Modeling Challenges for Credit Risk and Economic Forecasting

William Morokoff
Managing Director
Quantitative Analytics and Research Group

Forensic Analysis of Financial Data Workshop
IPAM, May 21, 2015

Agenda

Modeling Credit Risk and Economics: Data

Examples:

- **PD Models and Performance Measures**
- **Credit Scoring Models**
- **Text Mining Regulatory Filings**
- **Bond Issuance Forecasting**

Contributors

Giorgio Baldassari

Wenbo Cao*

Ronen Feldman**

Mark Haydutov

Charles Huang*

Saiyid Islam

Jun Torres

* Work done while at S&P Ratings Service (currently TIAA-CREF)

** Consultant for S&P, Professor, Hebrew University Business School

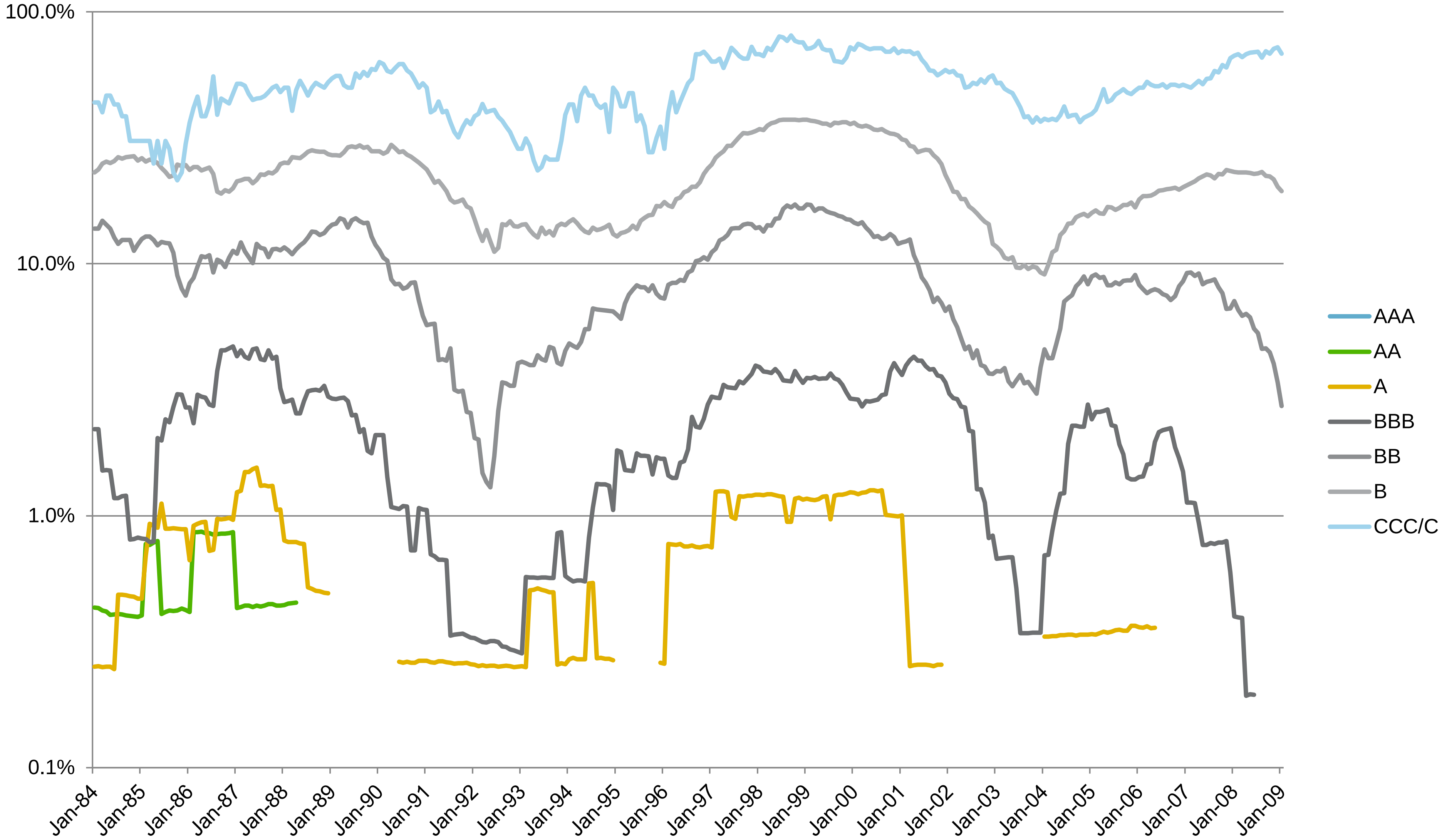
Modeling Driven by Data

The nature of data available for economic and credit risk analysis is changing, raising questions of how best to work with Big Data, Unstructured Data, and generally Different Data.

