International Journal of Economics and Finance; Vol. 6, No. 8; 2014
ISSN 1916-971X E-ISSN 1916-9728
Published by Canadian Center of Science and Education

Modeling Corporate Default Rates

Xiaoming Tong[1]

[1] Credit and Emerging Markets, Citigroup, New York, NY, USA

Correspondence: Xiaoming Tong, Credit and Emerging Markets, Citigroup, 388 Greenwich St., New York, NY 10013, USA. Tel: 1-646-285-8779. E-mail: katextong@gmail.com

Received: May 27, 2014 Accepted: June 9, 2014 Online Published: July 25, 2014

doi:10.5539/ijef.v6n8p1 URL: http://dx.doi.org/10.5539/ijef.v6n8p1

Abstract

In this paper, I propose a model for predicting annual one-year high yield default risk. My work is based on the earlier work of Hampden-Turner (2009). My model forecasts monthly default rates using four predictors, each with various lags: Libor 3-month/10-year Slope, U.S. Lending Survey, U.S. Funding Gap, and Gross Domestic Product Quarter-over-Quarter Growth. Forecasts of future corporate default rates are useful for evaluating the attractiveness of credit market investments and for estimating value-at-risk on credit portfolios. I present results of out-of-sample predictions of annual default rates. I also address some imperfections of the Hampden-Turner formulation through utilization of more rigorous selection of variable lags and a logistic transformation of predicted default rates. I demonstrate that estimates of future default probabilities are useful for predicting changes in high yield credit spreads.

Keyword: default, financial crisis

1. Introduction

In this paper I present a model for generating twelve consecutive monthly predictions of high yield default rates and compare out-of-sample model predictions with observed historical default rates. In addition, I demonstrate in historical testing that predicted default rates, but not trailing default rates, can predict directional moves in high yield credit spreads.

The paper begins by providing some perspective on historical default rates and presents a brief description of previous attempts to predict corporate default rates. I next describe in detail the original Hampden-Turner (HT) model. The HT model takes market prices and economic indicators as inputs and generates monthly default rate predictions for the subsequent twelve months. However, the model, in its original form, has some limitations. These include the possibility of generating negative default rates and relatively poor out-of-sample performance. I address these and other shortcomings of the HT model by developing a new model, called HT-2.0. This enhanced model includes a logistic transformation of predicted default rates to avoid negative values, imposes a penalty for large regression coefficients, and uses "pre-whitening" to improve selection of lags for the input variables. I then show the improvement in out-of-sample predictions of historical default rates using the HT-2.0 model.

To demonstrate the usefulness of models for estimating future default rates, I show how forecasts of default probabilities, but not current default rates, are useful for predicting subsequent changes in high yield credit spreads. I also use the model for my simulations of expected losses from default on credit portfolios.

Rating agencies typically calculate the current annual default rate as the percentage of high yield firms that have defaulted over the past twelve months. For example, Moody's Investors Service publishes monthly trailing high yield default rates calculated as the ratio of the number of firms rated below Baa/BBB that have defaulted during the trailing 12-month period to the total number of non-defaulted firms rated below Baa at the beginning of that period (Note 1). Since the universe of rated firms differs among the various rating agencies, it is not surprising that rating agencies typically report different trailing annual default rates. For example, Figure 2 displays annual default rates from Standard & Poor's (Vazza & Kraemer, 2013) and Moody's (Ou, Chu, In, & Metz, 2013) for all firms and for high yield firms only. Although default rates reported by Standard & Poor's appear to be slightly higher than those from Moody's, they typically rise and fall in tandem. This is illustrated graphically in Figure 3, which displays the speculative annual default rates reported by S&P and Moody's.

	All Firms		High Yield	
Year	S&P	Moody's	S&P	Moody's
1981	0.1	0.2	0.6	0.7
1982	1.2	0.7	4.4	2.3
1983	0.8	0.5	2.9	1.8
1984	0.9	0.5	3.3	1.7
1985	1.1	0.4	4.3	1.4
1986	1.7	0.9	5.7	2.8
1987	1.0	0.6	2.8	1.6
1988	1.4	0.8	3.8	1.9
1989	1.7	1.4	4.7	3.3
1990	2.7	2.3	8.1	6.3
1991	3.3	2.0	11.1	6.0
1992	1.5	0.8	6.1	2.7
1993	0.6	0.6	2.5	2.0
1994	0.6	0.3	2.1	1.0
1995	1.0	0.5	3.5	1.6
1996	0.5	0.2	1.8	0.6
1997	0.6	0.3	2.0	0.9
1998	1.3	0.8	3.7	2.0
1999	2.1	1.5	5.6	3.5
2000	2.5	2.2	6.2	5.0
2001	3.8	3.4	9.8	8.4
2002	3.5	2.3	9.3	5.8
2003	1.9	1.2	5.0	3.2
2004	0.8	0.4	2.0	1.2
2005	0.6	0.3	1.5	0.8
2006	0.5	0.3	1.1	0.8
2007	0.4	0.2	0.9	0.5
2008	1.8	1.5	3.6	3.0
2009	4.1	3.7	9.6	8.3
2010	1.2	0.7	2.9	1.6
2011	0.8	0.5	1.7	1.2
2012	1.1	0.7	2.5	1.5
Average	1.5	1.0	4.2	2.7
Median	1.1	0.7	3.6	1.9
Max	4.1	3.7	11.1	8.4
Min	0.1	0.2	0.6	0.5

Figure 1. Annual default rates from standard & poor's and moody's

Source: Standard & Poor's and Moody's Investors Service.

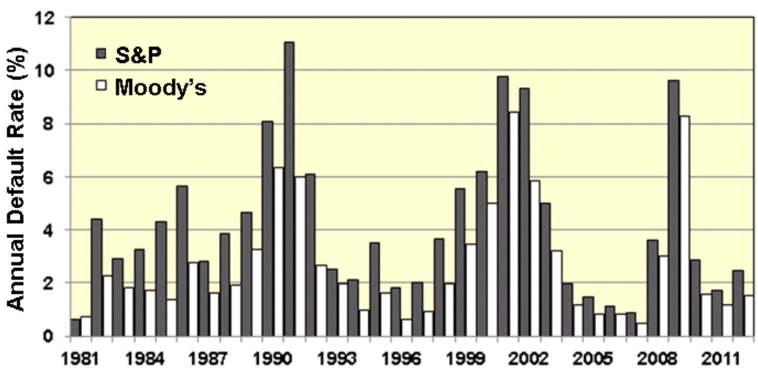

Figure 2. Annual trailing 12-month default rates from standard & poor's and moody's

Source: Standard & Poor's and Moody's Investor's Service.

Although trailing default rates are of some interest to investors, projections of future default rates are even more relevant for performance of corporate markets, particularly those in high-yield. Moreover, expected default rates are of interest to lenders, risk managers, and other counterparties to credit-based transactions. The goal is to generate accurate forecasts of monthly default rates for the next twelve months and to demonstrate their usefulness for making investment decisions.

