The 2010 Medici Summer School in Management Studies

William Greene
Department of Economics
Stern School of Business
Econometric Models When There Are Rare Events
Part 1: Introduction
Agenda

- Session 1
  - Rare and unusual events
  - Modeling
  - Probabilities and correlation
  - Regression
- Session 2
  - Linear and robust quantile regression
  - Robust quantile models for counts
  - Binary choice when the event is unusual
Rare Events

- What defines a “rare” event?
  - In retrospect
  - In prospect
  - In context
- By what construction is the definition useful?
Assigning Probabilities to Rare Events

Colliding Bullets
Assigning Probabilities

Colliding Economists
Low Probabilities and Unusual Events

- Meaning of “Probability”
  - Objective probabilities
  - Subjective probabilities
- What function do probabilities serve?
  - Understanding
  - Prediction
Assign a Probability?

For all the criticism BP executives may deserve, they are far from the only people to struggle with such low-probability, high-cost events. Nearly everyone does. “These are precisely the kinds of events that are hard for us as humans to get our hands around and react to rationally,”

On the other hand, when an unlikely event is all too easy to imagine, we often go in the opposite direction and overestimate the odds. After the 9/11 attacks, Americans canceled plane trips and took to the road.

Quotes from Spillonomics: Underestimating Risk
Modeling and Probabilities

- Events
- Experiments
  - Repeatable under the same conditions
- Probabilities
  - (Classical) Long run frequency interpretation of objective probabilities
  - (Bayesian) Introspective interpretation of subjective probabilities
Assigning probabilities to events requires a construct of a meaningful framework within which the experiment can be repeated.

Assigning infinitesimal probabilities to events in retrospect is not a productive exercise.
What is the principal risk of econometric technical analysis?

Technical analysis can be a very powerful tool, but it lacks the ability to predict future rare events that have never happened before. The reason for this is because econometric algorithms are based on market movements in past years. These models frequently do a fantastic job of modeling normal market gyrations, but cannot incorporate the impact of events that have never happened before. Because of this, over-reliance on technical analysis leaves investors susceptible to the impact of rare events that cause massive market disruptions.
Low Probability Events

- There are hundreds of examples of colliding bullets
  - War is no longer fought as it was at Gallipoli and Gettysburg
  - The probabilities are only meaningful in retrospect
- There is no well defined experiment to produce probabilities for the colliding economists
What is the point of studying rare events?

- Generalities about structures
- Reaction to rare events in general
- Looking backwards to the rare event (fighting the last war)
Conclusions: Unusual Events

- Life is a sequence of events that have zero probability of occurring, yet do occur.
- Assigning probabilities
- UNUSUAL EVENTS and low probabilities
  - Katrina, Upper Mississippi, Nashville
  - Financial meltdown
Part 2: Econometric Modeling
All models are wrong, but some are useful.


Regardless of who the quote belongs to, it is a classical example of making sense out of nonsense. Posted at http://www.anecdote.com.au/archives/2006/01/all_models_are.html by Claude Farah at June 17, 2008 1:50 PM.
In what sense is the model wrong?

- It is only descriptive of an objective reality; the model is not the reality
- The model is an abstraction of an objective reality and therefore, it only includes the most salient features of that reality
- The model is only probabilistic – the $R^2$ is not one. Life is only probabilistic.

By what construction is a model “wrong?”
A scientific theory is a mathematical model that describes and codifies the observations we make. A good theory will describe a large range of phenomena on the basis of a few simple postulates and will make definite predictions that can be tested.

What function does the model serve?

- Understanding outcomes in retrospect
- Predicting or assigning probabilities to outcomes in prospect
  - Allocate resources effectively
  - Defend against low probability, adverse events
Anticipating Rare Events: Can Acts of Terror, Use of Weapons of Mass Destruction or Other High Profile Acts Be Anticipated?

