Accepted Manuscript


Title: Inconsistencies in Bond Market Quotes: Is it the Wrong Model or the Wrong Data?

Author: Victor Lapshin

PII: S1877-7503(17)30552-5
DOI: http://dx.doi.org/doi:10.1016/j.jocs.2017.05.013
Reference: JOCS 682

To appear in:

Received date: 29-12-2016
Revised date: 17-4-2017
Accepted date: 10-5-2017

Please cite this article as: Victor Lapshin, Inconsistencies in Bond Market Quotes: Is it the Wrong Model or the Wrong Data?, *<![CDATA[Journal of Computational Science]]>* (2017), http://dx.doi.org/10.1016/j.jocs.2017.05.013

# Inconsistencies in Bond Market Quotes: Is it the Wrong Model or the Wrong Data?

Victor Lapshin

*National Research University Higher School of Economics, Moscow, Russia*

**Abstract**

We use the linear programming approach to quantify quote inconsistencies in risk-free bond markets. We present an algorithm to identify whether an inconsistency is probably due to the insufficient framework flexibility, the insufficient data quality, or the non-homogeneity of the dataset. In the latter case we study the problem of filtering out some instruments so that the remaining dataset be homogeneous. We show that the traditional filtering approach performs unacceptably poor and propose new algorithms. We find that the bonds, which get mispriced the most by a fitting algorithm, surprisingly are not the bonds, which cause the inconsistencies.

*Keywords:* quote inconsistency, data filtering, risk-free bonds, linear inequality system, approximate algorithm

*2000 MSC:* 91G80, 68W25, 90C90, 90-04, 90B99

## 1. Introduction

The notion of arbitrage plays a key role in the financial theory. Many, if not all asset pricing frameworks include a no-arbitrage supposition. However, when applied to the real data, some frameworks often result in arbitrage opportunities. This does not necessarily imply the existence of real-world arbitrage opportunities. Theoretical arbitrage opportunities (which we call inconsistencies in the paper to avoid the word 'arbitrage') may be caused by various factors, real examples of which will follow in the main body of the paper.

1. Insufficient framework flexibility. The framework might lack important

*Email address:* vlapshin@hse.ru (Victor Lapshin)

features like taxation, transaction costs or other hindrances. Or it could just be not accurate enough for this particular dataset.

2. Non-homogeneous data. The framework could be OK, but our assumptions about the data might be wrong. One particular case of assumption violation which we consider in this article is non-homogeneity of the dataset – the framework assumes that the dataset contains similar financial instruments, whereas in fact some of the instruments are fundamentally different and should not be part of this dataset.
3. Insufficient data quality. The data itself can be of poor quality – it might contain outliers, obsolete values, etc. This might result in the framework indicating that there are arbitrage opportunities while it is really the effect of data errors.
4. Real-world arbitrage opportunities in the market. This might happen eventually, but taxation, transaction costs, timing issues and other factors make it hard to identify for sure, especially only from historical data without a direct market access. It is also very easily confused with other cases mentioned above. In this paper we don't consider real-world arbitrage opportunities, as we are building an analytical tool for developing and validating pricing models, not a trading tool. We do not wish to make unfounded claims about real-world arbitrage opportunities. Our approach is able to detect real-world arbitrage opportunities among other inconsistencies. However, before claiming that a particular inconsistency is an arbitrage opportunity one has to account for taxes, transaction costs and other imperfections. This is too formidable a task for now, so we don't consider real-world arbitrage opportunities in this paper.

We study the inconsistencies in the government bond markets. We propose a simple approach to quantify the inconsistencies in the observed bond quotes with respect to a chosen pricing framework. We say that the bid and ask quotes of several bonds are consistent if within this pricing framework we can fit the data so that the predicted prices of all the bonds in the dataset lie between the observed bid and ask quotes of the respective bonds. The bid-ask spread is generally considered as a measure of observation errors in the price data, so we just require that the framework should be sufficiently flexible given the observation errors. To quantify the inconsistencies, we employ the notion of the tightness factor. The tightness factor is 2 if the bid-ask spreads of all bonds in the dataset have to be twice as large in order for the quotes to be consistent. This factor is 0.25 if the bid-ask spreads can be shrunk to be 4 times as narrow without losing consistency.

We also propose a way to distinguish between the possible inconsistency causes mentioned above. Data quality issues can be identified via analyzing the inconsistency patterns, because bad quotes usually happen at random. To distinguish between insufficient framework flexibility and non-homogeneity of the data, we solve the filtering problem – we filter out some instruments in a quasi-optimal way. If over time the same instruments tend to get filtered out, there is evidence that these instruments might be different in some way and should not be a part of this dataset at all. Of course, the ultimate decision should be made by a human. On the other hand, an erratic exclusion pattern is a strong indication that the framework is just not flexible enough for the dataset in question.

It turns out that the filtering problem mentioned above is not trivial, so we propose several algorithms and compare them with each other and with a traditional filtering algorithm using the real data. The experiments demonstrate the unacceptably poor performance of the traditional algorithm, which consists in excluding the most mispriced bonds one by one.

The main contribution of the paper is methodological. We propose a means to quantify the inconsistencies in bond quotes in an interpretable way and an algorithm to identify the probable cause of these inconsistencies via optimal filtering.

The presence of arbitrage in financial markets has been the subject of continuous research for many decades. The results depend greatly on the exact definition of arbitrage, i.e. on financial instruments and operations allowed, as well as on the financial market in question.

Our approach is after [5], but with several crucial extensions and modifications. The same modifications are attempted in [6], but in a slightly different way. Our approach is different in some details, but mainly in the presence of a financial interpretation. However, the main difference from the cited papers is that within approximately the same framework, we concentrate not on spotting the inconsistencies, but on what to do next – on quantifying them and identifying their probable cause. We propose an approach for filtering them out in an optimal way; for this, we study and compare several algorithms. We also use a richer dataset.

In [12] the authors proved that the no-arbitrage condition is equivalent to the existence of a positive pricing operator in a rather general setting. The notion of arbitrage bounds on the discount factors was introduced in [16] – a necessary condition for the existence of a pricing operator in the case of a risk-free bond market.

Empirical literature on arbitrage opportunities in stock and derivative markets is abundant, however fixed income markets have received far lesser attention. The US bond market was studied in [14]; they found that the US bond market assigns different values for coinciding cash flow patterns depending on the exact instruments making up the cash flow pattern. Among recent papers, [4] found the same anomaly in the UK bond market. However, there are few papers devoted to quantifying the arbitrage. Among these – [8, 5]. The primal-dual approach to bond market arbitrage with transaction costs dates back to [20, 13]. A more general primal-dual stochastic programming approach to arbitrage can be found in [18].

The remainder of the article is organized as follows. Section 2 formulates the pricing framework and the tightness factor, which is the key methodological element of the proposed approach. Section 3 presents the approximate filtering algorithms, which are also necessary for the approach. Section 4 tests the algorithms on the real data and justifies the choice of one of them for practical purposes. Section 5 presents the algorithm to classify inconsistencies and includes real-world examples of almost every inconsistency type mentioned in the algorithm. Section 6 discusses methodological applications to tasks such as data filtering and model validation.

## 2. The framework

In what follows, we talk about 'the framework'. We distinguish between a bond pricing framework (e.g. discounted cash flows with no credit risk and perfect liquidity) and a term structure fitting model within this framework, e.g. Nelson-Siegel or smoothing splines. The discounted cash flows framework, which we describe below in detail, is popular, but it is not always applicable. In [15] the author shows that defaultable coupon bonds should not be in general considered as portfolios of defaultable zero-coupon bonds and provides conditions as to when this representation holds. As examples of alternative frameworks, one could consider including nonzero credit risk and/or bond liquidity into account.

Note that we don't fix the exact term structure fitting model within the framework, i.e. we consider the most general case possible. Our approach could thus be termed 'model-free'.

We illustrate our approach on the most simple and widespread bond pricing framework – discounted cash flows. However, it is easily applicable to other pricing frameworks, such as assuming nonzero default risk taking bond liquidity into account. However, for illustration purposes we use the simplest alternative, which is discounted cash flows only.

In order to show the origin and the intuition behind the proposed notions, we briefly repeat the derivation of the discounted cash flows pricing framework while highlighting the role played by the bid-ask spreads. We sincerely hope that this brief excursion to the arbitrage – pricing duality will help applying our approach to more advanced frameworks along the same lines.