However, most work is still done on structured, time series data that must be carefully curated and may not be that big:

- Quarterly or Annual observations, e.g. GDP
- Rare events such as investment grade defaults
- Limited time history
- Long time horizons of interest – default risk over 10 years
- Need for human calibration/training

5 Year Rated Corporate Default Rates



A Few References

- **Hal Varian. Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives*: Vol. 28 No. 2 (Spring 2014)**
- **Floris van Ruth. Analysing and predicting short term dynamics in key macro-economic indicators. 34th International Symposium on Forecasting, Rotterdam, 2014.**
- **Craig Friedman and Sven Sandow. Utility-Based Learning from Data (Chapman & Hall/Crc: Machine Learning & Pattern Recognition), 2010.**

Probability of Default Modeling

Corporate PD Modeling

- **Goal: Estimate probability of default of a firm based on a firm's financial data and economic data, calibrating to best separate defaulters from survivors**
- **Challenges:**
  - Limited number of defaults. Consortiums often formed, but with inconsistent data collection standards changing through time.
  - Data often reported only annually,
  - Data may be restated.
  - Difficult to measure absolute performance
- **Methodology: Many methods used, but logistic regression (or generalizations) is commonly used.**

PD Modeling Process

Collect Firm and Market Data:

- Company specific financial ratios, debt levels, liquidity measures, …
- Macro-economic and market data
- Equity price, volatility, rank, etc.
- Need many years and many firms

Collect Default Data:

- Tag each firm observation as survivor or defaulter period T

Construct and Scale/Rank-Transform Factors

Calibrate Model:

- Factor selection (e.g. Greedy Forward)
- Parameter calibration (MLE)
- Out of sample evaluation – k-fold validation.

Measure Performance

- How well does model differentiate defaulters and survivors
- How well does model fit the observed data

Performance: Likelihood and Goodness of Fit

- **Log-Likelihood:**

$$L = \sum_{j=1}^{N_D} \log\left(PD_{i(j)}\right) + \sum_{k=1}^{N_S} \log\left(1 - PD_{i(k)}\right)$$

- **Likelihood Ratio Test: Determine if model A fits the data better than model B.**

$$R = 2\left(L_B - L_A\right) \qquad \mathrm{R} \sim \chi^2$$

- **Deviance Test: Does the model fit the data well?**

$$D = -2L \qquad D \sim \chi^2$$

- **Chi-Squared Test: Does the model fit the data well?**

$$\chi^2 = \sum_{j=1}^{N_D} \frac{1 - PD_{i(j)}}{PD_{i(j)}} + \sum_{k=1}^{N_S} \frac{PD_{i(k)}}{1 - PD_{i(k)}}$$