There have been several approaches to modeling future default rates. Most begin with the observation of the current 12-month trailing default rate. One approach to projecting the default rate for the next 12 months is to generate stochastic future default rates using a model that relies on mean-reverting properties calibrated to the historical properties of historical default rates. The key assumption of this type of model is that default rates follow certain stochastic process and therefore the time-series record of actual default rates is a sample path generated by that process. Although this approach can be useful for simulation purposes, it constrains us to use mean default rates as expected levels of future default rates. To derive a statistical model with predictive power, others have adopted an alternative approach that incorporates econometric factors leading default. Examples include linear models detailed in Fons (1991), HelIge and Kleiman (1996), Jonsson and Fridson (1996). These authors have identified macroeconomic variables of explanatory power whose effectiveness is evaluated by calculating root-mean-squared errors between predicted and obtained default rates. An alternative model proposed by Keenan, Sobehart and Hamilton (1999) incorporates the effect on default rates of changes in the universe of issuers, both in terms of their credit ratings and the time since they first came to market (the "aging effect"). Their model also captures macroeconomic conditions as measured by the industrial production index and interest rate variables. Finally, Hampden-Turner (2009) has developed statistical models to predict future default rates from one to twelve months using least-squares regression and vector autoregressive models.

Unfortunately, most existing models that show good performance, including the Hampden-Turner (HT) model, are validated in-sample. However, I find most studies do not evaluate out-of-sample forecasts over a suitably long period (e.g., cover an entire credit cycle), and it remains unclear how well these models perform, especially in periods of high default rates. In this paper, I first describe the Hampden-Turner model, pointing out its advantages and limitations. Then, building upon the HT framework, I apply statistical approaches to address that model's limitations, while also explaining out-of-sample validation on the enhanced model. Finally, I show how an accurate model of default prediction can provide useful information regarding the attractiveness of investment in high yield corporate debt.

2. The Hampden-Turner Default Model

Since my model takes its starting point from the Hampden-Turner formulation, I present that model briefly in this section. The HT default model fits and predicts monthly default rates using lagged versions of the following four predictors:

- Libor Slope: designated as *LIB*, which is the yield spread between 10-year and 3-month LIBOR rates divided by the term difference between 10 years and 3 months (i.e., 10.0-0.25) (Note 2).
- The U.S. Lending Survey: denoted *LS*. The U.S. Federal Reserve sends lending surveys quarterly to gather opinions from banks' senior loan officers on bank lending practices. The survey estimates the net percentage of domestic banks tightening standards for commercial and industrial loans to large and middle-market firms. Banks tighten loan standards when financial conditions are deteriorating or are expected to worsen, thus leading to tougher environment for high yield credits and higher default rates.
- The U.S. Funding Gap: denoted *FG*. FG is "the macroeconomic equivalent of final free cash flow. It is the net cash flow a company receives (or requires) after capital expenditures, dividend payments, mergers and acquisitions, and net equity issuance. FG is typically negative in a bull market, indicating that corporations' need to increase financing, and positive in a bear market, as consolidation occurs and spending is reduced.
- GDP Quarter-over-Quarter Growth: designated as *GDP*. GDP is the market value of all officially recognized final goods and services produced within a country in a year. GDP growth is indicative of strong economic conditions, portending good corporate performance and vice versa.

Of the four input variables, only *LIB* is collected monthly, while the other inputs are available quarterly. The HT model uses a simple linear interpolation to generate monthly values for variables *LS*, *FG*, and *GDP* to be paired with monthly values of *LIB* between their quarterly updates. For example, to interpolate the value of *GDP* for a given prediction month t one month after the last *GDP* update at t-1, I use GDP_{t-1} and the corresponding value of GDP_{t+2} that will be reported next. That is,

$$GDP_t = \frac{GDP_{t+2} - GDP_{t-1}}{3} + GDP_{t-1} \tag{1}$$

Similarly,

$$GDP_{t+1} = GDP_{t+2} - \frac{GDP_{t+2} - GDP_{t-1}}{3} \tag{2}$$

Because of the linear interpolation, calculation of monthly data at time t requires data ahead at time t+2. As a result, the predictor values at time t can only be used to predict default rates later than $t + 1$ or $t + 2$ unless it is a month of a GDP report (or other quarterly variables). In practice, this is not a problem as the model includes only monthly predictors with lags longer than two months.

HT's first default model fits a lagged regression using ordinary least-squares (OLS) to predict monthly default rates, which for month t I denoted as Γ_t, so that:

$$\Gamma_t = \beta_0 + \beta_1 \cdot LIB_{t-24} + \beta_2 \cdot LS_{t-12} + \beta_3 \cdot FG_{t-12} + \beta_4 \cdot GDP_{t-12} \tag{3}$$

Hampden-Turner reports that the U.S. Lending Survey is the most important predictor, leading the default rate by about 10-12 months. For example, Figure 4 plots historical normalized (i.e., converted to z-scores) values of the U.S. Lending Survey and monthly default rates from Moody's since late 1997. Clearly, the U.S. Lending Survey data not only lead default rates, but also appear to predict their normalized magnitudes well.

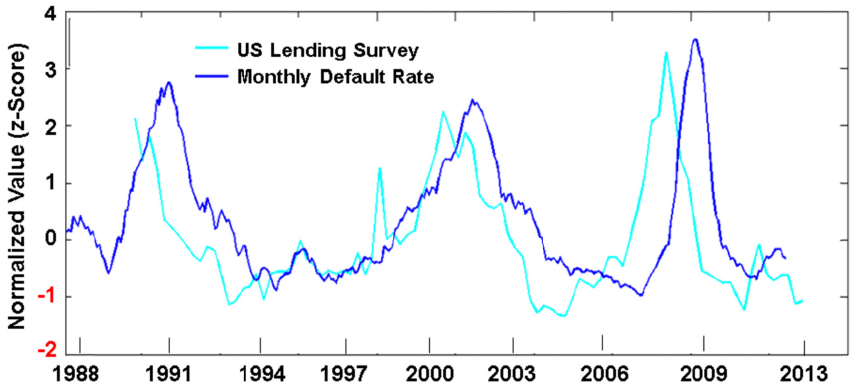

Figure 3. Time series comparison of normalized us lending survey and monthly default rates

Source: Moody's Investor's Service, the U.S. Federal Reserve.

In practice, each January the HT model generates twelve successive out-of-sample monthly default rates based on available data through the previous December. Each default rate for months = 1, …, 12 is calculated as:

$$\hat{\Gamma}_{t+i} = \hat{\beta}_0 + \hat{\beta}_1 \cdot LIB_{t+i-24} + \hat{\beta}_2 \cdot LS_{t+i-12} + \hat{\beta}_3 \cdot FG_{t+i-12} + \hat{\beta}_4 \cdot GDP_{t+i-12} \tag{4}$$

Figure 5 illustrates how the HT model is used in practice. That is, the figure shows that the model is trained on data up to the end of 2011 and then generates monthly predictions of trailing 12-month default rates for 2012 (the red line in Figure 5).

