Credit
Research
Initiative
Risk Management Institute, National University of Singapore

# Variable Selection with Big Data based on Zero Norm and via Sequential Monte Carlo

**Jin-Chuan Duan**

National University of Singapore and CriAT
(https://rmi.nus.edu.sg/duanjc)

(September 2018)

# Regulated linear regression

$$\boldsymbol{y} = \boldsymbol{X}\beta + \epsilon$$

where $\boldsymbol{y} = (y_1, \cdots, y_n)'$, and $\boldsymbol{X}$ denotes the $n$ observations of $p$ regressors. $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_k)'$ is the $k$-dimensional regression coefficients, and $\epsilon$ is $n$-dimensional *i.i.d.* normally distributed errors with mean 0 and variance $\sigma^2$.

**<u>Lasso:</u>** The penalized regression subject to the $l_1$-norm regularization:

$$\arg\min_{\boldsymbol{\beta}} ||\boldsymbol{y} - \boldsymbol{X}\beta||_{l_2}^2$$
$$s.t. \;\; ||\boldsymbol{\beta}||_{l_1} \leq t$$

Alternatively formulated as a Lagrangian problem:

$$\arg\min_{\boldsymbol{\beta}} ||\boldsymbol{y} - \boldsymbol{X}\beta||_{l_2}^2 + \lambda||\boldsymbol{\beta}||_{l_1}$$

**Zero-norm:** The penalized regression subject to the zero-norm regularization:

$$\begin{gathered}\arg\min_{\beta} ||\boldsymbol{y} - \boldsymbol{X}\beta||_{l_2}^2 \\ s.t.\ \ ||\beta||_{l_0} \leq p_s \leq p\end{gathered}$$

where $||\cdot||_{l_2}$ is the $l_2$-norm and $||\cdot||_{l_0}$ is the zero-norm, which counts the number of non-zero entries in $\beta$.

Alternatively formulated and targeted in this study:

$$\arg\max_{\{\boldsymbol{U}\in\boldsymbol{P}(p_s)\}} \exp\left\{-||\boldsymbol{y} - \boldsymbol{X}_{\boldsymbol{U}}\hat{\beta}(\boldsymbol{U})||_{l_2}^2\right\}$$

where $\boldsymbol{U} = (U_1, U_2, \cdots, U_{p_s}) \in \boldsymbol{P}(p_s) \equiv \{\boldsymbol{U} \in \boldsymbol{P}^{p_s} \& U_1 \neq U_2 \neq \cdots \neq U_{p_s}\}$, $\boldsymbol{P}^{p_s}$ and $\boldsymbol{P} = \{1, 2, \cdots, p\}$; $\boldsymbol{X}_{\boldsymbol{U}}$ denotes the sub-matrix of $\boldsymbol{X}$ whose columns correspond to $\boldsymbol{U}$; and $\hat{\beta}(\boldsymbol{U}) = (\boldsymbol{X}_{\boldsymbol{U}}'\boldsymbol{X}_{\boldsymbol{U}})^{-1}\boldsymbol{X}_{\boldsymbol{U}}'\boldsymbol{y}$, the optimal regression coefficients given $\boldsymbol{U}$.

# Zero-norm vs. $l_1$-norm penalty

- Zero-norm is intuitive and natural to variable selection, but it's NP-hard. Lasso and its variants can be efficiently solved with convex optimization.
- Lasso (Tibshirani, 1996) doesn't possess the oracle property. All regression coefficients are biased toward zero due to the $l_1$ penalty. (This can be fixed by, for example, adaptive Lasso of Zou (2006).
- The "Irrepresentable Condition" of Zhao and Yu (2006) states that "Lasso selects the true model consistently if and (almost) only if the predictors that are not in the true model are 'irrepresentable' by predictors that are in the true model. So, multicollinearity creates problems for Lasso.
- The $l_2$ norm is invariant to rotations but the $l_1$ penalty is not. Practically speaking, correlated variables may not be selected even though one of them is part of the true model.