A Scientific Perspective on Problems, Pitfalls and Prospective Solutions

Rare events vs. unusual events

- WARNING ... opinion

- It makes no sense to “model” rare events (black swans)
  - It makes no sense to model the colliding economists
- It does make sense to model unusual events
  - It is not useful (though it is interesting) to model the colliding bullets.
A black swan is a highly improbable event with three principal characteristics: It is unpredictable; it carries a massive impact; and, after the fact, we concoct an explanation that makes it appear less random, and more predictable, than it was. The astonishing success of Google was a black swan; so was 9/11.

Four hundred years ago, Francis Bacon warned that our minds are wired to deceive us. “Beware the fallacies into which undisciplined thinkers most easily fall. They are the real distorting prisms of human nature. Chief among them: Assuming more order than exists in chaotic nature.”
What is an unusual event?

- Outliers
  - The range of experience
  - The reach of available theories
  - The probability assigned to the outcome

- Events: What makes an event an outlier?
Prices paid at auction for Monet paintings vs. surface area (in logs)

\[ \log_{10}\text{Price} = a + b \log_{10}\text{Area} + e \]

- Not an outlier: Monet chose to paint a small painting.
- Possibly an outlier: Why was the price so low?
A sample of 27,326 German household-years:
Number of visits to the hospital in the survey year.
Histogram for HOSPVIS  NOBS =  27326, Too low: 0, Too high: 0

<table>
<thead>
<tr>
<th>Bin</th>
<th>Value of HOSPVIS</th>
<th>Frequency</th>
<th>Cumulative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>24931 ( .9124)</td>
<td>24931 ( .9124)</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1911 ( .0699)</td>
<td>26842 ( .9823)</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>323 ( .0118)</td>
<td>27165 ( .9941)</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>56 ( .0020)</td>
<td>27221 ( .9962)</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>25 ( .0009)</td>
<td>27246 ( .9971)</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>14 ( .0005)</td>
<td>27260 ( .9976)</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>11 ( .0004)</td>
<td>27271 ( .9980)</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>7 ( .0003)</td>
<td>27278 ( .9982)</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>2 ( .0001)</td>
<td>27280 ( .9983)</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>3 ( .0001)</td>
<td>27283 ( .9984)</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>12 ( .0004)</td>
<td>27295 ( .9989)</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>2 ( .0001)</td>
<td>27297 ( .9989)</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>3 ( .0001)</td>
<td>27300 ( .9990)</td>
</tr>
<tr>
<td>13</td>
<td>13</td>
<td>3 ( .0001)</td>
<td>27303 ( .9992)</td>
</tr>
<tr>
<td>14</td>
<td>14</td>
<td>9 ( .0003)</td>
<td>27312 ( .9995)</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
<td>2 ( .0001)</td>
<td>27314 ( .9996)</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td>11 ( .0004)</td>
<td>27325 ( .9999)</td>
</tr>
<tr>
<td>51</td>
<td>51</td>
<td>1 ( .0000)</td>
<td>27326 (1.0000)</td>
</tr>
</tbody>
</table>
A Poisson Model

Experiment: Pick an individual from the population and assign a probability to the observed outcome = number of visits.

- For the $K = 51$: $1/27326 = .0000366$? This will vastly overestimate the probability. Costly.

- By a Poisson model: $P(K) = \frac{\exp(-\mu)\mu^K}{K!}$, $\mu = E[K] = .138257$

$P(51) = .0000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000000084014506$
Conditional Modeling

- The colliding economists
  - No information
  - On the way to London on the same day
  - Theater plans that night
  - Habitually use the tube instead of taxis
  - Still probability essentially zero
  - Still not a replicable experiment to which probabilities should be attached

- The 51 visits to the hospital
  - No information
  - On kidney dialysis needing weekly treatments. \( P \to 1 \)
Wired’s Googly View of Models

- We have an infinite amount of data and infinite computing (search) capability.

- We no longer need models

- The logical limit of Newton’s view of the universe?
  - With enough initial conditions information, we can predict anything!
  - Implications for western religion. The notion of free will?
  - True randomness in the universe? Quantum mechanics
  - Newtonian uncertainty? The three body problem
  - Newtonian undertainty? Will the ad campaign work?

