Forecasting Investment-Grade Credit-Spreads.  
A Regularized Approach

Thiago De Oliveira Souza  
ECARES, SBS-EM, Université Libre de Bruxelles

ECARES working paper 2011-037
Forecasting investment-grade credit-spreads – A regularized approach

By Thiago de Oliveira Souza

It is common for banks to have liabilities attached to the Treasury’s rate and assets attached to a corporate rate. A change in the difference between these rates (i.e., a change in the credit-spread) impacts the banks’ balance sheet. In order to forecast this risk, I propose the use of (very) short estimation windows using the lasso estimation. The lasso shrinks some of the estimated coefficients to zero, improving their finite sample performance also allowing the use of smaller estimation windows. I compare the out-of-sample performance of several credit-spread forecasting models for each investment-grade credit-rating in the period between 2000 and 2011. Considering the 6 and 12 months forecasts of AAA-rated credit-spreads, the historical average outperforms, in terms of mean absolute prediction error, the Martingale and several other forecasting models. These models are based on the shape (level, slope and curvature) of the risky and risk-free yield curves, and based also on the spot, forward and average past yields. Considering all other credit-ratings, the forecasts given by the lasso tend to outperform those based on long estimation windows. Keywords: credit-spread, forecasting, out-of-sample, regularization, yield-curves, credit-rating.

- Preliminary version -

In this paper I compare the out-of-sample performance of credit-spread forecasting models for European investment-grade bonds of each major credit-rating. For that matter, I use the respective iBoxxx indices between Jan./00 and Mar/11. I compare the Mean Absolute Prediction Error (MAPE) of models based on current yields (similar to those in Krishnan, Ritchken and Thomson (2010)). Therefore, I extract the whole term structure of interest rates; the whole credit-spread yield curve; and the forward, spot and average rates. Moreover, I compare different sizes of rolling estimation windows with a recursive estimation. I also use the lasso estimation of Tibshirani (1996) to obtain more stable estimates of the parameters and more accurate forecasts. The use of the lasso is particularly important because of the small estimation windows that frequently contain less observations than regressors.

* ECARES - Solvay Brussels School of Economics and Management, Université libre de Bruxelles and School of Economics and Finance - Queen Mary, University of London. E-mail: t.souza@qmul.ac.uk. I wish to thank Marcelo Fernandes and David Veredas for helpful discussion and academic support. The usual disclaimer applies.
The empirical results, new in the literature, are the following: First, the lasso produces better out-of-sample forecasts of credit-spreads than the ordinary least squares estimation. This happens regardless of the sample size used in the estimation (rolling window or recursive estimation). It also happens for both periods before and after the 2008 credit crisis, and for all credit ratings. Secondly, using small rolling windows to estimate the models produces better overall results than a recursive estimation.\(^1\) Third, the historical average is the only forecast that outperforms the Martingale model for higher quality AAA-bonds. For lower quality BBB-bonds, the most regularized lasso forecast is the only that marginally outperforms the Martingale model.

It is common for banks and other financial institutions to have a significant part of its liabilities attached to the treasury’s rate (e.g., in saving accounts), and assets attached to a corporate rate (e.g., in the form of corporate loans). If the required return on corporate bonds increases, even without changing the return on treasury securities, there will be a negative impact on the bank’s balance sheet. This happens because the value of assets decrease while liabilities remain unchanged. This situation describes an increase in the “credit-spread”, generally defined as the difference between a base rate and another relevant rate. Through this paper I define it as the difference between a given corporate bond’s yield and the treasury security’s yield within the same maturity. Volatile yields, as observed in Figure 1, makes the task of monitoring risks even more important and difficult.

Figure 1 also illustrates that corporate bonds from different ratings follow different dynamics and should be modelled as such. Investment-grade bond yields are also particularly difficult to explain when compared to speculative-grade bonds. Collin-Dufresne, Goldstein and Martin (2001) find that investment-grade bonds behave like Treasury bonds, while speculative-grade bonds are more sensitive to stock returns and behave more like equity. Avramov, Jostova and Philipov (2007) examine the U.S. market and find that the model’s explanatory power differs across bond classes. The variation is partly due to the different role played by firm-level fundamentals such as volatility, leverage, and growth opportunities. These variables are particularly important in the speculative-grade segment, but play little role in the investment-grade segment. They also find that common factors capture twice as much variation in the low-grade tercile relative to the high-grade tercile. In addition, the high-grade bond tercile is more sensitive to the FED’s monetary policy. Hotchkiss and Jostova (2007) argue that bond trading volume (and hence, liquidity) can vary with credit rating and recent literature studying determinants of corporate credit spreads emphasizes the role of liquidity risk.\(^2\)

The literature on credit spread forecasting is by far less extense than the liter-

\(^1\) However, the relative performance of the lasso regression is better before or after the crisis depending on the credit-rating.