# What is sequential Monte Carlo (SMC)?

- SMC involves generating a sequence of Monte Carlo samples so that the final sample represents a target distribution.
- Bayesian posterior distribution is a motivating example where the posterior is known by the Bayes theorem to be proportional to the product of the prior distribution and likelihood function.
- The prior can be easily simulated by Monte Carlo, but the posterior is not. Hence, a sequence of Monte Carlo samples is needed to reach the posterior.
- Density tempering proposed in Del Moral, *et al* (2006)[1] is one of the two ways to do so. The other is through expanding data, which is a less robust procedure according to the Duan and Fulop (2015)[2].

[1] Sequential monte carlo samplers, Del Moral, P., A. Doucet, and A. Jasra (2006), *Journal of the Royal Statistical Society: Series B* 68(3), 411-436

[2] Density-Tempered Marginalized Sequential Monte Carlo Samplers, J.C. Duan and A. Fulop (2015), *Journal of Business and Economic Statistics* 33(2), 192-202

# A primer on Monte Carlo sampling

## Importance sampling

If one wants to sample according to a density $f(x;\theta)$, it can alternatively deploy an easy-to-sample density $g(x;\psi)$ where the support of $g(x;\psi)$ must contain that of $f(x;\theta)$ and apply the importance weight $w(x) = f(x;\theta)/g(x;\psi)$. For a sample $\{x_i, i = 1, 2\cdots, n\}$, the following weighted sample can be used:

$$(\boldsymbol{x}, \boldsymbol{w}) = \begin{pmatrix} x_1, & w(x_1)/\sum_{i=1}^n w(x_i) \\ x_2, & w(x_2)/\sum_{i=1}^n w(x_i) \\ \vdots & \vdots \\ x_n, & w(x_n)/\sum_{i=1}^n w(x_i) \end{pmatrix}$$

**Example**: generate a *t*-distributed random sample using a normal distribution with the same mean and a larger variance.

## Theoretical justification

$$E_f[H(X)] = E_g\left[H(X)\frac{f(X,\theta)}{g(X;\psi)}\right]$$

if the support of $g(X;\psi)$ contains the support of $f(X,\theta)$.

Since $E_g\left[\frac{f(X,\theta)}{g(X;\psi)}\right] = E_f(1) = 1$,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{f(x_i,\theta)}{g(x_i;\psi)} \approx 1 \quad \text{or} \quad \sum_{i=1}^{n} w(x_i) \approx n$$

Thus, importance sampling can be executed without the norming constant in the density/distribution function.

## Effective sample size

$$ESS = \frac{\left(\sum_{i=1}^{n} w_i\right)^2}{\sum_{i=1}^{n} w_i^2} \leq n,$$ and equals $n$ only with a constant weight.

## Resampling

Resample $(\boldsymbol{x}, \boldsymbol{w})$ to generate an equally-weighted sample

$$(\boldsymbol{x}^*, \mathbf{1}) = \begin{pmatrix} x_1^*, & 1/n \\ x_2^*, & 1/n \\ \vdots & \vdots \\ x_n^*, & 1/n \end{pmatrix}$$

which won't improve the sample quality because equal weights are obtained at the expense of a reduced support, i.e., the number of distinct sample points is less than $n$. But resampling is critical to SMC.

# What is the density-tempered SMC?

Density-termpered SMC for Bayesian estimation in a most general way[3]:

$$f_\gamma(\boldsymbol{\theta}; \boldsymbol{y}_{1:T}) \propto \left( \frac{\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y}_{1:T}) p(\boldsymbol{\theta})}{I(\boldsymbol{\theta})} \right)^\gamma \times I(\boldsymbol{\theta})$$

where $\boldsymbol{y}_{1:T}$ is the data set, $\boldsymbol{\theta}$ is the unknown model parameter, $\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y}_{1:T})$ is the likelihood (or pseudo-likelihood) function, $p(\boldsymbol{\theta})$ is the prior density/distribution, and $I(\boldsymbol{\theta})$ is the density/distribution of some initialization sampler whose support contains that of $\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y}_{1:T}) p(\boldsymbol{\theta})$.

Setting $p(\boldsymbol{\theta}) = 1$ gives rise to maximum likelihood estimation.