- Is the Googly view too optimistic? Try “googling” “rare events” econometrics
"rare events" econometrics - Google Search - Windows Internet Explorer

Managing Rare Events and Learning from the Unexpected (2010 Medici ... Mar 2, 2010 ... Econometric Models When There Are Rare Events - Morning session: William Greene Afternoon session: William Greene ... www.societyandorganizations.org/.../2010-medici-summer-school-managing-rare-events-and-learning-from-the-unexpected/ - Cached

ReLogit: Rare Events Logistic Regression
by L. Zeng - 2003
We study rare events data, binary dependent variables with dozens to thousands of ... from Choice Based Samples. Econometrica, Econometric Society, vol. ... ideas.repec.org/a/ssjs jes/08i02.html - Cached - Similar

Some Notes on Financial Econometrics - CXO Advisory
Financial econometrics gives empirical life (and death) to financial market models. ... "What is the best way to measure the likelihood of rare events and ... www.cxoadvisory.com/big.../some-notes-on-financial-econometrics/ - Cached

Rare Events and the Equity Premium
by RJ Barro - Cited by 53 - Related articles
Rare Events and the Equity Premium ... Conference on Econometrics & Mathematical Economics
Part 3: Probabilities and Correlation
Probabilities

- Events and measurement
  - Settings in which a sampling frame will be dominated by a few observations
  - The 80-20 rule

- Probability models: Assigning a probability distribution to outcomes with possibly infinite variance
The 80-20 “Rule”
The 80/20 Rule means that in anything a few (20 percent) are vital and many (80 percent) are trivial. In Pareto's case it meant 20 percent of the people owned 80 percent of the wealth. In Juran's initial work he identified 20 percent of the defects causing 80 percent of the problems. Project Managers know that 20 percent of the work (the first 10 percent and the last 10 percent) consume 80 percent of your time and resources. You can apply the 80/20 Rule to almost anything, from the science of management to the physical world.
“Thick Tailed” Distributions
Lognormal
Thick Tails?

- Distribution of income or wealth?
- Quite “normal”
- Log(x) ~ Normal(μ, σ²)
- Variance(x) = exp(σ² − 1)exp(2μ + σ²) – very finite.
- Not interesting enough
**Power Laws**

- \( f(x) = g(x) \ x^{-\beta}, \ x > x_{\text{min}} > 0 \)
  - \( \beta > 1 \)
  - \( g(x) \) such that \( \lim_{x \to \infty} \frac{g(tx)}{g(x)} = 1 \).

- Moments: \( \mathbb{E}[x^m] = \frac{\alpha - 1}{\alpha - 1 - m} x_{\text{min}} \)
  - Defined only for \( m \geq \beta - 1 \)
  - If \( \beta < 3 \), the variance is infinite
Theoretical Applications

- T distribution with less than 3 degrees of freedom
- Cauchy distribution – reciprocal of normal
- Yule-Simon discrete distribution

Yule–Simon PMF on a log-log scale. (Note that the function is only defined at integer values of k. The connecting lines do not indicate continuity.)
Empirical Applications

- Zipf’s law – occurrence of words
- Gutenberg-Richter law of earthquake magnitudes
- Richardson’s Law for the severity of violent conflicts
Pareto Distribution

\[ f(x) = \frac{\alpha x_{\text{min}}^\alpha}{x^{\alpha+1}}, \quad x \geq x_{\text{min}} \]

\[ F(x) = 1 - \left( \frac{x_{\text{min}}}{x} \right)^\alpha, \quad x \geq x_{\text{min}} \]

\[ E[x] = \frac{\alpha x_{\text{min}}}{\alpha - 1} \]

\[ \text{Var}[x] = \left( \frac{x_{\text{min}}}{\alpha - 1} \right)^2 \left( \frac{\alpha}{\alpha - 2} \right), \text{ infinite if } \alpha < 2 \]
Maximum Likelihood Estimation

ML Estimator of $x_{\text{min}} = \text{Min}(x_1, \ldots, x_N)$

ML Estimator of $\alpha = \frac{N}{\sum_{i=1}^{N} \left( \log x_i - \log x_{\text{min}} \right)}$
Movie Success: A Familiar Application
Implications of Power Law
What Do We Learn?