\(^2\) This literature is large and one of the debates is about the most appropriate measure of liquidity for corporate bond and equity pricing; see Mahanti et al. (2008) for a discussion.
Figure 1. The figure shows how the credit-spread (on the vertical axis), vary in time for indices composed by A- and AAA-rated spreads with maturity in 1 year.

ature on credit spread determination. Among the papers concerned with credit spread forecasting, a major contribution is the work of Krishnan, Ritchken and Thomson (2010). They show that credit-spread forecasts based on the current spot and forward rates can be significantly improved upon by using the information contained on the shape (level, slope and curvature) of the credit-spread and treasury yield curves. They also find that other macroeconomic, market-wide, and firm-specific risk variables do not significantly improve the estimation.

The results in Krishnan, Ritchken and Thomson (2010) find support in previous work. For example, Fama and Bliss (1987), Campbell and Shiller (1991), Cochrane and Piazzesi (2005) and Diebold and Li (2006) show that the current yield curve contains significant information on future yields.\(^3\)

Finally, we can understand the use of yield curves alone for forecasting purposes by the way agents form expectations. Intuitively, the prevailing interest rates at different maturities contain all the available information that is relevant to forecast credit-spreads in an efficient market. Litterman and Scheinkman (1991), for instance, find that the three most important factors driving the term-structure of interest rates are its level, steepness and curvature.

The results in the empirical section show that forecasts based on small estimation windows are more accurate than those based on long samples. One

\(^3\)The impact of changes in the spot rate on credit-spreads, can be understood from the work of Longstaff and Schwartz (1995) and confirmed by Duffee (1998): The static effect of a higher spot rate is to increase the drift of the firm value process. A higher drift reduces the probability of default, and in turn, reduces the credit spreads.
explanation is that using small estimation windows improves the forecasts in the presence of a temporary unobservable factor (or shock).

Although there is evidence that the credit-spread yield curve already contains all the information available to agents at a given point in time, there is still apparently a common factor driving credit spreads. Pedrosa and Roll (1998) document considerable co-movement of credit spread changes among index portfolios of bonds from various industry, quality, and maturity groups. Campbell and Taksler (2003) document a synchronous upward move in aggregate spreads and aggregate idiosyncratic volatility from 1990 to 2000. Finally, Collin-Dufresne, Goldstein and Martin (2001) report that after running a regression that explains about 25 percent of the credit-spread changes, the residuals from these regressions are highly cross-correlated. They also show that principal components analysis suggests that they are mostly driven by a single common factor that is not firm specific.

These results are not surprising, since theory predicts that all credit spreads should be affected by aggregate variables such as changes in the interest rate, changes in business climate, changes in market volatility, and so forth. But after controlling for these aggregate determinants, the systematic movement of credit spread changes still remains, and indeed, is the dominant factor (Collin-Dufresne, Goldstein and Martin (2001)).

The presence of this systematic movement suggests the use of small estimation windows to account for the unobservable temporary shocks. However, the long list of explanatory variables mentioned before makes the possibility of over-fitting the model an important issue. The finite sample properties of the estimators are also particularly relevant given the use of short estimation windows.

In this paper I use the lasso estimation of Tibshirani (1996) to avoid both problems mentioned above. The advantage is two-fold: First, it produces more parsimonious models because some of the estimated coefficients are set to zero and less coefficients allows for the use of smaller estimation windows. These windows can be even smaller than the original number of regressors. Secondly, the lasso estimation produces more stable coefficients in small samples by applying shrinkage to the solution. This improves the out-of-sample performance of the model.