The discounted cash flows pricing framework is based on the no-arbitrage principle and the unlimited short selling assumption together with the ability to hold cash without interest. We now recall the notions of arbitrage and sequential arbitrage from [5] and references therein with a slightly modified notation to better suit our exposition.

Consider $N$ traded bonds indexed by $k = 1, 2, ..., N$ with the corresponding present prices $P_k$. Let $F_{i,k}$ be the cash flow promised by the $k$-th bond at time $t_i,\ i = 1, 2, ..., n$. Without loss of generality we may consider $t_i$ to be the same for all bonds by introducing zero cash flows when necessary. We assume that $t_i$ are already expressed in year fractions using the day count conventions applicable.

We also assume that the promised future cash flows are viewed and priced by the market as guaranteed. This is usually the case for government bonds – even if they are not truly risk-free, the market is usually generally inclined to regard them as such. Another important assumption which we make is no trading restrictions. By the usual financial argument, if an arbitrage deal requiring short-selling a bond existed, a market participant having that particular bond in stock would be able to engage in this arbitrage – this would constitute a simple and easily feasible portfolio rebalancing for this participant.

**Definition 1.** *A market model is said to be arbitrage-free (respectively, sequential arbitrage-free) if for any portfolio vector $w$ such that $F^T w \succeq 0$ ($AF^T w \succeq 0$), where $x \succeq y$ means $x \geq y$ element-wise, but $x \neq y$, and*

$$A = \begin{pmatrix} 1 & 0 & 0 & ... & 0 \\ 1 & 1 & 0 & ... & 0 \\ & & & ... & \\ 1 & 1 & 1 & ... & 1 \end{pmatrix},$$

*we have $-P^T w < 0$.*

This defines arbitrage (resp. sequential arbitrage) as taking a position, which does not require any funds: $-P^T w$ is the initial net outcome of taking the position (negative when paying), which generates non-negative cash flows (resp. whose negative cash flows can be absorbed by previous positive cash flows).

We modify this definition to include the bid-ask spreads into consideration.

**Definition 2.** *A market model is said to be arbitrage-free (respectively, sequential arbitrage-free) if for any portfolio described by the vector of long positions $w^{long}$ and the vector of short positions $w^{short}$ such that $F^T(w^{long} - w^{short}) \succeq 0$ ($AF^T(w^{long} - w^{short}) \succeq 0$), we have $-a^T w^{long} + b^T w^{short} < 0$, where $a$ and $b$ are the vectors of ask and bid quotes of the corresponding bonds.*

As it turns out, this slight modification offers a significant simplification in the primal-dual approach to arbitrage of [5]. In what follows we consider sequential arbitrage. Moreover, we modify its definition slightly to conform to the financial intuition (this corresponds to the sequential arbitrage portfolio of the second type in [5]):

**Definition 3.** *A market model is said to be sequential arbitrage-free if for any number $\xi \geq 0$ and any portfolio described by the vector of long positions $w^{long}$*

*and the vector of short positions* $w^{short}$ *such that* $AF^T(w^{long} - w^{short}) \succeq -\xi 1$, *we have* $-a^T w^{long} + b^T w^{short} < \xi$. *Here* 1 *is the vector of ones.*

The only innovation in definition 3 is that the arbitrageur is now allowed to hold cash amount $\xi$ gained from the initial position taking, to cover future negative cash flows, whereas by definition 2 the arbitrageur must invest all proceedings from taking short positions into market instruments.

For example, the following scheme is an arbitrage by definition 3, but is not by the definition 2. Consider a single bond, which pays 1 at time $t = 1$ ($F = 1$) and is now traded at both bid and ask quotes of 2 ($a = b = 2$). Since we only have one time moment, $A = 1$ and the 'sequential arbitrage' part of definition 2 is redundant and coincides with the first part. To verify definition 2, we need to prove that for any numbers $w^{long} \geq 0$ and $w^{short} \geq 0$ such that $F^T(w^{long} - w^{short}) = 1 \cdot (w^{long} - w^{short}) > 0$ we have $-a^T w^{long} + b^T w^{short} = 2 \cdot (w^{short} - w^{long}) < 0$. This is of course true. Therefore, this market situation is not an arbitrage in terms of definition 2.

However, definition 3 classifies this as an arbitrage opportunity, because there exist $\xi = 1.5 \geq 0$, $w^{long} = 0$, $w^{short} = 1$ such that $AF^T(w^{long} - w^{short}) = 1 \cdot (-1) \geq -1.5$ and $-a^T w^{long} + b^T w^{short} = 2 \geq 1.5$.

In order to quantify the arbitrage, we consider the following optimization problem

$$\begin{cases} -a^T w^{long} + b^T w^{short} - \xi \to \max, \\ -AF^T(w^{long} - w^{short}) - \xi 1 \leq 0, \\ \delta^T(w^{long} + w^{short}) = 1, \\ w^{long}, w^{short}, \xi \geq 0, \end{cases} \tag{1}$$

the decision variables being $(w^{long}, w^{short}, \xi)$, and 1 being the vector of ones. It corresponds to finding the initial portfolio, which includes holding $\xi$ in cash, such that its initial financial outcome is maximal and its promised negative cash flows are absorbed by the preceding payments and the cash held. The normalizing condition $\delta^T(w^{long} + w^{short}) = 1$ is needed because without it the optimal value of this problem is either 0 (in case of no arbitrage) or $\infty$ (if even a slightest arbitrage opportunity exists, it may be scaled up infinitely).

There are no restrictions on the normalizing vector $\delta$, provided that $\delta > 0$, but we'll see later that it has a financial interpretation.

The dual problem to (1) is

$$\begin{cases} \gamma \to \min, \\ -FA^T c + \gamma\delta \geq -a, \\ FA^T c + \gamma\delta \geq b, \\ -1^T c \geq -1, \\ c \geq 0 \end{cases}$$

with the dual decision variables being $c$ and $\gamma$. Substituting new variables $d = A^T c$ and reverting the signs, we come to the following problem

$$\begin{cases} \gamma \to \min, \\ b - \gamma\delta \leq Fd \leq a + \gamma\delta, \\ 1 \geq d_1 \geq d_2 \geq ... \geq d_n \geq 0. \end{cases} \tag{2}$$

This problem also has a clear financial interpretation: the decision variables $d$ are the discount factors corresponding to the moments in time $t$. The introduction of $\xi$ was necessary to guarantee $1 \geq d_1$. We are minimizing the maximum mispricing cost $\gamma$ given that $\delta$ measures the cost of mispricing different bonds: small $\delta_k$ makes the optimal program replicate the price of the $k$-th bond closer to the bid-ask bounds. We can thus interpret $\delta_k$ as the observation error for the price of the $k$-th bond. So the most intuitive proxy for the price observation error is the bid-ask spread itself, since the fair price should lie between bid and ask quotes. We could let $\delta = a - b$, but for the later convenience we choose $\delta = \frac{1}{2}(a - b)$, which only affects the numerical values of $\gamma$.

Note that if some bonds have zero bid-ask spreads, then the corresponding elements of $\delta$ will be zero. It does not automatically result in the problem being infeasible, but it makes it harder and in general increases the value of $\gamma$, because the prices of bonds with zero bid-ask spreads have to be replicated exactly in any case with this choice of $\delta$, even if other bonds are allowed pricing discrepancies of $\pm\gamma \cdot \frac{1}{2}(a - b)$.

If this is not the desired effect of if one has better prior knowledge of data credibility or the associated transactional costs, it can easily be incorporated by modifying $\delta$. For example, one could search for minimal proportional transaction cost level eliminating the arbitrage in the data. $\delta$ would represent the relative transaction costs for various bonds in this case. In this example, $\delta = \delta_0 1 + \frac{1}{2}(a-b)$ could be interpreted as transactional costs $\delta_0$ per transaction for one bond in addition to the bid-ask spread.

Another very powerful interpretation of $\delta$ arises for $\delta = \frac{1}{2}(a-b)$. We now consider the quote data unreliable (this can happen if the quotes are non-commiting or asynchronous) and seek to measure the credibility of the price data by the effective bid-ask spread (that is, the spread allowing for the no-arbitrage condition to hold). So in this approach, instead of questioning the market efficiency we question the numerical data.

