

Portfolio Credit Risk

University of Chicago Masters in Financial Mathematics 36702

<https://uchicago.instructure.com/courses/48373>

Lecture 3

Thursday 6 April 2023

Expected LGD, vended estimates of PD and ρ , statistical tools, search and significance

Jon Frye

JonFrye@UChicago.edu

Probability and Statistics

Probability

- **Given a mechanism that produces random outputs,**
 - what is the probability that a particular set of outcomes comes out?
- **Example: The sum of spots on two six-sided dice.**
 - We make assumptions (each surface appears exactly $1/6$ of the time, independence, etc.) and use logic to arrive at answers. Very satisfying!

Statistics

- **Given some data, find models that can generate it.**
 - Simpler models have fewer things that can go wrong.
 - Richer models can match the data more closely.
 - We can test to see which model seems to work best, given the data sample.
 - We can never be sure that we have the best model. Very unsatisfying!

The portfolio default rate model

We imagine that a bank can accurately estimate PDs.

We do not know how the defaults of firms are connected.

- This is as dispiriting to write as it is to read.

Statistical analysis offers few alternatives.

- The t -copula inhibits a derivation like the Vasicek Distribution.
- We use the Gauss copula, and so do most banks.
 - A 1,000 firm portfolio has 499,500 $\rho_{i,j}$.
 - “The bigger the matrix, the less likely it is positive definite.”

We restrict the Gauss copula to the single factor model.

- The 499,500 values of $\rho_{i,j}$ are implied by 1,000 values of ρ_i .
 - Each Z_i is a linear function of Z and $\{X_i\}$. Therefore, $\{Z_i\}$ are jointly normal.
 - The correlation matrix of $\{Z_i\}$ is positive definite.

Questions or comments?

Questions or thoughts?

Week 3 topics

The t -copula: Slides 52-58 and 73-81 from Lecture 1

LGD and its expectation

Vended estimates of PD and ρ

Review and preview of statistical tools

How model search defeats indications of significance

LGD and its expectation

Definitions

Historical data

Expected LGD

Definitions

Loss *given* default

A firm declares bankruptcy to get protection from lawsuits.

- **Otherwise the defaulting firm could get sued by every firm it owes.**
- **The court determines who gets what, usually honoring seniority.**
 - **It is rare for every creditor to gain 100% recovery.**

If a firm is 90 days behind schedule on a bank loan, the bank has many options.

- **The bank might continue to seek greater recovery for years.**
 - **Eventually the bank determines it has obtained all it can.**
 - **It discounts all cash flows back to the time of default and computes an LGD.**

LGD is the lender's loss as a fraction of loan exposure.

Loans and bonds

Most forms of debt are loans or bonds.

- **Loans are private agreements between a firm and a “bank”,**
 - or other financial institution, hedge fund, etc.
 - Banks recognize a corporate default when the loan is 90 days past due.
 - LGDs on loans are private information.
- **Bonds are publicly traded promises to repay on a schedule.**
 - Bonds are considered to be in default if a payment is one day late.
 - LGD is usually calculated as $\text{LGD} = 1 - \text{post-default price} / \text{par}$.

Each debt instrument has a defined seniority.

- **In bankruptcy, the seniority of a debt determines the likelihood that the lender will obtain full repayment.**
 - The most-senior debt gets full repayment if possible.
 - If the firm has money is left over, the next-most-senior debt gets paid.
 - And so forth down the scale of seniority.
 - Net, more-senior debt tends to have lower LGD than junior debt.

The scale of seniority

Loans are senior to bonds.

A firm probably has multiple bonds outstanding.

- **The names of bonds usually reflect seniority.**
 - **“Senior Debentures” would be more senior.**
 - **“Junior Subordinated Notes” would be among the least senior.**

A bankruptcy judge tends to follow “strict seniority.”

- **The banks recover the largest fraction of their exposure.**
- **If there is money left over after the banks are paid, it goes to the holders of the most senior bonds.**
- **And so on down the scale of seniority.**

If there is no bankruptcy, loan recoveries remain private.

- **But the firm must act as it has promised in debt documents.**

Security

Some loans and some bonds have a “second way out”:

- The debt is secured with collateral.
 - It is like a consumer auto loan or home loan.
 - In the event of default, the debt holder obtains ownership of the identified collateral asset and sells the collateral to obtain partial or full recovery.
- If the collateral does not provide a full recovery, the remaining exposure has a defined place on the scale of seniority.

In a default,

- LGD is likely to be least for a senior secured bank loan, and
- LGD is likely to be greatest for a junior unsecured bond.

Looking across a large number of defaults,

- Bank loans usually enjoy substantial recovery.
- A “sub” bond might a small fraction of par, on average.

Strange things happen

A loan secured by a gasoline station defaults.

- The bank gets title to the station and the land under it.**
- The gasoline tanks had been leaking.**
 - Land under the station must be removed and replaced.**
 - The loss to the bank was several times the amount of the loan.**
 - LGD is greater than 100%.**

A loan secured by a different station defaults.

- The bank itself runs the gas station for months or years.**
- Surprisingly, the station becomes successful.**
 - When the bank sells the station, total recovery is greater than the amount of the original loan. LGD is less than zero.**

Historical data

Bonds and loans

For a publicly traded bond, a good measure of LGD is par (100%) minus the post-default price.

Publicly traded loans are scarce.

- **Therefore, banks need a different measure for loans.**

A bank estimates most LGDs by discounting cash flows to the date of the default.

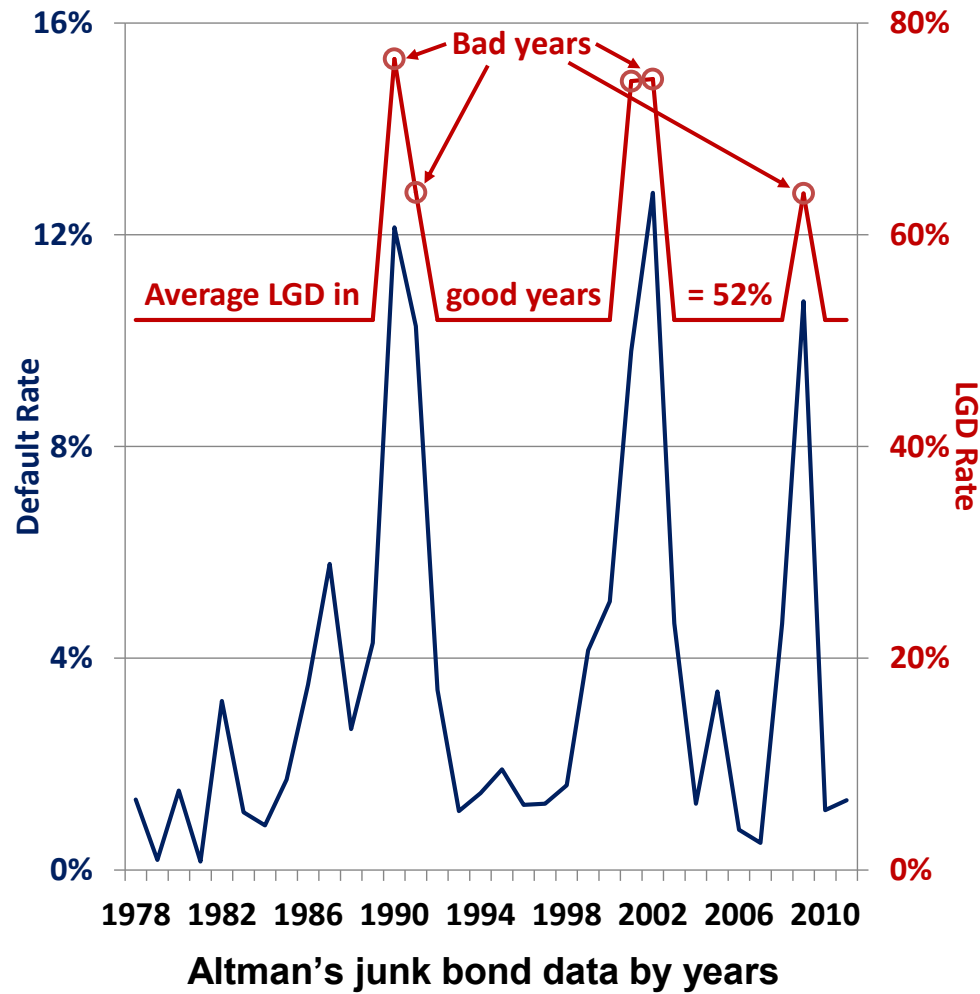
- **A bank need to keep track of the costs of recovery.**
- **It can take months or years to fully resolve a default.**
 - **And no observable rate of discount is clearly appropriate.**
 - **No other asset has risks like those of a defaulted bank loan.**