I. The model

For each maturity and credit-rating, I use the following model to forecast the respective credit-spreads:

\[ S_{t+h} = \alpha + \beta_1 S_t + \beta_2 \bar{S}_t + \beta_3 F^h_t + \beta_4 \theta_t + \beta_5 \theta_{t-1} + \epsilon_t, \]

where \( S_{t+h} \) is the credit-spread \( h \) periods in the future (at \( t+h \)); \( S_t \) is the credit-spread in time \( t \); \( \bar{S}_t \) is the past average credit-spread up to time \( t \); \( F^h_t \) is
the respective forward rate; $\theta_t^f = \{\theta_t^{f1}, \theta_t^{f2}, \theta_t^{f3}\}$ is a vector with the level, slope and curvature factors from the risk-free yield curve; $\theta_t^c = \{\theta_t^{c1}, \theta_t^{c2}, \theta_t^{c3}\}$ is the vector with the level, slope and curvature factors from the credit-spread yield curve, with $\theta_{t-1}^c$ being its lagged value and $\varepsilon_t$ an error term.

I use the technique in Diebold and Li (2006) to extract the level, slope and curvature factors from the current yields of different maturities, obtaining estimates for $\theta_t^f$ and $\theta_t^c$. I explain the procedure in detail on the next section. In the following one, I present the lasso estimation of Tibshirani (1996). I use the lasso estimation as an alternative to the ordinary least squares estimation of the the $\beta$ coefficients in model (1) for the reasons mentioned earlier.

### A. Extracting the yield curves: Diebold and Li 2006

For a given credit-rating, I fit the following model to the yields of different maturities

\[
y_t(n) = \theta_{1t} + \theta_{2t} F_{2}^{(n)} + \theta_{3t} F_{3}^{(n)},
\]

where $y_t(n)$ is the yield to maturity at time $t$, for maturity $n$ and

\[
F_{2}^{(n)} = \frac{(1 - e^{-\lambda_t n})}{\lambda_t n}
\]
\[
F_{3}^{(n)} = \frac{(1 - e^{-\lambda_t n})}{\lambda_t n} - e^{-\lambda_t n}
\]

The Nelson and Siegel (1987) functional form is a convenient and parsimonious three-component exponential approximation where the vector $\theta_t = \{\theta_{1t}, \theta_{2t}, \theta_{3t}\}$ describes the whole yield curve in time $t$. Diebold and Li (2006) interpret $\theta_{1t}$, $\theta_{2t}$, and $\theta_{3t}$ as three latent dynamic factors respectively corresponding to level, slope and curvature of the yield curve. They also show how these coefficients can be used to forecast the yield curve.

The function corresponds to a discount curve that begins at one at zero maturity and approaches zero at infinite maturity. The parameter $\lambda_t$ is a decay factor: small values produce slow decay and can better fit the curve at long maturities, while large values produce fast decay and can better fit the curve at short maturities; it also tells where the loading on $\theta_{3t}$ achieves its maximum.

Following Diebold and Li (2006), we fix the value of $\lambda_t^4$ and use ordinary least squares to estimate the vector $\hat{\theta}_t = \{\theta_{1t}, \theta_{2t}, \theta_{3t}\}$, given that, now, the values of the regressors are fixed.

\footnote{In this paper, differently from them, assuming that medium maturities happen between 2, 3 or 5 years and finding the value of $\lambda$ that maximizes the loading for these maturities.}
The data on the risk-free yields generates the time series for the parametric risk-free yield curve, $\hat{r}_t^{\text{rf}} = \{\hat{r}_t^{\text{rf}1}, \hat{r}_t^{\text{rf}2}, \hat{r}_t^{\text{rf}3}\}$. The data on credit-spread yields generate the time series for the parametric credit-spread yield curve, $\hat{c}_t^{\text{cs}} = \{\hat{c}_t^{\text{cs}1}, \hat{c}_t^{\text{cs}2}, \hat{c}_t^{\text{cs}3}\}$.

**B. The lasso estimation: Tibishirani 1996**

Given a set of input measurements $(x^i, y_i), i = 1, 2, ..., N$, where $x_i = (x_{i1}, ..., x_{ip})^T$ are the predictor variables and $y_i$ are the responses and letting $\hat{\beta} = (\hat{\beta}_1, ..., \hat{\beta}_p)^T$; the lasso estimate $(\hat{\alpha}, \hat{\beta})$ is defined by

$$
(\hat{\alpha}, \hat{\beta}) = \arg \min_{\beta} \sum_{i=1}^{N} (y_i - \alpha - \sum_{j=1}^{p} \beta_j x_{ij})^2 \\
\text{s.t.} \sum_{j=1}^{p} |\beta_j| \leq \delta
$$

(3)

The sum of squared residuals, as usually, is taken over the observations in the data set. The sum of the absolute values of the coefficients is taken over all $p$ values. The bound $\delta$ is a tuning parameter that we need to chose. It controls the amount of shrinkage applied to the estimates.