• Data are consistent with Pareto Law
• Variance is not finite
• Implication for investment in movies: Portfolio algorithms require finite variances

• A Remaining Challenge: Demonstrate that the data are not consistent with a distribution with finite variance such as lognormal.
A Bayesian Exercise on Default Probability

Estimate the Probability of an Event

N Bernoulli trials, s successes (Binomial)

\[ L(\theta; N, s) = \binom{N}{s} \theta^s (1 - \theta)^{N-s} \]
**Inference**

Classical: The MLE is \( \hat{\theta} = \frac{s}{N} \)

Bayesian: The posterior is

\[
L(\theta; N, s)P(\theta) = \frac{\theta^s(1-\theta)^{N-s}P(\theta)}{\int_0^1 \theta^s(1-\theta)^{N-s}P(\theta)d\theta}
\]

where \( P(\theta) \) is the prior information.

The estimator is the posterior mean

\[
E[\theta|N, s] = \frac{\int_0^1 \theta^{s+1}(1-\theta)^{N-s}P(\theta)}{\int_0^1 \theta^s(1-\theta)^{N-s}P(\theta)d\theta}
\]
The Bayesian Estimator

- The posterior distribution embodies all that is “believed” about the model.
  - Posterior = f(model|data)
    = Likelihood(θ,data) * prior(θ) / P(data)
- “Estimation” amounts to examining the characteristics of the posterior distribution(s).
  - Mean, variance
  - Distribution
  - Intervals containing specified probabilities
Priors and Posteriors

- Noninformative and Informative priors for estimation of parameters
  - Noninformative (diffuse) priors: How to incorporate the total lack of prior belief in the Bayesian estimator. The estimator becomes solely a function of the likelihood
  - Informative prior: Some prior information enters the estimator. The estimator mixes the information in the likelihood with the prior information.
Diffuse (Flat) Priors

E.g., the binomial example: \( L(\theta; N, s) = \binom{N}{s} \theta^s (1 - \theta)^{N-s} \)

Uninformative Prior (?): Uniform (flat) \( P(\theta) = 1, \ 0 \leq \theta \leq 1 \)

\[
P(\theta|N,s) = \frac{\binom{N}{s} \theta^s (1 - \theta)^{N-s} \times 1}{\int_0^1 \binom{N}{s} \theta^s (1 - \theta)^{N-s} \times 1 d\theta} = \frac{\theta^s (1 - \theta)^{N-s}}{\Gamma(N - s + 1)\Gamma(s + 1)}
\]

\[
= \frac{\Gamma(N + 2)}{\Gamma(s + 1)\Gamma(N - s + 1)} \theta^s (1 - \theta)^{N-s} = \text{a Beta distribution}
\]

Posterior mean = \( \frac{s+1}{(N-s+1)+(s+1)} = \frac{s+1}{N+2} \)

Posterior Mean > MLE. Why? The prior was informative.

(Prior mean = .5)
Mathematical device to produce a tractable posterior.

This is a typical application.

\[
L(\theta;N,s) = \frac{\Gamma(N+1)}{(s+1)\Gamma(N-s+1)} \theta^s (1-\theta)^{N-s}
\]

Use a conjugate beta prior, \( p(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} \)

Posterior = 
\[
\int_0^1 \left( \frac{\Gamma(N+2)}{(s+1)\Gamma(N-s+1)} \theta^s (1-\theta)^{N-s} \right) \left( \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} \right) d\theta
\]

\[
= \frac{\theta^{s+\alpha-1} (1-\theta)^{N-s+\beta-1}}{\int_0^1 \theta^{s+\alpha-1} (1-\theta)^{N-s+\beta-1} d\theta}
\]

Posterior mean = \( \frac{s+\alpha}{N+\alpha+\alpha} \) (we used \( \alpha = \beta = 1 \) before)
THE Question