Historical bond LGDs

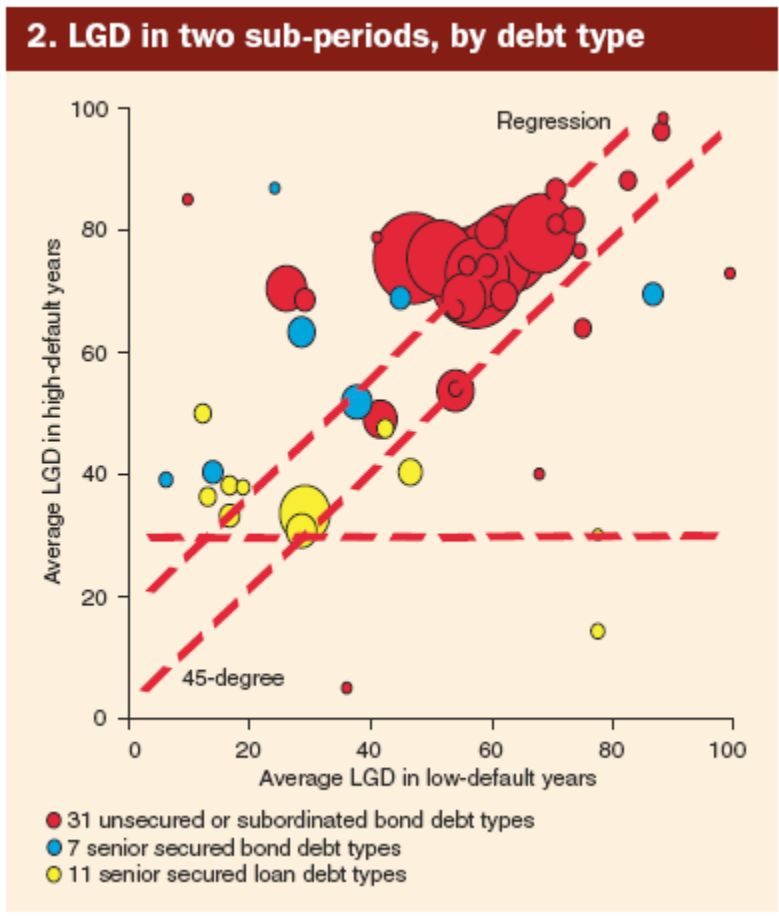
The history of bond LGDs shows three broad behaviors:

- Average LGD is elevated when the default rate is elevated.**
- The elevation is moderate.**
- The elevation is similar across bonds with different seniorities and levels of security.**

Positive, moderate response



Similar response



Moody's loans and bonds

Questions or comments?

ELGD: Expected loss given default

Orientation

A lender wants to estimate ELGD, a loan's expected LGD.

- Using data, take the average of historical LGD's of similar loans.
 - In the data there are many more defaults in 2009 than in most other years.
 - Each of these defaults produces an LGD.
 - On average, these LGDs are greater than in most other years.

When calculating ELGD, one must build in the same effect.

- The average of annual LGD rates would give the wrong answer,
 - because 2009 has many LGDs and it has elevated LGDs.

Suppose that a loan defaults in one of two periods.

- In a low-default period the loan is not likely to default, and it will tend to have a low LGD.
- In a high-default period the loan is more likely to default, and it will tend to have a high LGD.
- This affects the loan's ELGD.

Roundup of symbols

LGD: The loss given default of a loan is a random variable.

cLGD: Conditionally expected LGD

- What value of LGD should you expect in certain conditions?

ELGD: Expected LGD

- What value of LGD should you expect for a given loan?

| | Default | Loss given default |
|---------------------------|---------|--------------------|
| Observed random variable | DR | LGD |
| Conditional expectation | cPD | cLGD |
| Unconditional expectation | PD | ELGD |

Three identities for any loan

$$\text{Loss} = D * \text{LGD}$$

- If $D = 0$, then $\text{Loss} = 0$; if $D = 1$, then $\text{Loss} = \text{LGD}$.
 - Both Loss and LGD are expressed as fractions of the exposure amount.

$$\text{cLoss} = \text{cPD} * \text{cLGD}$$

- These quantities refer to particular conditions such as $Z = z$.

$$\text{EL} = \text{PD} * \text{ELGD}$$

- Proof:
$$\begin{aligned} E[\text{Loss}] &= E[D * \text{LGD}] \\ &= E[D] * E[\text{LGD}] = \text{PD} * \text{ELGD} \end{aligned}$$
 - “A conditional variable is independent of whatever it is conditioned on.”
 - Suppose you know that if A happens, then B will happen. Then you find out that A has happened. It is still the case that B will happen, right?

Loan expectations

Suppose a loan
has two states:

| | Prob | cPD | cLGD | cLoss |
|---------|------|-----|------|-------|
| State 1 | 2/3 | 0.1 | 0.3 | 0.03 |
| State 2 | 1/3 | 0.4 | 0.6 | 0.24 |

$$PD = .1 (2/3) + .4 (1/3) = .2$$

$$EL = .03 (2/3) + .24 (1/3) = .1$$

$$ELGD = EL / PD = .1 / .2 = .5$$

An equivalent calculation weights cLGD by both
probability and relative frequency:

$$\begin{aligned} ELGD &= .3 (2/3) (.1 / PD) + .6 (1/3) (.4 / PD) \\ &= .02 / PD + .08 / PD = .1 / .2 = .5 \end{aligned}$$

ELGD is also called “Default-weighted LGD.”

Expected cLGD

Suppose the same loan as before:

| | Prob | cPD | cLGD | cLoss |
|---------|------|-----|------|-------|
| State 1 | 2/3 | 0.1 | 0.3 | 0.03 |
| State 2 | 1/3 | 0.4 | 0.6 | 0.24 |

$$E [\underline{cLGD}] = .3 (2/3) + .6 (1/3) = .4$$

- This is sometimes called "time-weighted LGD."
 - Take average LGD each year, then take the average of averages.

Note that $E [\underline{cLGD}] < ELGD$. When cLGD is elevated, more defaulted loans are produced.

- Averaging over conditions produces $E [\underline{cLGD}]$.
- Averaging over loans produces ELGD.

Say it in math

Suppose you have a distribution of cPD, $f_{cPD}[r]$.