In this case, the optimal value $\gamma^*$ has the following nice interpretation – $1+\gamma^*$ is the factor by which one should widen the bid-ask spreads in order to ensure consistency of the dataset. Throughout the rest of the paper, we call this quantity the tightness factor and denote it by $\theta$. $\theta \leq 1$ means the dataset is consistent (one can even tighten the bid-ask spreads if $\theta < 1$) and $\theta > 1$ means the bid-ask spreads are too tight and need to be widened in order to get rid of the inconsistencies. The tightness factor has been introduced in [22] within a different (and more complicated) pricing framework. However our approach to it allows for more intuition.

The tightness factor may also provide a valuable insight to the data quality and market efficiency, as it turns out that $\theta > 1$ is quite a common situation for some markets. This of course does not mean that there are arbitrage opportunities, but suggests the need for either some preliminary data treatment or a more flexible pricing framework.

This paper deals with the following problem: given that the quotes are inconsistent, find the probable cause of it. To do so, we first search for the exact bonds 'responsible' for the inconsistency, i.e. identify the mispriced bonds. We state it as follows: given that $\theta > 1$, find the minimal set of bonds which, when excluded from the dataset, makes $\theta \leq 1$.

## 3. NP-equivalence and approximate algorithms

In the previous section, we have quantified the quote inconsistency via the tightness factor – the smallest widening of bid-ask spreads necessary. Doing so is one way of pre-processing the data to ensure its consistency. However, there is another way – we can filter out some instruments from the dataset in order for the remaining to be consistent. The problem is choosing which exact instruments should be filtered out. Presumably, we'd like to throw away as little data as we possibly can, so the problem becomes to remove the least possible number of bonds from the dataset to make it consistent.

In this section we show that this problem is NP-equivalent. To do so, we first transform this problem slightly. Note that $\theta > 1$ is equivalent to the infeasibility of the following set of constraints:

$$\begin{cases} b \leq FA^T c \leq a, \\ 1^T c \leq 1, \\ c \geq 0, \end{cases} \tag{3}$$

where $c_i = d_{i-1} - d_i$ assuming $d_0 = 1$. Now the problem states: remove the minimum number of double inequalities from the system (3) so that it become feasible. This problem looks like the known problem of finding the maximum feasible subsystem, which has been shown to be NP-hard in [21, 7]. In [2] it was shown that this property still holds for systems of homegeneous inequalities (both strict and not). Similar approximate algorithms were studied in [2, 3, 1]. [19] described an exact solution of a similar problem through integer programming subproblems, which was used in [10] to construct another approximate algorithm. Chinneck [9] proposed another approximate algorithm. A more detailed exposition of the history of this problem may be found in [11]. It also bears a certain resemblance to the problem of finding the maximum zero of a boolean function [17].

Unfortunately, our problem differs in details – it only allows elimination of double inequalities, but not the normalizing constraint or the positivity constraints, so these results for a general system are inapplicable; thus, a separate proof is in order. It is however not very enlightening, so we leave it in the appendix. Having proven that the exact solution is very hard to find, we may now resort to searching for an approximate solution.

We analyze the 'traditional' industry standard algorithm for this task and propose two other algorithms. We argue that a reasonable approximate algorithm must have a clear financial intuition behind it. Our algorithms are all greedy algorithms (iterative, making a locally optimal choice on each iteration), eliminating bonds from the dataset one by one until the remaining set becomes consistent. The difference lies in the heuristics used to determine the next bond to eliminate. These heuristics have to have a clear financial interpretation in order for the algorithm to be acceptable to the management or to the general public.

**Algorithm 1.** *(Traditional filtering approach) This is the obvious algorithm of choice for the data filtering problem and a de-facto industry standard for filtering the data. At each stage we fit some term structure model $M$ (e.g. a Nelson-Siegel model) to the bond price data. After that we eliminate the bond $k^*$ which is mispriced the most by the fit. In order for the algorithm to be consistent with our setting, the fitting and the choice of the most mispriced bond have to be carried out relative to the bid-ask spreads of the bonds:*

$$d^*(\cdot) = \arg \min_{d(\cdot)\in M} \sum_{k=1}^{N} \left( \frac{1}{a_k - b_k} \left[ \sum_{i=0}^{n} F_{i,k} d(t_i) - P_k \right] \right)^2 ;$$

$$k^* = \arg\max_k \frac{1}{a_k - b_k} \left| \sum_{i=0}^{n} F_{i,k} d^*(t_i) - P_k \right| .$$

**Algorithm 2.** *(Arbitrage filtering approach) At each stage we solve the primal problem (1) to get the arbitrage portfolio loadings $w^{short}, w^{long}$ and exclude the bond which has the largest weight in the arbitrage portfolio:*

$$k^* = \arg\max_k \left( w_k^{short} + w_k^{long} \right) .$$

*We also record all bonds having comparably large weights (say, 90% of the maximum) for the candidate search algorithm below and for plotting the inconsistency pattern.*

**Algorithm 3.** *(The steepest descent) At each stage we try excluding every bond and choose the one which, when excluded, decreases the tightness factor the most. To speed this search up, note that we only need to consider the bonds belonging to the arbitrage portfolio or, in the dual setting (2), for which the bid-ask bounds on the model prices are binding. We also record all bonds corresponding to comparably large decrease in the tightness factor for the candidate search algorithm below and for plotting the inconsistency pattern.*

Note that the algorithm 2 works in the primal space of arbitrage portfolios. The traditional algorithm 1 works in the dual space of discount factors and bond prices. Finaly, the algorithm 3 is primal/dual insensitive (since the optimal values for the primal and the dual problems are equal), instead it looks one step ahead at each iteration.

It turns out that the algorithm 3 often ends up with several elimination candidates. This happens because when two bond quotes are conflicting, after any one of them is removed the other ceases to impose binding constraints. For the clearness, we choose the one with the largest weight in the arbitrage portfolio like in the algorithm 2.

To assess the performance of the algorithms, we conduct a direct exhaustive search to find the true minimum number of bonds to be removed (this took a class of 25 relatively modern PCs about a week). Motivated by a heuristic insight, we also consider the two limited-domain search algorithms.

**Algorithm 4.** *(Restricted search) Perform exhaustive search, but only among the bonds* <u>*picked*</u> *on all steps by algorithms 2 or 3 above.*

**Algorithm 5.** *(Candidate search) Perform exhaustive search, but only among the bonds* <u>*considered*</u> *on all steps by the algorithm 3 (all bonds, whose removal at some stage lead to the same decrease in the tightness factor).*

## 4. Testing approximate algorithms

In this section we compare our approximate algorithms using real data on bond quotes. Bonds are usually traded over the counter (OTC) in developed markets, which makes the bid/ask quote data hard to aggregate and interpret. However, bonds are exchange-traded in a few countries, among which are Russia and China. Chinese data turned out to be quite unreliable (bid quotes exceeding ask quotes unfortunately were normal for this dataset) and therefore unfit for our purpose. We also do the calculations on government bonds of some major euro zone sovereign borrowers: Austria, Belgium, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Slovakia, Spain, and Portugal.

Russian bonds data is provided by Cbonds.info, a parter with the Moscow Exchange, and covers the period from January 2008 to March 2015. Euro zone bonds data is provided by Markit and comprises quotes of bonds making up the iBoxx index – these are the most liquid sovereign bonds denominated in EUR. It spans the period from January 2007 to December 2011, except for Greece (to June 2010) and Portugal (to July 2011). Both datasets include daily bid and ask quotes as well as all necessary bond specifications such as cash flow timetables.

Table 1 shows the percentage of days with inconsistencies along with the mean and maximum values of the tightness factor only for inconsistent days.

Table 1: Basic inconsistency statistics.

| Dataset | Inconsistent days | In % | Mean TF | Max TF |
|---|---|---|---|---|
| Austria | 0 | (0%) | n/a | 0.00 |
| Belgium | 154 | (12%) | 1.37 | 2.74 |
| Finland | 0 | (0%) | n/a | 0.00 |
| France | 36 | (3%) | 2.13 | 3.02 |
| Germany | 202 | (16%) | 1.83 | 15.99 |
| Greece | 71 | (8%) | 1.29 | 5.46 |
| Ireland | 0 | (0%) | n/a | 0.27 |
| Italy | 1266 | (99%) | 6.95 | 34.24 |
| The Netherlands | 31 | (2%) | 1.11 | 1.32 |
| Portugal | 0 | (0%) | n/a | 0.91 |
| Russia | 209 | (10%) | 6.58 | 158.04 |
| Slovakia | 0 | (0%) | n/a | 0.00 |
| Spain | 3 | (0%) | 3.40 | 7.85 |

Since no inconsistencies are found for Austria, Finland, Ireland, Portugal, and Slovakia, we exclude these datasets from the subsequent analysis. These datasets are perfectly consistent with the discounted cash flow pricing framework. Judging by the tightness factor, we are expecting the most interesting results from German, Italian and Russian datasets.

