

Оценка кривых доходностей на неполных данных

Виктор Лапшин, к.ф.-м.н.,

доцент НИУ ВШЭ, научный сотрудник Лаборатории по
финансовой инженерии и риск-менеджменту (FERM Lab).

Игорь Сорокин, аспирант ф-та ВМК МГУ им. М.В.
Ломоносова

- **Мотивация**
 - Теория
 - Данные
 - Примеры
- **Фильтрация некачественных данных**
 - Теория
 - Примеры
- **Восстановление пропущенных данных**
 - Теория
 - Примеры
- **Нерешённые задачи**
 - Примеры

- Задача оценки кривой бескупонной доходности:

- Теоретическая справедливая цена облигации

P_k – цена k -ой облигации

$$P_k = \sum_{i=0}^n F_{i,k} d(t_i)$$

$F_{i,k}$ – выплата по k -ой облигации в момент времени t_i

$d(t)$ – функция дисконтирования

- В случае непрерывного начисления процентов

$$d(t) = e^{-r(t)t}$$

- Теоретически проблемы хорошо известны:

- Количество неизвестных больше количества уравнений.

$$P_k = \sum_{i=0}^n F_{i,k} d(t_i)$$

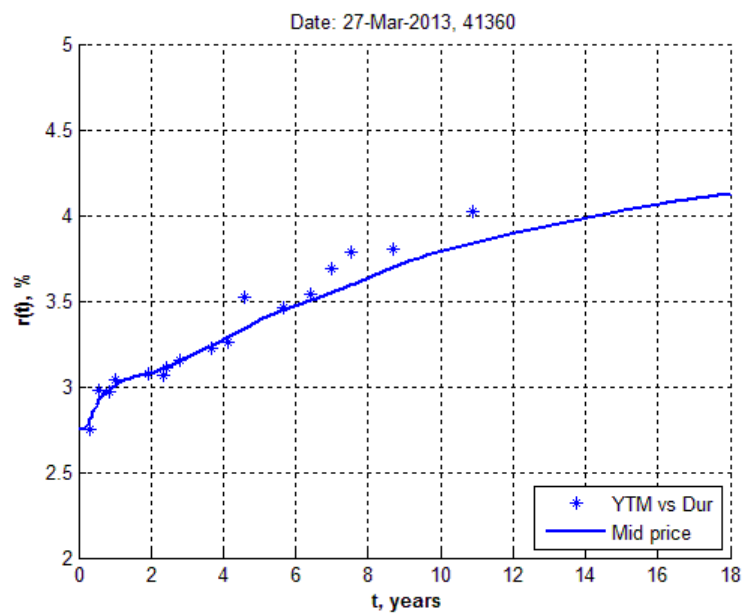
- ✓ Решение: параметрические и сплайновые методы.
- Неточность и недостоверность данных.
 - ✓ Фильтрация входных данных.
- Временное отсутствие данных (неликвидность рынка).
 - ✓ Восстановление пропущенных данных.

- Данные о ежедневных котировках закрытия государственных облигаций Польши на внебиржевом рынке. Предоставлены Cbonds.
- Встречаются все типичные проблемы.
- Далее будут показаны примеры особенностей данных и способов их обработки.

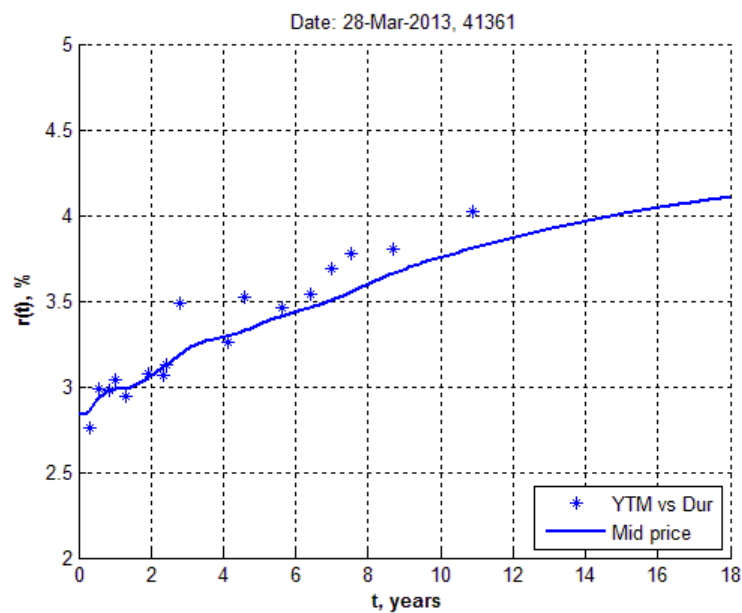
- Данные могут быть недостоверными:
 - Нерыночные сделки
 - Манипулирование рынком
 - Технологический сбой / программная ошибка
- Расчёт по данным, не отражающим объективную рыночную ситуацию, скорее всего, тоже не будет её отражать:

Garbage In – Garbage Out

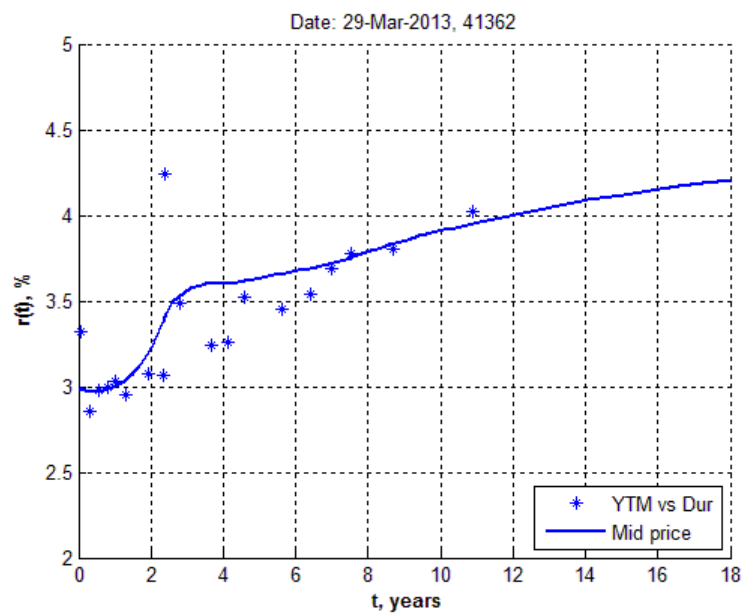
Кривая доходности



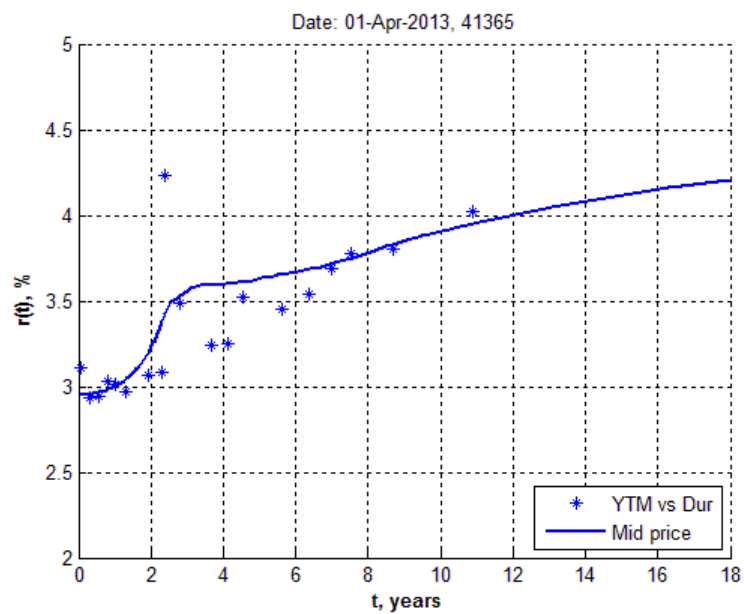
Кривая доходности



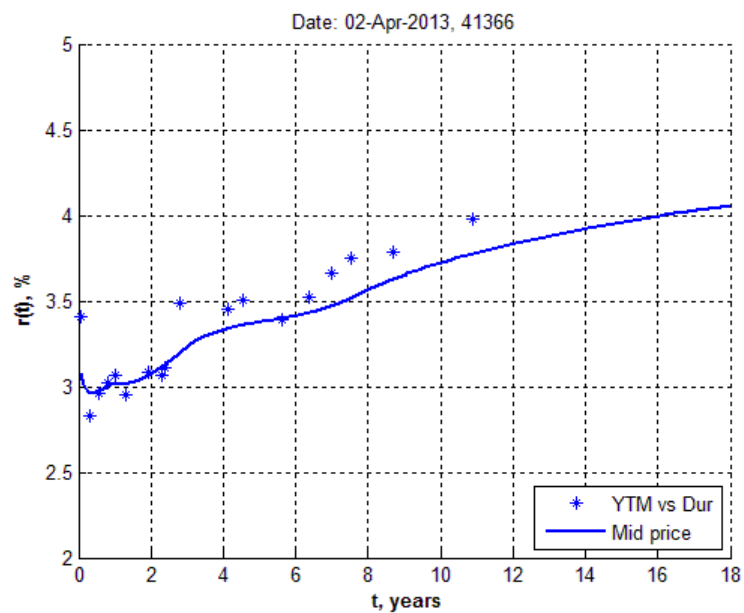
Кривая доходности



Кривая доходности



Кривая доходности



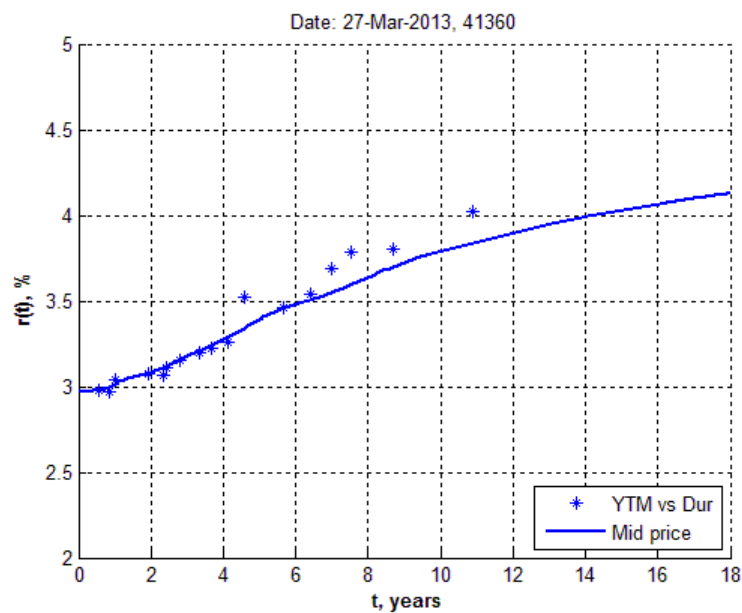
Фильтрация входных данных

- Простейший пример фильтрации — «правило трёх сигм»:
 - Отбрасывать все наблюдения, которые не укладываются в интервал
 $[m - 3\sigma; m + 3\sigma]$
 - Если пересчитывать параметры каждый день, будет простейшая динамическая фильтрация.
- Подходы, основанные на модели.
 - Динамическая модель:
$$dX_t = a_t(X, t)dt + b_t(X, t)dW_t$$
- Теория качества данных (Credibility Theory).