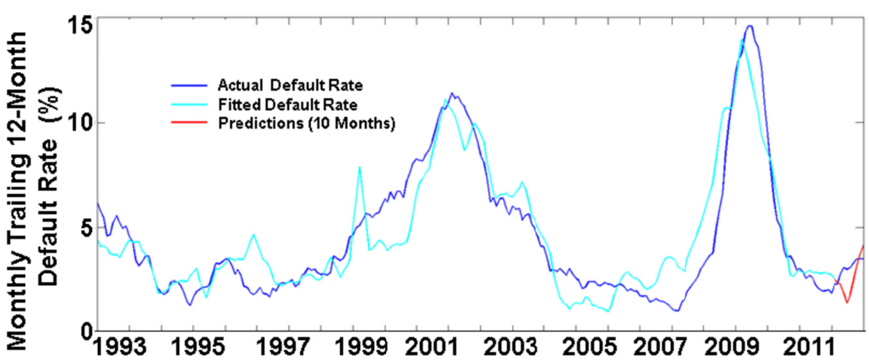

Figure 4. Actual default rates (dark blue) along with fitted and predicted monthly default rates via Hampden-Turner's lagged OLS regression

Hampden-Turner proposed a second model to predict annual default rates where he uses vector autoregressive (VAR) to derive coefficients on the four variables above as well as on the lagged default rate. That is, VAR is used to derive the coefficients for each monthly prediction of Γ_t.

$$\begin{aligned}\Gamma_t = {} & \alpha_1 \cdot \Gamma_{t-8} + \alpha_2 \cdot \Gamma_{t-12} + \alpha_3 \cdot \Gamma_{t-24} + \alpha_4 \cdot \Gamma_{t-48} + \\ & \beta_1 \cdot LIB_{t-4} + \beta_2 \cdot LIB_{t-8} + \beta_3 \cdot LIB_{t-12} + \beta_4 \cdot LIB_{t-24} + \\ & \gamma_1 \cdot LS_{t-4} + \gamma_2 \cdot LS_{t-8} + \gamma_3 \cdot LS_{t-12} + \gamma_4 \cdot LS_{t-24} + \\ & \delta_1 \cdot FG_{t-4} + \delta_2 \cdot FG_{t-8} + \delta_3 \cdot FG_{t-12} + \delta_4 \cdot FG_{t-24} + \\ & \phi_1 \cdot GDP_{t-4} + \phi_2 \cdot GDP_{t-8} + \phi_3 \cdot GPD_{t-12} + \phi_4 \cdot GDP_{t-24} + \omega\end{aligned} \tag{5}$$

Since the longest lag is for default rates 49 months prior, the first month of prediction requires 49 months of previous default rates (i.e., $t \geq 49$). Unlike the lagged linear regression, the minimum lag in the VAR formula is four months. Therefore, in order to make 12 out-of-sample predictions of r_t, I need predictions of LIB, LS, FG as well as GDP at least for months five through twelve. HT proposes to fit a VAR model to each of these time series. For example, $\hat{LIB}_{t+i}$ $(1 \leq i \leq 4)$ can be predicted by fitting the following VAR up to time t:

$$\begin{aligned}LIB_t = {} & \alpha_1 \cdot DR_{t-8} + \alpha_2 \cdot DR_{t-12} + \alpha_3 \cdot DR_{t-24} + \alpha_4 \cdot DR_{t-48} + \\ & b_1 \cdot LIB_{t-4} + b_2 \cdot LIB_{t-8} + b_3 \cdot LIB_{t-12} + b_4 \cdot LIB_{t-24} + \\ & c_1 \cdot LS_{t-4} + c_2 \cdot LS_{t-8} + c_3 \cdot LS_{t-12} + c_4 \cdot LS_{t-24} + \\ & d_1 \cdot FG_{t-4} + d_2 \cdot FG_{t-8} + d_3 \cdot FG_{t-12} + d_4 \cdot FG_{t-24} + \\ & e_1 \cdot GDP_{t-4} + e_2 \cdot GDP_{t-8} + e_3 \cdot GPD_{t-12} + e_4 \cdot GDP_{t-24} + f\end{aligned} \tag{6}$$

Then, to predict LIB_{t+j} $(5 \leq j \leq 12)$, HT treats $\hat{LIB}_{t+i}$ $(1 \leq i \leq 4)$ as already observed data and applies the VAR formula in Equation 6 iteratively to generate estimates of *LIB* for months five through twelve. Otherwise, the general training procedure for the VAR model is similar to that for the OLS predictions in Figure 4 for which the model is trained up to the end of each year and used to predict trailing 12-month default rates for each month in the following year. The resulting estimated trailing 12-month default rates from the VAR model and actual default rates appear in Figure 6.

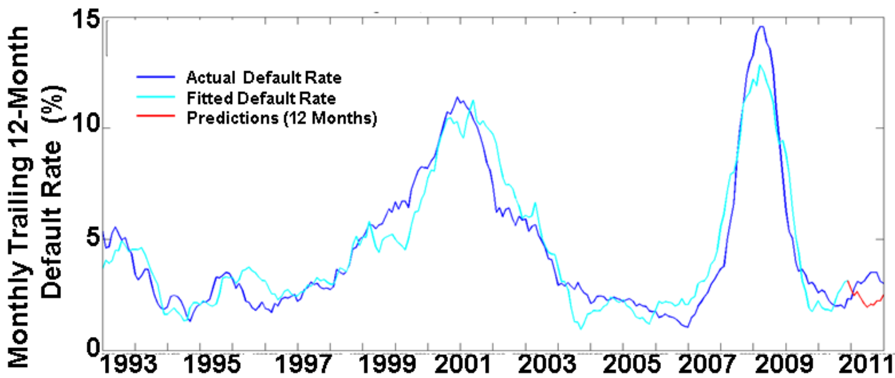

Figure 5. Actual default rates (dark blue) along with fitted and predicted monthly default rates via Hampden-Turner's vector autoregressive model

3. Issues with the Hampden-Turner OLS and VAR Models

There are two statistical issues related with the linear and VAR regression models as implemented in the Hampden-Turner model. First, I find that both models may give rise to negative default rate predictions. In addition, I observed that although the VAR model performs well fitting default data in sample, it exhibited poor out-of-sample prediction performance. For example, Figure 7 shows actual default rates from 1996 to 2013 and out-of-sample predictions from the VAR model (left panel) and the OLS regression (right panel). Neither the VAR or the OLS model capture the actual annual rates well and both models predict negative annual default rates in year 2006. Also, comparison of out-of-sample performance in Figure 7 shows that the simple lagged regression performs better than the VAR model at predicting annual default rates.