Once again we should note that an inconsistency is not necessarily a real-world arbitrage (one could hardly expect as much in a highly liquid German bond market). But theoretical arbitrage opportunities expose that a framework with such assumptions should not be used for this data. As we already men-

Table 2: Basic optimal solution statistics.

| Dataset | Inconsistent days | Mean N | Max N |
|---|---|---|---|
| Belgium | 154 | 1.00 | 1 |
| France | 36 | 1.00 | 1 |
| Germany | 202 | 1.02 | 2 |
| Greece | 71 | 1.01 | 2 |
| Italy | 1266 | 2.23 | 5 |
| The Netherlands | 31 | 1.00 | 1 |
| Russia | 209 | 1.12 | 3 |

tioned in the introduction, there are several possible explanations for it, which will be discussed later in greater detail.

Table 2 shows the statistics for $N$ – the exact minimum number of bonds to be eliminated, conditioned on it being greater than 0.

Table 3 shows the performance statistics of all methods on the least consistent dataset – Italian. It reports the number of dates for which each of the algorithms could not find the optimal $N$, and only for these dates it also reports the mean and the maximum excess $N$. Results on the remaining datasets are reported in the appendix.

Table 3: Performance statistics for approximate algorithms. Dataset: Italy.

| Algorithm | Non-optimal days | In % | Mean excess $N$ | Max excess $N$ |
|---|---|---|---|---|
| Traditional (No 1) | 1253 | (99%) | 8.36 | 37 |
| Arbitrage (No 2) | 707 | (56%) | 1.35 | 4 |
| Steepest Descent (No 3) | 262 | (21%) | 1.19 | 4 |
| Restricted Search (No 4) | 44 | (3%) | 1.05 | 2 |
| Candidate Search (No 5) | 14 | (1%) | 1.00 | 1 |

As we can see, the traditional filtering algorithm 1 is virtually incapable of identifying the bonds to be filtered out. So regardless of this algorithm's appealing financial interpretation, it is obvious that the bonds which get mispriced by the term structure fitting algorithm are not the bonds, which are actually introducing the inconsistencies.

We also see that the steepest descent algorithm 3 on average performs better than the arbitrage algorithm 2. However, there are several exception dates when it actually performs worse.

It should also be noted that both restricted search algorithms 4 and 5, while able to find the exact solution in the 'lightweight' settings (see the appendix for the reports on the remaining datasets), could not always reach the optimum in the Italian case. This means that the approximate algorithms 3 and 2 don't necessarily even consider all the bonds constituting the optimal solution.

### *4.1. Comparison with previous findings*

An anonymous referee noted that our results seem to contradict some of the previously documented findings, especially on the German market. However, we argue that in fact there is no contradiction, because in our framework (we actually share the same framework with [6]) the results greatly depend on the exact dataset used and on the details of data preprocessing. To illustrate this, we compare the setting of our work with the one of [6], which deals with a similar problem for approximately the same dataset.

Table 4 presents a brief comparison of the data and the results; we only list the differences crucial for the German market to save the space.

As we can see, the two studies differ significantly in both the instruments in the dataset and the preprocessing. Therefore it is not surprising that the percentage of the inconsistencies spotted differ significantly. We expect this to be the case in comparisons with other studies too.

Let us now turn to the main objective of the paper – identifying the probable cause for the inconsistencies. The algorithms described above will play a crucial role in this task.

## 5. Possible causes for inconsistency

We start this section by listing the possible reasons for the dataset to be inconsistent. These are all well-known causes. We continue with presenting some informal tests to choose the most probable one from this list.

Table 4: Comparison with [6]: datasets and results.

| Criterion | In [6] | In this paper |
|---|---|---|
| Data provider | Reuters | Markit |
| Bonds selection criteria | Drop the bond if its price is exactly the same as the day before (a liquidity proxy) | The bonds constituting the Markit iBoxx (EUR) index (the inclusion criteria are based on liquidity) |
| Nature of the data | Contributor quotes, Reuters' own composite valuation service | Contributor quotes |
| Data preprocessing by vendor | Drop the highest and the lowest values and then average the rest | A complicated algorithm[1] |
| Data preprocessing by authors | Manual outlier removal | None |
| Arbitrage spotted in | About 75% of trading days | About 15% of trading days |

### *5.1. List of possible causes*

1. **Data errors.** It is not uncommon even for commercial data sources to eventually contain wrong data, either because of technical problems or due to human errors. Data errors are by definition random, relatively rare and usually don't depend on the data.
   On the other hand, if data errors tend to happen often and/or in predictable patterns, e.g. for some instruments the data is less reliable than for other, then the issue should probably not be classified as 'data errors'. Predictable patterns for data errors can (and often should) be incorporated within a model. For example, we could introduce more aggressive filtering or search for a feature which correlates with the observed errors to correct the pricing model.

[1] See Markit Bond Price Consolidation Rules at https://products.markit.com/indices/

2. **A heterogeneous dataset.** Data selection problems resulting in heterogeneous datasets and inconsistencies may come in various subtypes. We discuss some of them below.
   (a) **'Outsider' bonds.** The dataset may be non-homogeneous with several 'outsider' bonds in it. These bonds are usually few and were not meant to be included in this dataset (either a priori or after additional consideration). For example, we could have thought that including a single callable bond in the dataset would not harm, but if in reality this bond turned out to be sufficiently different, then we have a clear case of an 'outsider' bond.
   (b) **Heterogeneous market.** In this case the 'outsiders' are sufficiently numerous and may be considered constituting a separate distinct part of the market with its own different pricing rules. More exactly, there is a specific instrument feature allowing us to split the dataset into two or more consistent parts. For example, in many countries government bonds are priced differently than corporate bonds even with comparable level of risk due to taxation differences or regulatory preference. Therefore, a dataset consisting of government and corporate bonds, even with comparable credit risk and liquidity, would still be heterogeneous. The 'taxation regime' feature of the bonds in the dataset would generate a splitting criterion.
3. **Insufficient framework flexibility.** The flexibility of the framework should ideally be appropriate for the data. An inconsistency does not necessarily mean that the framework is bad. It just means that this particular framework is not flexible enough to allow models which explain this particular data. One should distinguish between two kinds of insufficient flexibility (we thank an anonymous referee for pointing this out).
   (a) **An insufficiently precise fitting model within a sufficiently flexible pricing framework.** An example could be using a 4-parameter Nelson-Siegel model to fit the term structure of interest rates where a 6-parameter Svensson model would be appropriate or using a spline with too few knots (given that the model with higher

download/products/guides/Markit_iBoxx_EUR_Bond_Price_Consolidation_Rules.pdf number of knots works just fine). In this case, the framework is flexible enough, but the chosen fitting model is too restrictive. We don't consider this case in our paper, because we don't work with term structure fitting models here, we only consider general pricing frameworks.

(b) **An insufficiently flexible bond pricing framework.** This is the case where our pricing framework is missing an important instrument feature, which does not enter the pricing equation, but does in fact significantly influence the market prices. As a consequence, no fitting model within this framework is able to fit the prices, because this feature is lost. For example, we could be ignoring credit risk or bond liquidity in pricing, e.g. considering 'off the run' and 'on the run' bonds together without reflecting this difference in the pricing.

*5.2. A guide to classifying the inconsistencies*

One cannot strictly prove that any of our explanations reflect the true situation, but they are at least plausible and offer a wide range of inconsistency interpretations. Here we propose a simple guide to better inform the choice. Note that this guide is not meant as an automated procedure. It should rather be regarded as a decision support algorithm.

We start by introducing the inconsistency pattern of a panel dataset. Figure 1 shows an example of an inconsistency pattern. The vertical axis marks the bond ID number, and the horizontal marks the date. A point $(x, y)$ means that the bond $y$ is considered for exclusion from the dataset on the date $x$ at some point. Patterns are expected to show up as long horizontal lines (not necessarily continuous) meaning that the same bond is getting attention over and over again on multiple dates.

[Figure 1 about here.]