[3] Estimating Distance-to-Default with a Sector-Specific Liability Adjustment via Sequential Monte Carlo, Duan, J.-C. and C. W. T. Wang (2017), in Applied Quantitative Finance, eds. C. Chen, W.K. Hardle and L. Overbeck, Springer.

- When $\gamma = 0$, then $f_0(\boldsymbol{\theta}; \boldsymbol{y}_{1:T}) = l(\boldsymbol{\theta})$.
- When $\gamma = 1$, $f_1(\boldsymbol{\theta}; \boldsymbol{y}_{1:T}) \propto \mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y}_{1:T}) p(\boldsymbol{\theta})$, i.e., the posterior distribution under Bayesian or $\propto \mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y}_{1:T})$, i.e., the likelihood function under frequentist.
- For statistical analyses, the distribution of $\boldsymbol{\theta}$, i.e., $f_1(\boldsymbol{\theta}; \boldsymbol{y}_{1:T})$, is the issue, one can apply importance sampling to generate $\boldsymbol{\theta}$ according to $l(\boldsymbol{\theta})$ and apply the importance weight: $\frac{\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y}_{1:T}) p(\boldsymbol{\theta})}{l(\boldsymbol{\theta})}$ (Bayesian) or $\frac{\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y}_{1:T})}{l(\boldsymbol{\theta})}$ (frequentist).
- One can use a Bayesian computation technique without being a Bayesian statistician. In fact, The density-tempered SMC can serve as an optimization method without any statistical inference purpose.

- However, the ESS is likely to be very small because importance weights are highly variable. So, one must temper the importance weight in a controlled fashion with a sequence of $\gamma$ from 0 to 1; that is a density-tempered SMC.

- Controlling the ESS is an excellent way to perform density-tempered SMC. This is done by looking for a $\gamma$ value at which the ESS is no less than, say, $n/2$. Sequentially, one in effect applies the incremental weight ($\gamma_j > \gamma_{j-1}$):

$$\frac{f_{\gamma_j}(\boldsymbol{\theta}; \boldsymbol{y}_{1:T})}{f_{\gamma_{j-1}}(\boldsymbol{\theta}; \boldsymbol{y}_{1:T})} \propto \left( \frac{\mathcal{L}(\boldsymbol{\theta}; \boldsymbol{y}_{1:T}) p(\boldsymbol{\theta})}{I(\boldsymbol{\theta})} \right)^{\gamma_j - \gamma_{j-1}}$$

# Critical steps in the density-tempered SMC

**Step 1: Initial sampling**
Find a reasonable initialization sampler whose density/distribution is $I(\theta)$, i.e., $\gamma_0 = 0$. Note that the support of $I(\theta)$ must be large enough, and beyond which the initialization sampler is arbitrary except affecting efficiency.

In a true Bayesian context, the prior distribution can in principle be the initialization sampler, which will simplify the importance weight to the likelihood value, i.e., $\mathcal{L}(\theta; \boldsymbol{y}_{1:T})$. However, the prior is likely a poor initialization sampler. Why?

**Step 2: Finding next $\gamma$**
Evaluate the tempered incremental importance weight:

$$\left(\frac{\mathcal{L}(\theta; \boldsymbol{y}_{1:T})p(\theta)}{I(\theta)}\right)^{\gamma-\gamma_{j-1}}$$

to find some $\gamma^*$ at which the ESS is greater than a threshold value.

## Step 3: Resampling and support boosting

Resample according to the incremental weights to equalize weights. It is essential to boost the support before moving forward because the empirical support will shrink due to resampling. Support boosting can be accomplished by applying several Metropolis-Hastings moves to randomly sampled sub-blocks to achieve, say, an average cumulative acceptance rate of at least 500%.

