

DOI: 10.26794/2587-5671-2023-27-6-44-53

УДК 336.763.331(045)

JEL G12, E43

Обработка пропусков в рыночных данных на примере задачи оценки кривой доходностей облигаций

М.С. Макушкин, В.А. Лапшин

НИУ ВШЭ, Москва, Россия

АННОТАЦИЯ

Пропуски в рыночных данных — частая проблема в финансовых исследованиях, которой уделяется относительно мало внимания на практике. Обработка пропусков, как правило, выполняется эвристически и является лишь вспомогательным шагом исследований. **Целью** нашей работы является разработка практических рекомендаций по работе с пропусками в рыночных данных. Проблема иллюстрируется на примере задачи оценки временной структуры процентных ставок на российском рынке государственных облигаций. Мы сравниваем три различных **метода** заполнения пропусков в данных — заполнение последним значением, фильтр Калмана и ЕМ-алгоритм — с простой стратегией удаления пропусков. Мы приходим к **выводу**, что эффект от заполнения пропусков на качество оценки кривой зависит от чувствительности модели кривой доходностей к рыночным данным. Для слабо чувствительных к данным моделей, например для модели кривой доходностей Нельсона-Зигеля, эффект от заполнения пропусков минимален. Для более чувствительных моделей кривой, таких как бутстрэп, за счет заполнения пропусков удается достичь статистически значимого улучшения качества оценки срочной структуры процентных ставок. При этом данный результат не зависит от способа заполнения пропусков. И простой метод заполнения последним значением, и более сложный ЕМ-алгоритм дают схожие результаты. **Рекомендация** исследования состоит том, что на практике при оценке кривой доходностей в условиях неполных данных необходимо либо использовать слабо чувствительные к данным параметрические модели кривой доходностей, либо заполнять пропуски в данных перед использованием чувствительных моделей.

Ключевые слова: кривая бескупонной доходности; срочная структура процентных ставок; рынок облигаций; метод Нельсона-Зигеля; ликвидность рынка; пропуски в данных; развивающиеся рынки

Для цитирования: Макушкин М.С., Лапшин В.А. Обработка пропусков в рыночных данных на примере задачи оценки кривой доходностей облигаций. *Финансы: теория и практика*. 2023;27(6):44-53. DOI: 10.26794/2587-5671-2023-27-6-44-53

ORIGINAL PAPER

Treatment of Missing Market Data: Case of Bond Yield Curve Estimation

M.S. Makushkin, V.A. Lapshin

National Research University Higher School of Economics, Moscow, Russia

ABSTRACT

Missing observations in market data is a frequent problem in financial studies. The problem of missing data is often overlooked in practice. Missing data is mostly treated using ad hoc methods or just ignored. Our **goal** is to develop practical recommendations for treatment of missing observations in financial data. We illustrate the issue with an example of yield curve estimation on Russian bond market. We compare three **methods** of missing data imputation — last observation carried forward, Kalman filtering and EM-algorithm — with a simple strategy of ignoring missing observations. We **conclude** that the impact of data imputation on the quality of yield curve estimation depends on model sensitivity to the market data. For non-sensitive models, such as Nelson-Siegel yield curve model, final effect is insignificant. For more sensitive models, such as bootstrapping, missing data imputation allows to increase the quality of yield curve estimation. However, the **result** does not depend on the chosen data imputation method. Both simple last observation carried forward method and more advanced EM-algorithm lead to similar final results. Therefore, when estimating yield curves on the illiquid markets with missing market data, we **recommend** to use either simple non-sensitive to the data parametric models of yield curve or to impute missing data before using more advanced and sensitive yield curve models.

Keywords: yield curve; term structure of interest rates; bond market; Nelson-Siegel method; liquidity level; missing data; emerging markets

For citation: Makushkin M.S., Lapshin V.A. Treatment of missing market data: Case of bond yield curve estimation. *Finance: Theory and Practice*. 2023;27(6):44-53. (In Russ.). DOI: 10.26794/2587-5671-2023-27-6-44-53

© Макушкин М.С., Лапшин В.А., 2023

ВВЕДЕНИЕ

Пропуски в данных — частая проблема во многих эмпирических исследованиях. Встречается она и в финансах, особенно когда речь идет о развивающихся, низколиквидных рынках. Наличие пропусков в данных осложняет оценку финансовых моделей, а также может искажать получаемые выводы. Пропуски в рыночных данных могут возникать по разным причинам, например, вследствие низкой ликвидности инструмента, цензурирования данных, фильтрации атипичных наблюдений.

На развитых рынках с высокой ликвидностью проблема пропусков в данных, как правило, игнорируется, поскольку пропусков в данных мало. Фокус чаще делается на модельных вопросах, чем на вопросах качества данных.

На рынках с низкой ликвидностью пропусков в данных больше. Простое удаление пропусков может привести к потере важной информации. Поэтому исследователи часто предварительно обрабатывают пропуски. Выбор метода обработки пропусков при этом осуществляется эвристически, а сама обработка данных является вспомогательным шагом на пути к ответу на основные исследовательские вопросы.

В работе мы подробнее изучаем вопрос обработки пропусков в рыночных данных применительно к задаче оценки бескупонной кривой доходностей облигаций. Нашей целью является разработка практических рекомендаций по работе с пропусками в рыночных данных при оценке временной структуры процентных ставок.

Для иллюстрации проблемы и ее возможных решений на практике мы используем данные по торгам российскими облигациями федерального займа (ОФЗ). Мы исследуем, как заполнение пропусков в данных по торгам влияет на качество оценки кривой доходностей. Это актуальная задача, поскольку оценка кривой доходностей на развивающихся рынках часто бывает осложнена наличием пропусков в рыночных данных [1].

Новизна работы заключается в применении статистических инструментов из области анализа неполных данных к задачам финансовой инженерии. Мы показываем, что вопросы, связанные с качеством и полнотой данных, на развивающихся рынках требуют не меньшего внимания, чем вопросы выбора финансовой модели.