As we can see, the primal algorithm 2 exhibits more stability by choosing the same bonds every time, while the steepest descent algorithm 3 produces more chaotic results. Therefore in what follows we consider only the output of the more stable primal algorithm, because for our purpose stability is a desired property.

The inconsistency pattern is the key tool in identifying the probable cause. The identification algorithm begins with characterizing the pattern to one of the types below.

- **Occasional groups of vertically aligned dots.** This behaviour is indicative of data errors or major market turmoils. To make the distinction, one has to closely inspect the numerical data and the market situation for each specific date corresponding to each of the vertically aligned groups.

  In either case, a simple solution would be dropping these dates from the dataset entirely.

- **Occasional seemingly random dots.** This is indicative of either individual data errors or insufficient framework flexibility. To make the distinction, look at the values for the tightness factor:

  - **High values** (greater than 2–3, but easily over 10) are indicative of data errors. A simple solution would be to drop this particular quote for this particular day.
  - **Low values** (about 1–2) could be indicative of moderate data quality, possibly due to low liquidity. It could also mean insufficient framework flexibility; in that case the dots would probably be a little more numerous. A simple solution would be to ignore these inconsistencies or (if the use case permits doing so) alter the input data, e.g. widen the bid-ask spreads.

- **Long horizontal lines** (usually almost continuous, but not necessarily). There are several possible explanations for this, but in any case there is always an easy solution of dropping some of the highlighted bonds from the dataset. Note that if two bonds produce long lines in an inconsistency pattern, actually only one of them needs to be removed from the dataset, because when for example two similar bonds have conflicting quotes, removing either one gets rid of the conflict automatically rendering the other one perfectly consistent. Therefore, all highlighted bonds must be studied, but some of them are expected to be 'normal' while others are expected to differ.

- **Data errors or program errors.** Inspect the bond description for these bonds. The description might be wrong or might be imported or used with errors. This happens surprisingly often.
- **'Outsider' bonds.** Think hard whether there is any reason to exclude some of these bonds from the dataset, whether they are in any sense different from the majority of the remaining bonds. Note that some of the 'outsider' bonds might not get spotted by this approach, especially if the dataset is not large. This is because 'outsider' bonds can only be spotted by our approach if there is sufficient amount of 'normal' bonds to contradict them. If for example all bonds in the dataset are short-term and the only long-term bond is somehow different, there is a large change that it won't cause any inconsistencies, simply because it is the only long-term bond, so there is nothing to contradict.

• **Short horizontal lines or random dots aligned horizontally**. This is the least straightforward case. The next step would be to identify the 'suspicious' bonds, i.e. those which get highlighted often enough. To do so, count the percentage of days on which each bond was highlighted, sort them by that percentage and consider the top of the list. Search for a feature which could separate the majority (certainly, not all) of the bonds at the top of the list from the majority (not necessarily all) of the remaining ones. Don't forget to check the 'on the run' feature, which in our experience seems to explain most of the short lines.

  - **There is such a feature**. The choice is now arbitrary. You can either say that the framework is not flexible enough, because that feature should really be incorporated into the pricing equation. Or you can say that your dataset must be comprised of the bonds selected (among other things) by that feature – this is especially useful when the required generalization of the pricing framework is not obvious and requires separate research.
  - **Such a feature is not found**. Unfortunately, this does not mean that there is no such feature. It might not be in plain sight. Or it

could be dynamic and change over time (such as a bond being 'on the run'). It could also lie in a completely other dimension, which is considered by the market, but is not included in the framework (e.g. eligibility for pension funds' investments or Sharia compliance). If the feature is still not found, we can offer possible explanations based on the values of the tightness factor.

* **Relatively low tightness factor**. This may be indicative of hidden costs (e.g. transaction costs). The approximate size of these costs relative to the bid-ask spread is given by the tightness factor.
* **Relatively high tightness factor**. In this case we are forced to conclude that market data is not reliable enough. The market could be highly illiquid, the data could be faulty. Non-committing quotes or hidden dealer/broker commissions sometimes also result in this, because that could lead to abnormally narrow bid-ask spreads, thus making the tightness factor unreasonably high.

Of course, sometimes these patterns overlap. For example, a cross-like inconsistency pattern would most probably be indicative of one day with data errors and an 'outsider' bond in the dataset.

*5.3. Examples*

In this sections we give real-life examples for the most interesting of the classification branches listed in the previous subsection.

**Spain**

[Figure 2 about here.]

This pattern features several vertically aligned dots on the right and some stray dots on the left. Upon further examination, the vertically aligned points are not a result of data errors (our data comes pre-filtered), but rather of a sudden market movement on 8 Aug 2011. The dots at the left have low tightness factor values (1.02 and 1.3) and are therefore most likely due to relatively low liquidity or just bad luck.

**Belgium**

[Figure 3 about here.]

This pattern features two relatively long horizontal lines. Therefore, we closely examine bonds N4 and N27. It turns out that the bond N27 had been issued very shortly before the dates corresponding to the horizontal line, on 18 Mar 2009. More importantly, its maturity coincides exactly with that of bond N4. Having two bonds in the same dataset with exactly same maturity date, one of them being 'on the run' (newly issued) is probably not a good idea. Therefore, our conclusion is that this is the case of an 'outsider bond' (unless we wish to alter our framework to incorporate the 'on the run' premium).

**Italy**

[Figure 4 about here.]

First of all, this pattern features several long lines. Upon closer examination, we discover that these are mostly caused by several eurobonds, which should have been filtered out. After we remove them from the dataset, the remaining bonds have a much clearer inconsistency pattern.

[Figure 5 about here.]

There are now several short lines. Bonds N48 and N64 mature within several days from each other, but N64 is 'on the run' in the end of 2008. Bonds N18 and N25 mature about a year apart, but the overall time to maturity is more than 15 years, so this might also contribute to the fact that they occasionally conflict during hard times (note that this contradiction only appears in the end of 2008 and in the end of 2011 – these were the unrestful times for the markets). Finally, there are vertically grouped dots around January 2012, which corresponds to a market turmoil at the end of 2011 – beginning of 2012.

**Germany**

[Figure 6 about here.]

The German inconsistency pattern is unexpectedly full of inconsistencies. Let's consider them one by one. Bonds N41 and N42 are inconsistent, but the latter is 'on the run'. The same applies to bonds N68 and N87 – N68 is 'on the run'. It is also not surprising that there is some inconsistency during the two turmoil periods: end of 2008 and end of 2011. Bid-ask spreads for German bonds are very narrow, so any disbalance is readily picked up by the tightness factor. Finally, there are two inconsistent bonds, whose inconsistency we are not able to explain. Bonds N12 and N13 both mature at the same date (1 Jul 2009) and have similar coupons (4% and 4.5% annually). In all other observable features they are quite alike. However, there is a simple portfolio (we cross-examined it by hand) resulting in arbitrage profits being made. Since the corresponding tightness factors are relatively low (about $1.1 - 1.2$), we conclude that this might be indicative of hidden transaction costs or commissions. These costs can be roughly estimated as

$$\frac{ask - bid}{2}(\gamma - 1).$$

In our situation this amounts to about 0.2 basis points, which is a reasonable value for a hidden cost.

**Russia**

[Figure 7 about here.]

Russian inconsistency pattern features several long dotted horizontal lines. Upon further examination of the highlighted bonds, we discover that in every conflicting group of bonds, one of them happens to be amortized. Moreover, if we split the dataset into two separate datasets: amortized and non-amortized bonds, the resulting datasets will exhibit significantly higher consistency levels with slight inconsistencies only in crisis days, which makes us believe that Russian bond market actually has two pricing mechanisms: one for non-amortized bonds and another one (more obscure) for amortized bonds. We have been able to confirm this conclusion in private communications with some market practitioners.

## 6. Discussion

There are numerous possible reasons for the inconsistencies to arise without the possibility for the real-world arbitrage, but the main two are data issues (data errors, dataset inhomogeneity, illiquidity, etc) or framework issues (excessive simplicity, omitted factors, wrong assumptions, etc). We propose an approach to quantify the inconsistencies and to determine their probable cause.

One should note that in fact we only test whether the data is consistent with the chosen pricing framework, so at the first glance there should be no formal way of telling which of the two is really wrong (or maybe both). That is ultimately true, however our approach gives the researcher strong hints as to where to direct his/her attention in the first place. The inconsistency pattern introduced in this paper helps develop the intuition necessary. Our approach is not designed for automated decision making, but it offers several new features for a human decision maker to consider.