Suppose that cLGD is a function of cPD, $cLGD = g[cPD]$.

$$E[cLGD] = \int_0^1 g[r] f_{cPD}[r] dr$$

$$E[LGD] = EL/PD = \frac{1}{PD} \int_0^1 r g[r] f_{cPD}[r] dr$$

Introduction to LGD: Summary

In a default, a firm can't pay what it owes.

- **Usually, one or more of the lenders experiences loss.**
 - **Anticipating this, lenders want security and seniority.**

Each loan has specified Seniority (its place on the scale of seniority) and security (“second way out”).

- **These influence a loan's expected LGD.**
 - **Still, the loss on each defaulted loan is highly random.**

Three identities for any given loan:

- **$\text{Loss} = D * \text{LGD}$**
- **$\text{cLoss} = \text{cPD} * \text{cLGD}$**
- **$\text{EL} = \text{PD} * \text{ELGD}$**

Questions? Comments?

Vended estimates of PD and ρ

Vended estimates of PD and ρ

PD_i and $\rho_{i,j}$ are parameters in a credit portfolio model.

- PD_i is the probability that firm i defaults.
- $\rho_{i,j}$ is a transformation of the probability that both i and j default.

It would be strange and wonderful if these default probabilities could be estimated without using default data.

- Therefore, this section is somewhat strange and wonderful.

Asset returns and PD_i

Moody's EDF estimate of PD_i

The Moody's EDF is famous, so if you work in this area you will need to know something about it.

- EDF = Estimated Default Frequency = another name for PD.
 - Other vendors have similar methods.

Moody's starts with an attractive intuition:

- A loan contains the option to default.
- A risk-neutral probability is part of option price theory.
- Therefore, ignoring risk preference, it might be possible to estimate the probability of default using option theory.

Moody's does not tell exactly how they calculate EDF.

- And their method changes over time.
- But here's my impression in stylized terms.

Applying option theory

The owners of a corporation have limited liability.

- The most they can lose is their investment.

Stylizing this, the owners of the firm have an option:

- They can pay the debt holders in money.
- They can give debt holders the firm's assets.

The firm's equity provides an option on its assets.

- The "strike price" is the amount of the firm's liabilities.
 - Moody's measures this as short-term debt plus half of long-term debt.
- The expiration of the option would be debt maturity.
 - Moody's might assume one year in all cases.

→ → → What is the probability that the equity holder exercises the option to default?

Prob [option exercise]

In the case of stocks, $d2 = \frac{\text{Log}[S/X] + (r - \frac{\sigma^2}{2})\tau}{\sigma \sqrt{\tau}}$;

(Risk neutral) Probability of call exercise = $\Phi [d2]$

Analogous to $d2$ is the firm's "distance to default":

$$DD = \frac{\text{Log}[A/D]}{\text{Annual volatility of assets}}$$

where A = Value of firm's assets

D = "default point"

Sorry to keep re-using the letter D, but that's what Moody's uses.

Probability of put exercise $\approx 1 - \Phi [DD] = \Phi [- DD]$

The two unknowns in a DD

$$DD = \frac{\text{Log}[A/D]}{\text{Annual volatility of assets}}$$

Two unknowns in the formula for DD:

- A = the value of assets
- The volatility of assets.
 - The default point, D, is considered known:
 - short-term debt plus half of long-term debt.

Neither asset value nor asset volatility is directly observable.

Moody's infers them from an iteration...

Inference by double iteration

Stock market capitalization depends on:

- liabilities and the value and volatility of assets.

The volatility of market cap. also depends on:

- liabilities and the value and volatility of assets.

So, some setting of (asset value, asset volatility) best produces the observed values of (market capitalization, market capitalization volatility).

Then, $PD = EDF = 1 - \Phi[DD] = \Phi\left[\frac{-\text{Log}\left[\frac{A}{D}\right]}{\text{asset vol.}}\right]$.

There is something you should know about this...

It gives a very poor answer!

1- $\Phi[DD]$ gives a very poor estimate of PD.

- **Consider a firm with $DD = 6 = 6$ SDs from the default point.**
 - **$1 - \Phi[6] = 0.000000001$**
 - **Moody's finds 42,000 instances of $DD \approx 6.0$ in its data base.**
 - **Option theory says there should have been about 0.00004 defaults.**
 - **But there were 17 defaults. The forecast is wrong by a factor of 400,00.**

Something is exceedingly wrong here.

- **Maybe options theory does not apply well to default.**
- **Maybe quantities are being poorly measured.**
- **Maybe the assumptions of stability are wrong.**
 - **When firms get in trouble, the trouble compounds and vol goes up.**
- **Maybe expiration is not 1 year.**
- **Maybe D is not short term + half of long term.**
- **Maybe asset returns do not have a LogNormal Distribution.**

Questions? Comments?

The EDF work-around

Moody's does not take $EDF = 1 - \Phi [DD]$.

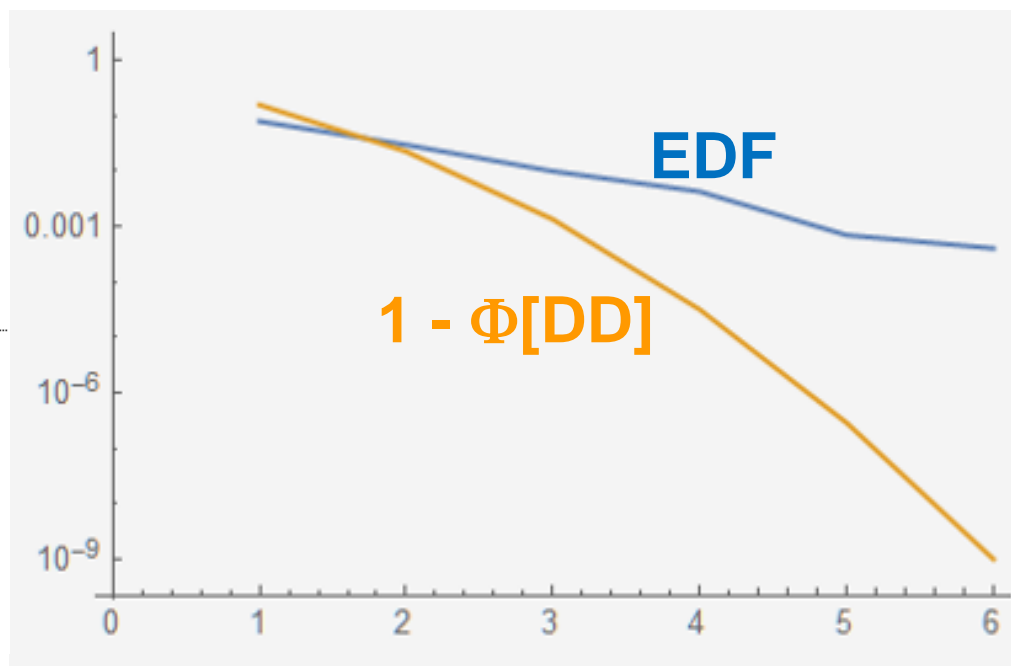
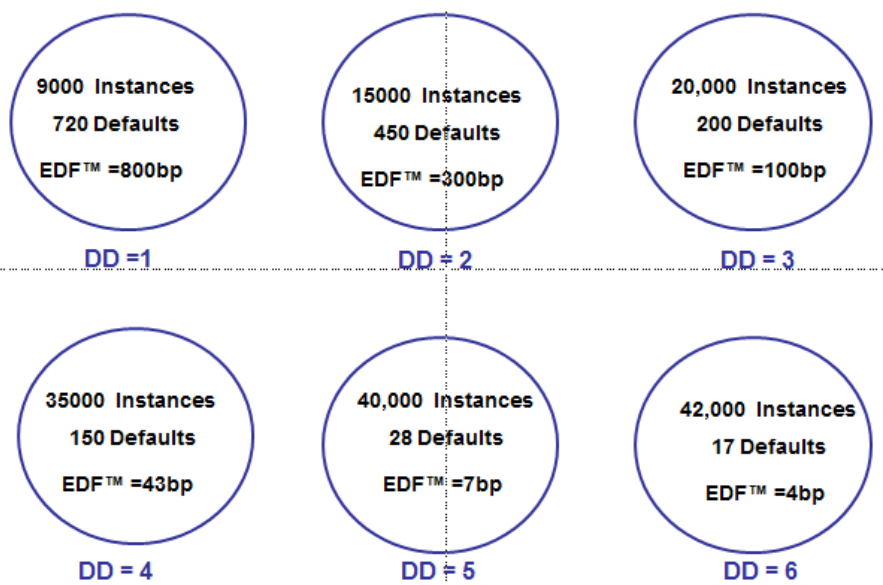
Instead, the EDF of a firm is set equal to the average historical default rate of firms having the same DD.