ОБЗОР ЛИТЕРАТУРЫ

В статистической науке существует отдельно направление, посвященное формальному анализу пропусков в данных [2–4]. Пропуски в данных опасны по двум причинам. Во-первых, они могут приводить к смещенным оценкам параметров модели.

Во-вторых, они увеличивают стандартную ошибку коэффициентов модели и уменьшают мощность статистических тестов [5].

Безопасно удалить пропуски из анализа можно только в том случае, если они совершенно случайны, т.е. не зависят ни от собственных ненаблюдаемых значений, ни от значений других наблюдений. Если пропуски просто случайны, т.е. зависят не от собственного пропущенного значения, а от значений других параметров, то пропущенные данные можно восстановить с помощью условного заполнения. Если пропуски не случайны, то для их заполнения необходимо знать процесс, генерирующий пропуски [3].

В рыночных данных могут встречаться как случайные, так и неслучайные пропуски. Неслучайные пропуски могут возникать, например, в ситуации, когда торги акцией или облигацией приостанавливаются при резком падении их цены (пример — приостановка торгов на российском рынке в феврале-марте 2022 г.). Такие случаи необходимо анализировать отдельно, уделяя особое внимание причинам возникновения пропусков.

Со случайными же пропусками, наоборот, можно эффективно работать. Это можно сделать двумя способами. Первый вариант — адаптировать модель под пропуски в данных. Это более корректный вариант, однако он усложняет модель, зависит от ее свойств и спецификации и не является универсальным. Применительно к задаче оценки кривой доходностей смотрите, например, работы [6–8].

Второй вариант работы с пропусками — предварительное заполнение пропусков в данных. Плюсы подхода заключаются в возможности сохранить простую исходную модель, а также использовать обработанные данные для других целей. В рамках нашего исследования мы рассматриваем именно этот более общий вариант.

На практике наиболее частым решением проблемы является удаление пропусков [9]. Это простое решение, но оно ведет к потере части важной информации. В некоторых ситуациях может оказаться полезным заполнение пропусков. Общий обзор возможных методов приведен в работах [10, 11]. Применительно к финансовым задачам в литературе встречаются как продвинутые методы заполнения (ЕМ-алгоритм [12, 13], байесовская модель с использованием методов Монте-Карло для марковских цепей [14]), так и более простые (заполнение последним значением [15] и последним значением со взвешиванием [16]). При этом метод заполнения пропусков в основном выбирается без специального обоснования. В перечисленных работах не обсуждается, как метод заполнения пропусков влияет на получаемые результаты. Само заполнение

пропусков упоминается вскользь и является промежуточным шагом на пути к ответу на основные исследовательские вопросы. В связи с этим важно обратить более детальное внимание на проблему пропусков в рыночных данных и сравнить методы их обработки применительно к финансовым задачам.

МЕТОДОЛОГИЯ

Модели оценки кривой бескупонных доходностей

Мы иллюстрируем важность проблемы обработки пропусков в рыночных данных на примере задачи оценки кривой доходностей на рынке облигаций. Кривая доходностей показывает зависимость стоимости денег от срочности. Она нужна для множества практических задач — от макроэкономического прогнозирования до риск-менеджмента и оценки финансовых инструментов. На развивающихся рынках оценка кривой доходностей часто оказывается осложнена наличием пропусков в данных [1]. Поэтому задача оценки кривой доходностей на данных с пропусками выглядит важной и актуальной.

Существует множество моделей для оценки кривой бескупонных доходностей. Все они базируются на общем принципе — информация о ненаблюдаемой кривой бескупонных доходностей извлекается из наблюдаемой рыночной информации, например, из цен купонных облигаций. Обзор моделей кривых доходностей можно найти в работе [17].

При выборе модели срочной структуры процентных ставок необходимо учитывать количество и качество доступных для калибровки данных [18]. Модель должна быть настолько сложной и точной, насколько это позволяют доступные данные. Там, где количество доступных данных ограничено, должны применяться консервативные предположения и более простые модели. Поэтому применительно к рыночным данным с пропусками целесообразно рассмотреть две модели кривой доходностей с разной степенью чувствительности к качеству данных. Мы выбираем достаточно простую, но стабильную модель кривой доходностей Нельсона-Зигеля [19] и более гибкую, но менее стабильную технику бутстрэпа процентных ставок [20].

Модель Нельсона-Зигеля в факторной спецификации [21] описывает доходность $y(\tau)$ на срочность τ следующим уравнением:

$$y(\tau) = \beta_1 + \beta_2 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) + \beta_3 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) + \varepsilon(\tau), \quad (1)$$

где $\theta = \{\beta_1, \beta_2, \beta_3, \lambda\}$ — вектор параметров модели; $\varepsilon(\tau)$ — некоторая ошибка.

Фактически ставка $y(\tau)$ в модели является взвешенной суммой трех факторов, где весами выступа-

ют коэффициенты модели $\beta_1, \beta_2, \beta_3$. Параметры $\beta_1, \beta_2, \beta_3$ можно интерпретировать как уровень, наклон и «горб» кривой доходностей (level, slope, curvature). Четвертый параметр λ описывает положение «горба» на графике.

Модель Нельсона-Зигеля хорошо описывает эмпирически наблюдаемые формы кривой доходностей. Поэтому ее в том или ином виде используют многие финансовые организации¹. В частности, модель кривой доходностей Московской Биржи (G-curve²) базируется на варианте кривой доходностей Нельсона-Зигеля. Модель также популярна на развивающихся рынках и рынках с низкой ликвидностью, поскольку имеет всего четыре параметра, которые могут быть удобно оценены в условиях небольшого доступного набора рыночных данных [22, 23].

Оценка параметров моделей может осуществляться за счет минимизации средней квадратичной ошибки переоценки доходностей облигаций либо непосредственно на основе цен облигаций. Описание технических особенностей оценки модели можно найти, например, в работе [24]. Мы выбираем вариант оценки модели на основе цен облигаций, поскольку на рынке ОФЗ бескупонные доходности напрямую не наблюдаются.