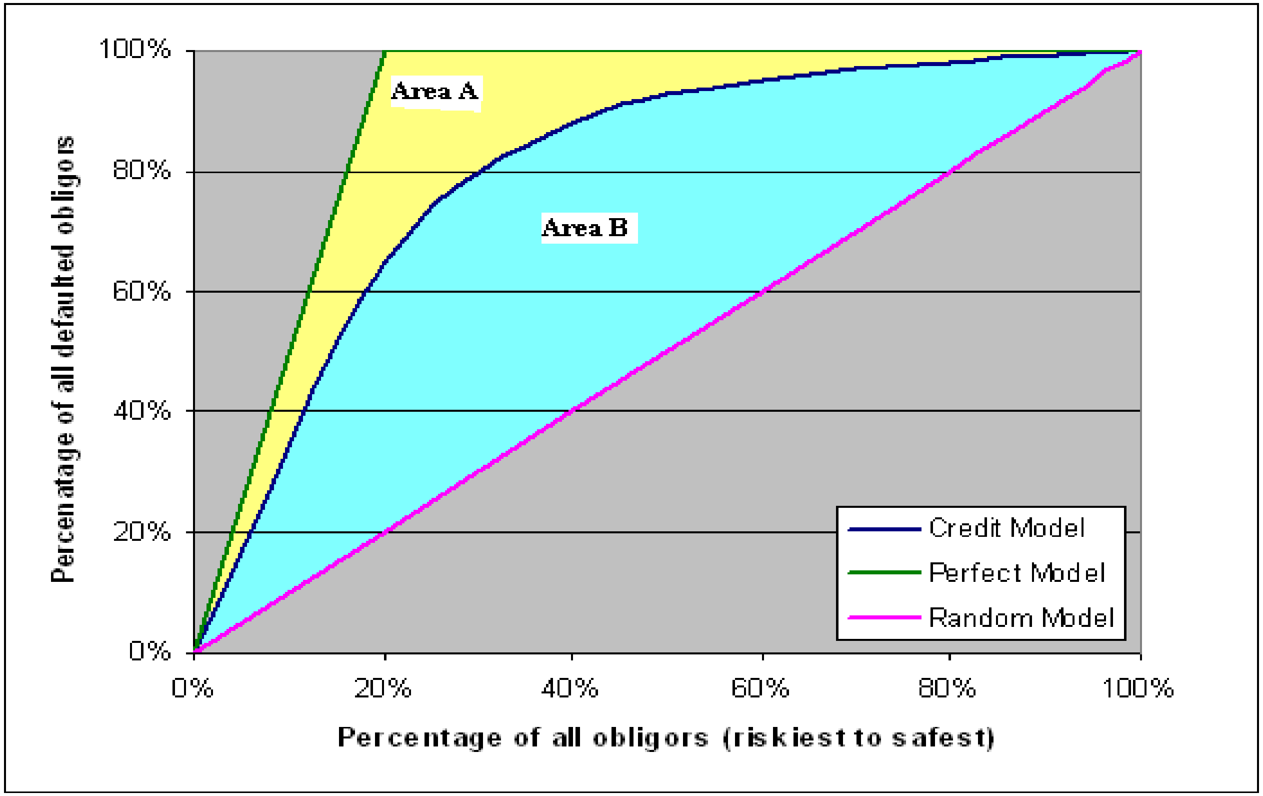
- **Hosmer Lemeshow Test:**

$$H = \sum_{g=1}^{G} \frac{\left(O_g - N_g \pi_g\right)^2}{N_g \pi_g \left(1 - \pi_g\right)}$$

Performance Measurement: Accuracy Ratio (AR)

- **Sort firms/assets/obligors from riskiest to safest as predicted by the credit model (x-axis) and plot against fraction of all defaulted obligors.**

- **Accuracy Ratio = B / (B + A)**



Note: The terms Gini Coefficient and Accuracy Ratio are often used interchangeably in credit modeling literature

Accuracy Ratio Calculation

$N =$ Number of sorted company PD observations: $PD_i \geq PD_{i+1}$

$M =$ Number of observed defaults

$X_i = 1$ if default, 0 if no default

$$AR_N = \frac{\frac{M}{N} - 2\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{i}{N}\right)X_i - \frac{M}{2N^2}\right]}{\frac{M}{N}\left(1 - \frac{M}{N}\right)}$$

Now take the limit as N goes to infinity.

Limiting AR

- Assume that there exists a distribution function F(PD)
- Data set is considered as a sample of PDs from F(PD) iid.
- PDs can be considered random variables.
- For this calculation, we assume that each sampled PD is a true PD, i.e. the probability that the issuer defaults is exactly PD, so X_i is a random variable and

$$E\left(X_i\right) = PD_i$$

- In the limit as N goes to infinity,

$$\frac{M}{N} \to \overline{PD} = E\left(PD\right) = \int_0^1 PD \; dF(PD) = 1 - \int_0^1 F(y)dy$$

Limiting AR

$$AR_{\infty} = \frac{\overline{PD} - 2\int_0^1 \left(\int_0^{F^{-1}(z)} y \; dF(y) \right) dz}{\overline{PD}\;(1 - \overline{PD})}$$

With some calculus:

$$AR_{\infty} = \frac{1 - \int_0^1 F^2(y)dy - \overline{PD}}{\overline{PD}\;\left(1 - \overline{PD}\right)}$$

Observations on Limiting AR

- If {PD_i} and the associated {X_i} are considered random variables, then AR_N is a random variable and

$$E\left(AR_N\right) = AR_\infty + O\left(1/N\right)$$

$$\sigma^2\left(AR_N\right) = O\left(1/N\right)$$

- Full distribution of AR_N can be computed with simulation.