- If a firm has $DD = 6$, its EDF would be $17 / 42,000 = 0.04\%$
 - An EDF uses DD to find historical analogs of current firms.
 - Then it uses historical defaults to calibrate EDFs.

Years ago, I found this on the Moody's website....

Calibrating EDF to DD

Moody's quantifies EDFs with historical default rates, not with the theoretical value of $1 - \Phi[DD]$.



EDF summary

Option theory gives very bad estimates of PD.

- A firm with low DD defaults about half of $1 - \Phi[DD]$.
- A firm with high DD defaults much greater than $1 - \Phi[DD]$.

Moody's recalibrates with historical data.

- An EDF of a firm today is the historical default rate among firms sharing the value of DD.
 - On average, EDFs tend to track the historical data. On average.

At least, that's what I think.

- Moody's changes its procedures over time.
- They do not say exactly what they do.
 - They sell EDFs; they don't sell how to calculate EDF's.
- So, I don't know for sure.

Questions? Comments?

Asset return correlation and $\rho_{i,j}$

Correlation in the credit model

"Correlation" refers to the correlation, $\rho_{i,j}$, between the latent variables $\{Z_i\}$ that are responsible for default.

- We also call it “credit correlation.”
- When we fit the Vasicek distribution of the default rate to data, we estimate the correlation between latent variables.
 - We can’t estimate the correlation between two particular firms this way.
 - Instead, we estimate a uniform value for many pairs of firms.

The result is the uniform value of ρ that would best explain the pattern of joint defaults in the data.

Asset return correlation

Many practitioners assume that correlation ($\rho_{i,j}$) is equal to the correlation between firm asset returns ($r_{i,j}$).

- This requires historical estimates of firm asset values.
 - As with estimating *PDs*, this requires de-leveraging market cap.

We saw that firm asset values and volatilities imply systematically wrong values of PD.

- Why would the correlation between asset returns imply good values of $\rho_{i,j}$?

When Moody's calculates EDF, it calibrates to default data.

- There is no recalibration of correlations.
 - If there were, we show it should be toward zero.

Why use asset return correlation

Robert Merton gave this argument years ago:

- Today, the value of a firm's assets is known.
- Within the current period,
 - The rate of return of the firm's assets has a normal distribution.
 - At the end of the period, the firm owes a known amount to its lenders.
 - If the firm does not pay the debt in cash, the lenders gain ownership.
 - The firm will be able to get the cash if the value of its assets is greater than the value of its debt.
 - The firm will be unable to get the cash if its asset return is too low.
- Same things hold for a second firm.
 - Its asset returns are jointly normal with the first firm.
- The probability that both firms default depends on the correlation between their asset returns.

It is a nice story

If the story were true, then we wouldn't have introduced the idea of latent variables responsible for default.

- We would have said, “Asset returns are responsible for default.”

The less you know, the easier the story is to accept.

- The value of a firm's assets is not known.
- Asset rate of return is not normal.
- There is no current period that ends.
- Corporate debt never comes due all at once.
- The lenders don't take the firm.
- The value of a firm's assets will not be known.
- RadioShack did get the money to pay its debts for a while.

If you were estimating something that has consequences, you might like to add realism.

Adding realism

Merton says that default and asset shortfall are the same.

- **Whenever a firm is in asset shortfall, it defaults on debt.**
- **Whenever a firm defaults on debt, it is in asset shortfall.**

But the link from asset return to default is not perfect.

- **Borrowers pay when they have enough money, not when their assets have returned enough.**
- **Bank lenders choose whether to declare an event of default.**
- **Nobody knows how much a firm is worth in any case.**

The connection between asset returns and defaults is noisy. The noise makes credit correlation less than asset return correlation.

What does the data say?

| Source Study | Data Source | Results | Correlation from default data is less than correlation from asset data |
|--|--|--|--|
| Gordy (2002) Cespedes (2000) Hamerle <i>et al.</i> (2003a) Hamerle <i>et al.</i> (2003b) Frey <i>et al.</i> (2001) Frey & McNeil (2003) Dietsch & Petey (2004) Jobst & de Servigny (2004) Duellmann & Scheule (2003) Jakubik (2006) | S&P Moody's S&P 1982 - 1999 UBS S&P 1981 - 2000 Coface 1994 - 2001 AK 1997 - 2001 S&P 1981 - 2003 DB 1987 - 2000 BF 1988 - 2003 | 1.5% - 12.5% 10% max of 2.3% 0.4% - 6.04% 2.6%, 3.8%, 9.21% 3.4% - 6.4% 0.12% - 10.72% intra 14.6%, inter 4.7% 0.5% - 6.4% 5.7% | |
| Duellmann <i>et al.</i> (2006) KMV (2001) Fitch (2005) Lopez (2002) | KMV Undisclosed Equity KMV Software | 10.1% 9.46% - 19.98% intra 24.09%, inter 20.92% 11.25% | |

Source: Chernih, Vanduffel, Henrad, 2006

Correlation and asset correlation

Years ago, Moody's subscribers got month-by-month estimates of the asset values of every rated firm.

I used them to imply the asset return correlation of every pair of firms over every 3-month period.