Let $\beta_j^0$ be the coefficients obtained by ordinary least squares and $\delta_0 = \Sigma |\beta_j^0|$. Values of $\delta < \delta_0$ will result in shrinkage of the coefficients towards 0, with some of them being exactly equal to zero. Choosing a value of $\delta = \delta_0/2$, for instance, has the same approximate effect of finding the best subset of size $p/2$. Another important feature of the model is that the design matrix does not need to be full rank, which allows for estimation samples to be smaller than the number of regressors.

The lasso estimation has two important consequences: sparsity (i.e., coefficients exactly equal to 0) and shrinkage. Sparsity allows for less coefficients to be estimated, and therefore it allows the use of smaller estimation samples. Shrinkage improves the stability of the estimates as I explain in the next section, in which I also explain what affects the optimal value of $\delta^*$. The section that follows after that shows how to choose the coefficient $\delta$ empirically, using cross-validation.

**Shrinkage estimation.** — The basic idea in shrinkage estimation is that it is possible to reduce an estimator’s variance by averaging it with a given constant that, by definition, has no variance. This can be done at the expense of including some bias on the estimation. The goal to correctly apply these estimators is to find the optimum balance between bias and variance.

James and Stein (1961) are the ones that come up with the idea of shrinkage estimation. They note that for $N \geq 3$ independent normal random variables, the vector of sample means, $\hat{\mu}$, is dominated in terms of joint mean-squared error by a convex combination of these means and a constant, $\mu_0$, resulting in the estimator:
\[ \mu_s = \delta \mu_0 + (1 - \delta) \bar{\mu}, \]

for \( \delta \in (0, 1) \). This estimator shrinks the sample mean towards a common value that is normally chosen to be the "grand mean" across all variables.

They show that the optimal balance between bias and variance, for a mean-squared error loss, is achieved for an optimal shrinkage factor \( \delta^* \). This is given by

\[ \delta^* = \min \left[ 1, \frac{(N - 2)/T}{(\bar{\mu} - \mu_0)/\Sigma^{-1}(\bar{\mu} - \mu_0)} \right]. \]

As expected, the optimum shrinkage decreases in the sample size \( T \) and dispersion of the sample means \( \bar{\mu} \) from the grand mean \( \mu_0 \), increasing in the number of means \( N \).

Equation 5 intuitively shows what affects the optimal choice of \( \delta^* \), but does not help in the determination of its value because it needs the very unknown parameters that should be estimated. One alternative to choose \( \delta^* \) is to use cross-validation.

2-fold cross-validation. — This is the simplest variation of \( k \)-fold cross-validation. We, first, randomly assign the data points (all observations in the estimation window) into two subsets (folds). Using only one of these subsets we compute the lasso estimation on a grid of values of \( \delta \).

We then estimate the out-of-sample mean squared error of the regressions corresponding to each \( \delta \) in the grid, using the second subset as out-of-sample observations. We change the order of the in-sample and out-of-sample subsets and repeat the process. Next, for each value of \( \delta \) on the grid, we average the estimates of the out-of-sample mean squared errors obtained in each of the two estimation rounds. Finally, we choose the value \( \delta^* \) that minimizes this average.

II. Empirical exercise

A. The data set

The data set with monthly frequency spans the period between January/2000 and March/2011. It contains credit-spreads and Treasury yields for maturities of 3, 6 and 9 months, and 1, 2, 3, 5, 7, 10 and 15 years. The credit-spread data corresponds to the iBoxx European Corporate Bond Index, obtained from UBS Delta. I collect them individually for each investment-grade rating "A", "AA", "AAA" and "BBB". The treasury yields correspond to the Euribor for maturities
of 3, 6, 9, and 12 months obtained from Reuters. For longer maturities, I use a government bond yield benchmark obtained from the ECB.

B. Estimation

I compare five different 6 months ahead forecasts of the 1-year credit-spreads: Martingale, Mean, OLS, lasso-Min and lasso-1se. I perform the analysis exclusively out-of-sample, comparing the models in terms of mean absolute prediction error. I split the sample in Sep./2007, before and after the credit crisis, and also individually report the results for each period. Finally, I use estimation windows of different sizes to observe their impact over the forecasting accuracy.