Where does the prior come from?
Kiefer’s Application

- Loan portfolio
- Basel II rules for capital requirements mandates determining “risk” of loan portfolio

- Known: \( \theta \) is about .01 - extremely small

- How to determine the prior? Ask some experts.
Using the Expert Information

- Restrict range of $\theta$ with a beta distribution over $[a,b]$ instead of $[0,1]$.

$$p(\theta) = \frac{\Gamma(\alpha + \beta)}{(b-a)\Gamma(\alpha)\Gamma(\beta)} \left( \frac{(\theta - a)}{(b-a)} \right)^{\alpha-1} \left( \frac{b - \theta}{(b-a)} \right)^{\beta-1}$$
A Crucial Assumption

• Will use the binomial model

• Assumes defaults are independent across trials. (Was this true in the recent financial meltdown?)

• Constant default probability across trials. We will reconsider this later.
Statistical Inference

\[
P(\theta \mid s, e) = \frac{P(s \mid \theta, e)P(\theta \mid e)}{P(s \mid e)} = \frac{P(\text{defaults} \mid \text{rate, experts})P(\text{rate} \mid \text{experts})}{P(\text{defaults} \mid \text{experts})}
\]
An Expert Opinion of a Portfolio

The minimum value for the default probability was 0.0001 (one basis point). The expert reported that a value above 0.035 would occur with probability less than 10%, and an absolute upper bound was 0.3. The upper bound was discussed: the expert thought probabilities in the upper tail of his distribution were extremely unlikely, but he did not want to rule out the possibility that the rates were much higher than anticipated (prudence?). Quartiles were assessed by asking the expert to consider the value at which larger or smaller values would be equiprobable given the value was less than the median, then given the value was more than the median. The median value was 0.01. The former was 0.0075. The latter, the 0.75 quartile, was assessed at 0.0125. The expert seemed to be thinking in terms of a normal distribution, perhaps using, informally, a central limit theorem combined with long experience with this category of assets.
Incorporate Expert Opinion in the Prior

This set of answers is more than enough information to determine a 4-parameter beta distribution. I used a method of moments to fit parametric probability statements to the expert assessments. The moments I used were squared differences relative to the target values, for example \((a - 0.0001)/0.0001)^2\). The support points were quite well-determined for a range of \(\{\alpha, \beta\}\) pairs at the assessed values \(\{a, b\} = [0.0001, 0.3]\). Thus, the expert was able to determine these parameter values consistently with his probability assessments.
### Posterior Means

Default probabilities—location and precision, $n = 500$

<table>
<thead>
<tr>
<th>$r$</th>
<th>$\bar{\theta}$</th>
<th>$\theta_m$</th>
<th>$\hat{\theta}$</th>
<th>$\sigma_\theta$</th>
<th>$\sigma_{\hat{\theta}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.0063</td>
<td>0.0081</td>
<td>0.000</td>
<td>0.0022</td>
<td>0 (1)</td>
</tr>
<tr>
<td>1</td>
<td>0.0071</td>
<td>0.0092</td>
<td>0.002</td>
<td>0.0023</td>
<td>0.0020</td>
</tr>
<tr>
<td>2</td>
<td>0.0079</td>
<td>0.0103</td>
<td>0.004</td>
<td>0.0025</td>
<td>0.0028</td>
</tr>
<tr>
<td>3</td>
<td>0.0086</td>
<td>0.0114</td>
<td>0.006</td>
<td>0.0026</td>
<td>0.0035</td>
</tr>
<tr>
<td>4</td>
<td>0.0094</td>
<td>0.0125</td>
<td>0.008</td>
<td>0.0027</td>
<td>0.0040</td>
</tr>
<tr>
<td>5</td>
<td>0.0102</td>
<td>0.0136</td>
<td>0.010</td>
<td>0.0028</td>
<td>0.0044</td>
</tr>
<tr>
<td>6</td>
<td>0.0109</td>
<td>0.0147</td>
<td>0.012</td>
<td>0.0029</td>
<td>0.0049</td>
</tr>
<tr>
<td>7</td>
<td>0.0117</td>
<td>0.0158</td>
<td>0.014</td>
<td>0.0030</td>
<td>0.0053</td>
</tr>
<tr>
<td>8</td>
<td>0.0125</td>
<td>0.0169</td>
<td>0.016</td>
<td>0.0031</td>
<td>0.0056</td>
</tr>
<tr>
<td>9</td>
<td>0.0132</td>
<td>0.0180</td>
<td>0.018</td>
<td>0.0032</td>
<td>0.0060</td>
</tr>
<tr>
<td>20</td>
<td>0.0215</td>
<td>0.0296</td>
<td>0.040</td>
<td>0.0040</td>
<td>0.0088</td>
</tr>
<tr>
<td>50</td>
<td>0.0431</td>
<td>0.0425</td>
<td>0.100</td>
<td>0.0053</td>
<td>0.0134</td>
</tr>
<tr>
<td>100</td>
<td>0.0753</td>
<td>0.0749</td>
<td>0.200</td>
<td>0.0065</td>
<td>0.0179</td>
</tr>
<tr>
<td>200</td>
<td>0.1267</td>
<td>0.1266</td>
<td>0.400</td>
<td>0.0069</td>
<td>0.0219</td>
</tr>
</tbody>
</table>
What Do We Learn