- I adjusted for small sample size to get unbiased estimates.**

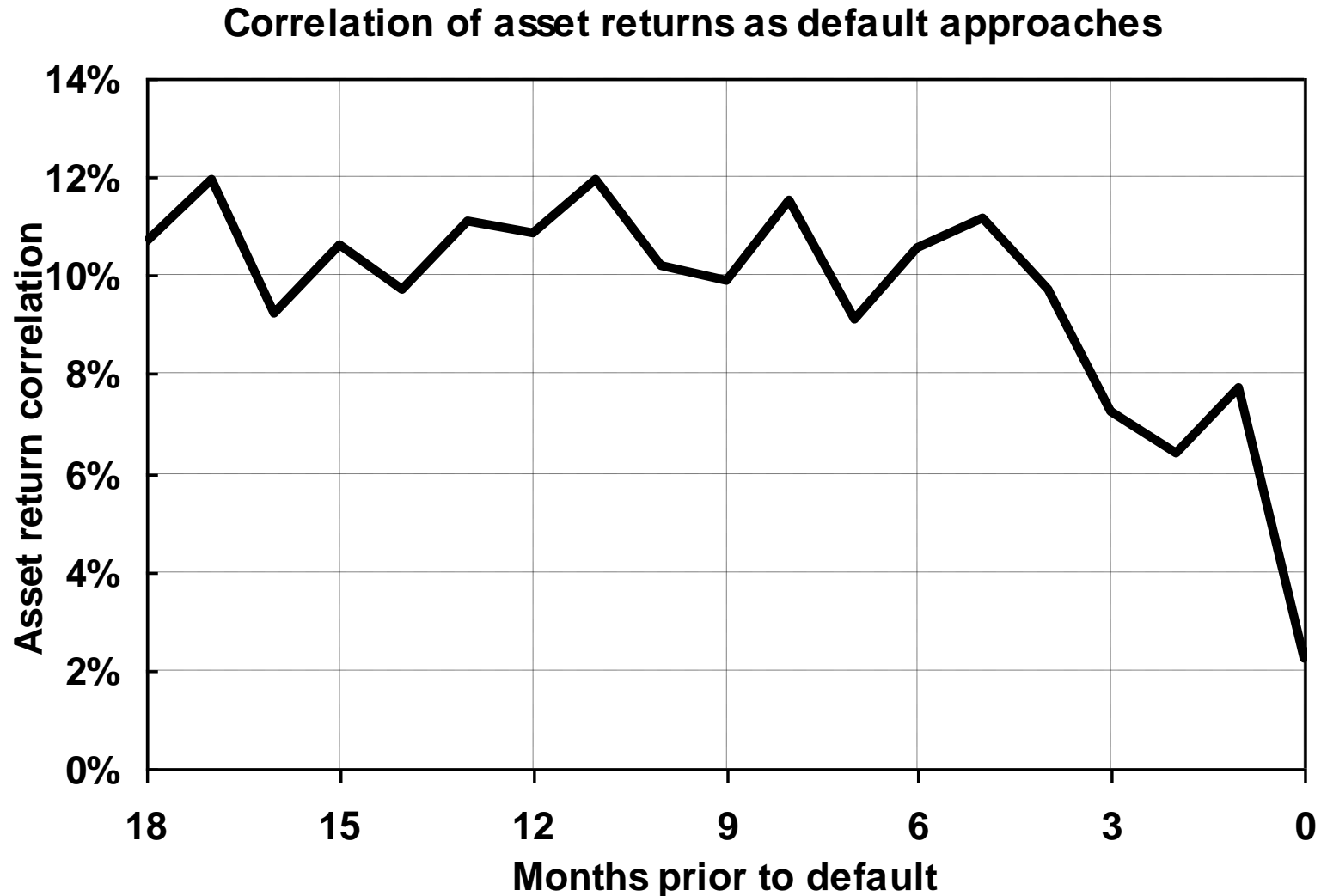
I calculated the “profile” ahead of each default:

- For each month prior to a default, I found the average asset return correlation of the defaulting firm with every other firm.**

Then I averaged the profiles.

- This gave the average asset return correlation of a firm that soon defaults with other firms. Correlation is not stable.**

Before default, *r changes*



Interpretation

A year before a firm defaults, its average asset return correlation with other firms is in the range 10%-12%.

- This is consistent with published asset correlation estimates.

As default approaches, average asset return correlation falls into the range of 6%-10%.

- This is broadly consistent with the credit correlation estimates calibrated to rates of joint default.

To predict future joint defaults over the one-year time frame, the value of $\rho_{i,j}$ must be closer to zero than $r_{i,j}$.

The values of $\rho_{i,j}$ could be calibrated to history by adjusting the $r_{i,j}$ to reflect historical data,

- just as data adjusts $1 - \Phi [DD]$ to EDF.

Questions? Comments?

Review and preview of tools

Review and preview of tools

Maximum likelihood estimation

Hypothesis testing

p-value

How model search defeats indications of significance

Maximum likelihood estimation

MLE works like this:

- 1. Write a symbolic PDF for each data record.**
 - The symbols represent parameters to be estimated.
- 2. Suppose that the data records are independent.**
 - The PDF of the data sample equals the product of the PDFs of the data records.
- 3. Find parameter values that maximize the PDF.**
 - Compared to other sets of parameter values, this one places greatest probability on the data set.
 - The function being maximized is called the likelihood.
 - The maximizing values are estimates called MLE's.

An example of MLE

The investigator believes that a variable R obeys the Vasicek distribution with a value of ρ equal to 0.15.

- Only the value of PD remains to be estimated.**

The data set consists of a single record: $r = 0.01$.

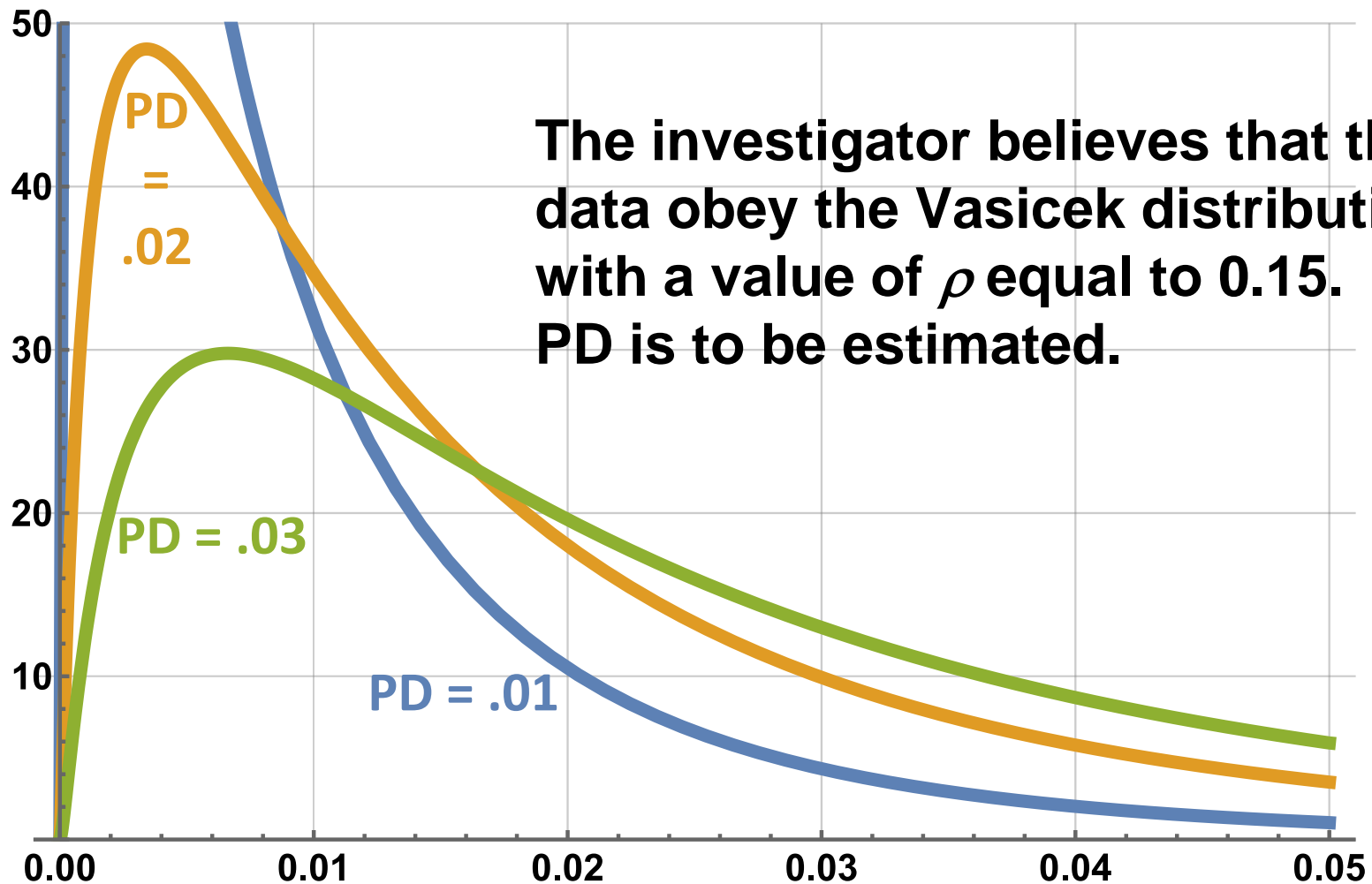
- A single data record makes easy illustrations with diagrams.**

The next slide shows three possible PDFs.