Observations on Limiting AR

Conclusion: Even for a perfect PD model, as a performance measure for the model, Accuracy Ratio is a noisy (sample-size dependent) estimate of a quantity that depends only on the nature of the population, not the quality of the model.

Credit Scoring Models

Credit Scores Modeled on Ratings

- **Goal: Estimate probability of a firm's credit score (rating) based on financial, economic and country data by calibrating model to a rated set of firms (agency or private credit assessments).**
- **Challenges:**
  - Rated universe may relatively small for some sectors or regions.
  - Nature of unrated firms may be different than calibration set.
  - Expert judgment can be difficult to model.
- **Methodologies:**
  - Linear Regression
  - Ordered Logistical Regression
  - Proximal Support Vector Machine
  - Exponential Density Model

Exponential Density Model (CreditModel)

- **Factors:** $\left(x_{1,i},\ldots,x_{m,i}\right)$ **are rank-transformed financial ratios for firm *i* .**
- **Credit Score:** $y_i \in \left(1,\ldots,18\right)$ **(AAA to CCC).**
- **Prior Distribution:** $p_0\left(y\right)$ **empirical distribution of ratings (e.g. % of companies rated BBB).**
- **Model:**

$$p(y\,|\,x,\beta) = \frac{\exp\left(\beta^T f\left(x,y\right)\right)}{c\left(x,\beta\right)}\,p_0\left(y\right)$$

- **Simplified Constraint:**

$$\beta^T\left(f\left(x,y\right) - f\left(x,y+1\right)\right) \geq 0$$

- **Credit Score** $= E\left(y\,|\,x\right) = y^T p\left(y\,|\,x,\beta^*\right)$

Advantages of Exponential Density

- Positive probabilities can be obtained for states not observed in training sample.
- Monotonic property in each factor can be explicitly enforced through constrained optimization.
- Previous versions of CreditModel used a SVM framework, but the non-linearity in the classifier led to occasional unintuitive results.
- Output of model is probability distribution, from which score can be estimated as the mean (or other statistic).

Automated Processing of Regulatory Filings

Real Time Analysis of SEC Filings (Sherlock)

- **Sherlock automatically flags text that potentially indicates credit deterioration problems.**
- **Sherlock processes incoming SEC filings to help prioritize analysis efforts.**
- **The tool allows us to identify, process and analyze disclosures that could affect credit quality as quickly as possible as information becomes available.**

Sherlock Methodology

- **Sherlock pre-processes SEC filings (10-Q, 10-K, 8-K, etc.) based on certain data rules.**
- **Each paragraph is converted to a set of terms and an associated sparse vector representation.**
  - The length of the vector is the size of the dictionary.
  - The value of the i-th component is the frequency of the i-th term in the paragraph.
- **Samples: Analysts have manually assessed many paragraphs and categorized them into 32 risk categories. These sample paragraphs have also been converted into vector representation.**

Sherlock Methodology

- **Hierarchical Classification Method (training):**
  - A binary linear support vector machine (*SVM*) is trained to separate samples with risk tags from samples without risk tags.
  - A multi-class SVM is trained to classify samples (with risk tags) into different risk categories using one vs. all approach.
- **Scores for NoTag and Risk Categories:**
  - w is the parameter for NoTag vs Tag and w_i is the parameter of the i_th risk category vs others. x is a sample vector.
  - $<w,x>$ greater than -0.2 is classified as 'Tag'.
  - Given 'Tag', the i_th risk category score is