Let $\mathcal{M}_{\boldsymbol{\theta}}$ denote some statistics of the sample for $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$ the proposed new parameter value based on some independent proposal density/distribution, $h(\boldsymbol{\theta}^* \,|\mathcal{M}_{\boldsymbol{\theta}}\,)$. Then, the MH acceptance probability is

$$\alpha_{\gamma_j}\{\boldsymbol{\theta} \Rightarrow \boldsymbol{\theta}^*\} = \min\left(1, \frac{f_{\gamma_j}(\boldsymbol{\theta}^*; \boldsymbol{y}_{1:T})}{f_{\gamma_j}(\boldsymbol{\theta}; \boldsymbol{y}_{1:T})} \frac{h(\boldsymbol{\theta}\,|\mathcal{M}_{\boldsymbol{\theta}}\,)}{h(\boldsymbol{\theta}^*\,|\mathcal{M}_{\boldsymbol{\theta}}\,)}\right)$$

Note: The proposal sampler is independent, because $\mathcal{M}_{\boldsymbol{\theta}}$ reflects the whole sample instead of any individual element.

## Step 4: Set $\gamma_j = \gamma^*$, and return to Step 2 to advance to next $\gamma^*$

# Advantages of SMC over MCMC

- The target distribution (intermediate or final) is the stationary solution to the Markov kernel defined by the acceptance probability. The MH move is only used to boost the support, which contrasts with its usage in MCMC where burn-in is required to ensure a "correctly" distributed sample by convergence of the Markov process defined by the kernel.
- A natural independent proposal can be constructed from the SMC sample, and a high acceptance rate is expected. This is in sharp contrast to the MH move used in MCMC where a natural proposal with a high acceptance probability is lacking.
- By design, the SMC sample representing the target distribution has independent elements whereas the MCMC sample after burn-in often has highly dependent elements. An independent sample naturally delivers a better statistical result than a dependent sample of equal size.

# Advantages of SMC over gradient-based methods

- SMC can handle models with many parameters, say, several hundred.
- The likelihood function may be discontinuous in parameters; for example, the likelihood function can only be evaluated with a particle filter.
- Real-time updating parameter values to time $T+1$ is straightforward and highly efficient by assigning the following incremental weight to the SMC sample obtained up to time $T$:

$$\left(\frac{\mathcal{L}(\boldsymbol{\theta};\boldsymbol{y}_{1:T+1})p(\boldsymbol{\theta})}{\mathcal{L}(\boldsymbol{\theta};\boldsymbol{y}_{1:T})p(\boldsymbol{\theta})}\right)^{\gamma}=\left(\frac{\mathcal{L}(\boldsymbol{\theta};\boldsymbol{y}_{1:T+1})}{\mathcal{L}(\boldsymbol{\theta};\boldsymbol{y}_{1:T})}\right)^{\gamma}$$

  Tempering this incremental weight is usually not needed because one extra data point unlikely reduces the ESS significantly.

- Resampling and support boosting are needed to raise the ESS.

# Accelerating SMC

## *k*-fold duplication[4]

1. After reaching $\gamma = 1$ with the target sample of size *n*, duplicate the sample *k* times to obtain a sample of size *kn*. Note that this sample's ESS remains unchanged.
2. Run the MH move to boost the empirical support from *n* to *kn*.
3. *k*-fold duplication is much more efficient than running the regular SMC directly with a sample size of *kn*, because tempering steps can be completely skipped.

[4] Non-Gaussian Bridge Sampling with an Application, J.C. Duan and C. Zhang (2016), National University of Singapore working paper

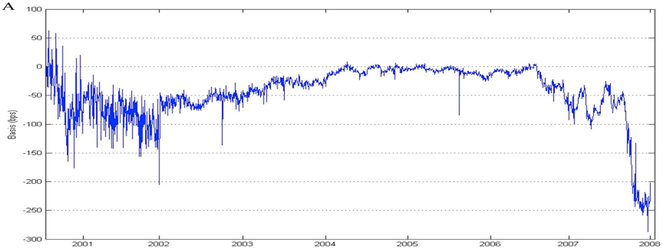
# Example: CDS usage and availability

- Corporate CDS are widely used in benchmarking for accounting and credit risk management purposes.

- Availability of liquid CDS is rather limited, and arguably no more than 500 corporate names worldwide. Out of necessity, users either confine themselves to this limited set of CDS or simply resort to aggregates derived from those liquid ones in different industry/rating combinations produced by, say, Markit.