- Each PDF is Vasicek with $\rho = 0.15$.**
- The PDFs have differing values of PD: 0.01, 0.02, 0.03.**

3 possible PDFs

$$R \sim \text{Vasicek}[\text{PD}, \rho = 0.15]$$



The sole data record is $r = 0.01$

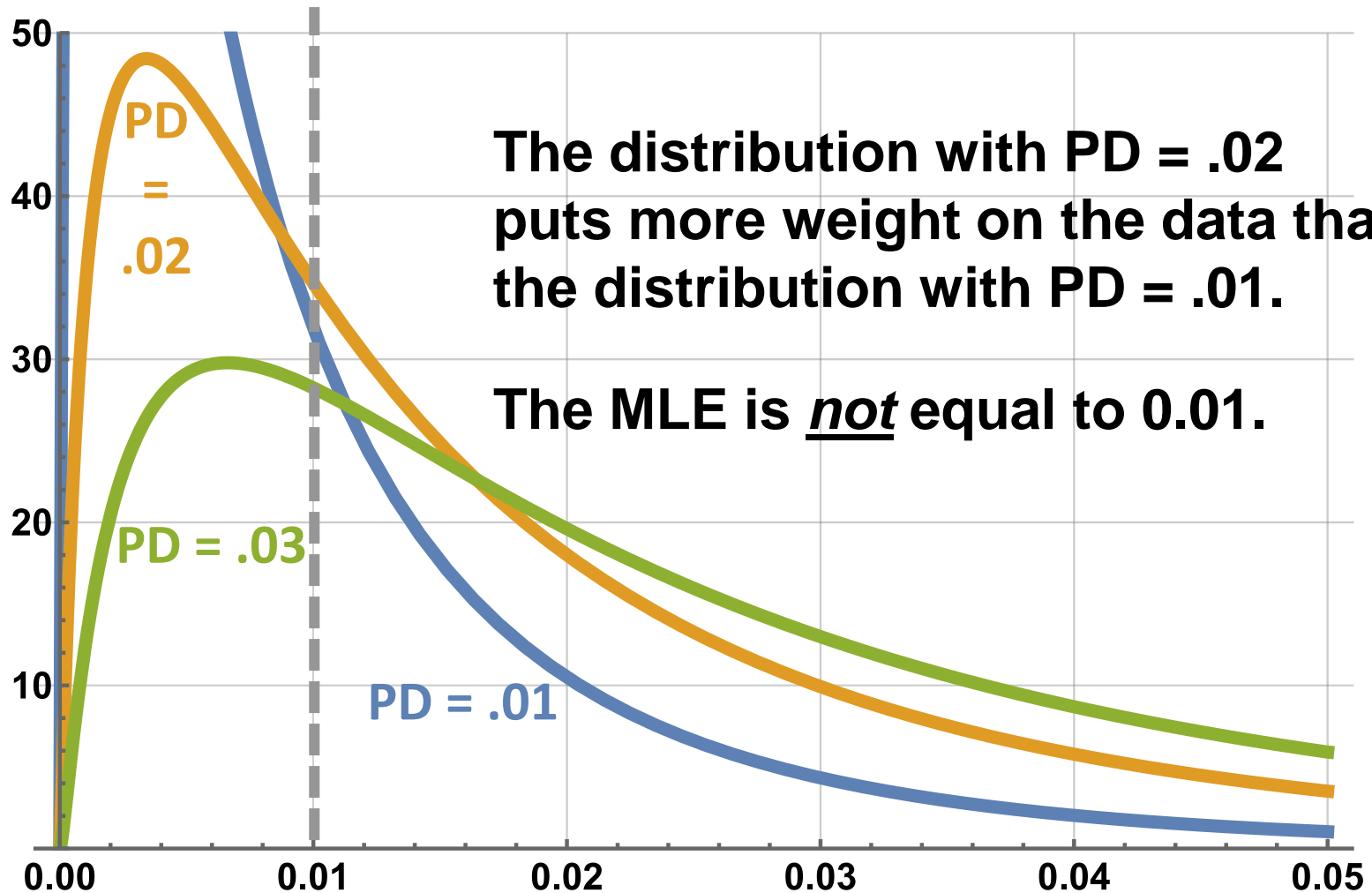
Therefore, an unbiased estimator of PD is $PD = 0.01$.

- This is unbiased because the expected value of a data point is the expected value of the distribution.
 - And the expected value of a Vasicek distribution is PD.

The next diagram shows that the MLE of PD is not 0.01.

- The distribution with $PD = 0.02$ places greater probability density on $r = 0.01$ than does the distribution with $PD = 0.01$.
 - A distribution with a mean different from 0.01 places more weight on the data set than does the distribution with mean equal to 0.01.
 - A distribution with a biased estimate of PD produces the data more frequently than a distribution with an unbiased estimate of PD.

The sole data record is $r = .01$



Economists and risk people

Economists care about parameter estimates.