The Martingale model is a benchmark and simply consists in forecasting the future credit-spread as the present credit-spread. Using the same notation as before, this means

\[ \hat{S}_{t+h} = S_t, \]

where \( \hat{S}_{t+h} \) represents the credit-spread forecast \( h \) months in the future.

The Mean model consists in using the historical mean up to time \( t \), \( \hat{S}_t = 1/n \sum_{i=1}^t S_t \), as a forecast of the future credit-spread, and gives

\[ \hat{S}_{t+h} = \tilde{S}_t. \]

The other three forecasts (the OLS, the lasso-Min and the lasso-1se) are based on the same model given in equation 1, so they all result in the same form of forecast:

\[ \hat{S}_{t+h} = \hat{\alpha} + \hat{\beta}_1 S_t + \hat{\beta}_2 \hat{S}_t + \hat{\beta}_3 \hat{r}_t^h + \hat{\beta}_4 \hat{c}_t + \hat{\beta}_5 \hat{r}_t^f + \hat{\beta}_6 \hat{r}_{t-1}, \]

The difference between them is only on the estimation technique to obtain the \( \beta \) coefficients. The yield curve factors, \( \hat{\theta}_t^f \) and \( \hat{\theta}_t^c \), on the other hand, are obtained by ordinary least squares as described earlier using the Diebold and Li (2006) methodology.

The OLS model corresponds to the forecasts obtained from the traditional ordinary least squares estimates of the \( \beta \) coefficients. The lasso-Min and the lasso-1se models correspond to forecasts obtained from the lasso estimation of the \( \beta \) coefficients in model 1, using the estimation procedure in 3 for different amounts of shrinkage. For the lasso-Min forecast, I choose \( \delta = \delta_{CV} \) in equation 3 by 2-fold cross-validation. In lasso-1se, \( \delta = \delta_{1se} \) is the smallest value that still produces a mean squared prediction error not statistically different from the one generated by \( \delta_{CV} \) in the estimation sample.
Table 1—Mean absolute error for the 6 months ahead forecasts of BBB-rated credit-spreads during the period between January/2000 and March/2011, using rolling estimation windows.

The models are described as follows: **Martingale** (equation 6): \( \hat{S}_{t+h} = S_t \); **Mean** (equation 7): \( \hat{S}_{t+h} = \bar{S}_t \); **OLS**, **lasso-Min** and **lasso-1se**: all based in equation 8, with different estimation procedures. **OLS**: (unrestricted) ordinary least squares. **Lasso-Min** and the **lasso-1se**: Lasso estimation described in equation 3. **Lasso-Min**: penalization chosen by cross-validation; **Lasso-1se**: penalization chosen as the smallest value that still produces a mean squared prediction error not statistically different from the one generated by cross-validation.

Intuitively, using \( \delta_{1se} \) generates estimates that are even more regularized (i.e., with a larger amount of shrinking) than the one obtained by cross validation. I select the value of \( \delta_{1se} \), however, in such a way that the mean squared error associated with it is still statistically not different from the one associated with \( \delta_{CV} \).

Regarding the past information used in estimating the model, I use two approaches. The first is to ignore the information in the data for periods longer than 1 year. This results in 6 monthly data points\(^5\) to be used in the model estimation at each point in time and I call it the **Rolling window (6 months)** estimation. I also report the performance of a **Recursive** approach, in which the data set increases with time because all the data prior to a given point in time is used in the estimation of the coefficients on that point.

**C. Results**

**BBB credit-spreads.** — Table 1 shows the mean absolute prediction error for BBB credit-spreads with one year to maturity. We see that the lasso-1se performs marginally better than the Martingale, and this is due to its better results after Oct./2007. We also see that the lasso-Min outperforms the Martingale before Oct./2007, but not after, resulting in overall less accurate forecasts over the whole period. We can also see that the lasso estimates outperform the OLS in all situations and that the average performs particularly well after Oct./2007 in relation to the others.

However, the performance of the OLS should be evaluated in the presence of a longer sample, and this is what we see in Table 2. This table reports the same

\(^5\)Given that I forecast returns 6 months ahead, the coefficients used in Jan/01 to forecast Jul/01 are estimated using data in the pairs of months Jan-Jul, Feb-Ago, Mar-Sep ,..., Jun-Dec.
Table 2—Mean absolute error for the 6 months ahead model forecasts of BBB-rated credit-spreads during the period between January/2000 and March/2011, using a recursive estimation window. The models are described as follows: Martingale (equation 6): $S_{t+h} = S_t$; Mean (equation 7): $S_{t+h} = S_t$; OLS, lasso-Min and lasso-1se: all based in equation 8, with different estimation procedures. OLS: (unrestricted) ordinary least squares. Lasso-Min and the lasso-1se: lasso estimation described in equation 3. Lasso-Min: penalization chosen by cross-validation; Lasso-1se: penalization chosen as the smallest value that still produces a mean squared prediction error not statistically different from the one generated by cross-validation.

