница в качественных оценках	25%	Красный	0%
несогласованность в оценках	25%	Желтый	13%
Разница между количественными/качественными оценками	50%	Зелёный	50%
	100%		<b>63%</b>

Итоговый вес, Q	Пороговое значение		Итог
63%	Зелёный	75%	Желтый
	Желтый	50%	☹️

Агрегирование на более высоких уровнях осуществляется аналогично.

## Осуществление процедуры контроля

---

- Осуществление процедуры контроля представляет собой периодическое проведение проверок с использованием специальных инструментов, предназначенных для проведения таких проверок в автоматическом режиме.
- В особых случаях процедура контроля может также проводиться вручную, то есть если уровень воздействия фактора оценивается как слишком высокий с точки зрения затрат в сравнении с прибылью; такие проверки могут проводиться также в качестве промежуточного решения в ходе внедрения соответствующего IT решения.
- Процедура контроля включает в себя :
  - Сбор данных (сбор подлежащей проверке информации из различных источников данных, как внутренних, так и внешних, в соответствии с бизнес-спецификациями)
  - Расчет метрик (расчеты с использованием собранных данных, цель которых — выведение каждой из требуемых метрик.)
  - Сохранение данных (сохранение рассчитанных метрик для сравнения с текущими данными в будущем и отслеживания формирующихся тенденций)

# Периодичность проведения процедуры контроля и взаимодействие участников процесса

## ■ Периодичность

Кредитная организация планирует и непрерывно осуществляет работу по обеспечению качества данных. Не реже одного раза в год: проведение аудита качества данных и определение эффективности мер контроля качества данных.

Периодичность, с которой осуществляется контроль, зависит от типологии рейтинговой модели, частоты обновления рейтингов и от особенностей процессов внутри банка. Как правило, расчет контрольных метрик осуществляется на ежеквартальной (ежемесячной) основе. Факторы, которые могут оказать влияние на частоту расчета:

- Резко меняющиеся внешние условия
- Изменение процессов или внедрение новых ИТ решений

## ■ Взаимодействие участников процесса контроля

В банке должен существовать порядок взаимодействия органов управления, подразделений и должностных лиц банка, устанавливающий их полномочия, ответственность и подотчетность по вопросам обеспечения качества данных;

### процедуры реагирования и исправления

- Анализ проблемы
- Определение причин
- Определение решений
- Внедрение решений



Вовлеченность подразделений можно прописать лишь на общем уровне, всегда необходим детальный анализ проблемы

Спасибо за внимание!

---