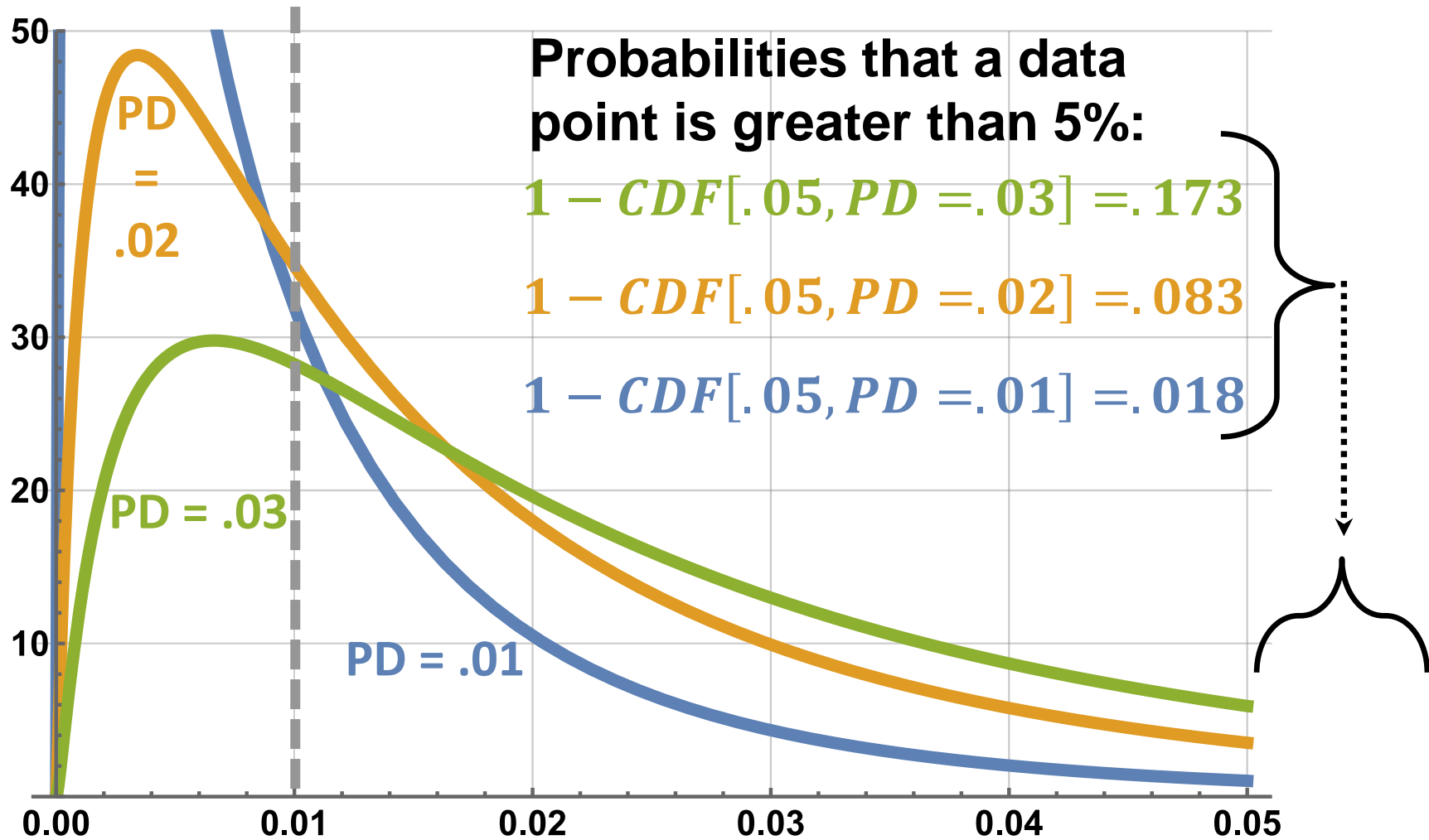
- **Their focus is the value of a parameter.**

Risk people care about distributions.

- **A distribution with mean 2% is more likely to produce the historical data than a distribution with mean equal to 1%.**
 - **Risk people prefer the distribution that produces the data most readily.**
 - **This distribution has a mean parameter that is different from the data average. So what?**
 - **Risk people probably don't care much about the mean anyway.**

It is worth noting that it really matters (to risk people, anyway) which distribution is chosen...

Distributions are much different



Questions or comments

The likelihood function

The likelihood function compares the 3 distributions to each other and to many other distributions.

Until now, we've plotted probability density functions.

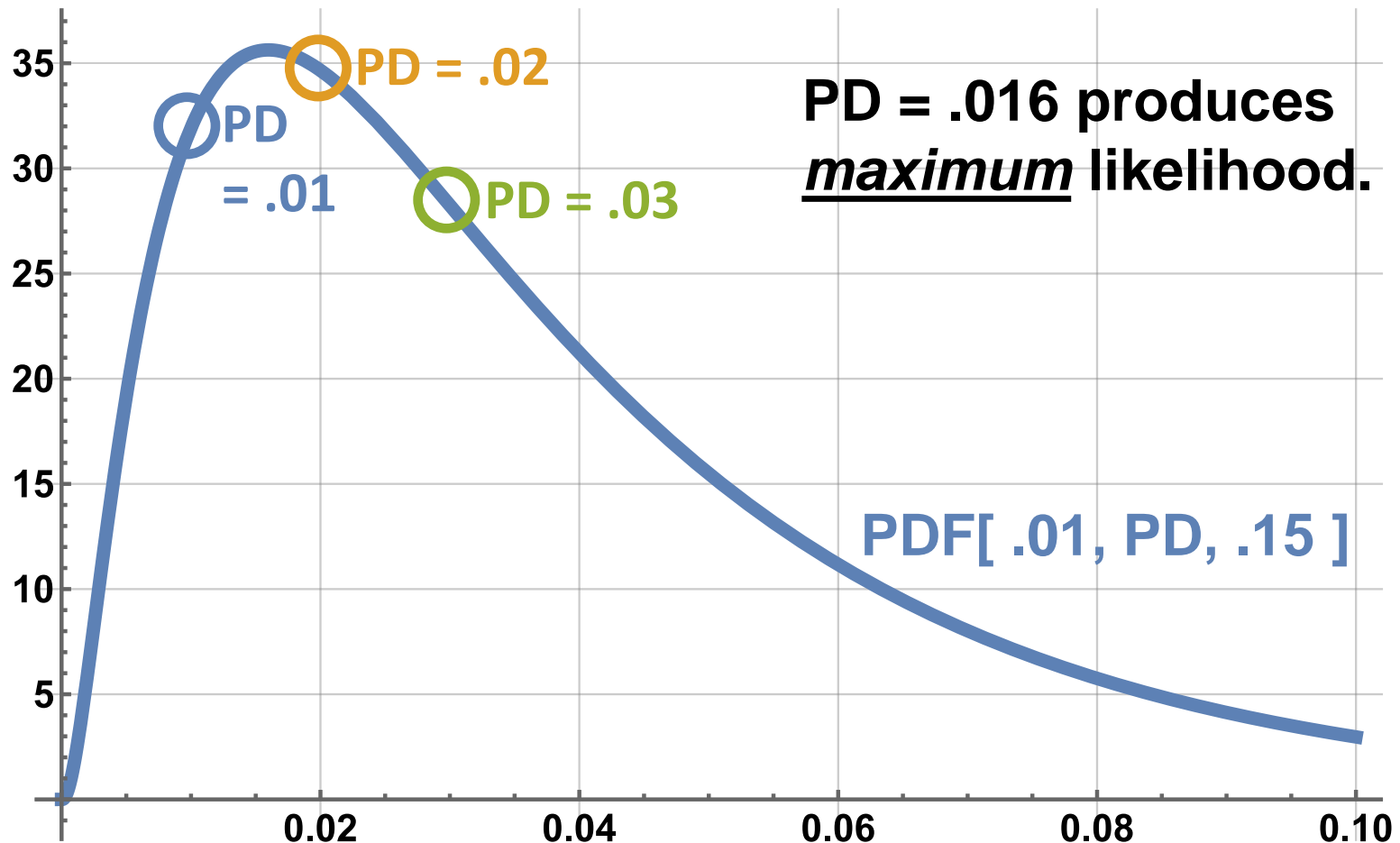
- For example, the Vasicek: $PDF_{vas}[R, PD, \rho]$.
- We assume values for PD and ρ .
- We plot “probability density” as a function of R .
- The area underneath is 1.0.

The arguments of a likelihood function are parameters.

- In this case, it is $PDF_{vasicek}[r = .01, PD, \rho = 0.15]$. and given the data point, $r = 0.1$, we plot likelihood as a function of the parameter PD .
- The area underneath is not 1.0...

The likelihood function

The three distributions are different; the likelihoods have quite similar values. Lack of data!



The MLE

Every point on the likelihood function corresponds to a different Vasicek distribution with $\rho = 0.15$.

Among them, the one with $PD = 0.016$ places greatest probability density on the data sample.

- If the population has $PD = 0.016$, the sample would be produced more often than if PD had a different value.

The data might be the rare product of a distribution with its PD far from 0.016.