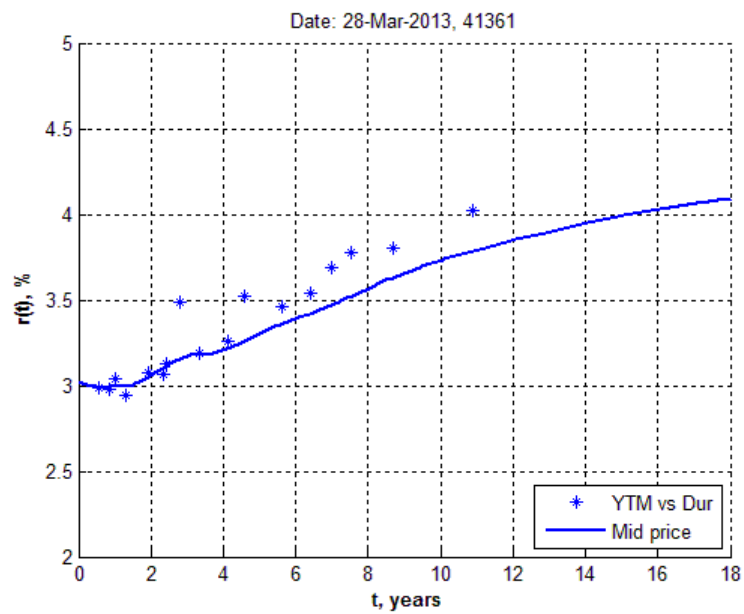
Кредит доверия

- Изначально каждая цифра снабжается некоторым кредитом доверия z .
- Он модифицируется с учётом сопутствующих показателей:
 - Нетипично низкие или высокие объёмы понижают кредит доверия.
 - Нетипично узкие или широкие bid-ask спреды понижают кредит доверия.
 - ...
- Затем кредит доверия переводится в «вероятность» q : от 0 до 1.
 - Либо отбрасываем данные при $q < c$, где c — пороговое значение.
 - Либо учитываем их с весом q .