# Bond yield spread to predict CDS?

CDS and corresponding bond yield spreads obey a parity relationship in a long run, i.e., co-integrated (Blanco, Brennan and Marsh, 2005, *J of Finance*; Zhu, 2006, *J of Fin Serv Research*). Their difference has been shown to be mostly negative with an average of 40 bps for individual corporate (Kim, Li and Zhang, 2017, *J of Futures Markets*). This parity relationship cannot be used to predict CDS for firms without traded bonds.



Equally-weighted investment grade CDS-bond basis, taken from Figure 2 of Kim, Li and Zhang, 2017, *Journal of Futures Markets*

# Duan (2017) “Proxy CDS Curves for Individual Corporates Globally”

- Develop a robust predictive regression that can generate proxy CDS curves for all exchange-traded corporates globally. The predictive regression has an $R^2$ over 80% for a large set of USD-denominated CDS spanning over 15 years.
- Use Actuarial Spread, a quantity closest to CDS spread, as a key predictor. Daily updated Actuarial Spreads on all exchange-traded firms (around 35,000) are computed by the Credit Research Initiative (CRI) at the National University of Singapore, and made freely available on its website. So, the proxy CDS curves can be operational on a real time basis.
- A large set of potential regressors is considered. Regressor selection utilizes a proprietary zero-norm penalty regression software developed by CriAT.

# The CDS data

- The USD-denominated CDS market prices are the Bloomberg computed CDS averages with end-of-day set to 6:00pm EST (New York time) on a monthly frequency starting in August 2001 all the way to February 2017.
- 405 corporate names include 309 US firms and others from 21 economies. The sample covers all 10 industries according the Bloomberg Industry Classification System with Financial being the largest containing 73 firms and Diversified being the smallest having 4 firms.
- The five CDS tenors are fairly equally distributed where 354 firms with 1-year, 319 with 2-year, 356 with 3-year, 314 with 4-year and 404 with 5-year.

- The post-crisis sample (the 2008 global financial crisis defined as September end of 2008 and afterwards) contains 395 firms, whereas the pre-crisis sample has 244 firms.
- The post CDS Big Bang (April 2009) sample has 374 firms whereas the pre CDS Big Bang sample covers 372 firms.
- The CDS sample contains 141,918 observations in total with 118,559 being investment-grade and the rest being the high-yield.
- The sample contains 92 data points on CDS referencing subordinated debt, and all are 5-year tenor with Shinshei Bank, a Japanese financial institution, as the reference entity.

# The 29 explanatory variables and their single-regressor $R^2$

| | $R^2$ | **Mean** | **Std** | **Max** | **Min** |
|---|---|---|---|---|---|
| **CDS(bps)** | | 150.0377 | 328.3817 | 9592.2010 | 1.2350 |
| **logCDS** | | 4.2253 | 1.1715 | 9.1687 | 0.2111 |
| **Regressor** | | | | | |
| logAS | 0.4860 | 2.1944 | 1.8994 | 9.6152 | -11.9240 |
| logASlevel | 0.4718 | 2.3773 | 1.7463 | 8.3655 | -10.2444 |
| DTDlevel | 0.3819 | 5.5199 | 3.0600 | 20.1084 | -1.1757 |
| SIGMA | 0.3769 | 0.0794 | 0.0555 | 0.9492 | 0.0233 |
| logIndustryCCI | 0.3023 | 2.8066 | 0.4200 | 3.8135 | 1.1936 |
| logCountryCCI | 0.2894 | 2.8950 | 0.6612 | 5.1790 | -0.8228 |
| isHY | 0.2745 | 0.1646 | 0.3708 | 1 | 0 |
| SIZElevel | 0.2616 | 3.5422 | 1.4390 | 8.1375 | -2.2648 |
| VIX | 0.1589 | 21.6953 | 9.6061 | 59.8900 | 10.4200 |
| NI/TAlevel | 0.1566 | 0.0038 | 0.0059 | 0.0761 | -0.0603 |
| Tenor-1y | 0.0800 | 0.1800 | 0.3842 | 1 | 0 |
| 3mRateUS | 0.0556 | 0.6785 | 1.3026 | 5.1239 | -0.0203 |