Если даны цены купонных облигаций P , то задача оценки параметров кривой доходности описывается следующими уравнениями:

$$\hat{\theta} = \arg \min_{\theta} \sum_i^N (\hat{P}_i(\theta) - P_i)^2, \quad (2)$$

$$\hat{P}_i = \sum_{j=1}^{J_i} CF_{ij} e^{-y(\tau_{ij})\tau_{ij}}, \quad (3)$$

где j — порядковый номер потока; J — общее число потоков; $y(\tau_{ij})$ — бескупонная ставка, соответствующая времени τ_{ij} до выплаты потока CF_{ij} по облигации i . Фактически модель оценивается как классическая регрессия, поэтому все нестабильности в данных сглаживаются на этапе оценки.

Бутстрэп — метод, более чувствительный к данным. Он подразумевает последовательный расчет бескупонных процентных ставок в порядке увеличения срочности облигаций. Зная цены облигаций на разные срочности, с помощью бутстрэпа можно итеративно найти бескупонные процентные ставки. Полученная кривая идеально воспроизводит исходные рыночные данные.

¹ Bank of International Settlements. Zero-coupon yield curves: Technical documentation. BIS Papers. 2005;(25).

² MOEX (2021). Zero-coupon Yield Curve for Sovereign Bonds. <https://www.moex.com/a3642> (дата обращения: 09.11.2022).

Формально логику бутстрэпа можно записать следующим образом. Предположим, мы имеем набор облигаций $P_i, i \in 1 \dots N$ на срочности $\tau \in 1 \dots J$. Тогда стоимость каждой облигации можно представить как сумму ее дисконтированных будущих потоков:

$$\begin{cases} \hat{P}_1 = CF_{11}e^{y(\tau_{11})\tau_{11}} \\ \hat{P}_2 = CF_{21}e^{y(\tau_{21})\tau_{21}} + CF_{22}e^{y(\tau_{22})\tau_{22}} \\ \dots \\ \hat{P}_N = CF_{N1}e^{y(\tau_{N1})\tau_{N1}} + CF_{N2}e^{y(\tau_{N2})\tau_{N2}} + \dots + CF_{NJ}e^{y(\tau_{NJ})\tau_{NJ}} \end{cases} \quad (4)$$

Систему уравнений (4) можно решить итеративным методом, подставляя найденные бескупонные доходности из первых уравнений в последующие.

В реальности набор доступных облигаций P ограничен. Требуется предположение о форме кривой доходностей на участках между срочностями доступных бумаг. Мы используем базовое предположение о кусочно-постоянной форме кривой доходностей, однако можно предположить и более сложные зависимости [25].

Итоговая процентная кривая $y(\tau)$ зависит от каждой точки рыночных данных и идеально воспроизводит начальные данные. Любой шум в исходных данных может существенно исказить форму кривой, а пропуски в данных сделают ее более «ступенчатой» [26]. Поэтому бутстрэп обычно используют на развитых ликвидных рынках, где торгуется достаточное количество облигаций, а структурные неэффективности сведены к минимуму.

Методы заполнения пропусков в данных

Мы сравниваем две стратегии работы с пропусками — простое удаление пропусков и заполнение пропусков. Первый вариант выступает в роли базовой стратегии, поскольку он наиболее часто встречается в финансовых исследованиях, оперирующих данными с пропусками. Второй вариант — более продвинутый. Он встречается в литературе реже, но иногда позволяет улучшить качество оценки модели. Мы изучаем, насколько стратегия заполнения пропусков позволяет улучшить качество оценки кривой доходностей по сравнению со стратегией удаления пропусков.

Мы выбрали три метода заполнения пропусков: 1) простой эвристический метод заполнения последним значением; 2) фильтр Калмана, учитывающий предыдущую динамику наблюдений; 3) ЕМ-алгоритм, учитывающий совместную динамику данных. Такой выбор методов позволяет сравнить, насколько сильно усложнение модели заполнения пропусков влияет на качество оценки кривой.

Простейшим методом заполнения пропусков является заполнение последним значением. Этот метод

прост в реализации, но использует сильную предположку о том, что в условиях отсутствия данных предыдущее наблюдение является наилучшей оценкой отсутствующего значения.

Фильтр Калмана — более продвинутый вариант заполнения пропусков. При заполнении пропусков этот метод учитывает предыдущую динамику наблюдений. На основе зашумленных наблюдений фильтр Калмана оценивает ненаблюдаемое состояние системы. Затем с помощью полученной оценки состояния системы можно оценить возможные значения пропущенных наблюдений. Технические аспекты метода приведены в работе [27].

ЕМ-алгоритм помимо предыдущей динамики самого восстанавливаемого наблюдения учитывает также динамику соседних наблюдений и ковариационную структуру данных. Это итеративный алгоритм, состоящий из двух шагов. На Е-шаге вычисляется ожидаемое значение (expectation) вектора ненаблюдаемых переменных на основе первого приближения коэффициентов модели, оцененных с помощью доступной информации. На М-шаге решается задача максимизации правдоподобия (maximization) и находится следующее приближение вектора параметров модели. С его помощью вновь оцениваются ненаблюдаемые значения и т.д. [28].

Сравнение качества оценки

Мы сравниваем качество оценки кривой доходностей с помощью кросс-валидации. Суть метода заключается в том, что для тестирования по очереди используются все наблюдения. Это полезно в условиях малого количества тестовых данных. Каждое наблюдение поочередно откладывают, оценивают модель без него, а затем рассчитывают ошибку для отложенного тестового наблюдения. Процедура выполняется итеративно для каждого наблюдения в выборке, а итоговый результат затем усредняется [29].

При расчете ошибки на кросс-валидации мы используем только реальные наблюдения. Восстановленные данные таким образом нужны лишь для более точной оценки кривой. Их не нужно учитывать в расчете ошибки, так как итоговая цель состоит в повышении качества воспроизведения реальных наблюдений, а не восстановленных.