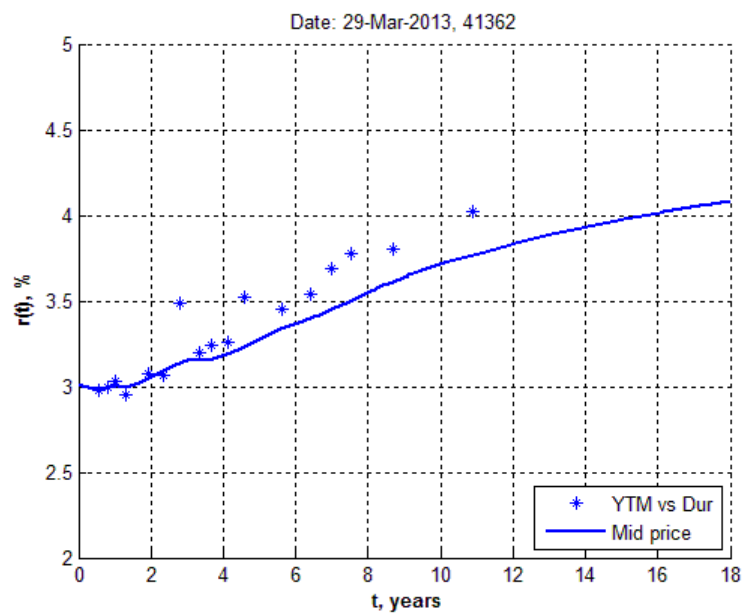
Кривая доходности



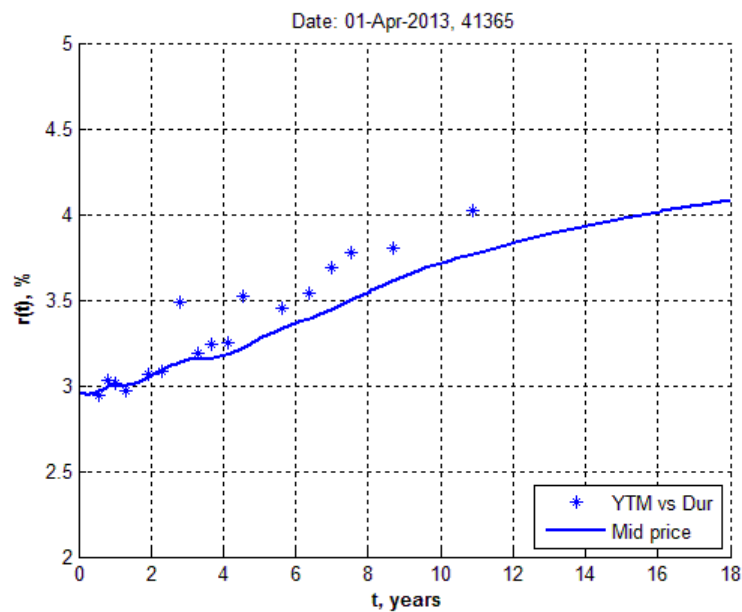
Кривая доходности



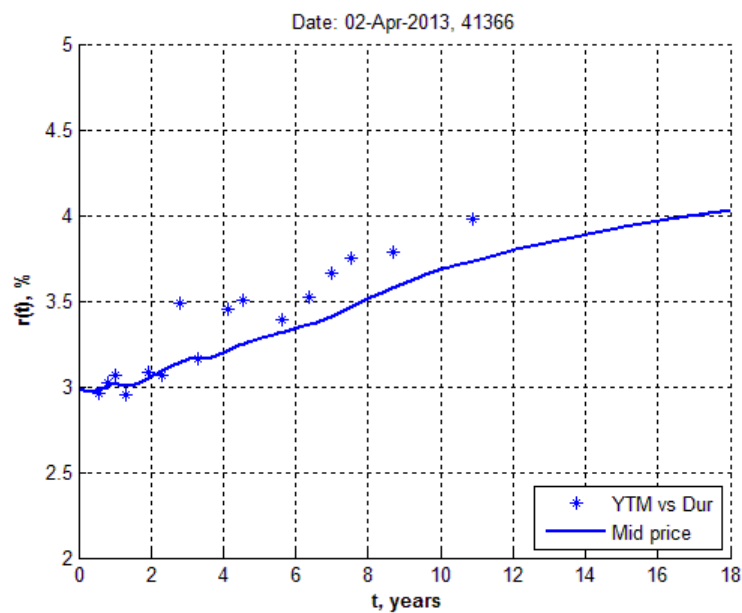
Кривая доходности



Кривая доходности

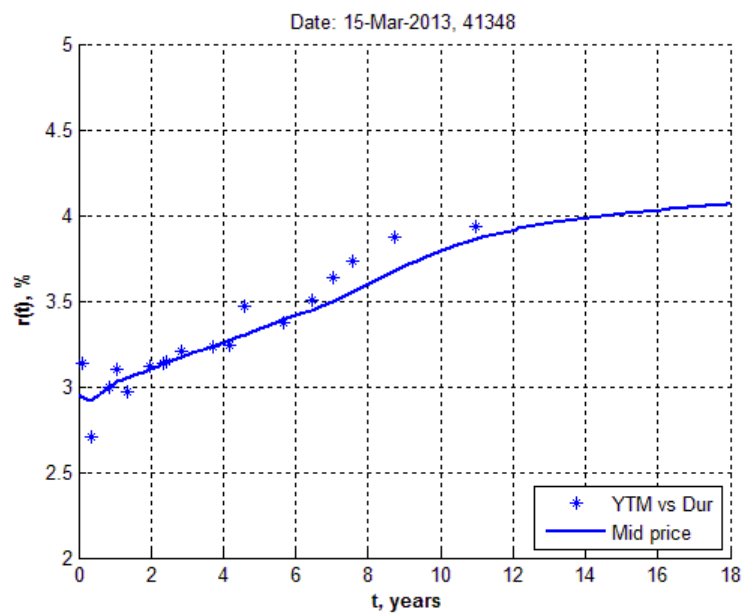


Кривая доходности

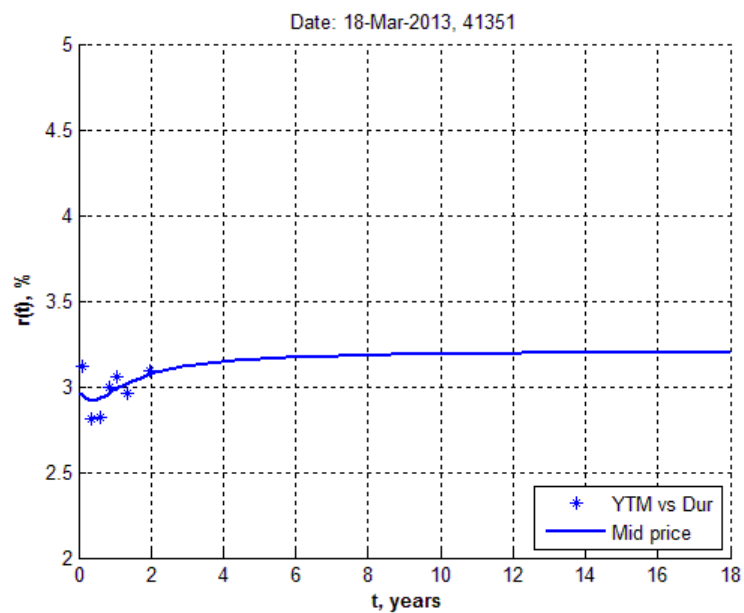


- Необходимые данные могут отсутствовать или быть доступными частично.
 - Низкая ликвидность рынка.
 - Технологический сбой / программная ошибка.
 - Данные были ранее признаны недостоверными.
- Если исключать из расчётов случаи, когда не хватает хотя бы одной цифры, может ничего не остаться.
- Использовать только доступные данные не стоит.