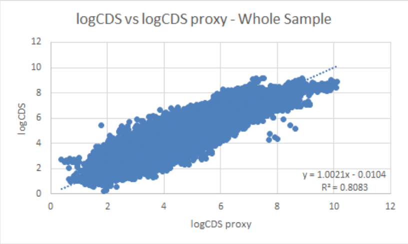
| | $R^2$ | Mean | Std | Max | Min |
|---|---|---|---|---|---|
| TL/TA | 0.0534 | 0.6690 | 0.1789 | 2.0325 | 0.1206 |
| postCrisis | 0.0404 | 0.7871 | 0.4093 | 1 | 0 |
| logAStrend | 0.0333 | -0.1830 | 0.6830 | 2.4696 | -8.1830 |
| DTDtrend | 0.0268 | 0.1090 | 1.3557 | 6.1350 | -7.0466 |
| SIZEtrend | 0.0216 | -0.0037 | 0.1794 | 1.6448 | -1.8962 |
| 3mRateEcon | 0.0155 | 0.9587 | 1.7389 | 23.7700 | -0.0800 |
| Tenor-4y | 0.0116 | 0.1462 | 0.3533 | 1 | 0 |
| SwapSpread5vs1 | 0.0113 | 1.1097 | 0.6374 | 2.7300 | -0.3562 |
| isFinancial | 0.0092 | 0.1467 | 0.3538 | 1 | 0 |
| Tenor-2y | 0.0038 | 0.1440 | 0.3511 | 1 | 0 |
| NI/TAtrend | 0.0016 | 0.0000 | 0.0066 | 0.1044 | -0.1459 |
| isSub | 0.0012 | 0.0006 | 0.0255 | 1 | 0 |
| Tenor-3y | 0.0011 | 0.1884 | 0.3910 | 1 | 0 |
| CASH/TAlevel | 0.0011 | 0.0910 | 0.1092 | 0.9785 | 0.0000 |
| postBigBang | 0.0008 | 0.7144 | 0.4517 | 1 | 0 |
| CASH/TAtrend | 0.0003 | 0.0011 | 0.0294 | 0.4829 | -0.3337 |
| isUS | 0.0000 | 0.8585 | 0.3485 | 1 | 0 |

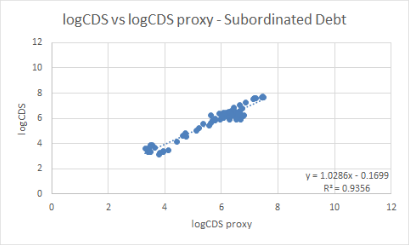
# Selection of 24 regressors (out of 442) based on the subsample of 3,000 observations and then applied to the whole sample

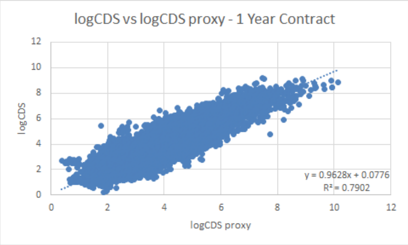
| | Subsample | | Whole Sample | |
|---|---|---|---|---|
| **Regressor** | **Estimate** | ***t*-Stat** | **Estimate** | ***t*-Stat** |
| Intercept | 1.7731 | 12.71 | 1.4930 | 72.69 |
| logAS*logASlevel | 0.0109 | 5.81 | 0.0125 | 42.39 |
| logAS*logIindustryCCI | 0.0748 | 22.10 | 0.0705 | 137.96 |
| logAStrend*SwapSpread5vs1 | -0.1079 | -8.51 | -0.1158 | -62.79 |
| DTDtrend*isHY | 0.1741 | 6.91 | 0.0580 | 16.21 |
| SIGMA | 2.0830 | 8.11 | 2.1699 | 56.24 |
| SIGMA*isSub | 5.6355 | 11.07 | 5.7308 | 13.37 |
| SIZElevel | -0.1532 | -16.50 | -0.1466 | -110.27 |
| SIZElevel*isUS | -0.0676 | -8.98 | -0.0637 | -59.28 |
| SIZElevel*Tenor2y | -0.0481 | -6.25 | -0.0523 | -46.82 |
| TL/TA*isHY | 0.5957 | 14.27 | 0.6614 | 106.67 |