ДАННЫЕ

Мы используем цены закрытия торгов по стандартным ОФЗ с фиксированным купоном в период с мая 2012 по декабрь 2015 г. (примерно 1000 наблюдений). Мы не включаем в выборку ОФЗ с амортизацией номинала, плавающим купоном и привязкой

к инфляции, так как принципы их ценообразования отличаются от стандартных купонных ОФЗ [30]. Цены были получены с аналитической платформы Финам.

Были также собраны и изучены более актуальные данные по торгам ОФЗ вплоть до 2022 г. Однако доля пропусков в этих данных мала, поэтому они были бы нерепрезентативны для исследования. В связи с этим мы ограничиваемся рассмотрением более старых данных.

Альтернативным решением могла бы стать генерация пропусков в данных, однако такой подход подразумевает искусственный, экзогенно заданный процесс возникновения пропусков, что может искажать выводы исследования. Поэтому мы жертвуем актуальностью данных в обмен на сохранение реальной и неискаженной структуры пропусков. Такой подход при этом не отменяет практической значимости получаемых результатов. Они все еще могут быть применены к другим рынкам облигаций, где проблема пропусков в данных по-прежнему актуальна — к рынку корпоративного долга, а также к менее ликвидным рынкам суверенных облигаций, где пропуски в данных — все еще частое явление.

Средняя доля в отобранных для анализа данных составляет 10%. При этом пропусков больше в начале выборки (в среднем 30% в 2012 г.). Затем доля пропусков снижается до 20%. К концу 2015 г. в выборке практически не остается пропусков. Большое количество пропусков сосредоточено около выходных и праздничных дней (посленовогодние дни, майские праздники). Пропуски распределены относительно случайно по отдельным ценным бумагам. Длинные серии пропусков отсутствуют. Обычно их длина не превышает двух-трех дней. Некоторые ценные бумаги более «склонны» к появлению пропусков, но в целом они не сильно выделяются из выборки.

Мы преобразуем данные следующим образом. Из цен закрытия мы рассчитываем купонные доходности. Далее заполняем пропуски в доходностях с помощью моделей, описанных ранее. Затем мы возвращаемся от обработанных доходностей снова к ценам облигаций. Итоговая кривая доходностей строится уже именно на основе новых цен. Для наблюдений, по которым нет пропусков, такой переход ничего не меняет, так как доходность и цена связаны однозначно. Таким образом все реальные наблюдения остаются нетронутыми, а пропущенные — восстанавливаются.

Переход от цен к доходностям и обратно необходим, чтобы учесть pull-to-par эффект (схождение стоимости облигации к номиналу по мере приближения к погашению). Кроме того, некоторые методы заполнения пропусков требуют нормально распределенных дан-

ных, например ЕМ-алгоритм. Распределение доходностей ближе к нормальному, чем распределение цен облигаций.

Для применения ЕМ-алгоритма мы дополнительно рассчитываем среднюю арифметическую доходность по всем ценным бумагам за каждый день. Это необходимо, потому что размерность выборки не фиксирована. Одни ценные бумаги гасятся, другие, наоборот, выпускаются. В результате достоверно оценить ковариацию между доходностями отдельных ценных бумаг оказывается сложно. Расчет ковариации со средней доходностью вместо ковариационной матрицы доходностей всех бумаг позволяет преодолеть эту проблему. Разумеется, такое упрощение ведет к потере части информации. Фактически при заполнении пропусков ЕМ-алгоритмом мы учитываем только связь наблюдения с общим «уровнем» кривой доходностей. Однако такое упрощение видится допустимым, так как параллельные сдвиги объясняют большую часть динамики КБД [31].

РЕЗУЛЬТАТЫ

На основе описанной методики были заполнены пропуски в данных по торгам облигациями. На *рис. 1, 2* приведено графическое сравнение разных способов заполнения пропусков в доходностях и ценах на примере ОФЗ 25065.

При заполнении пропусков в доходностях результаты могут сильно отличаться. Заполнение последним значением (выколотые точки) выглядит наиболее грубым. Оно создает в данных длинные постоянные участки. Заполнение фильтром Калмана (крестики) работает несколько лучше. Модель учитывает предыдущую динамику доходности, поэтому константные участки заменяются на наклонные. Наиболее правдоподобные результаты достигаются при заполнении с помощью ЕМ-алгоритма (плюсы). Фактически динамика средней доходности по всем ценным бумагам переносится на восстанавливаемую бумагу.

При переходе к ценам разница между методами заполнения пропусков становится менее заметной. Это связано с тем, что по мере уменьшения срока до погашения чувствительность облигации к изменению доходности уменьшается, а ее цена стремится к номиналу.

На данных с заполненными пропусками мы затем оценили кривые бескупонных доходностей с помощью модели Нельсона-Зигеля и бутстрэпа и рассчитали среднюю абсолютную ошибку переоценки бумаг на кросс-валидации. Сравнив значения ошибки, мы рассчитали долю дней, в которые заполнение пропусков позволило улучшить качество оценки кривой по сравнению с простым удалением пропусков. При