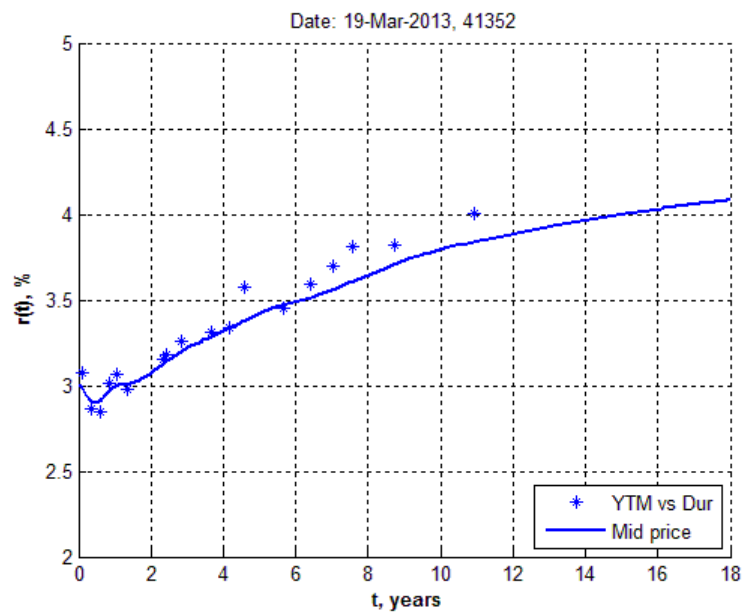
Кривая доходности



Кривая доходности



Кривая доходности



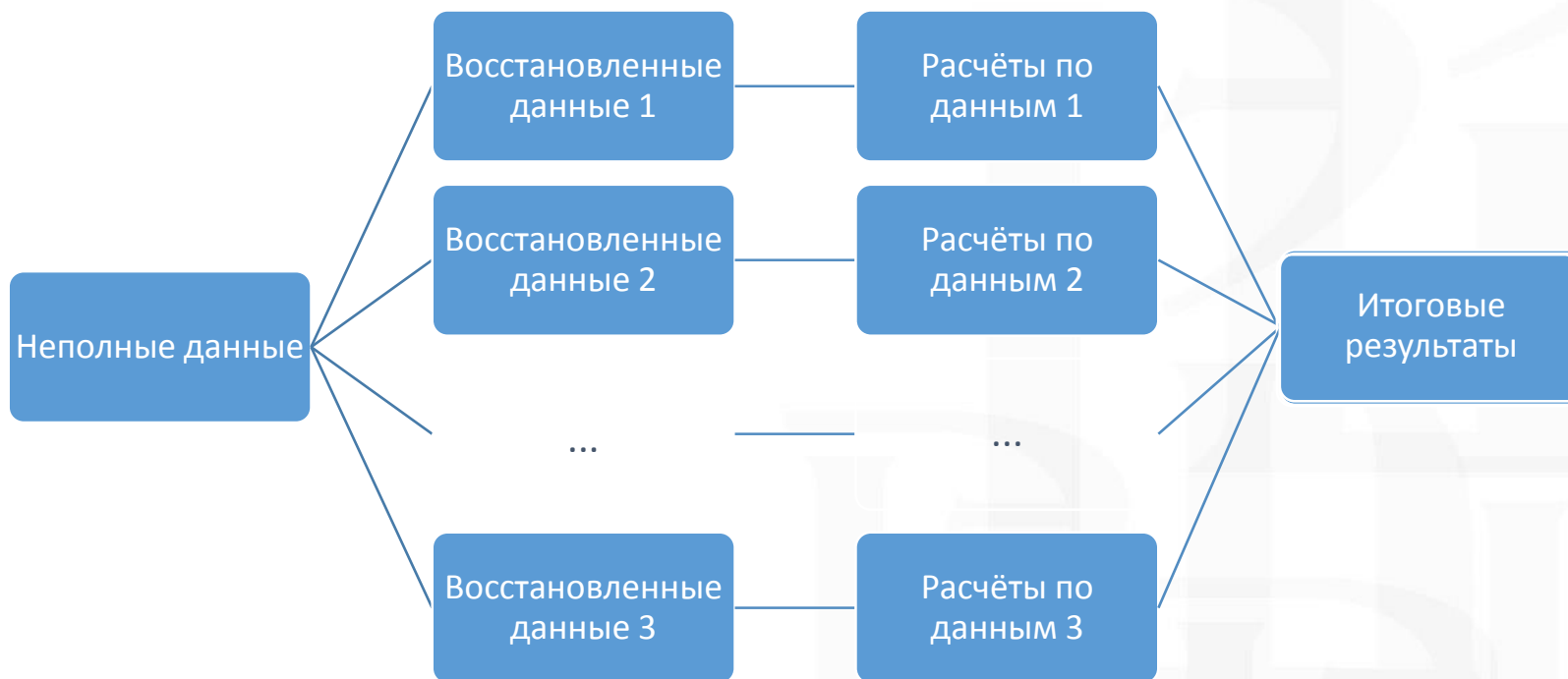
- Можно заполнить пропуски в данных. Как?
 - Заполнить предыдущим значением.
 - Линейная интерполяция.
 - EM-алгоритм: оценить параметры и заполнить математическими ожиданиями при оценённых параметрах.
 - Множественное восстановление (Multiple Imputation): оценить параметры и заполнить случайными числами из соответствующего распределения. Повторить несколько раз.
 - Одна из конкретных реализаций — методы Монте-Карло для марковских цепей (Markov Chain Monte-Carlo; MCMC).

- EM = Expectation – Maximization
- Нужно математическое ожидание пропущенных данных. Это легко, если мы знаем параметры процесса.
- Для оценки параметров процесса методом максимального правдоподобия нужны полные данные.
- Итеративный процесс: выбрать начальные приближения.
 - E: при текущих параметрах вычислить ожидания пропущенных данных.
 - M: при текущих значениях пропущенных данных оценить параметры.
 - Повторить до сходимости. Обычно 4-7 итераций.