<table>
<thead>
<tr>
<th>Period</th>
<th>Martingale</th>
<th>Mean</th>
<th>OLS</th>
<th>lasso-Min</th>
<th>lasso-1se</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan-02</td>
<td>Sep-07</td>
<td>37.1</td>
<td>63.8</td>
<td>55.7</td>
<td>50.5</td>
</tr>
<tr>
<td>Oct-07</td>
<td>Sep-10</td>
<td>129</td>
<td>81</td>
<td>114</td>
<td>129</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td>69</td>
<td>70</td>
<td>76</td>
<td>77</td>
</tr>
</tbody>
</table>

The forecasting accuracy of the recursive OLS, however, is still lower than the lasso-1se using rolling windows. In fact its overall performance is also lower than the recursive lasso-1se.

AAA credit-spreads. — Table 3 shows the performance of the models in forecasting the credit-spread of AAA bonds using rolling windows in the estimation. Comparing these results with the OLS results in Table 4, we can see that both lasso estimates (lasso-Min and lasso-1se) outperform the OLS regardless of the period considered, and regardless of the estimation window used. Although an interesting result, this is not enough to outperform the Martingale model.

It is a known fact that higher grade investment bonds are more difficult to forecast, and this shows in the results of Table 3 and Table 4. The news are the fact that the Mean model performed so well, and much better than the Martingale, which also explains why long estimation windows also improve the results for AAA credit-spreads.

Theory predicts yields to revert back to a long-run equilibrium, therefore the result is not surprising. What was not clear, however, was that reversion to a long-run equilibrium can happen at horizons as short as 6 months.

Creating an investment strategy. — We can use the predictability of the AAA credit-spread reported in Table 3 to create investment strategies. In this context, 6Considering even later starting points does not favor the OLS.
Table 3—Mean absolute error for the 6 months ahead model forecasts of AAA-rated credit-spreads during the period between January/2000 and March/2011, using rolling estimation windows. The models are described as follows: Martingale (equation 6): $\hat{S}_{t+h} = S_t$; Mean (equation 7): $\hat{S}_{t+h} = \bar{S}_t$; OLS, lasso-Min and lasso-1se: all based in equation 8, with different estimation procedures. OLS: (irrestricted) ordinary least squares. Lasso-Min and the lasso-1se: Lasso estimation described in equation 3. Lasso-Min: penalization chosen by cross-validation; Lasso-1se: penalization chosen as the smallest value that still produces a mean squared prediction error not statistically different from the one generated by cross-validation.

<table>
<thead>
<tr>
<th>Period</th>
<th>Martingale</th>
<th>Mean</th>
<th>OLS</th>
<th>lasso-Min</th>
<th>lasso-1se</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan-02</td>
<td>Sep-07</td>
<td>5.1</td>
<td>3.9</td>
<td>18.0</td>
<td>5.2</td>
</tr>
<tr>
<td>Oct-07</td>
<td>Sep-10</td>
<td>74</td>
<td>42</td>
<td>95</td>
<td>80</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td>29</td>
<td>17</td>
<td>44</td>
<td>31</td>
</tr>
</tbody>
</table>

Table 4—Mean absolute error for the 6 months ahead model forecasts of AAA-rated credit-spreads during the period between January/2000 and March/2011, using a recursive estimation window. The models are described as follows: Martingale (equation 6): $\hat{S}_{t+h} = S_t$; Mean (equation 7): $\hat{S}_{t+h} = \bar{S}_t$; OLS, lasso-Min and lasso-1se: all based in equation 8, with different estimation procedures. OLS: (irrestricted) ordinary least squares. Lasso-Min and the lasso-1se: Lasso estimation described in equation 3. Lasso-Min: penalization chosen by cross-validation; Lasso-1se: penalization chosen as the smallest value that still produces a mean squared prediction error not statistically different from the one generated by cross-validation.