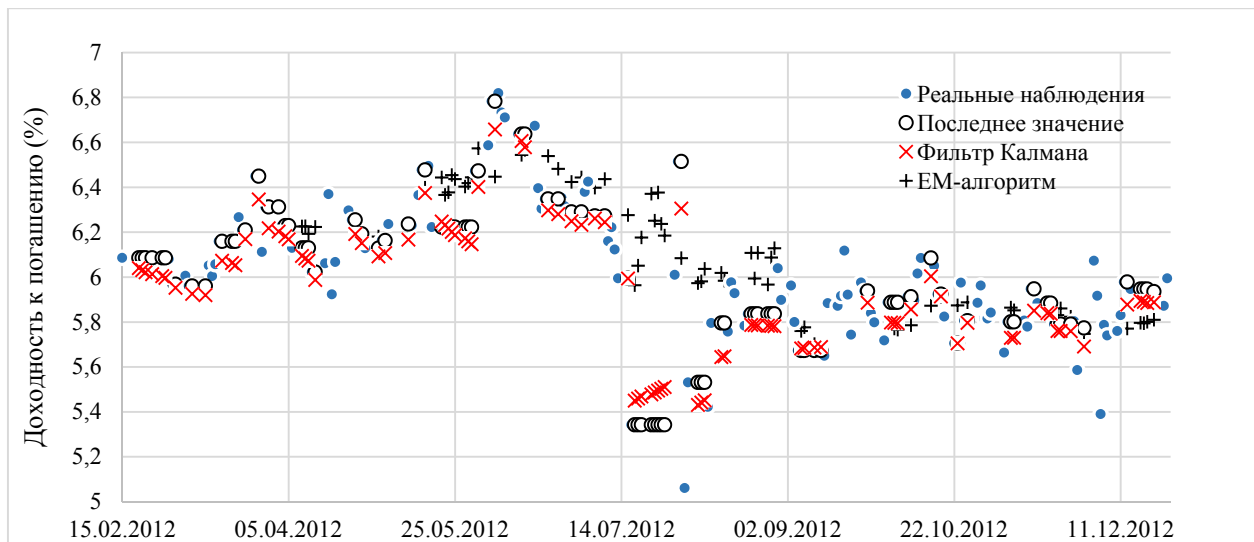


Рис. 1 / Fig. 1. Заполнение пропусков в доходностях на примере ОФЗ 25065 / An Example of Processing Yields Gaps for FLB 25065

Источник / Source: расчеты авторов / Author's calculations.

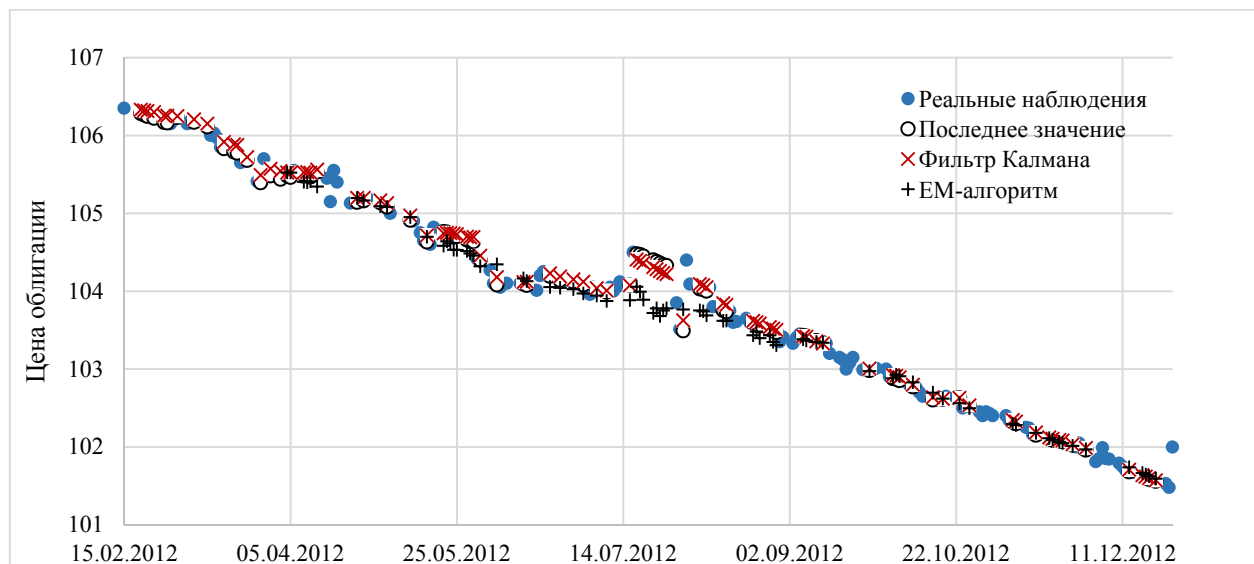


Рис. 2 / Fig. 2. Заполнение пропусков в ценах на примере ОФЗ 25065 / An Example of Processing Prices Gaps for FLB 25065

Источник / Source: расчеты авторов / Author's calculations.

этом учитывались только те дни, в которые были зафиксированы пропуски.

Стратегия заполнения пропусков позволяет улучшить качество оценки кривой доходностей методом бутстрэпа по сравнению с простой стратегией удаления пропусков (см. таблицу). Дней, когда заполнение пропусков позволило улучшить качество, оказывается примерно 65%. Этот показатель чуть выше для случая заполнения пропусков с помощью ЕМ-алгоритма, однако в целом различия между конкретными методами заполнения пропусков незначительны. Полученное улучшение качества от заполнения про-

пусков статистически значимо на 95%-ном уровне уверенности.

Для параметрической модели Нельсона-Зигеля стратегия заполнения пропусков не дает значительного улучшения качества оценки кривой. Дней, когда восстановление данных помогло улучшить качество оценки, и дней, когда кривую лучше оценивать лишь по имеющимся данным, оказывается примерно поровну.

Разный эффект от заполнения пропусков на качество оценки кривой доходностей с помощью модели Нельсона-Зигеля и с помощью бутстрэпа можно объяс-

Доля дней, когда рассматриваемый метод заполнения пропусков дает лучшее качество оценки кривой, чем простое удаление пропусков / Percentage of the Days when Processing Data Gaps Provides a Better Yield Curve Fit than Simple Removal Processing of Gaps

	Бутстрэп / Bootstrapping		Модель Нельсона-Зигеля / Nelson-Siegel Mode	
	Удаление пропусков / Removal of gaps	Заполнение пропусков / Processing data gaps	Удаление пропусков / Removal of gaps	Заполнение пропусков / Processing data gaps
Последнее значение	35%	65%	52%	48%
Фильтр Калмана	34%	66%	59%	41%
ЕМ-алгоритм	31%	69%	49%	51%

Источник / Source: расчеты авторов / Author's calculations.