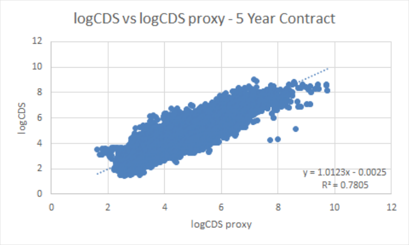
| | Subsample | | Whole Sample | |
|---|---|---|---|---|
| | Estimate | *t*-Stat | Estimate | *t*-Stat |
| logCountryCCI | 1.0672 | 11.72 | 1.2269 | 90.83 |
| logCountryCCI$^2$ | -0.1503 | -8.48 | -0.1770 | -66.67 |
| logCountryCCI*SIZEtrend | -0.1644 | -8.96 | -0.1514 | -56.96 |
| logIindustryCCI*postCrisis | 0.2376 | 7.56 | 0.2371 | 51.10 |
| 3mRateUS | -0.2393 | -12.44 | -0.2242 | -78.22 |
| 3mRateUS*isFinancial | -0.1220 | -6.98 | -0.1147 | -43.11 |
| 3mRateUS*isHY | 0.1468 | 6.30 | 0.1228 | 36.94 |
| 3mRateEcon*DTDlevel | 0.0120 | 7.14 | 0.0109 | 43.36 |
| 3mRateEcon*postBigBang | 0.0792 | 7.15 | 0.0762 | 43.34 |
| SwapSpread5vs1$^2$ | -0.1354 | -16.81 | -0.1238 | -104.09 |
| VIX | 0.0487 | 13.91 | 0.0486 | 92.31 |
| VIX*postCrisis | -0.0364 | -9.70 | -0.0369 | -65.60 |
| Tenor3y | -0.1616 | -6.41 | -0.1457 | -40.38 |
| $R^2$ | 81.96% | | 80.89% | |
| Sample Size | 3,000 | | 141,918 | |
| BIC | -3,889.73 | | -189,642.28 | |