• Using sample information to infer the population parameter

• How to use outside information in a Bayesian update

• Default is an unusual (not rare) event. How to learn about probabilities of this unusual event?
  • Sample information and simulation
  • Incorporation of experts’ accumulated information

• Incorporation of information subject to great uncertainty in a policy decision – the capital requirements of Basel 2 regulated banks.
Correlation

- Pearson Correlation: Correlation of a variable $y$ with $E[y|x]$
Correlation with Respect to Unusual Values

- Does the notion make any sense?
- Spearman correlations are essentially unrelated
- **Copula functions** are used to build models of correlation of extreme values
Copulas

- A cousin to the bivariate distribution: Describes a measure of dependence among random variables – but not the familiar correlation.

- Copula function: \( C(u,v) = C(F_X^{-1}(u), F_Y^{-1}(v), \theta) \)
  - \( F_X(x) \) and \( F_Y(y) \) = cdf’s of \( X \) and \( Y \)
  - \( \theta \) induces the dependence between the variables

- Certain copula functions provide a more general type of association between two random variables than does the bivariate density with the correlation.
Copula Families

Frank Copula  \[ C(u,v,\theta) = \frac{-1}{\theta} \log \left\{ 1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{(e^{-\theta} - 1)} \right\} \]

Plackett Copula  \[ C(u,v,\theta) = \frac{[1+(\theta-1)(u+v)]-\sqrt{[1+(\theta-1)(u+v)]^2 - 4uv\theta(\theta-1)}}{2(\theta-1)} \]

Gaussian Copula  \[ C(u,v,\theta) = \Phi[\Phi^{-1}(u), \Phi^{-1}(v), \theta)] \]

All three are symmetric and concentrate probability in the center of the distribution so they are not suitable for modeling extreme events whose distributions have long thick tails.
WIRED MAGAZINE: 17.03

TECH BIZ : IT

Recipe for Disaster: The Formula That Killed Wall Street

By Felix Salmon  02.23.09

In the mid-'80s, Wall Street turned to the quants—brainy financial engineers—to invent new ways to boost profits. Their methods for minting money worked brilliantly... until one of them devastated the global economy.

Photo: Jim Krantz/Gallery Stock
Here's what killed your 401(k)  

David X. Li's Gaussian copula function as first published in 2000. Investors exploited it as a quick—and fatally flawed—way to assess risk. A shorter version appears on this month's cover of Wired.