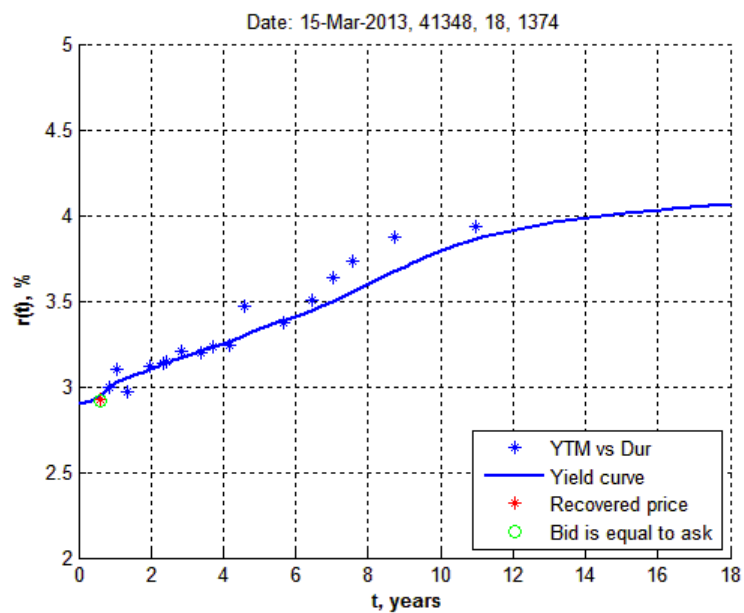
Markov Chain Monte-Carlo

- EM: вычисление средних пропущенных значений и наиболее вероятных параметров. Усреднение.
- Вместо этого будем восстанавливать случайные значения в качестве пропущенных и случайные же в качестве параметров.
- Формула Байеса: $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$
- Итеративный процесс: выберем начальные приближения.
 - При текущих параметрах выбрать случайные пропущенные значения с распределением $P(X_{mis} | X_{obs}, \Theta)$
 - При текущих пропущенных значений выбрать случайные параметры с распределением $P(\Theta | X_{mis}, X_{obs})$
- Результат — распределение, а не число.