For example, the cases of Russia and Italy above are really quite similar from the formal point of view. The system raises a 'red flag' and a human studies the highlighted data. In the Italian case the highlighted data is then found to be mistakenly included in the dataset, so it's clearly a data selection error. In the Russian case the highlighted data is rightfully a part of the dataset, but we can easily devise an interpretable and meaningful criterion to isolate the highlighted data. Note that instead we could regard this as a framework deficiency – the framework did not account for the 'being an amortized bond' feature and did not incorporate it in the pricing process, whereas the market clearly did that. Splitting the market into two subsegments is merely a way to avoid having to improve the pricing framework by effectively eliminating the feature in question

Inconsistencies can also be used to justify the need for data filtering. Preliminary data treatment is omnipresent in finance, but the need for it is rarely formally justified. Even less common in empirical studies are justifications of the necessary amount of filtering. Data filtering is usually conducted via an ad hoc method with little or no justification to its amount and its necessity in general. As well known in theory, excessive data filtering can easily alter the perceived distributions in the data (e.g. by filtering out 35% outliers we can easily turn a super-heavy-tails Cauchy distribution, which does not even have a finite expectation, into a nice normal distribution; less severe transformations require less aggressive filtering – this process is known as von Neumann elimination). Therefore, justifying the need for data filtering and the exact amount of filtering to perform is a crucial methodological point.

Our paper gives an easy and elegant solution to this problem. Filtering the data until it becomes consistent provides a methodologically sound basis for determining the necessary amount of filtering. No-arbitrage condition is a widely accepted criterion of data consistency and soundness.

With respect to the filtering application it is very important to underline the result of testing different filtering algorithms on the real data. Interestingly, the traditional algorithm of eliminating the bond with the highest ratio of the pricing discrepancy to the bid-ask spread performs prohibitively poor.

**Acknowledgments**

Comments and suggestions from two anonymous referees have greatly improved this paper. The author would also like to thank Marat Kurbangaleev for useful comments and fruitful discussions, Sergey Smirnov, Stavros Zenios, Nick Andreev, and Alexander Chigodaev for valuable comments. Support from the Basic Research Program of the National Research University Higher School of Economics is gratefully acknowledged.

[1] Amaldi, E., Bruglieri, M., Casale, G., 2008. A two-phase relaxation-based heuristic for the maximum feasible subsystem problem. Computers and Operations Research 35 (5), 1465–1482.

[2] Amaldi, E., Kann, V., 1995. The complexity and approximability of finding maximum feasible subsystems of linear relations. Theoretical Computer Science 147 (1-2), 181–210.

[3] Amaldi, E., Pfetsch, M., Trotter Jr, L., 2003. On the maximum feasible subsystem problem, IISs and IIS-hypergraphs. Mathematical Programming 95 (3), 533–554.

[4] Armitage, S., Chakravarty, S. P., Hodgkinson, L., Wells, J., 2012. Are there arbitrage gaps in the UK gilt strips market? Journal of Banking and Finance 36 (11), 3080–3090.

[5] Balbás, a., López, S., 2008. Sequential arbitrage measurements and interest rate envelopes. Journal of Optimization Theory and Applications 138 (3), 361–374.

[6] Balbás, A., Peng, Y., 2015. Sequential arbitrage measurement in bond markets: Theory and empirical applications in the Euro-zone. Tech. rep., Universidad Carlos III de Madrid, Madrid.

[7] Chakravarti, N., 1994. Some results concerning post-infeasibility analysis. European Journal of Operational Research 73 (1), 139–143.

[8] Chen, Z., Knez, P. J., 1995. Measurement of market integration and arbitrage. Review of Financial Studies 8 (2), 287–325.

[9] Chinneck, J., 1996. An effective polynomial-time heuristic for the minimum-cardinality IIS set-covering problem. Annals of Mathematics and Artificial Intelligence 17 (1), 127–144.

[10] Chinneck, J., 2001. Fast heuristics for the maximum feasible subsystem problem. INFORMS Journal on Computing 13 (3), 210–223.

[11] Chinneck, J., 2007. Feasibility and infeasibility in optimization: algorithms and computational methods. Springer Verlag.

[12] Dermody, J. C., Prisman, E. Z., 1993. No arbitrage and valuation in markets with realistic transaction costs. Journal of Financial and Quantitative Analysis 28 (1), 65–80.

[13] Ehrhardt, M. C., 1989. A new linear programming approach to bond portfolio management: a comment. Journal of Financial and Quantitative Analysis 24 (4), 533–537.

[14] Grinblatt, M., Longstaff, F. A., 2000. Financial innovation and the role of derivative securities: an empirical analysis of the Treasury STRIPS program. The Journal of Finance 25 (3), 1415.

[15] Jarrow, R. A., 2004. Risky coupon bonds as a portfolio of zero-coupon bonds. Finance Research Letters 1 (2), 100–105.

[16] Jaschke, S. R., 1998. Arbitrage bounds for the term structure of interest rates. Finance and Stochastics 40 (1), 29–40.

[17] Katerinochkina, N., 1975. Search for maximal upper zero of logic algebra monotone function. Doklady Akademii Nauk SSSR 224 (3), 557–560.

[18] King, A. J., 2002. Duality and martingales: a stochastic programming. Mathematical Programming 562 (March 1998), 543–562.

[19] Parker, M., Ryan, J., 1996. Finding the minimum weight IIS cover of an infeasible system of linear inequalities. Annals of Mathematics and Artificial Intelligence 17 (1), 107–126.

[20] Ronn, E. I., 1987. A new linear programming approach to bond portfolio management. Journal of Financial and Quantitative Analysis 22 (4), 439–466.

[21] Sankaran, J., 1993. A note on resolving infeasibility in linear programs by constraint relaxation. Operations Research Letters 13 (1), 19–20.

[22] Smirnov, S. N., Lapshin, V. A., Kurbangaleev, M. Z., 2016. Deriving implied risk-free interest rates from bond and cds quotes: a model-independent approach. Optimization and Engineering.
URL `http://dx.doi.org/10.1007/s11081-016-9333-2`

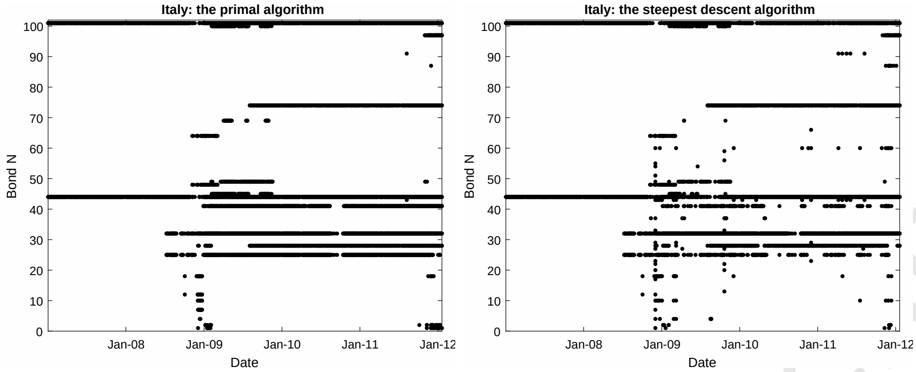


Figure 1: The inconsistency patterns. Dataset: Italy.

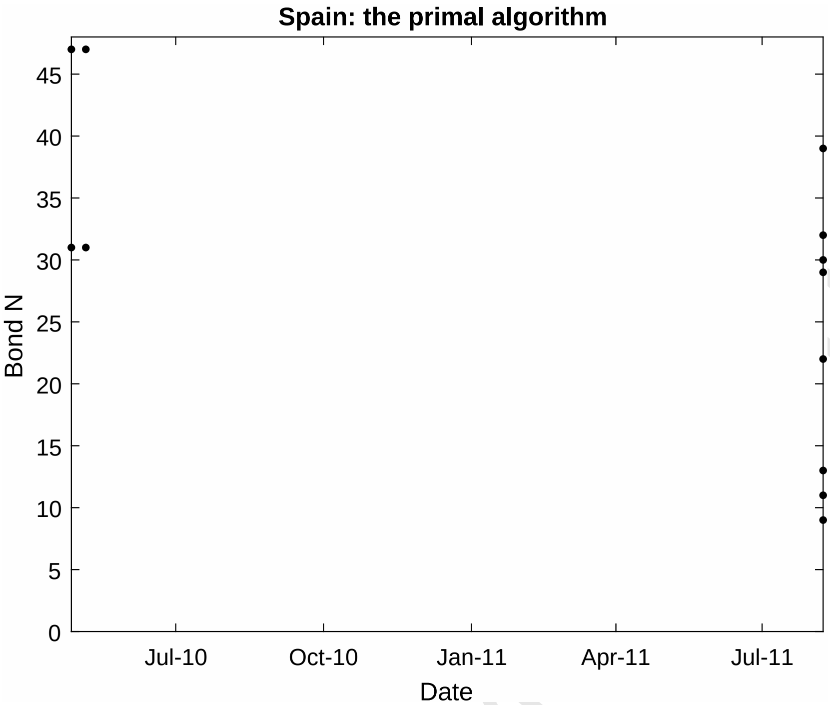


Figure 2: The inconsistency pattern. Dataset: Spain.