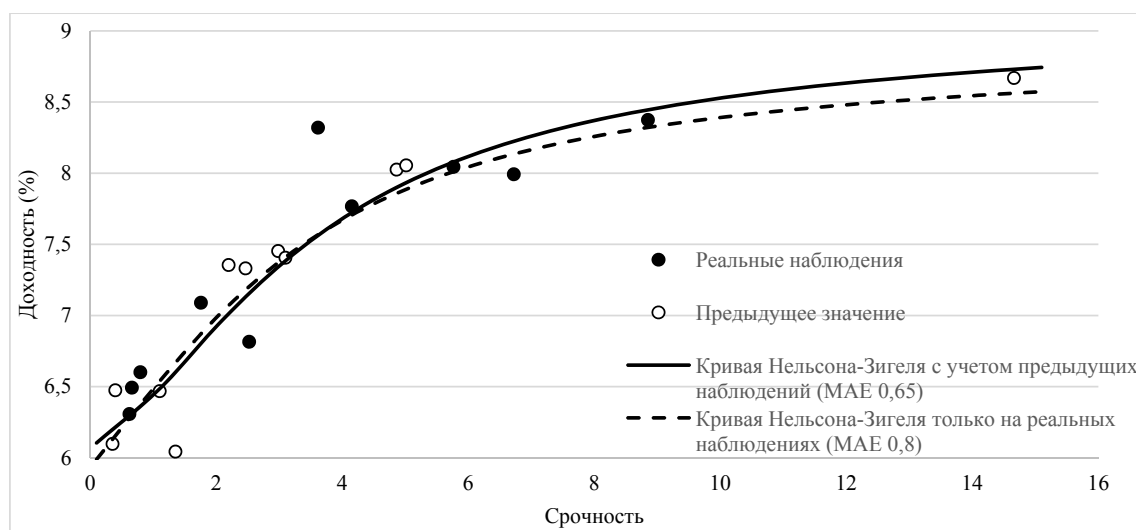


Рис. 3 / Fig. 3. Оценка кривой Нельсона-Зигеля по российским ОФЗ за 9 июня 2012 г. по реальным наблюдениям и по данным с заполненными пропусками / Nelson-Siegel Yield Curve for Russian FLB Constructed Using Only Real Data and Using Both Real and Gaps Data as of June 9, 2022

Источник / Source: расчеты авторов / Author's calculations.

снить разной чувствительностью моделей к рыночным данным. В модели Нельсона-Зигеля всего 4 параметра. Их можно достаточно хорошо оценить и по доступным наблюдениям. Добавление к, условно, 16 реальным точкам еще двух-трех восстановленных не окажет значительного влияния на результат. Бутстрэп же, наоборот, зависит от каждой точки рыночных данных. Добавление даже одного наблюдения взамен пропущенного позволяет сделать кривую доходностей менее «ступенчатой».

Пример улучшения качества оценки кривой доходностей приведен на рис. 3, 4 на основе данных о торгах за 9 июня 2012 г. Это предпраздничный день, поэтому торговая активность была ниже, а пропусков в данных больше. Кривая доходностей была оценена

сначала только на реальных наблюдениях, а потом на комбинации реальных данных и заполненных пропусков. И для модели Нельсона-Зигеля (рис. 3), и для бутстрэпа (рис. 4) заполнение пропусков позволило точнее описать срочную структуру процентных ставок, особенно на дальнем конце. Средняя абсолютная ошибка на кросс-валидации была значимо уменьшена (значения ошибки приведены на рисунках).

ВЫВОДЫ

Было проведено сравнение различных методов работы с пропусками в рыночных данных. Результаты проиллюстрированы на примере задачи оценки кривой бескупонной доходностей на российском рынке. Мы показали, что простое игнорирование

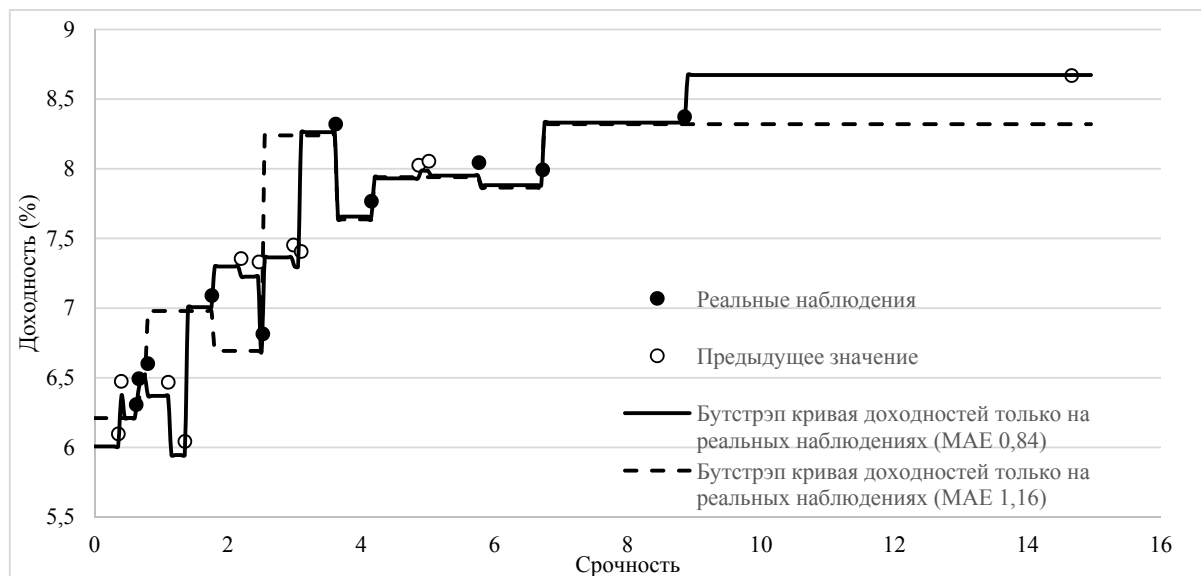


Рис. 4 / Fig. 4. Оценка кривой бескупонной доходностей методом бутстрэпа по российским ОФЗ за 9 июня 2012 г. по реальным наблюдениям и по данным с заполненными пропусками / Bootstrapped Zero-Coupon Yield Curve for Russian FLB Constructed Using Only Real Data and Using Both Real and Gaps Data as of June 9, 2022

Источник / Source: расчеты авторов / Author's calculations.