# Performance of the proxy CDS model in predicting market price of CDS in different subcategories

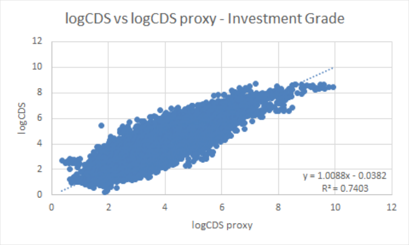


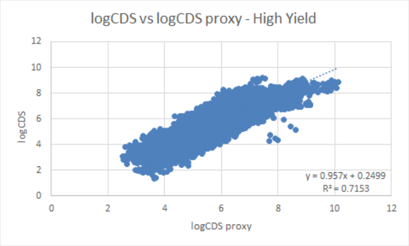


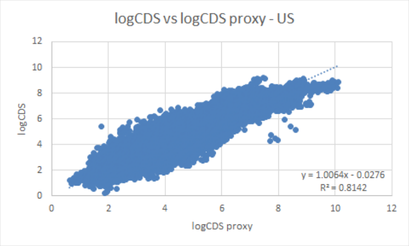


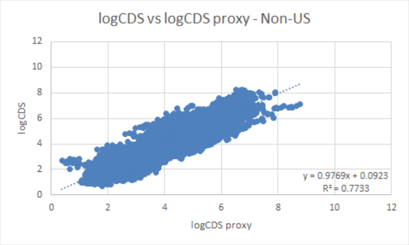


# Application: proxy CDS curves

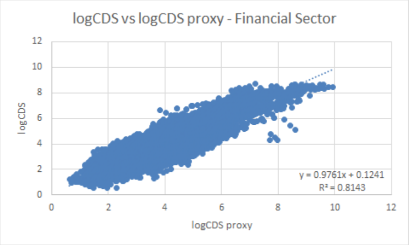

logCDS vs logCDS proxy - Investment Grade


logCDS vs logCDS proxy - High Yield

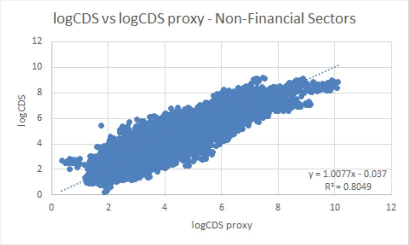

logCDS vs logCDS proxy - US


logCDS vs logCDS proxy - Non-US

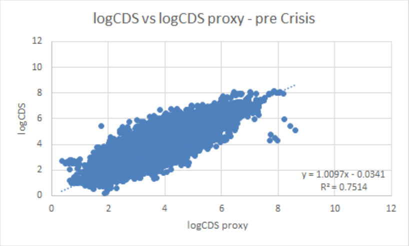
# Application: proxy CDS curves



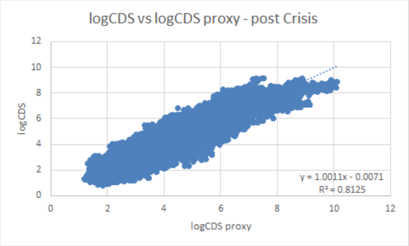
logCDS vs logCDS proxy - Financial Sector



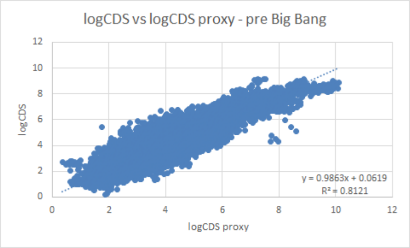
logCDS vs logCDS proxy - Non-Financial Sectors

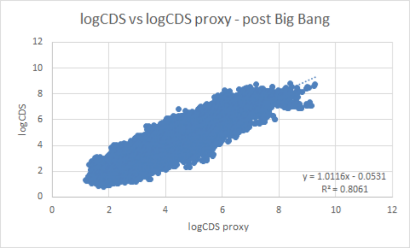


logCDS vs logCDS proxy - pre Crisis



logCDS vs logCDS proxy - post Crisis


logCDS vs logCDS proxy - pre Big Bang
logCDS
logCDS proxy
0
2
4
6
8
10
12
y = 0.9863x + 0.0619
R² = 0.8121


logCDS vs logCDS proxy - post Big Bang
logCDS
logCDS proxy
0
2
4
6
8
10
12
y = 1.0116x - 0.0531
R² = 0.8061

# $R^2$ of the proxy CDS model for different subcategories

| | $R^2$ | # of Corporates | # of Data |
|---|---|---|---|
| **Whole sample** | 80.89% | 405 | 141,918 |
| **US** | 81.41% | 309 | 121,840 |
| **Non-US** | 77.29% | 96 | 20,078 |
| **Financial** | 81.36% | 73 | 20,818 |
| **Non-Financial** | 80.48% | 332 | 121,100 |
| **Investment grade** | 74.02% | 370 | 118,559 |
| **High yield** | 71.38% | 138 | 23,359 |
| **Senior debt** | 80.79% | 404 | 141,826 |
| **Subordinated debt** | 93.49% | 1 | 92 |
| **Pre-financial crisis** | 75.14% | 244 | 17,696 |
| **Post-financial crisis** | 81.25% | 395 | 124,222 |
| **Pre-CDS Big Bang** | 81.20% | 372 | 40,529 |
| **Post-CDS Big Bang** | 80.60% | 374 | 101,389 |
| **Tenor(1 year)** | 78.72% | 354 | 25,548 |
| **Tenor(2 years)** | 79.93% | 319 | 20,432 |
| **Tenor(3 years)** | 78.54% | 356 | 26,740 |
| **Tenor(4 years)** | 77.17% | 314 | 20,750 |
| **Tenor(5 years)** | 77.78% | 404 | 48,448 |