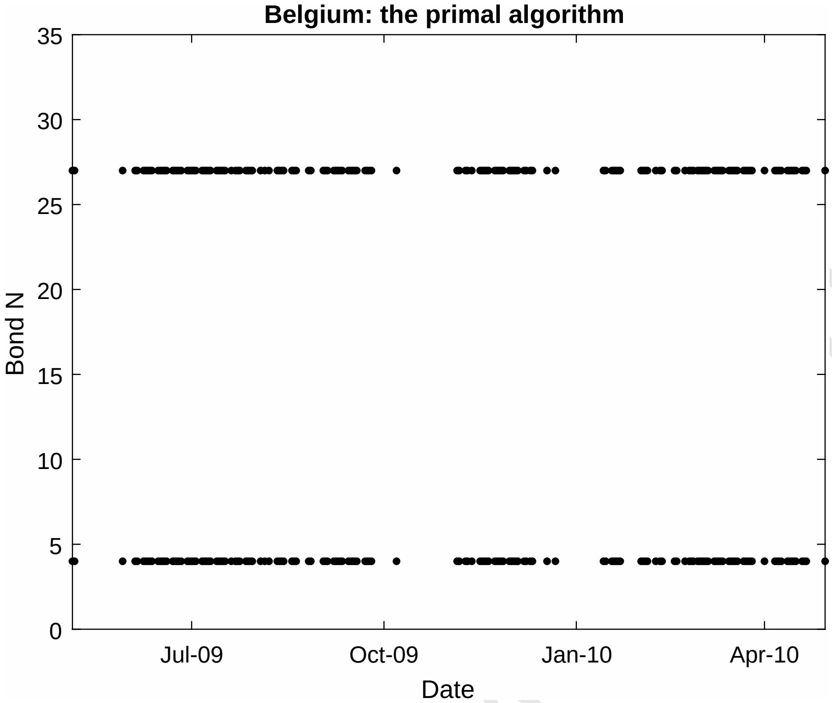


Figure 3: The inconsistency pattern. Dataset: Belgium.

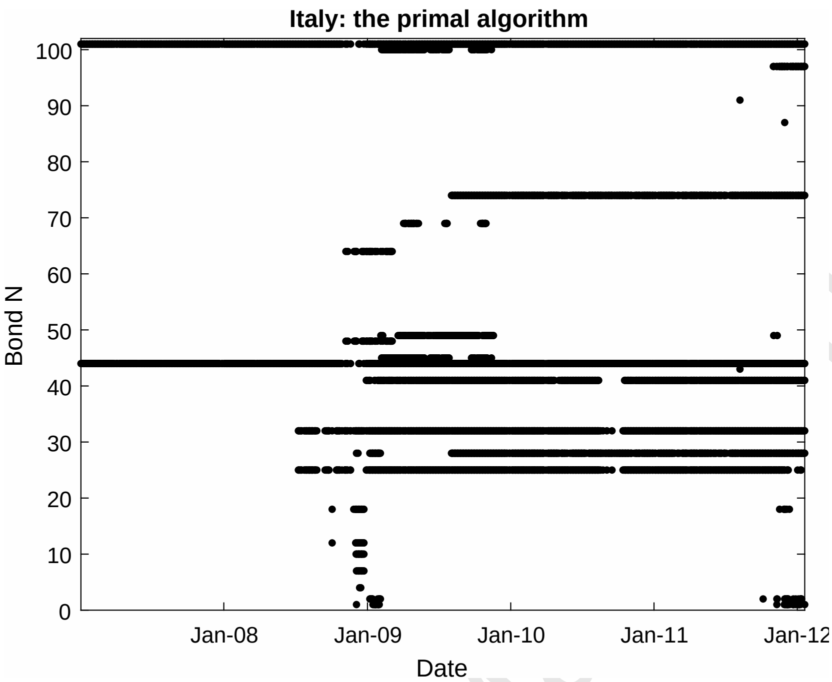


Figure 4: The inconsistency pattern. Dataset: Italy.

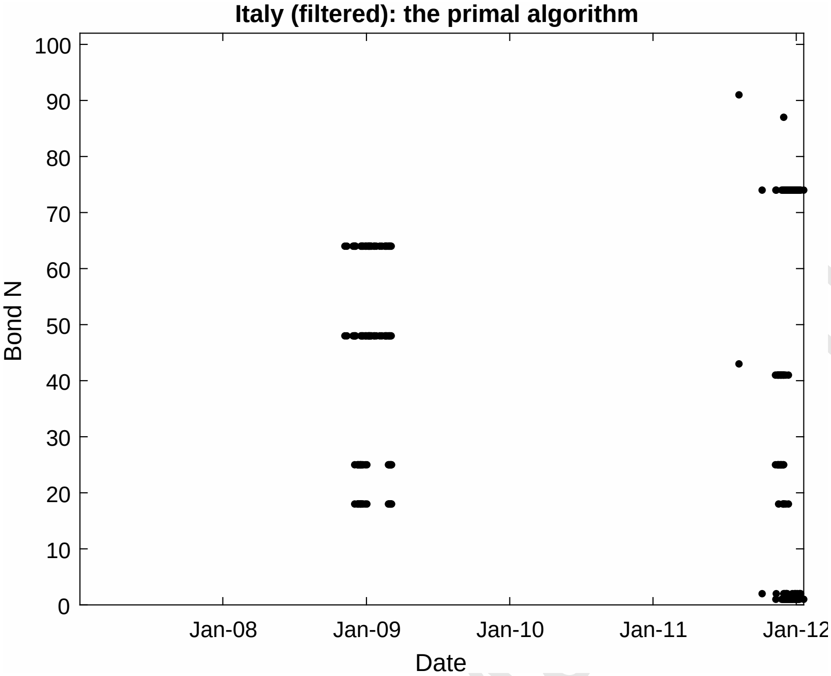


Figure 5: The inconsistency pattern. Dataset: Italy (filtered).

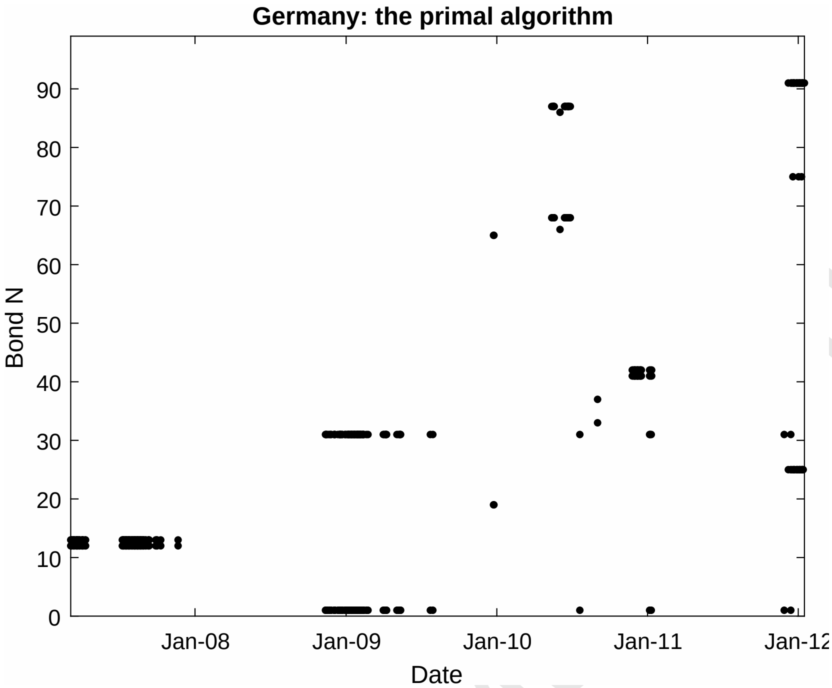


Figure 6: The inconsistency pattern. Dataset: Germany.