$$\max\left(\min\left(< w_i, x >, 2\right), -2\right) / 4 + 0.5$$

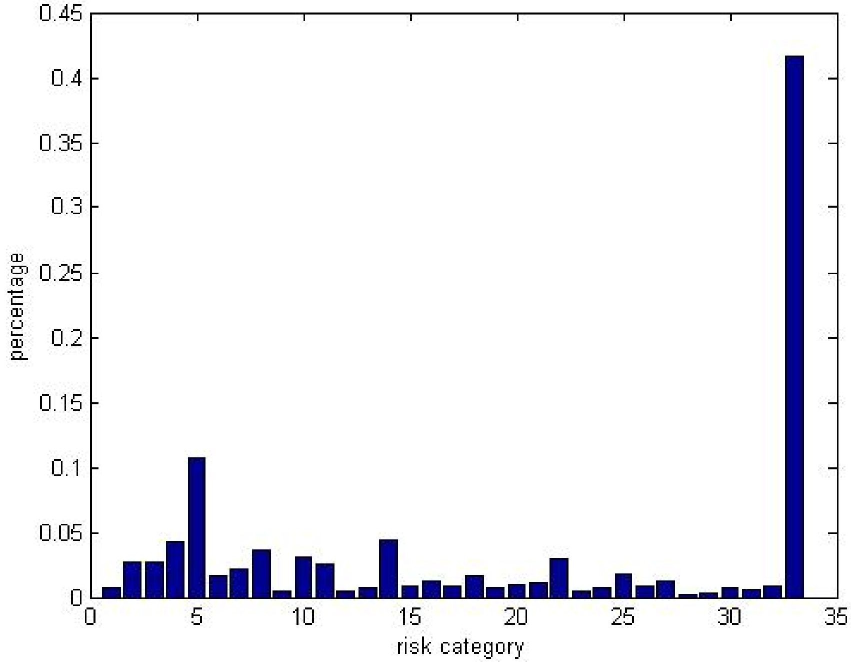
  - The larger scores will imply the bigger possibility of the sample in the risk category.

Sherlock Model Performance

- **For each category, we randomly split samples into training and test data sets (80% vs. 20%).**
- **Parameters are estimated with training samples.**
- **Performance is measured for test samples.**
  - $$\text{Recall} = \frac{\#\text{ of correctly classified tagged samples}}{\#\text{ of tagged samples}} \times 100;$$
  - $$\text{False negative rate} = \frac{\#\text{ of tagged samples being classified as NoTag}}{\#\text{ of tagged samples}} \times 100;$$
  - $$\text{False positive rate} = \frac{\#\text{ of notag samples being classified as tagged}}{\#\text{ of notag samples}} \times 100.$$
- **We repeated the procedure 20 times.**

Sherlock Model Performance

- **33 categories**
  - 32 Risk
  - 1 NoTag
- **14990 samples**
- **6247 NoTag Samples**
- **For 32 risk categories**
  - Max: 1600 samples
  - Min: 31 samples
  - Mean: 273.22 samples



	Mean	Std
recall	87.84	0.60
false negative rate	3.30	0.50
false alarm rate	14.88	1.46

Challenges for Model Calibration

- **Across analysts, tags may not be consistently identified, and this is difficult to detect.**
- **Correctly labeling a large set of samples is time consuming and labor intensive. Further work: is it possible to classify ‘easy-to-identify’ samples and use analysts for ‘hard-to-identify’ samples?**
- **The model maximizes the recall rate and minimizes the false negative rate. However, this causes a large number of NoTag samples to be labeled by the model as Tag sample. This requires analysts to review a large set of alerts.**

Bond Issuance Forecast

Two Methods for Issuance Forecasting

- **Goal: Forecast US corporate and financial institution bond issuance for the next four quarters.**
- **Motivation:**
  - Research – Report on market and economic trends that impact issuance in the credit markets across various sectors.
  - Business Operations Planning – Rating Agency revenue is highly tied to bond issuance. Forecasts are useful for budget planning.
- **Two approaches studied:**
  - Random Forest method using 11 years of quarterly data on 59 factors, potentially lagged.
  - Classic OLS regression that incorporates economic variable forecasts calibrated on up to 25 years of data.

Random Forest Method

- Different models built for investment grade and speculative grade credits using 11 years of data.
- For each model, 500 decision trees, selected to be different and minimize correlation, were sampled.
- Each tree contained ~35 variables, with 8-10 randomly selected variables considered at each split.
- ‘Out-of-bag’ method used for cross-validation with 70/30 training/testing for each tree.
- 10 months ‘out-of-time’ testing conducted.