пропусков может приводить к искаженным оценкам модели кривой доходностей. Напротив, заполнение пропусков способно улучшить качество оценки кривой доходностей по сравнению со случаем удаления пропущенных наблюдений из выборки.

Эффект от заполнения пропусков в данных на качество оценки кривой доходностей зависит от выбранной модели кривой. Для параметрической модели Нельсона-Зигеля положительный эффект от заполнения пропусков минимален. Для бутстрэпа фиксируется статистически значимое улучшение качества оценки при заполнении пропусков в данных.

Наблюдаемое различие мы связываем с разной степенью чувствительности моделей кривой доходностей к рыночным данным. Параметрическая модель Нельсона-Зигеля может быть эффективно оценена и по небольшому числу наблюдений с пропусками. Для бутстрэпа же каждое дополнительное наблюдение, в том числе заполненное, имеет значение.

На практике при выборе стратегии обработки пропусков мы рекомендуем учитывать степень чувствительности финансовой модели к рыночным данным. Если используются чувствительные модели, полезно будет предварительно заполнить пропуски в данных. Конкретный способ обработки пропусков применительно к задаче оценки кривой доходностей оказывает не так значим. И простое заполнение последним значением, и более сложные методы заполнения пропусков на основе фильтра Калмана или ЕМ-алгоритма дают схожий результат. Если же нет возможности полноценно обработать пропуски в данных, тогда необходимо использовать более простые и менее чувствительные к данным модели.

Выводы работы могут быть полезны для оценки кривой бескупонных доходностей на рынках с низкой ликвидностью и для других финансовых исследований, работающих с неполными рыночными данными.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

1. Макушкин М.С., Лапшин В.А. Кривые доходностей на низколиквидных рынках облигаций: особенности оценки. *Экономический журнал Высшей школы экономики*. 2021;25(2):177–195. DOI: 10.17323/1813–8691–2021–25–2–177–195
Makushkin M., Lapshin V. Yield curve estimation in illiquid bond markets. *Ekonomicheskii zhurnal Vyshei shkoly ekonomiki = The HSE Economic Journal*. 2021;25(2):177–195. (In Russ.). DOI: 10.17323/1813–8691–2021–25–2–177–195
2. Haitovsky Y. Missing data in regression analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1968;30(1):67–82. DOI: 10.1111/j.2517–6161.1968.tb01507.x
3. Rubin D.B. Inference and missing data. *Biometrika*. 1976;63(3):581–592. DOI: 10.1093/biomet/63.3.581

4. Little R.J.A. Missing-data adjustments in large surveys. *Journal of Business & Economic Statistics*. 1988;6(3):287–296. DOI: 10.1080/07350015.1988.10509663
5. Newman D.A. Missing data: Five practical guidelines. *Organizational Research Methods*. 2014;17(4):372–411. DOI: 10.1177/1094428114548590
6. Cortazar G., Schwartz E.S., Naranjo L.F. Term-structure estimation in markets with infrequent trading. *International Journal of Finance & Economics*. 2007;12(4):353–369. DOI: 10.1002/ijfe.317
7. Poghosyan K., Poghosyan A. Yield curve estimation and forecasting in Armenia. *Armenian Journal of Economics*. 2019;(4):1–19. URL: https://www.aea.am/files/papers/a2019_1.pdf
8. Nagy K. Term structure estimation with missing data: Application for emerging markets. *The Quarterly Review of Economics and Finance*. 2020;75:347–360. DOI: 10.1016/j.qref.2019.04.002
9. Kofman P., Sharpe I.G. Using multiple imputation in the analysis of incomplete observations in finance. *Journal of Financial Econometrics*. 2003;1(2):216–249. DOI: 10.1093/jjfinec/nbg013
10. Briggs A., Clark T., Wolstenholme J., Clarke P. Missing... presumed at random: Cost-analysis of incomplete data. *Health Economics*. 2003;12(5):377–392. DOI: 10.1002/hec.766
11. Graham J.W. Missing data analysis: Making it work in the real world. *Annual Review of Psychology*. 2009;60:549–576. DOI: 10.1146/annurev.psych.58.110405.085530
12. Warga A. Bond returns, liquidity, and missing data. *Journal of Financial and Quantitative Analysis*. 1992;27(4):605–617. DOI: 10.2307/2331143
13. Zhu H. An empirical comparison of credit spreads between the bond market and the credit default swap market. *Journal of Financial Services Research*. 2006;29(3):211–235. DOI: 10.1007/s10693-006-7626-x
14. Косьяненко А.В. Опыт восстановления пропущенной рыночной информации на основе Байесовского подхода. Финансовая инженерия, риск-менеджмент и актуарная наука. Препринт WP16/2007/02. М.: ГУ ВШЭ; 2007. 32 с. URL: https://www.hse.ru/data/2010/05/07/1217275563/WP16_2007_02.pdf
Kos'yanenko A. V. Missing market data imputation using Bayesian inference. Financial engineering, risk management and actuarial science. Preprint WP16/2007/02. Moscow: SU HSE; 2007. 32 p. (In Russ.).
15. Теплова Т.В., Буданова Д.М. Эффективность ценообразования на российском рынке корпоративных облигаций. *Вестник Московского университета. Серия 6: Экономика*. 2017;(4):3–28. DOI: 10.38050/01300105201741
Teplova T., Budanova D. Pricing effectiveness at the Russian corporate bonds market. *Vestnik Moskovskogo universiteta. Seriya 6: Ekonomika = Moscow University Economics Bulletin*. 2017;(4):3–28. (In Russ.). DOI: 10.38050/01300105201741
16. Хакимжанов С., Мустафин Е., Кубенбаев О., Атабек Д. Построение кривой доходности на рынке с низкой ликвидностью. *Деньги и Кредит*. 2019;78(4):71–98. DOI: 10.31477/rjmf.201904.71
Khakimzhanov S., Mustafin Y., Kubenbayev O., Atabek D. Constructing a yield curve in a market with low liquidity. *Russian Journal of Money and Finance*. 2019;78(4):71–98. DOI: 10.31477/rjmf.201904.71 (In Russ.: *Den'gi i kredit*. 2019;78(4):71–98. DOI: 10.31477/rjmf.201904.71).
17. Лапшин В.А., Терещенко М.Ю. Выбор модели срочной структуры процентных ставок на основе ее свойств. *Корпоративные Финансы*. 2018;12(2):171–187. DOI: 10.17323/j.jcfr.2073-0438.12.2.2018.171-187
Lapshin V., Tereshchenko M. The choice of the model of the term structure of interest rates on the basis of its properties. *Korporativnye finansy = Journal of Corporate Finance Research*. 2018;12(2):171–187. (In Russ.). DOI: 10.17323/j.jcfr.2073-0438.12.2.2018.171-187
18. Курбангалеев М.З., Лапшин В.А. Анализ требований к оценке срочной структуры безрисковых ставок в финансовых задачах. *Управление финансовыми рисками*. 2015;(1):50–60.
Kurbangaleev M., Lapshin V. An analysis of the risk-free term structure estimation requirements posed by various financial problems. *Upravlenie finansovymi riskami = Financial Risk Management Journal*. 2015;(1):50–60. (In Russ.).
19. Nelson C.R., Siegel A.F. Parsimonious modeling of yield curves. *The Journal of Business*. 1987;60(4):473–489. DOI: 10.1086/296409
20. Fama E. F., Bliss R. R. The information in long-maturity forward rates. *The American Economic Review*. 1987;77(4):680–692.
21. Diebold F. X., Li C. Forecasting the term structure of government bond yields. *Journal of Econometrics*. 2006;130(2):337–364. DOI: 10.1016/j.jeconom.2005.03.005
22. Zoricic D., Orsag S. Parametric yield curve modeling in an illiquid and undeveloped financial market. *UTMS Journal of Economics*. 2013;4(3)234–252. URL: https://www.utmsjoe.mk/files/Vol.%204%20No.%203/1-1B-Zoricic-Orsag_-_Croatia.pdf