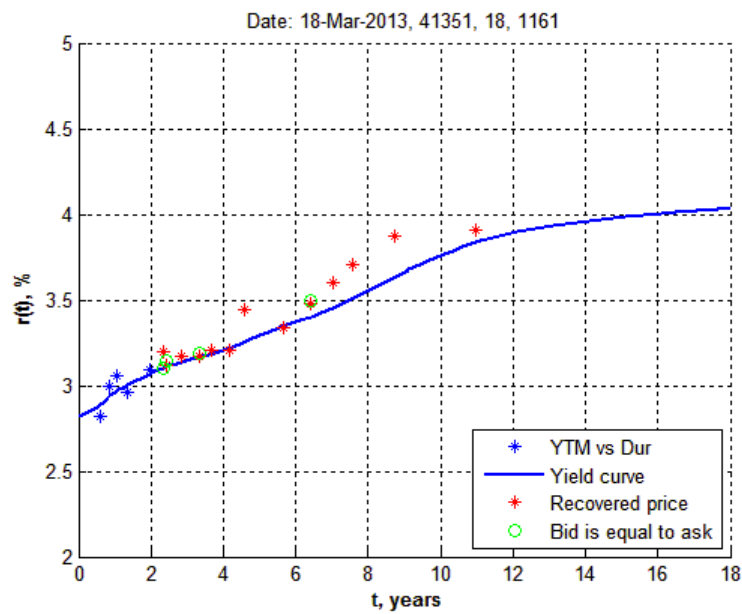
Схема Multiple Imputation



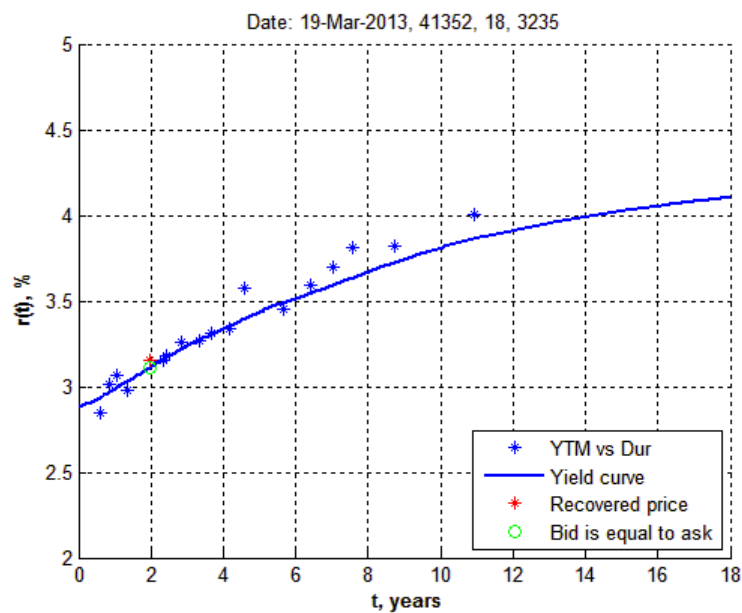
Кривая доходности



Кривая доходности

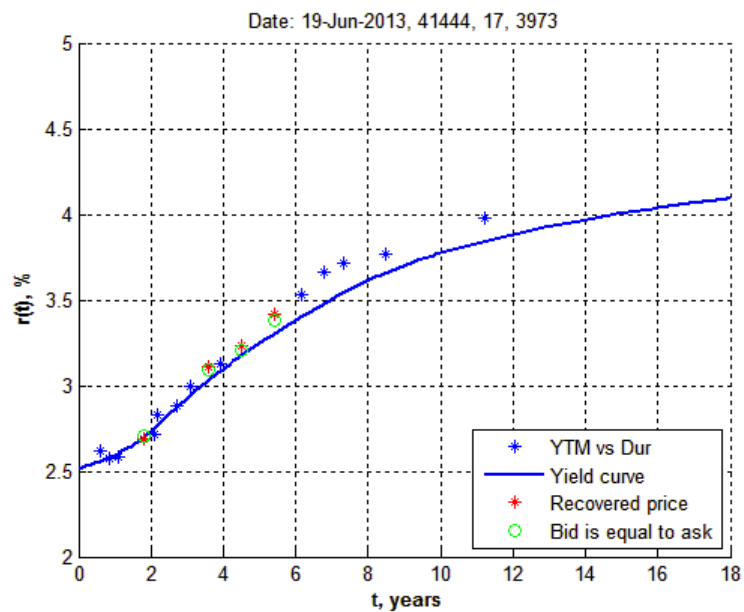


Кривая доходности

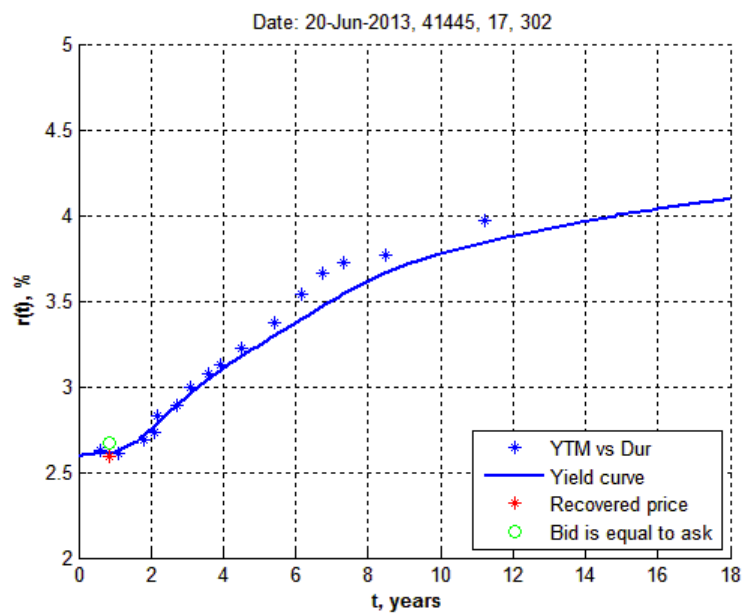


- К сожалению, наш подход — не панацея.
- Не всегда возможно определить, является ли значение выбросом или полезными данными. Даже эксперту.
- Подход на основе марковских цепей достаточно мощен и универсален, но не очевиден и требует затрат вычислительной мощности как на восстановление, так и на последующие расчёты.
- Тем не менее, он открывает уникальные перспективы в оценке рисков.