<table>
<thead>
<tr>
<th>Period</th>
<th>Martingale</th>
<th>Mean</th>
<th>OLS</th>
<th>lasso-Min</th>
<th>lasso-1se</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan-02</td>
<td>Sep-07</td>
<td>5.1</td>
<td>3.9</td>
<td>6.3</td>
<td>4.2</td>
</tr>
<tr>
<td>Oct-07</td>
<td>Sep-10</td>
<td>74</td>
<td>42</td>
<td>171</td>
<td>142</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td>29</td>
<td>17</td>
<td>68</td>
<td>51</td>
</tr>
</tbody>
</table>
section, we analyse the risk and profitability of such strategies. The basic idea is that when the spot rate is below the historical rate, the investor should "buy" the spread, since the spread is expected to widen, and vice-versa.

In general lines, "buying" the spread means to short-sell the Treasury security (borrow at the Euribor rate in this case), and buy the AAA-bond index in the present. Then, later in the future, sell the bond index and use the money to close the short position in the Treasury security (paying back the loan). And vice-versa for "selling" the spread.

This strategy results in a gross return equal to the difference in the spreads:

\[ R_{t+h} = S_{t+h} - S_t. \]

**Forecasting 6-month credit spreads.** — There are several ways to use the predictability of returns to create such strategies. One of them involves the forecast of 6-month credit-spreads. This results in a forecast of the 6-month spread, six months in the future, giving the expected 1-year spread:

\[ E_t [S_{t+12}^{12}] = (1 + S_t^{6})(1 + E_t [S_{t+6}^{6}])), \]

where \( E_t [\cdot] \) is the expectation operator in time \( t \), \( S_{t+12}^{12} \) is the 12-months credit-spread (from 6-months securities) in time \( t + 12 \), \( S_t^{6} \) is the 6-month credit-spread in time \( t \), and \( S_{t+6}^{6} \) is the 6-month credit-spread in time \( t + 6 \).

\( E_t [S_{t+12}^{12}] \) should be compared to the 1-year spread in the market, \( S_t^{12} \). If \( E_t [S_{t+12}^{12}] > S_t^{12} \), then we buy the spread. Otherwise, we sell it.

Buying the spread in this case means to first buy the 1-year bond index, and short-sell the 6-months treasury security in the present. Six months later, the first treasury security matures and we short-sell again a 6-month treasury security to close our first position. In practice, it means to borrow paying the Euribor rate at the first six months, and refinancing the loan six months later with the new Euribor rate that was forecasted before while holding the bond index portfolio for that time. After one year, we sell the bond index, and buy back the treasury bond (i.e., pay back the loan).

This strategy produces gross returns similar to the one in equation 9, but substituting \( S_{t+12}^{12} = (1 + S_t^{6})(1 + S_{t+6}^{6}) \):

\[ R_{t+12} = (1 + S_t^{6})(1 + S_{t+6}^{6}) - S_t^{12}. \]

In table 5 we see the results of using this strategy. The table reports return, standard deviation and Sharpe ratios obtained with the strategy. Although positive, the overall Sharpe ratio of 0.17 is small. Even before the credit crisis, a Sharpe ratio of 0.44 is not large either, being comparable to the average around 0.5 reported for the U.S. stock market (e.g., Campbell (2003)). This suggests that
Table 5—Return, standard deviation and Sharpe ratio of a strategy based on the predictability of the 6-month credit-spread of AAA-bonds: "buying" the spread when the spot rate is below its historical value and vice-versa. The return (given in equation 11) is given by

\[ R_{t+12} = (1 + S_{t}^6)(1 + S_{t+6}^6) - S_{t}^{12}, \]

where, \( S_{t}^6 \) is the 6-month credit-spread in time \( t \), and \( S_{t+6}^6 \) is the 6-month credit-spread in time \( t + 6 \) and \( S_{t}^{12} \) is the 12-month spread in time \( t \).

<table>
<thead>
<tr>
<th>Period</th>
<th>Return</th>
<th>Stand. Dev.</th>
<th>Sharpe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan-02</td>
<td>3.9</td>
<td>23.5</td>
<td>0.17</td>
</tr>
<tr>
<td>Mar-08</td>
<td>9.3</td>
<td>40.9</td>
<td>0.25</td>
</tr>
<tr>
<td>Mar-08</td>
<td>1.3</td>
<td>3.0</td>
<td>0.44</td>
</tr>
<tr>
<td>Sep-10</td>
<td>9.3</td>
<td>40.9</td>
<td>0.25</td>
</tr>
<tr>
<td>Sep-10</td>
<td>3.9</td>
<td>23.5</td>
<td>0.17</td>
</tr>
</tbody>
</table>

even expecting positive returns, the associated risk may be too high. Assuming however that banks are risk neutral, only expected returns matter and because the strategy is profitable it should be implemented.