<table>
<thead>
<tr>
<th>Probability</th>
<th>Survival times</th>
<th>Equality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specifically, this is a joint default probability—the likelihood that any two members of the pool (A and B) will both default. It’s what investors are looking for, and the rest of the formula provides the answer.</td>
<td>The amount of time between now and when A and B can be expected to default. Li took the idea from a concept in actuarial science that charts what happens to someone's life expectancy when their spouse dies.</td>
<td>A dangerously precise concept, since it leaves no room for error. Clean equations help both quants and their managers forget that the real world contains a surprising amount of uncertainty, fuzziness, and precariousness.</td>
</tr>
<tr>
<td><strong>Copula</strong></td>
<td><strong>Distribution functions</strong></td>
<td><strong>Gamma</strong></td>
</tr>
<tr>
<td>This couples (hence the Latin term copula) the individual probabilities associated with A and B to come up with a single number. Errors here massively increase the risk of the whole equation blowing up.</td>
<td>The probabilities of how long A and B are likely to survive. Since these are not certainties, they can be dangerous: Small miscalculations may leave you facing much more risk than the formula indicates.</td>
<td>The all-powerful correlation parameter, which reduces correlation to a single constant—something that should be highly improbable, if not impossible. This is the magic number that made Li's copula function irresistible.</td>
</tr>
</tbody>
</table>
David Li’s Copula Function

In 2000, while working at JPMorgan Chase, Li published a paper in *The Journal of Fixed Income* titled "On Default Correlation: A Copula Function Approach." (In statistics, a copula is used to couple the behavior of two or more variables.) Using some relatively simple math—by Wall Street standards, anyway—Li came up with an ingenious way to model default correlation without even looking at historical default data. Instead, he used market data about the prices of instruments known as credit default swaps.

Li’s copula function was used to price hundreds of billions of dollars' worth of CDOs filled with mortgages. And because the copula function used CDS prices to calculate correlation, it was forced to confine itself to looking at the period of time when those credit default swaps had been in existence: less than a decade, a period when house prices soared. Naturally, default correlations were very low in those years. But when the mortgage boom ended abruptly and home values started falling across the country, correlations soared.
What do we learn from this?

• Wired magazine argues that the copula model was inadequate, or inappropriate for the data to which it was applied.

• The wrong model?
  • Simple unconditional, parametric correlations
  • Relationships among variables

• Invalid model assumptions: Gaussian
Part 4: Regression
Conditional Mean Function
Describing a relationship between variables

Auction Prices for Monet Paintings

$log(price) = 5.255 + 1.251 \log(area)$

$R^2 = 0.3336$
Linear Regression Model

• $y(i) = \alpha + \beta x(i) + \varepsilon(i)$

• What is the “model?”
  • Assumptions of the regression (zero mean, linearity, etc.?)
    • The conditional mean assumption?
    • The conditional mean?
What is the regression for?

• Understanding: What are the important parameters of the relationship?

• Prediction: Use the estimated model to predict the outcome variable
Least Squares Regression

Minimize \( \sum_{i=1}^{N} (y_i - \alpha - \beta x_i)^2 \)

\[
b = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N} (x_i - \bar{x})^2}
\]
Questions about Least Squares

- Heavily influenced by outlying observations

- Estimates the conditional mean function. Perhaps the conditional mean function is not the function of interest.
Outliers?

- What is an outlier
- Outside the normal range predicted by the regression?
  - (What is “normal?”)
  - Outside the reach of the model?
  - Obviously strange?
- AN OUTLIER IS AN OUTLIER BECAUSE THE OBSERVER DECLARES IT TO BE AN OUTLIER.
  - There is no appropriate objective measure.
Prices paid at auction for Monet paintings vs. surface area (in logs)

$logPrice = a + b \logArea + e$

- **Not an outlier:** Monet chose to paint a small painting.
- **Possibly an outlier:** Why was the price so low?
Mechanically Remove Outliers?
Removing Outliers Creates Outliers