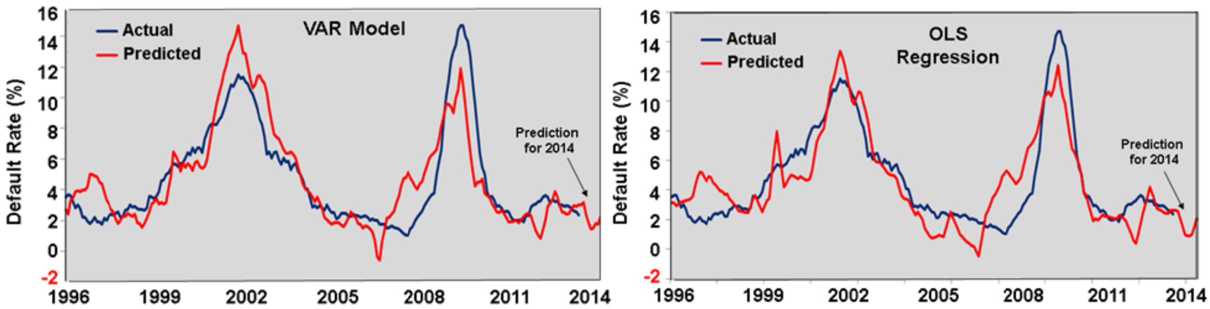

Figure 6. Predicted (red) and Actual (blue) default rates: left: predictions from the VAR model; Right: predictions form OLS regression

Another issue with the HT models involves lag selection. As discussed before, if the minimum lag is less than twelve months, one cannot make twelve monthly out-of- sample default rate predictions from the available data. This motivated HT to build the VAR model for each predictor so that smaller lags can be used in the formula. For the simple lagged linear regression, small lags become a limitation of the model. For example, Hampden-Turner used the cross-correlation function (CCF) to determine the lag for the OLS model. As shown in the left panel of Figure 8, the estimated cross-correlation between *LS* and *DR* peaks around a lag of 10 to12 months. Hence choosing a lag of twelve months for *LS* is convenient and reasonable–I can produce a total of 12 predictions, and I do not lose much predictive power (with respect to, say, a lag of 10 or 11).

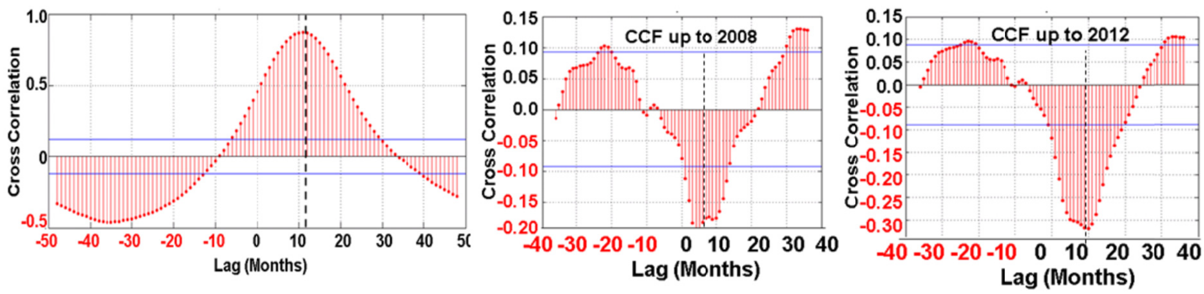

Figure 7. Left: sample cross-correlation function (CCF)–US lending survey versus monthly default rates; sample CCF for GDP QoQ growth versus monthly default rates up to 2008 (middle) and 2012 (right)

Although picking the lag for the Lending Survey when predicting default was straightforward, Hampden-Turner claims that, for the other predictors, picking the lag is “not always this easy.” For example, the middle and right panels of Figure 8 show CCFs for GDP quarter-over-quarter growth versus the monthly default rates using data up to 2008 and 2012, respectively. The CCFs in those panels peak at very different lags, (i.e., four months and ten months), suggesting that using a fixed lag for all the periods may not be appropriate. In addition, both panels show that a lag of twelve months for the GDP is not the optimal choice, apart from its ability to make twelve monthly predictions. Thus for the lagged linear regression, I must chose between the optimal lag and one that can make a sufficient number of predictions. On the other hand, the VAR model does not fully address the problem of lag selection. The VAR model involves considerably more lags that are arbitrarily selected, and how to carry out the selection procedure for the main model as well as for all the predictors is clearly not obvious. In the following sections, I propose a quick fix to the lag selection dilemma along with a discussion of more advanced methods.

4. Lagged Regression Model, HT-2.0

I propose alternative methods to address the issues raised in the previous section for the HT models. First, I opt not to use the VAR model due to (1) its propensity to overfit the data and (2) the fact that it uses predicted results as model inputs that may degrade accuracy of predictions. (I need to check that all the predictor time series are stationary, see Appendix I for a discussion.) Also, to ensure that that predicted default rates are not less than zero, I adopt a simple logistic transformation on the default rates. That is, instead of regressing default rates on the other four predictors, I transform them first via the logit function:

$$\widetilde{DR} = logit(DR\%) = log\frac{DR\%}{1-DR\%} \quad (7)$$

Thus, the simple lagged regression is converted to:

$$\widetilde{DR_t} = \tilde{\beta}_0 + \tilde{\beta}_1 \cdot LIB_{t-24} + \tilde{\beta}_2 \cdot LS_{t-12} + \tilde{\beta}_3 \cdot FG_{t-12} + \tilde{\beta}_4 \cdot GDP_{t-12} \quad (8)$$

To get predictions for the original default rate, I can use the inverse logit transformation $DR = 100 \cdot e^{\widetilde{DR}} / (e^{\widetilde{DR}} + 1)$, which is confined to the range between (0,100). However, because of the exponential term in Equation 8, the predictions from the model are not satisfactory when the underlying regression becomes unstable. For example, the left panel of Figure 9 shows out-of-sample annual default rate predictions from 1996 to 2013 for the lagged regression with logistic transformation of default rates. Although all predictions are now greater than zero, the model has large errors in estimating default rates, particularly from 2001–2002.

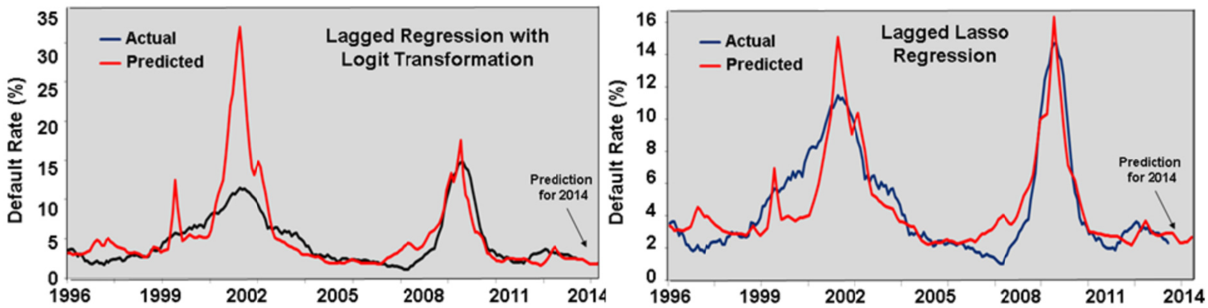

Figure 8. Predicted and Observed Default Rates from Lagged Regression with Logistic Transformation of Default Rates (Left Panel) and Lagged Lasso Regression (Right Panel)Source: Moody's Investor's Service