Forecasting 1-year credit spreads. — The results do not change much if we use forecasts of the 1-year credit-spread. In this case, the construction of a real strategy would ideally involve maturities of 18 months, as in six months the original spreads will become the forecasted 12-months spreads.

However, as an exercise to compare with the previous results in table 5, we present, in table 6, the performance of a theoretical strategy that buys the 1-year credit spread at period 1, and sells it back six months later (and vice-versa). We can interpret these results as over-the-counter contracts where the interest rate is attached to the 1-year credit-spreads.

The returns in this case will be given as:

\[ R_{t+6} = S_{t+6}^{12} - S_{t}^{12}. \]

As we see in table 6 the results are very similar to the ones in table 5, and the conclusions are equivalent.

III. Conclusion and future research

There are two main conclusions from the present study: The first is that it is possible to create a strategy that yields positive expected returns based on the predictability of credit-spreads among AAA-rated bonds. The second is that it is better to use small estimation windows to forecast credit-spreads because small windows are better to capture changing regimes. The consequence is that the
Table 6—Return, standard deviation and Sharpe ratio of a strategy based on the predictability of the 12-month credit-spread of AAA-bonds: "buying" the spread when the spot rate is below its historical value and vice-versa. The return (given in equation 12) is given by $R_{t+6} = S_{t+6}^{12} - S_t^{12}$, where $S_{t+6}^{12}$ is the 12-month credit-spread in time $t+6$, and $S_t^{12}$ is the 12-month credit-spread in time $t$.

<table>
<thead>
<tr>
<th>Period</th>
<th>Return</th>
<th>Stand. Dev.</th>
<th>Sharpe</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan-02 Mar-08</td>
<td>2.4</td>
<td>5.7</td>
<td>0.42</td>
</tr>
<tr>
<td>Mar-08 Sep-10</td>
<td>14.9</td>
<td>85.5</td>
<td>0.17</td>
</tr>
<tr>
<td>Jan-02 Sep-10</td>
<td>6.5</td>
<td>48.8</td>
<td>0.13</td>
</tr>
</tbody>
</table>

lasso estimation results in better forecasts when compared to the ordinary least squares estimation based on larger estimation samples.

Regarding the strategy based on the mean reversion of AAA-rated credit-spreads, we may use the standard deviation of the returns obtained as a measure of risk. Computing the corresponding Sharpe ratio, we observe that the strategy has a performance similar to a buy and hold investment in the U.S. stock market. Campbell (2003) reports a stock market Sharpe ratio of around 0.5 in the U.S. for instance. The strategy’s performance is worse during the market turmoil of 2007 but still positive. This is clearly an advantage over the stock market. A possible drawback, however, are the transaction costs that I did not consider. Portfolios with small returns, such as the ones obtained here, are particularly sensitive to these costs.

The use of smaller estimation windows has the advantage of better capturing changing regimes, or temporary shocks. However, the challenge is the trade off between the size of the estimation window and a large number of regressors, as it is the case in the present formulation. To address this issue, I propose the use of regularization, the lasso, to improve the finite sample properties of the estimators. The results are, in fact, better than the ones obtained from the traditional ordinary least squares estimation based on a large estimation sample. This happens for all credit-ratings examined, but only for BBB-rated credit-spreads the result is good enough to give (marginally) better forecasts than the Martingale model.

There are a few future research steps. The first is to consider if the metrics used to evaluate the forecasting accuracy is in fact appropriate. For instance, the work of Patton and Timmermann (2007) shows that depending on the criterion used to evaluate forecasts, the mean absolute prediction error, or even the widely used mean squared prediction error can be misleading.

Another step is to improve the estimation of the coefficients. For instance, the lasso estimation has the disadvantage of having only one parameter to control both variable selection and shrinkage of the coefficients. There are alternatives...
such as the "adaptive lasso" of Zou (2006), which is similar to the Lasso but with adaptive weights used to penalize each regression coefficient separately. Another alternative is the "relaxed lasso" of Meinshausen (2007), that has two parameters: one parameter to control variable selection and the other to control shrinkage of the selected coefficients.

REFERENCES