Were they really outliers?
**Did It Matter?**

<table>
<thead>
<tr>
<th></th>
<th>Standard</th>
<th>Prob.</th>
<th>Mean of X</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOGPRICE</td>
<td>Coefficient</td>
<td>Error</td>
<td>t</td>
</tr>
<tr>
<td>------------------</td>
<td>-------------------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>Constant</td>
<td>-8.52560***</td>
<td>.60753</td>
<td>-14.03</td>
</tr>
<tr>
<td>LOGAREA</td>
<td>1.32609***</td>
<td>.09060</td>
<td>14.64</td>
</tr>
</tbody>
</table>

Based on all 430 observations

<table>
<thead>
<tr>
<th></th>
<th>Standard</th>
<th>Prob.</th>
<th>Mean of X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-8.20921***</td>
<td>.44429</td>
<td>-18.48</td>
</tr>
<tr>
<td>LOGAREA</td>
<td>1.25760***</td>
<td>.06598</td>
<td>19.06</td>
</tr>
</tbody>
</table>

Based on 316 observations remaining after 26 iterations
Is it the estimator or the model?

• Least Squares Regression
  • Estimates the conditional mean
  • Does not account for heteroscedasticity
  • Influenced by outlying observations(?)

• Consider other aspects of the distribution:
  Conditional quantiles
Quantile Regression

- \( Q(y|\text{quantile},x) = \alpha + \beta x(i) \) (Computed as the solution to a linear programming problem)
- Makes no assumption about heteroscedasticity
- Not affected by transformation
- Is not influenced by extreme observations
<table>
<thead>
<tr>
<th>LOGPRICE</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>Prob.</th>
<th>Mean of X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-6.03172</td>
<td>.....(Fixed Parameter) ..... Quantile=.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOGAREA</td>
<td>1.17902</td>
<td></td>
<td></td>
<td>6.68007</td>
</tr>
<tr>
<td>Constant</td>
<td>-6.21923</td>
<td>.....(Fixed Parameter) ..... Quantile=.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOGAREA</td>
<td>1.12034</td>
<td></td>
<td></td>
<td>6.68007</td>
</tr>
<tr>
<td>Constant</td>
<td>-8.76379</td>
<td>.....(Fixed Parameter) ..... Quantile=.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOGAREA</td>
<td>1.44794</td>
<td></td>
<td></td>
<td>6.68007</td>
</tr>
<tr>
<td>Constant</td>
<td>-8.48896</td>
<td>.....(Fixed Parameter) ..... Quantile=.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOGAREA</td>
<td>1.35468</td>
<td></td>
<td></td>
<td>6.68007</td>
</tr>
<tr>
<td>Constant</td>
<td>-8.18627</td>
<td>.....(Fixed Parameter) ..... Quantile=.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOGAREA</td>
<td>1.20066</td>
<td></td>
<td></td>
<td>6.68007</td>
</tr>
<tr>
<td>Constant</td>
<td>-8.02616</td>
<td>.....(Fixed Parameter) ..... Quantile=.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOGAREA</td>
<td>1.20313</td>
<td></td>
<td></td>
<td>6.68007</td>
</tr>
<tr>
<td>Constant</td>
<td>-9.12630</td>
<td>.....(Fixed Parameter) ..... Quantile=.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOGAREA</td>
<td>1.32223</td>
<td></td>
<td></td>
<td>6.68007</td>
</tr>
<tr>
<td>Constant</td>
<td>-9.65846</td>
<td>.....(Fixed Parameter) ..... Quantile=.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOGAREA</td>
<td>1.36271</td>
<td></td>
<td></td>
<td>6.68007</td>
</tr>
<tr>
<td>Constant</td>
<td>-11.7723</td>
<td>.....(Fixed Parameter) ..... Quantile=.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOGAREA</td>
<td>1.60832</td>
<td></td>
<td></td>
<td>6.68007</td>
</tr>
</tbody>
</table>
Nonparametric Regression

- Drop linearity assumption
- No normality assumption
- No assumption of homoscedasticity
Nonparametric “Regression”
What Did We Learn?

• Basic Linear Regression Analysis
  • Essential assumptions
  • Reasons to consider other approaches
• Dealing with outliers
• Quantile regression: Other features of the relationship
• Nonparametric regression: Minimal assumptions; essentially qualitative results