To improve model performance, I impose a Lasso $\mathcal{L}_1$ penalty on the lagged logit-transformed regression so that the regression coefficients $\tilde{\beta}_1$, $\tilde{\beta}_2$, $\tilde{\beta}_3$ and $\tilde{\beta}_4$ in Equation 8 do not become overly large (Note 3). (Details for the Lasso $\mathcal{L}_1$ penalty appear in Appendix II.) Note that the coefficients in Equation 8 are determined by finding:

$$\hat{\beta} = \arg\min_{\beta} \sum_{t=25}^{T} (\widetilde{DR}_t - \tilde{\beta}_0 - \tilde{\beta}_1 \cdot LIB_{t-24} - \cdots - \tilde{\beta}_4 \cdot GDP_{t-12})^2 \quad (9)$$

The Lasso finds the fitted coefficients by

$$\hat{\beta}^{LASSO} = \arg\min_{\beta} \left\{ \frac{1}{2} \sum_{t=25}^{T} (\widetilde{DR}_t - \tilde{\beta}_0 - \tilde{\beta}_1 \cdot LIB_{t-24} - \cdots - \tilde{\beta}_4 \cdot GDP_{t-12})^2 + \lambda \left(\left|\tilde{\beta}_1\right| + \left|\tilde{\beta}_2\right| + \left|\tilde{\beta}_3\right| + \left|\tilde{\beta}_4\right| \right) \right\} \quad (10)$$

To address the issue of instability in lagged variable relationships, I first obtain the best lag from the data set and run the transformed regression with Lasso penalty as in Equation 10 until I cannot make further out-of-sample predictions. Then I rerun Equation 8 with minimum lag at twelve months to get the rest of the monthly default predictions. For example, say the lag selected for GDP is seven months. I use the seven-month lag to make the first seven predictions. Then I switch to a twelve-month lag to make the next five predictions. The improvement in out-of-sample results is evident in the right panel of Figure 9 which shows the predicted default rates using the lagged Lasso regression. I call my lagged Lasso regression model HT-2.0, as an extension of the original Hampden-Turner model.

5. Selecting Lags

In the previous sections I have used the absolute peak in CCF to choose the lags for the input variables in the regressions. One striking feature of these CCF plots in Figure 8 is that the estimated CCFs seem to be highly persistent. That is, there are numerous lags that cluster around the peak value, making it difficult, and perhaps even misleading, to choose the best lag relationship on that basis. The cause of this clustering is typically attributed to the autocorrelation structure in individual series. Here I used the pre-whitening technique introduced by Jenkins and Watts (1968) to clarify the lagged relationships (Note 4). See Appendix III for the systems approach (Chatfield, 2004) I use that is based on pre-whitening.

For example, to study the lagged relationship between monthly default rates and U.S. Lending Survey data up to the end of 2012, I first plot the sample autocorrelation function (ACF) and partial autocorrelation function (PACF) of the US Lending Survey time series as shown in Figure 10. The ACF plot in the left panel displays a damped sine-cosine pattern and PACF cutoff before lag 10, indicative of an autoregressive process, *AR*(*p*), with complex roots. By using the Bayesian Information Criterion (BIC) (Note 5) as the model selection criterion, I choose the maximum lag p = 8 and arrive at an *AR*(8) model that only includes lags at 1 and 8 months. If I denote B as the backshift operator (i.e., $BX_t = X_{t-1}$), the estimated model can be written as:

$$(1-1.055B+0.0914B^8)(LS_t-\mu)=w_t \tag{11}$$

Pre-whitened LS_t is just the residual w_t from the *AR*(8) model. The left and middle panels of Figure 11 confirm that the residual time series w_t does not show large deviations from the standard normal distribution.

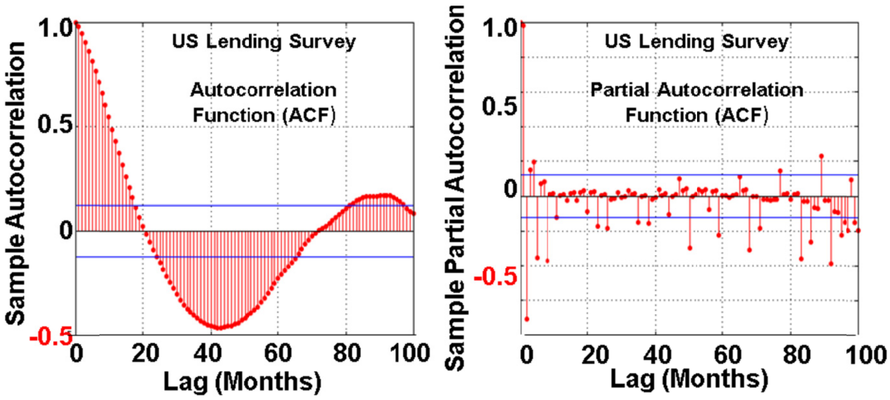

Figure 9. Sample autocorrelation function and partial autocorrelation function of US lending survey with future default rates

Next, I filter the monthly default rate DR_t using the same *AR*(8) model for LS_t and obtain

$$(1-1.055B+0.0914B^8)DR_t=d_t \tag{12}$$

Finally, in the right panel of Figure 11 I plot the sample CCF of d_t versus w_t, which differs dramatically from the CCF shown in Figure 8. I conclude that lags of LS_t at lags of 9, 11, 12, and 13 should be tried as predictors of DR_t and selected via BIC. Interestingly, lag 10 does not appear to be significant anymore in the new CCF plot. Figure 11 also shows why the VAR model for *LS* may be overfitting by including previous lags of *DR*; after all, *LS* is leading *DR*.

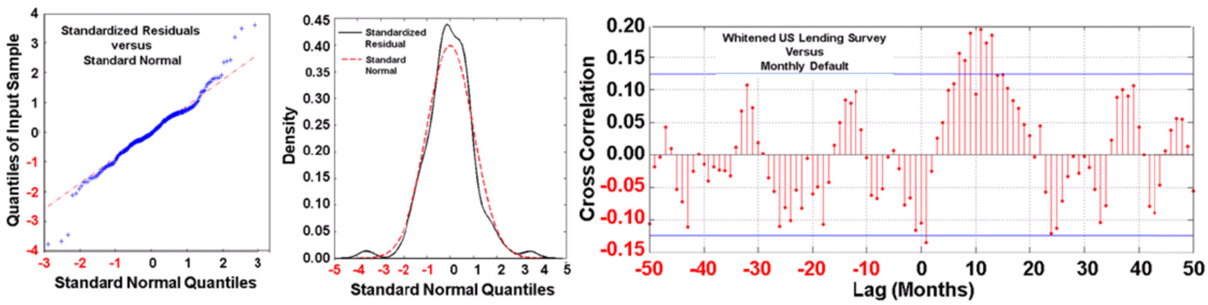

Figure 10. U.S. whitened lending survey: quantile-quantile plot (left); density plot (middle); and CCF Function (right)

6. Why Predict the Default Rate?

One may question the utility of developing a model to predict the overall corporate default rate. After all, the rating agencies publish forecasts of default rates at least on an annual basis. The main reason for building my own model to predict default is that I wish to test the ability of monthly out-of-sample forecasts of annual default rates to signal price moves in the corporate credit markets. I present results of my initial tests of that ability in this section. In addition, I need monthly estimates of default rates as inputs to my stochastic model of credit-cycle-dependent rating transitions (see BenzschalI, Lee, & Li, 2012). Finally, I wish to gain insight into the factors that underlie overall default rates and developing models to forecast default rates is one way to improve that understanding.