Random Forest Results and Questions

Top Factors Selected:

✓Employment Picture
✓Slope of Yield Curve
✓Treasury Curve
✓LIBOR Rates
✓Corp. Debt Structure Index
✓Business Growth
✓Stock Market Dynamics
✓Number of CUSIP Applications
✓Consumer Confidence

	Difference ($M)	Difference (%)
2013Q3	(22,000)	-12.4%
2013Q4	11,000	7.0%
2014Q1	(3,000)	-1.9%
2014Q2	(20,000)	-10.4%
2014Q3	43,000	33.6%
Average	1,800	3.2%

- **How robust is model through time in different regimes?**
- **How can results be more easily attributed to various factors?**

Regression Model

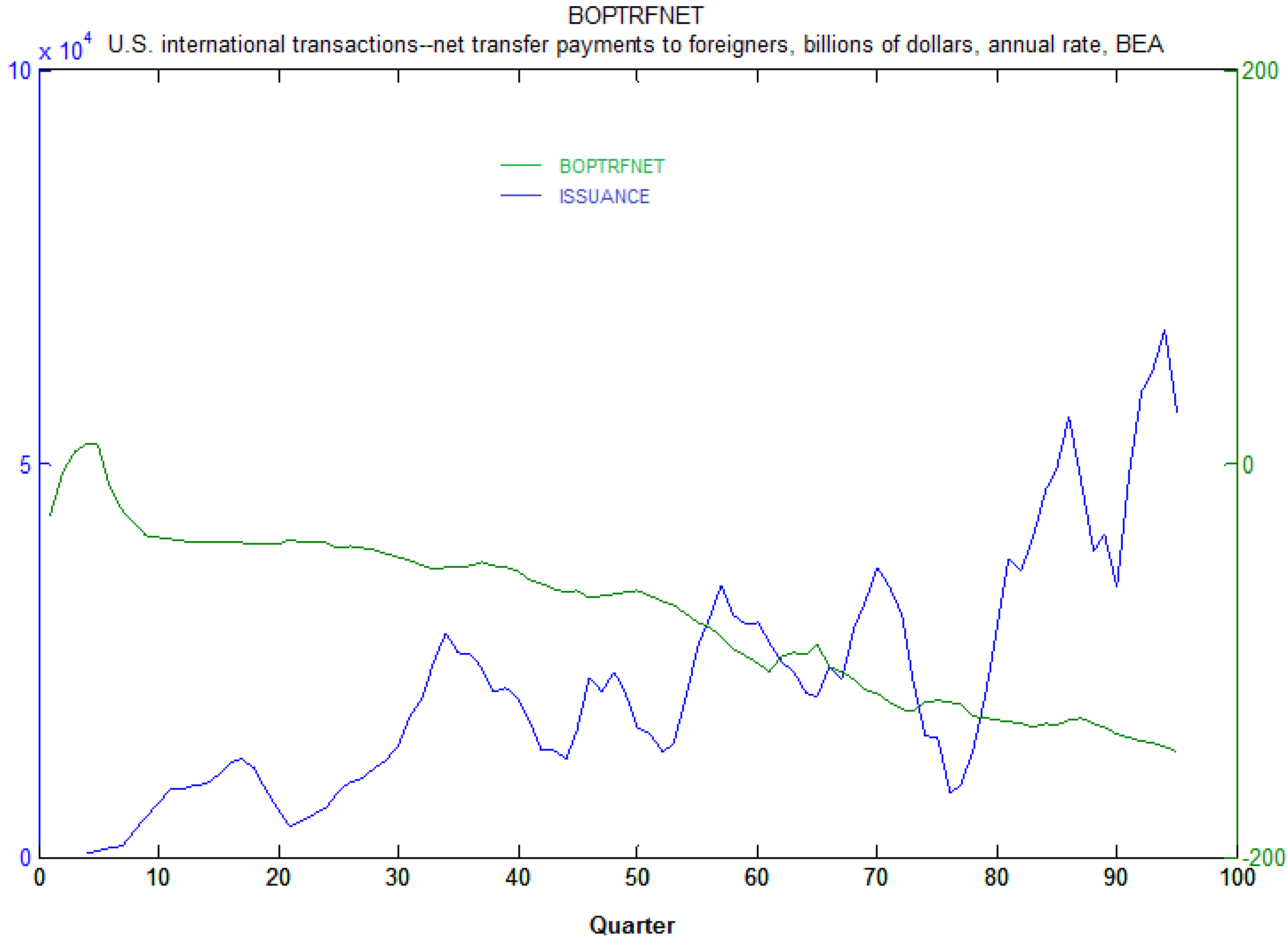
Data

- Time-series (quarterly) of macroeconomic data from Global Insight and Fed Flow of Funds
- Market data such as historical equity index and VIX levels
- (Independent) forecasts of macroeconomic factors from S&P economists