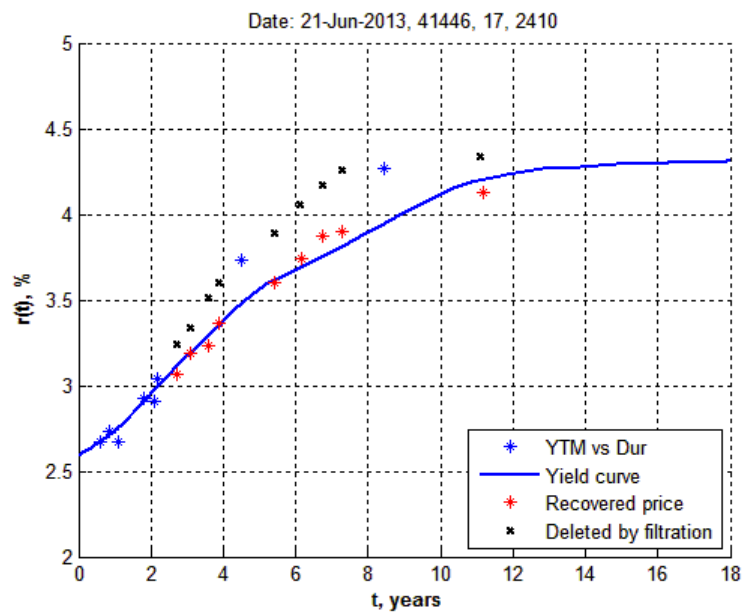
Кривая доходности



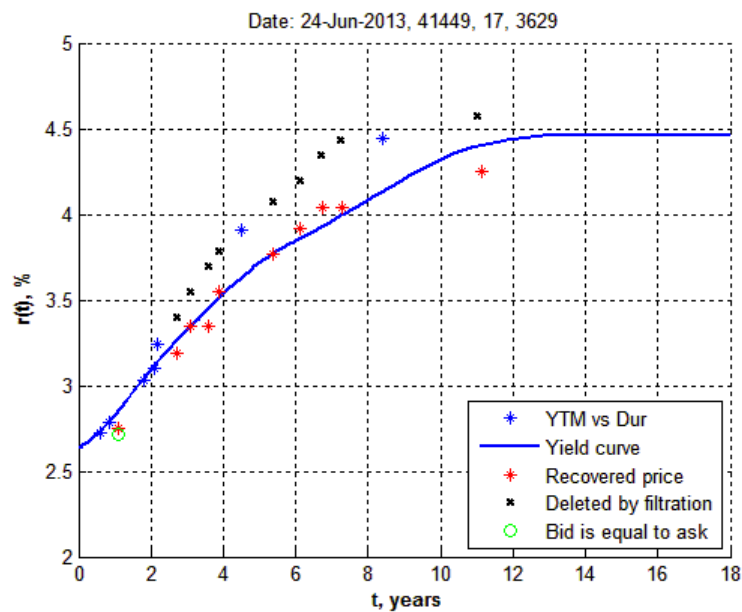
Кривая доходности



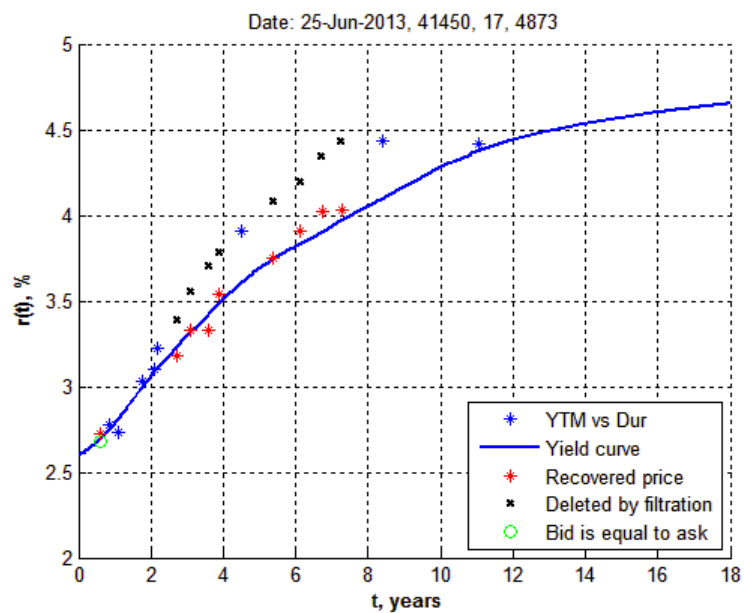
Кривая доходности



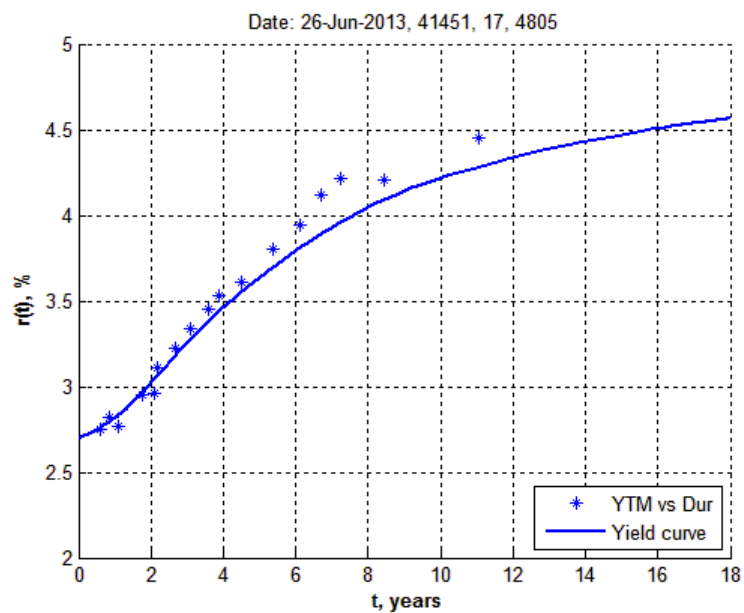
Кривая доходности



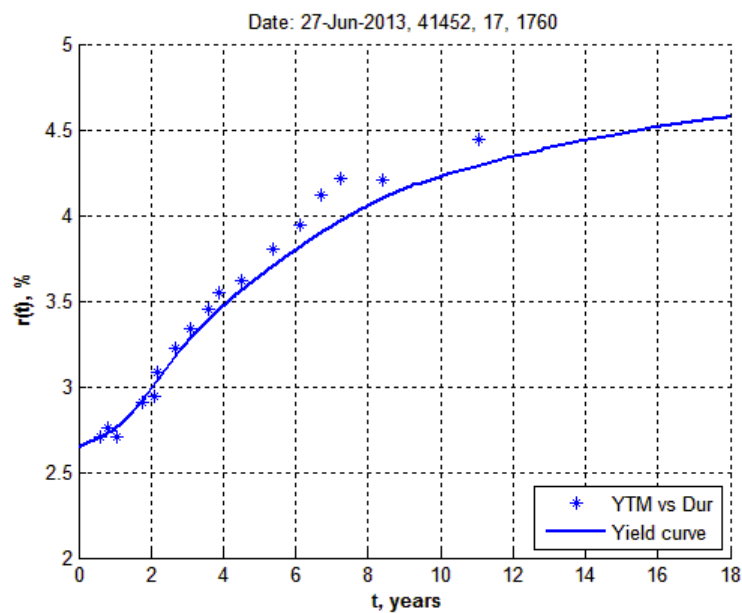
Кривая доходности



Кривая доходности



Кривая доходности



- Входные данные заслуживают огромного внимания.
 - Чем сложнее модель, тем большего внимания заслуживают данные.
- Фильтрация данных — больше искусство, чем наука.
 - В отличие от восстановления, которое, скорее, больше наука.
- Все пользуются одними и теми же данными, но цели у всех разные.
 - Всем нужна разная обработка данных.
- Любая модель нуждается в эксперте.



NATIONAL RESEARCH
UNIVERSITY

Thank you
for your attention!

vlapshin@hse.ru