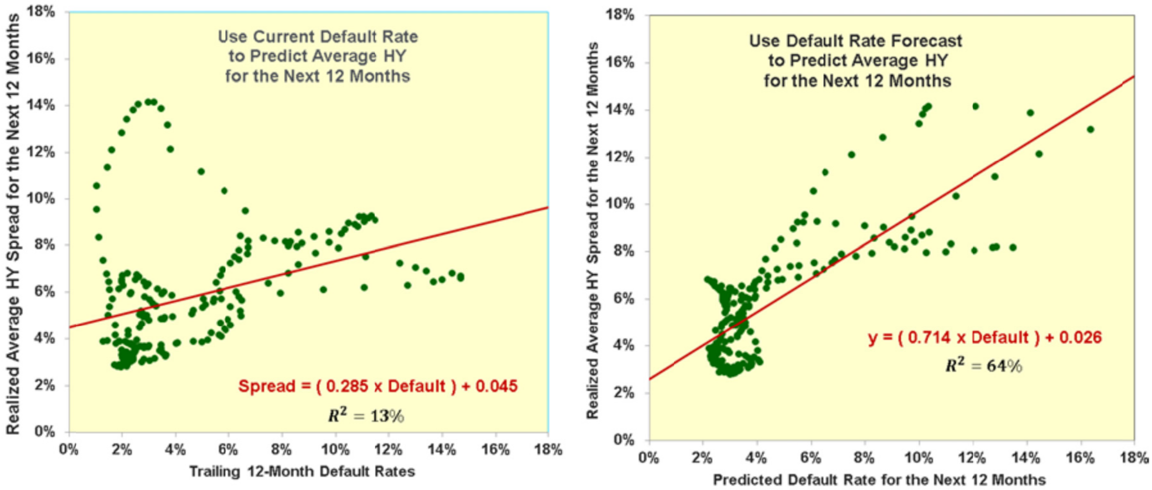

Figure 11. Relationship between annual changes in high yield spreads versus 12-month trailing default rates (left) and predicted default rates (right), Jan-95 to Jun-13

To illustrate the use of the HT-2.0 model to predict high yield credit spreads, consider Figure 12. The figure displays annual changes in average high yield corporate bond spreads versus either the trailing 12-month default rate (left panel), called the current default rate, or the predicted default rate (right panel). To construct those plots, each month I determine either the trailing or predicted 12-month default rates and their associated changes in average high yield corporate bond spreads over the subsequent 12 months. These are the points on the graphs in the left and right panels of Figure 12, respectively. The predicted default rate has an R^2=0.64 over the period from January 1995 to June 2013, whereas the current default rate has an R^2=0.13.

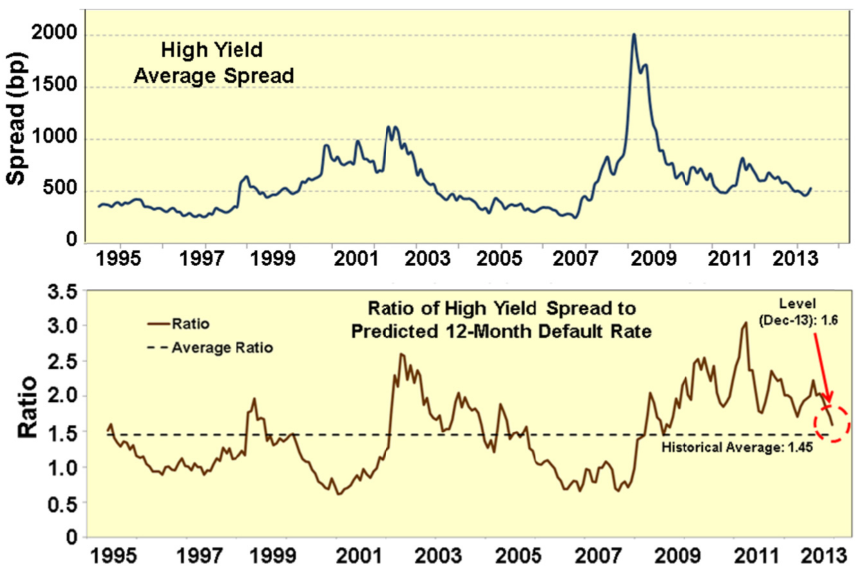

Figure 12. Top: average high yield corporate bond spreads; bottom: ratios of high yield spreads to predicted default rates, 1995–2013

Despite the advantage of predicted default rates over trailing default rates for estimating changes in high yield spreads over the next year, neither predictor is satisfactory. A more relevant question for potential investors in high yield bonds is, "How much yield will I receive for taking on a given level of default risk?" For example, even if default rates are relatively high, an investor may be well compensated by outsized yield spreads to Treasuries (e.g., think 2009). Conversely, if spreads are tight, defaults may be low and investors may still earn attractive returns owing to few defaults. Consider the historical series of average high yield corporate bond spreads in the top panel of Figure 13. Clearly, spread levels vary widely over the cycle. However, the absolute level of spreads does not indicate whether investment in high yield is attractive nor provides reliable signals regarding the future direction of spread moves. A large determinant of those returns depends on the expected default rate over the investment horizon.

To determine if the ratio of high yield spreads to default provides useful information to investors, I first plotted ratios of average high yield spreads to predicted default rates from the HT-2.0 model since 1995. These appear in the lower panel of Figure 13 (Note 6). The assumption is that the ratio of the current high yield spread to predicted default rate is indicative of the attractiveness of high yield returns. Consider first the left-hand panels of Figure 14. The upper panel shows one-year changes in high yield spreads as a function of ratios of the current high yield average spreads to trailing 12-month default rates. In that plot, the green circles represent changes in spreads when the spread-to-default ratios are below average, with the orange squares plotting changes when the ratios are above average. Points are determined monthly, with the vertical blue line showing the average spread-to-default ratio over the period from 1995 to 2013. The scatterplot reveals that ratios using the current trailing 12-month default rate have little ability to forecast high yield spreads one year later. This is confirmed in the bar chart in the lower left-hand panel that presents probabilities of spreads widening or tightening if the ratio of high yield spreads to trailing default rates are above average or below average, respectively. That is, probabilities of spreads widening or tightening are independent of the ratio of high yield spread to current default rate (i.e., probabilities are roughly 50% for all ratios), falling at or near the dashed chance performance line.