23. Chou J.-H., Su Y.-S., Tang H.-W., Chen C.-Y. Fitting the term structure of interest rates in illiquid market: Taiwan experience. *Investment Management and Financial Innovations*. 2009;6(1):101–116. URL: https://www.researchgate.net/publication/288118114_Fitting_the_term_structure_of_interest_rates_in_illiquid_market_Taiwan_experience
24. Wahlstrøm R.R., Paraschiv F., Schürle M.A. Comparative analysis of parsimonious yield curve models with focus on the Nelson-Siegel, Svensson and Bliss versions. *Computational Economics*. 2022;59(3):967–1004. DOI: 10.1007/s10614-021-10113-w
25. Hagan P.S., West G. Interpolation methods for curve construction. *Applied Mathematical Finance*. 2006;13(2):89–129. DOI: 10.1080/13504860500396032
26. Авдеева О.А., Цыплаков А.А. Метод адаптивного оценивания срочной структуры процентных ставок. *Экономический журнал Высшей школы экономики*. 2015;19(4):609–639.
Avdeeva O., Tsyplakov A. A method for adaptive estimation of the term structure of interest rates. *Ekonomicheskii zhurnal Vysshei shkoly ekonomiki = The HSE Economic Journal*. 2015;19(4):609–639. (In Russ.).
27. Kalman R. A new approach to linear filtering and prediction problems. *Transactions of the AMSE. Journal of Basic Engineering*. 1960;82(1):35–45. DOI: 10.1115/1.3662552
28. Dempster A.P., Laird N.M., Rubin D.B. Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1977;39(1):1–22. DOI: 10.1111/j.2517-6161.1977.tb01600.x
29. Lapshin V., Sohatskaya S. Choosing the weighting coefficients for estimating the term structure from sovereign bonds. *International Review of Economics & Finance*. 2020;70:635–648. DOI: 10.1016/j.iref.2020.08.011
30. Курбангалеев М.З., Лапшин В.А., Шепелева И.С. Согласованность котировок государственных облигаций России. *Управленческий учет и финансы*. 2016;(1):40–51.
Kurbangaleev M., Lapshin V., Shepeleva I. Consistency of Russian sovereign bond prices. *Upravlencheskii uchet i finansy = Management Accounting and Finance Journal*. 2016;(1):40–51. (In Russ.).
31. Falkenstein E., Hanweck J. Minimizing basis risk from non-parallel shifts in the yield curve. Part II: Principal components. *The Journal of Fixed Income*. 1997;7(1):85–90. DOI: 10.3905/jfi.1997.408199

ИНФОРМАЦИЯ ОБ АВТОРАХ / ABOUT THE AUTHORS



Михаил Сергеевич Макушкин — аспирант, Базовая кафедра инфраструктуры финансовых рынков, НИУ ВШЭ, Москва, Россия
Mikhail S. Makushkin — postgraduate student, Department of Financial Market Infrastructure, National Research University Higher School of Economics, Moscow, Russia
<https://orcid.org/0000-0002-3491-1561>
 Автор для корреспонденции / Corresponding author:
mikhailmakushkin@gmail.com



Виктор Александрович Лапшин — кандидат физико-математических наук, доцент, Лаборатория по финансовой инженерии и риск-менеджменту, Школа финансов, НИУ ВШЭ, Москва, Россия
Victor A. Lapshin — Cand. Sci. (Phys.-Math.), Assoc. Prof., Financial Engineering and Risk Management Laboratory, School of Finance, National Research University Higher School of Economics, Moscow, Russia
<https://orcid.org/0000-0002-9396-4161>
vlapshin@hse.ru

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.
 Conflicts of Interest Statement: The authors have no conflicts of interest to declare.

Статья поступила в редакцию 22.08.2022; после рецензирования 26.10.2022; принята к публикации 26.11.2022.
 Авторы прочитали и одобрили окончательный вариант рукописи.
 The article was submitted on 22.08.2022; revised on 26.10.2022 and accepted for publication on 26.11.2022.
 The authors read and approved the final version of the manuscript.