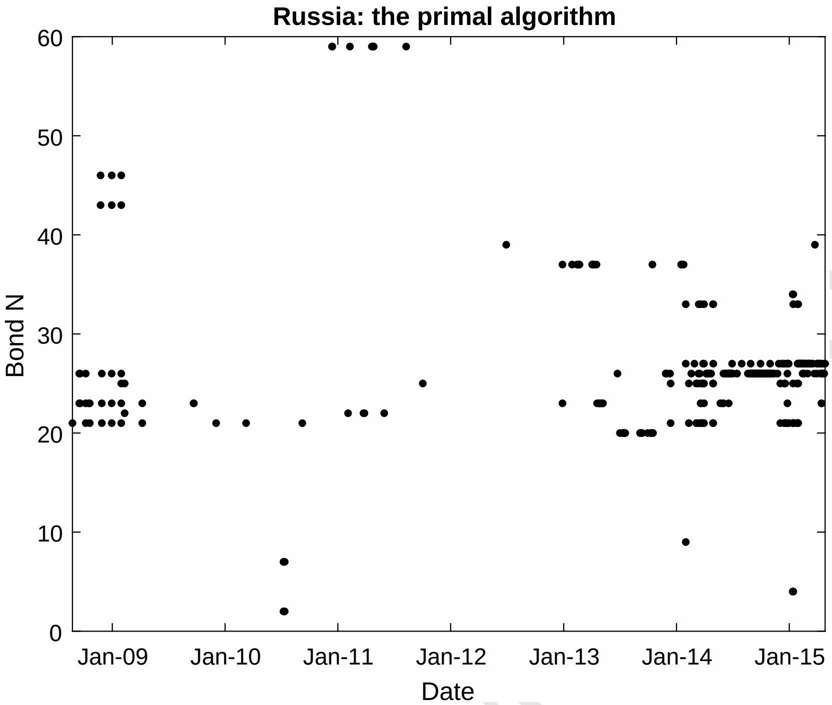


Figure 7: The inconsistency pattern. Dataset: Russia.

### AppendixA. Proof of the NP-equivalency

To prove NP-equivalency we use the following yes/no problem statement: given $k$, is there a feasible subsystem of (3) containing $k$ double inequalities?

**Theorem 1.** *The problem of verifying this property is NP-complete.*

*Proof.* It's obvious that the problem belongs to NP – it suffices to check all the inequalities, the certificate being the feasible vector $c$ and the indices of the remaining double inequalities (as usual, we assume that arithmetic operations are quick enough).

For the 'complete' part, we now reduce an arbitrary binary integer programming problem to the form (3). Consider a binary integer programming problem: find the vector $x \in \{0,1\}^n$ such that $Bx \leq c$, where $B \in \{-1,0,1\}^{m\times n}$, $c \in \mathbb{Z}^m$. We can restate it as follows: is there a feasible subsystem of

$$\begin{cases} Bx \leq c, \\ 0 \leq x_1 \leq 0,\ 1 \leq x_1 \leq 1, \\ 0 \leq x_2 \leq 0,\ 1 \leq x_2 \leq 1, \\ \dots \\ 0 \leq x_n \leq 0,\ 1 \leq x_n \leq 1, \end{cases} \tag{A.1}$$

containing all $m$ inequalities of the $Bx \leq c$ and exactly $n$ double inequalities. If there is, its solution is also the solution of the original binary integer programming problem and vice versa.

We now transform this system to look like (3). Let $y = c - Bx$. We can rewrite (A.1) as

$$\begin{cases} \mu \leq Ay \leq \nu, \\ y \geq 0, \end{cases}$$

where $A, \mu, \nu$ should be chosen so that $Ac - \mu = b$, $Ac - \nu = a$, $AB = T$, where $a \leq Tx \leq b$ is the matrix form of the double inequalities part of (A.1). $AB = T$ is equivalent to $B^T A^T = T^T$ and it has a solution $A^T$ if and only if the rank of $B^T$ is equal to the rank of $[B^T, T^T]$. For the latter it suffices that the rank of $B$ be $n$. In general, this is not the case, but we can always augment the initial binary integer programming problem with a sufficient number of inequalities with the corresponding $c_k$ large enough for the inequalities to be satisfied for any $x$.

We also need one additional modification to introduce the $1^T c \leq 1$ constraint. Let $c = \frac{y}{n}, \mu_1 = \frac{\mu}{n}, \nu_1 = \frac{\nu}{n}$. Now the transformed problem is

$$\begin{cases} \mu_1 \leq Ac \leq \nu_1, \\ c \geq 0, \\ 1^T c \leq 1. \end{cases}$$

The last inequality is not binding due to the nature of the inequalities. We have thus shown that the binary integer programming problem may be reduced to our yes/no problem in polynomial time, which finishes the proof. □

## AppendixB. Filtering results for other datasets

The descriptive statistics for the performance of all five filtering algorithms on the Italian dataset has been reported in Table 3. This appendix reports the same for all the other datasets.

Table B.5: Performance statistics for approximate algorithms. The remaining datasets.

| Dataset | Algorithm | Non-optimal days | In % | Mean excess $N$ | Max excess $N$ |
|---|---|---|---|---|---|
| Belgium | Traditional (No 1) | 105 | (68%) | 3.31 | 11 |
| | Primal (No 2) | 0 | (0%) | 0 | 0 |
| | Steepest Descent (No 3) | 0 | (0%) | 0 | 0 |
| | Restricted Search (No 4) | 0 | (0%) | 0 | 0 |
| | Candidate Search (No 5) | 0 | (0%) | 0 | 0 |
| France | Traditional (No 1) | 36 | (100%) | 5.67 | 12 |
| | Primal (No 2) | 0 | (0%) | 0 | 0 |
| | Steepest Descent (No 3) | 0 | (0%) | 0 | 0 |
| | Restricted Search (No 4) | 0 | (0%) | 0 | 0 |
| | Candidate Search (No 5) | 0 | (0%) | 0 | 0 |
| Greece | Traditional (No 1) | 67 | (94%) | 4.67 | 9 |
| | Primal (No 2) | 0 | (0%) | 0 | 0 |
| | Steepest Descent (No 3) | 0 | (0%) | 0 | 0 |
| | Restricted Search (No 4) | 0 | (0%) | 0 | 0 |
| | Candidate Search (No 5) | 0 | (0%) | 0 | 0 |
| Germany | Traditional (No 1) | 179 | (89%) | 11.83 | 26 |
| | Primal (No 2) | 3 | (1%) | 1.00 | 1 |
| | Steepest Descent (No 3) | 0 | (0%) | 0 | 0 |
| | Restricted Search (No 4) | 0 | (0%) | 0 | 0 |
| | Candidate Search (No 5) | 0 | (0%) | 0 | 0 |
| The Netherlands | Traditional (No 1) | 5 | (16%) | 1.20 | 2 |
| | Primal (No 2) | 0 | (0%) | 0 | 0 |
| | Steepest Descent (No 3) | 0 | (0%) | 0 | 0 |
| | Restricted Search (No 4) | 0 | (0%) | 0 | 0 |
| | Candidate Search (No 5) | 0 | (0%) | 0 | 0 |
| Russia | Traditional (No 1) | 136 | (65%) | 3.57 | 17 |
| | Primal (No 2) | 18 | (9%) | 1.22 | 2 |
| | Steepest Descent (No 3) | 2 | (1%) | 1.00 | 1 |
| | Restricted Search (No 4) | 0 | (0%) | 0 | 0 |
| | Candidate Search (No 5) | 0 | (0%) | 0 | 0 |

Victor Lapshin is currently an associate professor at the Faculty of Economic Sciences of National Research University Higher School of Economics, Moscow, Russia. He has a PhD in mathematical modelling. His research interests are Risk Management and Mathematical Modelling in Financial Markets.

We propose a measure of bond quotes inconsistency given a pricing framework.

Quote inconsistencies may be due to data errors, dataset or framework problems.

We give a guide to identify the most probable cause for the inconsistencies.

Filtering out the most mispriced bonds does not help eliminating the inconsistencies.