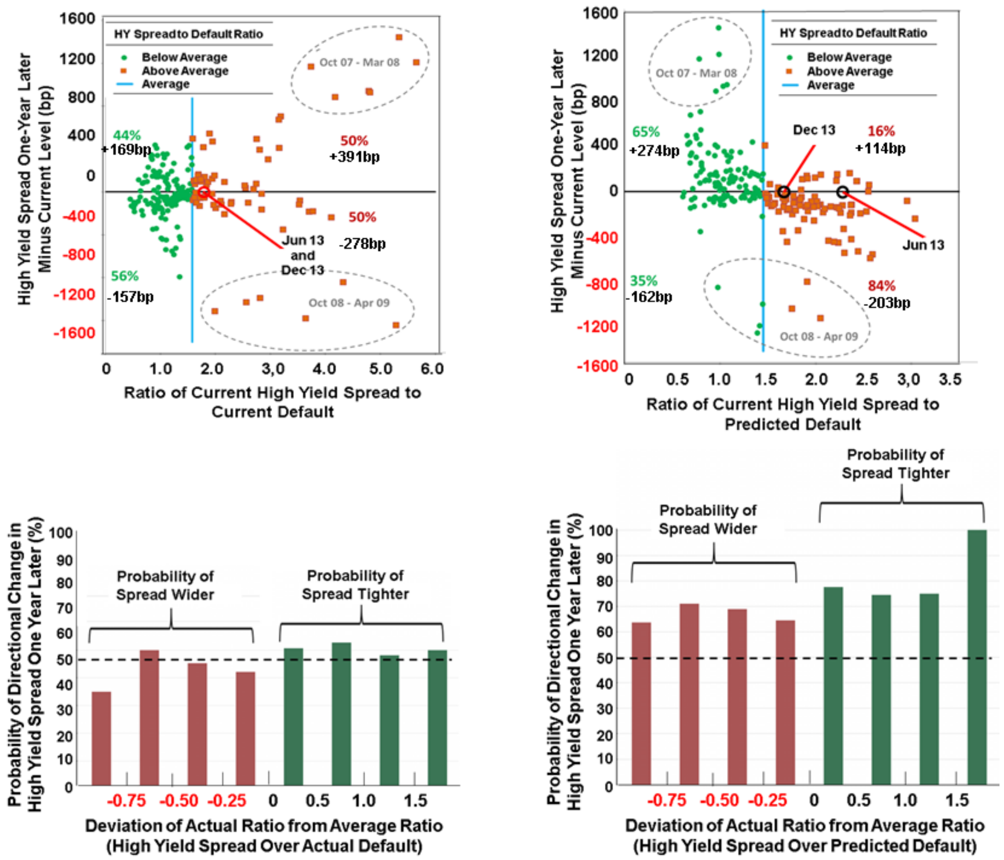

Figure 13. Comparison of predicted changes in high yield spreads based on current default rates (left panels) or predicted default rates (right panel). percentages of falling in each cell and average spread changes are also shown

In contrast to the results using ratios of spreads to trailing default rates, the panels on the right in Figure 14 demonstrate that ratios of spreads-to-default using predicted 12-month default rates are highly related to one-year changes in high yield spreads. That is, when ratios of high yield spreads to predicted default rates are above average, high yield spreads one year later are tighter 85% of the time. When ratios are below average, spreads are wider 65% of the time. The histogram of percentages of spreads widening or tightening as a function of the spread-to-default ratio in the lower right panel of Figure 14 confirms the above-chance performance over the entire range of ratios. The figure also reveals that the size of the ratio has little effect on directional accuracy, except if the ratio is above or below zero. In particular, notice how the directional changes in spreads reverse from widening to tightening on either side of the average spread-to-predicted default ratio. Finally, note that when there are "errors" in the signal from the spread-default ratio (i.e., spreads tightening when ratios are below average and vice versa), average "losses" are smaller than average gains when "correct." For example, when the spread-to-default ratio is above average, the average spread tightening is 203bp, whereas when spreads rise, they rise only 114bp on average. In fact, given the spread-to-predicted default ratio in December 2013 indicated by the circle in upper right panel of Figure 14, the expected spread tightening by December 2014 is:

$$\left[84\% \times (-203bp)\right] + [16\% \times (+114bps)] = -152bps$$

Similarly, if the ratio of high yield spreads to predicted default is less than average, then historical analysis suggests that the average high yield spread will widen by:

$$\left[65\% \times (+247bp)\right] + [15\% \times (-162bps)] = +121bps$$

The results presented in Figure 13 and Figure 14 are intended to demonstrate the usefulness of modeling predicted default rates. I will continue to update my twelve-month default predictions on a monthly basis and use the model for my simulations of expected losses from default on credit portfolios.

7. Summary

I described a model for predicting 12-month default rates and examined its performance in out-of-sample testing since 1994. The default forecasting model is called HT-2.0 as it is an extension of the model first described by Hampden-Turner (2009). HT-2.0 takes market prices and economic indicators as inputs and generates monthly default rate predictions for the next twelve months. The paper begins by providing some perspective on historical default rates and includes a brief discussion of previous attempts to forecast corporate default rates. I next described in detail the original Hampden-Turner model and addressed some limitations of that model. These include the possibility of the original model to generate negative default rates and that, when I use the model to generate out-of- sample predictions, I observe relatively poor performance. I overcame the shortcomings of the original HT model in HT-2.0 by adopting a logistic transformation on predicted default rates to avoid negative values, by imposing a penalty for large regression coefficients, and by using statistical "pre-whitening" to improve estimates of optimal lags for the input variables. I then show the improvement in out-of-sample predictions of historical default rates using the HT-2.0 model.

To demonstrate the usefulness of estimating future default rates, I show how forecasts of future default probabilities, but not current default rates, are useful for predicting subsequent changes in high yield credit spreads.

References

Chatfield, C. (2004). *Analysis of Time Series: An Introduction*. Chapman and Hall.

Dickey, D., & Fuller, W. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association, 74*(366), 427–431. http://dx.doi.org/10.2307/2286348

Fons, J. (1991). *An Approach to Forecasting Default Rates*. Moody's Special Paper, August.

Hampden-Turner, M. (2009). *From Vicious to Virtuous? Corporate Default Rate Forecasting*. Citi.

Hastie, T., Tibshirani, R., & Friedman, J. (2008). *The Elements of Statistical Learning: Data Mining*. Inference, and Prediction, Springer.

HelIge, J., & Kleiman, P. (1997). Understanding Aggregate Default Rates of High Yield Bonds. *The Journal of Fixed Income, 7*(1), 55–61. http://dx.doi.org/10.3905/jfi.1997.408202

Jenkins, G., & Watts, D. (1968). *Spectral Analysis and Its Applications*. Holden-Day.

Jonsson, G., & Fridson, M. (1996). Forecasting Default Rates on High Yield Bonds. *The Journal of Fixed Income, 6*, 69–77. http://dx.doi.org/10.3905/jfi.1996.408166

Keenan, S., & Hamilton. (1999). *Predicting Default Rates: A Forecasting Model for Moody's Issuer-Based*

Default Rates. Moody's Investors Service, August 1999. Retrieved from http://ssrn.com/abstract=1020303

Ou, S., Chu, D., In, B., & Metz, A. (2013). Annual Default Study: Corporate Default and Recovery Rates, 1920–2012. Moody's Investors Service.

Schwarz, G. E. (1978). Estimating the Dimension of a Model. *Annals of Statistics, 6*(2), 461–464. http://dx.doi.org/10.1214/aos/1176344136

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society, Series B, 58*(1), 267–288.

Vazza, D., & Kraemer, N. (2012). Annual Global Corporate Default Study and Rating Transitions, Standard and Poor's.