Selection of Variables/Factors

- Roughly 1800 variables considered
- About 70% eliminated based on "economic intuition" / "visual inspection"
- Another 25% eliminated through correlation analysis
- Either Low correlation with issuance or High correlation with other factors
- Single factor regressions on issuance to eliminate non-significant factors
- Further regressions in combinations of two/three/more factors, including lagged and transformed (logs, powers) variables, and eliminating insignificant ones

Visual Confirmation of Unrelated Factor

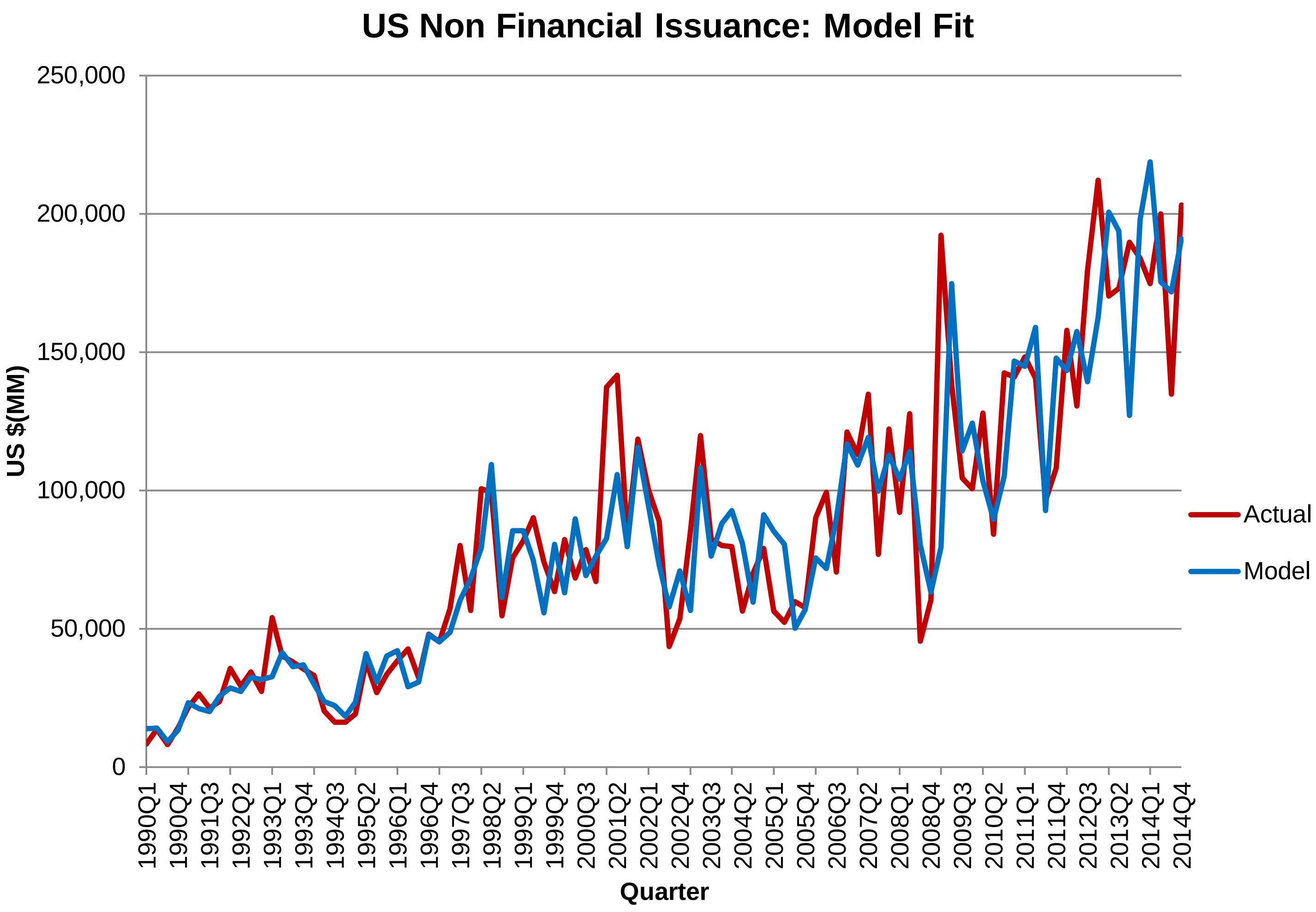


Regression Model for US Corporates

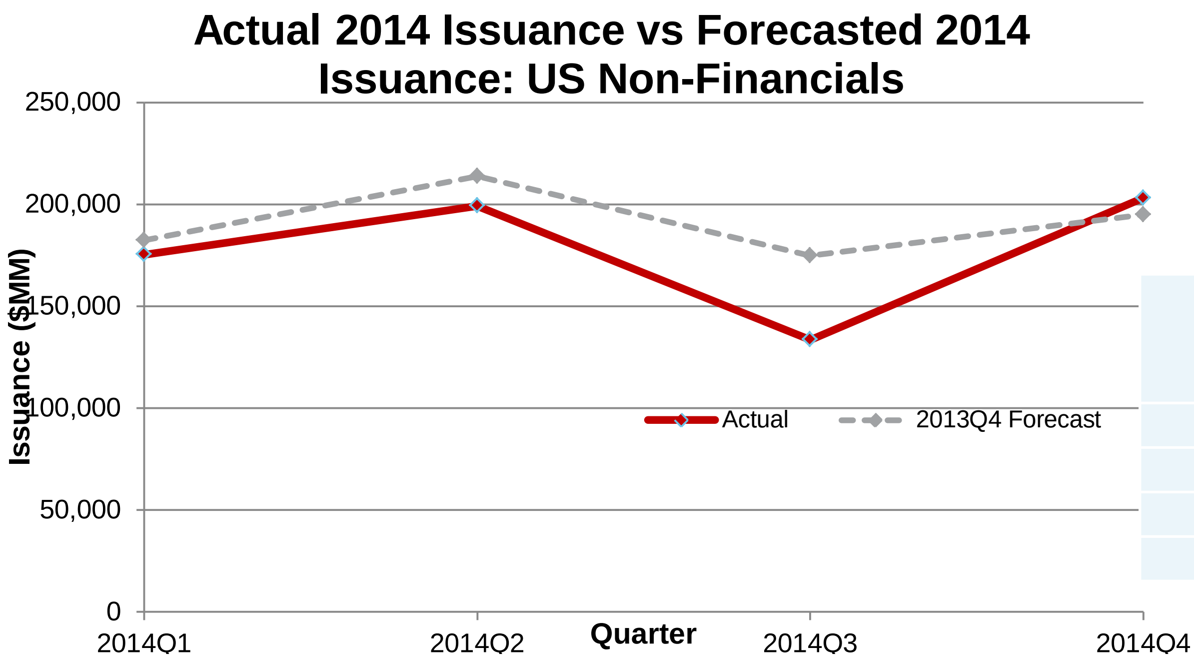
Parameter	Standardized Coefficient	t Stat
First lag of quarterly issuance	0.53	6.16
Second lag of quarterly issuance	0.25	2.98
Log GDP	0.20	2.72
Change in the 10-Year Treasury Yield	-0.13	-2.94
S&P 500 Growth Rate	0.13	3.97
Change in Weekly Hours in Durable Manufacturing	-0.12	-3.49
Change in the VIX Index	-0.10	-2.49
Third Quarter Seasonality Dummy	-0.14	-3.58

Adjusted R Square: 89.1%

Regression Model for US Corporates



US Corporates: 2014 Forecast vs Actual



Quarter	Actual Issuance	Issuance Forecasted in Q4 2013
2014Q1	175,238	182,373
2014Q2	199,130	213,813
2014Q3	133,578	174,811
2014Q4	203,188	194,738

- **Is calibration to actual historical performance or historical forecasts better?**
- **Should calibration be to actual as-of date data or revised data?**

Thank You

William Morokoff
Head of Quantitative Analytics
T: 212.438.4828
William.morokoff@standardandpoors.com

STANDARD & POOR'S RATINGS SERVICES
McGRAW HILL FINANCIAL