Notes

Note 1. High yield firms are those rated below Baa by Moody's and BBB-by Standard & Poor's with investment grade firms having higher ratings by each agency.

Note 2. Note that the Hampden-Turner (2009) documentation claims to use the number of high-yield issuers to the total number of issuers, but the actual implementation used the LIBOR slope instead.

Note 3. The Lasso method is one example of regularization methods designed to prevent overfitting by penalizing extreme parameter values. Regularization introduces a second factor, in Equation 10, which shrinks regression coefficients. For a technical discussion of the Lasso method, see Tibshirani (1996) and Appendix II.

Note 4. The term "pre-whitening" is used in construction of lagged regressions to minimize the effects of co-movements in the x- and y-variables in the lag analysis on the cross-correlation function.

Note 5. The Bayesian Information Criterion (BIC), developed by Schwartz (1978), measures the variance reduction provided by the addition of each variable to the model. In addition, it imposes a penalty for having too many variables, thereby guarding against overfitting the data.

Note 6. The units for default in the numerator of the ratio in Figure 13 are in percent. For the high yield spread in the denominator, we use the spread in basis points. For example, for a predicted default rate of 2.3% and a current high yield spread of 230bp, we calculate the ratio as

Note 7. For example, the most salient condition for stationary is that the mean of the series does not change over time. For example, an upward or downward drift in the series would imply non-stationary. There are other considerations as well.

Note 8. The Dickey–Fuller test (Dickey and Fuller, 1979) is used to determine whether a unit root is present in an autoregressive model.

Appendix A

Time Series Regression

To perform time series regressions, it is crucial to determine whether or not the series under consideration are stationary over time (Note 7). That is, regressing non-stationary series on non-stationary series may lead to spuriously significant regressions. To test if a time series is stationary, I use the Augmented Dickey- Fuller (ADF) test (Note 8).

If I have identified a time series input to be non-stationary, I need to be careful including such series as a predictor. Some type of transformation should be applied to make the series more stationary. For example, the raw GDP time series may have an upward trend over the long run, so it is better to use, say, the quarterly GDP growth instead. This is the approach that I took in the model. I can also apply transformations such as first-order difference (for example, the monthly change in a predictor variable), or de-trending a series to avoid predictions that appear to fit well in-sample but will have low predictive power in practice.

Appendix B

Lasso $\mathcal{L}_1$ Penalty. Linear Regression with $\mathcal{L}_1$ Penalty—An Introduction to the Lasso

For a linear regression model with dependent variable y_i (i = 1, ... , N), input variables x_{ij} (j = 1, ... , p), and coefficients $\beta_0, \beta_1, \dots, \beta_p$, the ordinary least squares (OLS) estimate of $\beta_0, \beta_1, \dots, \beta_p$ is given by:

$$\hat{\beta}^{OLS} = \arg\min_{\beta} \sum_{i=1}^{N} (y_{i-\beta_0-} \sum_{j=1}^{p} x_{ij}\beta_j)^2 \tag{13}$$

Although the fitted coefficients of the OLS regression are unbiased, they may suffer from ill-posed conditions and complicated correlation structure among variables. In addition, if my focus is on time-series regression, the fitted coefficients may be unstable over time, leading to unintuitive changes in the fitted values over different training periods. Modern regularization techniques such as the lasso ($\mathcal{L}_1$) attempt to overcome these problems by seeking a sparse solution by inclusion of a $\mathcal{L}_1$ penalty for the coefficients:

$$\hat{\beta}^{lasso} = \arg\min_{\beta} \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 \tag{14}$$

$$\text{subject to } \sum_{j=1}^{p} \left|\beta_j\right| \le t$$

Rewriting Equation AII-2 in an equivalent Lagrangian form, we have

$$\hat{\beta}^{lasso} = \arg\min_{\beta} \left\{ \frac{1}{2} \sum_{i=1}^{N} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \left|\beta_j\right| \right\} \tag{15}$$

Since the $\mathcal{L}_1$ constraint makes the solution to the optimization problem nonlinear in y_i, no closed-form solution exists. However, one can use quadratic programming to compute the lasso solution, and efficient algorithms are developed to generate the entire path of solutions with varying λ (Hastie, Tibshirani and Friedman, 2008).

Note that t or λ, the so called the shrinkage parameter, plays an important role in controlling the size of the fitted coefficients. For example, if t is sufficiently large, the penalty almost has no effect and the lasso estimate is essentially the OLS estimate. On the other hand, if t is relatively small, then some of the coefficients would be shrunk towards 0. In the extreme case when t=0, all the coefficients will be shrunk to 0 in the optimization routine.

Because lasso tends to shrink many coefficients to 0, it automatically performs variable selection.

Appendix C

Pre-Whitening. Spurious Correlation and Pre-Whitening for Lagged Regression

To identify the lags in a lagged time-series regression, a common approach is to examine the plot of the Cross-Correlation Function (CCF) between the response time series Y_t and input series X_t. Unfortunately, the CCF between Y_t and X_t may be influenced by the autocorrelation structure of these two series. That is, significant cross-correlation between two unrelated time series may be observed as an artifact of strong autocorrelation in both series. To avoid reading a spurious relationship between the two time-series, I adopt the "pre-whitening" strategy as outlined in the following steps:

1. Determine a time series model for the X variable. For example, for stationary X_t with mean 0, fit an (*ARMA*(p, q) model to X_t.

$$X_t = \sum_{i=1}^{p} \varphi_i X_{t-i} + w_t + \sum_{j=1}^{q} \theta_j w_{t-j} \tag{16}$$

where w_t is a white noise series (i.e., independent and identically distributed random series with mean 0 and finite variance). Alternatively, one can simplify the above formulation by the backshift operator B as:

$$\varphi(B) X_t = \theta(B) w_t \tag{17}$$

where $\varphi(B)$ and $\theta(B)$ are polynomials in B In other words, one can "whiten" the original series X_t by applying the filter $\varphi(B)/\theta(B)$:

$$\frac{\varphi(B)}{\theta(B)} X_t = w_t \tag{18}$$

The fitted AMRA parameters then lead to the residual $\tilde{w}_t$:

$$\frac{\hat{\varphi}(B)}{\hat{\theta}(B)} X_t = \tilde{\tilde{w}}_t \tag{19}$$

One can check the normality assumption by a Quantile-Quantile plot on the residuals. It is not crucial, and often not practical to find the exact time series model for *X*, but filtering *X* to an approximate white noise series is necessary.

2. Since *Y* and *X* are assumed to have a lagged linear relationship, I filter the *Y* variable similarly by the estimated whitening filter from the first step. That is

$$\frac{\hat{\varphi}(B)}{\hat{\theta}(B)} Y_t = \tilde{d}_t \tag{20}$$

Note that, in practice, it is easier to filter by AR coefficients only (i.e., $\theta(B) = 1$), so for convenience we may want to check if fitting *X* with an AR process suffices during the first step.

3. Examine the CCF plot $\tilde{w}_t$ and $\tilde{d}_t$ and identify possible lags for the lagged relationship between *Y* and *X*.

Copyrights