- We don't know and we can't tell.
 - PD is what it is with probability 1.0 and 0.0 otherwise.
- Nothing can tell us $Prob[PD = 0.016]$.

MLE—five data points

Suppose R has a Vasicek PDF:

$$f[R, PD, \rho] = \frac{\sqrt{1-\rho}}{\sqrt{\rho} \phi[\Phi^{-1}[R]]} \phi \left[\frac{\sqrt{1-\rho} \Phi^{-1}[R] - \Phi^{-1}[PD]}{\sqrt{\rho}} \right]$$

Suppose there are five data records: 1%,2%,3%,4%,5%.

- If the records are independent draws,
 - then the PDF of the data set is the product of five Vasicek PDFs.

The likelihood function for these data is $L [PD, \rho] =$

$$f[.01, PD, \rho] f[.02, PD, \rho] f[.03, PD, \rho] f[.04, PD, \rho] f[.05, PD, \rho]$$

As a practical matter, one often maximizes $\text{Log}[L [PD, \rho]]$.

Simultaneous MLEs are $PD = 3.02\%$, $\rho = 5.45\%$.

- Check for yourself and see.

Questions? Comments?

Hypothesis testing

Two models are each fit to a data set.

- A simple “Null” model fits the data to a certain degree.
- A richer “Alternative” model has more parameters.
 - Setting the parameters to certain values reproduces the Null model.
 - Setting them to their MLE’s provides a better fit than the Null.

Only if the Alternative fits significantly better is the Null Hypothesis rejected.

- The model being tested is the Null Hypothesis.
 - If a Null Hypothesis survives numerous tests, it can become a modeling standard.
 - If it is rejected numerous times, it should be improved.

We prefer the simpler hypothesis

We prefer the Null until we are convinced that it is wrong.

- William of Occam said the simpler idea is more likely to be true.
 - Today, no one depends on “truth”. We wouldn’t know it if we saw it.
- A modern person prefers the Null because it is simple.
 - A more complicated model has more things that can go wrong.
 - It is more likely to contain false effects. This is called Type 1 Error.
 - This is like superstition in daily life. It is best to get rid of it.
- It is very likely that the Null fails to contain something it should.
 - This is called Type 2 error.
 - This is less serious because we don’t expect to know everything right now.

The test knows that the Alternative always fits better.

- It prefers the Null unless the Alternative fits significantly better.

Putting in a little math

Each hypothesis places a certain probability on the data set.

- **MLE maximizes this probability for each hypothesis in turn.**
 - **Because it is more flexible, the alternative puts more probability.**

The “likelihood ratio” is the ratio of these two probabilities.

Twice the log of the likelihood ratio has a known distribution!

- **This is “Wilks’ Theorem.” It depends on two conditions:**
 - **There is an asymptotic amount of data. (There isn’t! Finite data set.)**
 - **The Null Hypothesis is true. (It isn’t! All models are false.)**

If twice the log likelihood ratio is in the tail of its distribution, then the test rejects the Null.

- **Tests assume that nothing rare ever happens.**
- **The test says nothing about the Alternative. It wasn’t tested.**

Wilks' Theorem

Define $L_0 = \text{Max [PDF [Data | Null hypothesis]]}$
 $L_1 = \text{Max [PDF [Data | Alternative hypothesis]]}$
 $k = \text{number of extra parameters in the Alternative}$

If the Null hypothesis is true, and if the number of independent records rises without limit, then the distribution of $D = -2 \text{ Log [} L_0 / L_1 \text{]}$ approaches χ^2_k .

- Too bad for us, people symbolize this statistic by "D".
 - D is positive and equals $2 (\text{Log [} L_1/L_0 \text{]})$.

Decision criterion

$D = 2 \log [L_1 / L_0]$ is asymptotically distributed χ^2_k .

- In practice, D is assumed distributed χ^2_k for the data at hand.
 - The variance in the finite sample is greater than the asymptotic variance.

Suppose that the D statistic is a tail observation. Either:

- The Null Hypothesis is true and something unlikely happened.
- The Null Hypothesis is not true.

Statisticians make the second choice:

- If D is a tail observation, then the test rejects the Null.
 - The data sample is not infinite, and this leads to fatter tails as always.
 - So there are more rejections than the stated “size” of the test.

A common criterion is the 95th percentile.

- Size = 5%: Critical values of D are 1df = 3.84; 2df = 5.99.
 - If $2 \log [L_1 / L_0] = 3.84$, then $L_1 / L_0 = \exp[1.92] = 6.82$.
 - The alternative must place 6.82 times more probability on the data.

Testing the three distributions

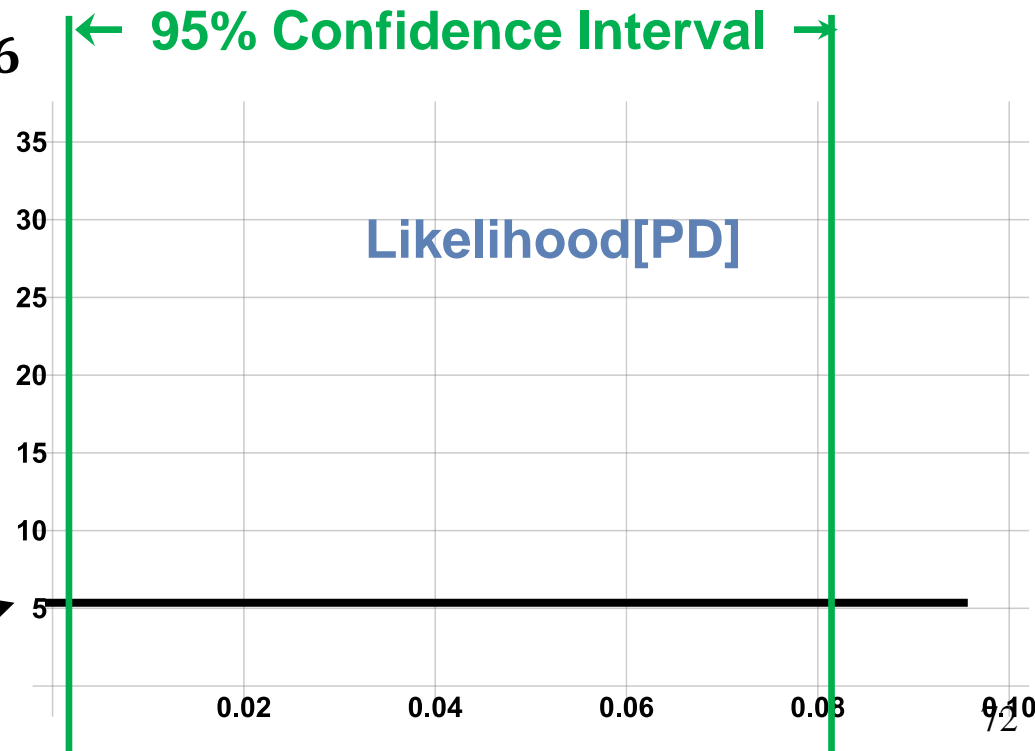
From the likelihood function,

- $L_1[PD = MLE = .016] = 35.63$
- $L_0[PD = .01] = 31.9; D = .22$
- $L_0[PD = .02] = 34.7; D = .06$
- $L_0[PD = .03] = 28.2; D = .46$

No hypothesis would be rejected.

For a test of size 5%, the critical value of L_0 is

$$\frac{L_1}{Exp[1.92]} = \frac{35.63}{6.82} = 5.22$$



Questions? Comments?

The 5-data-record sample

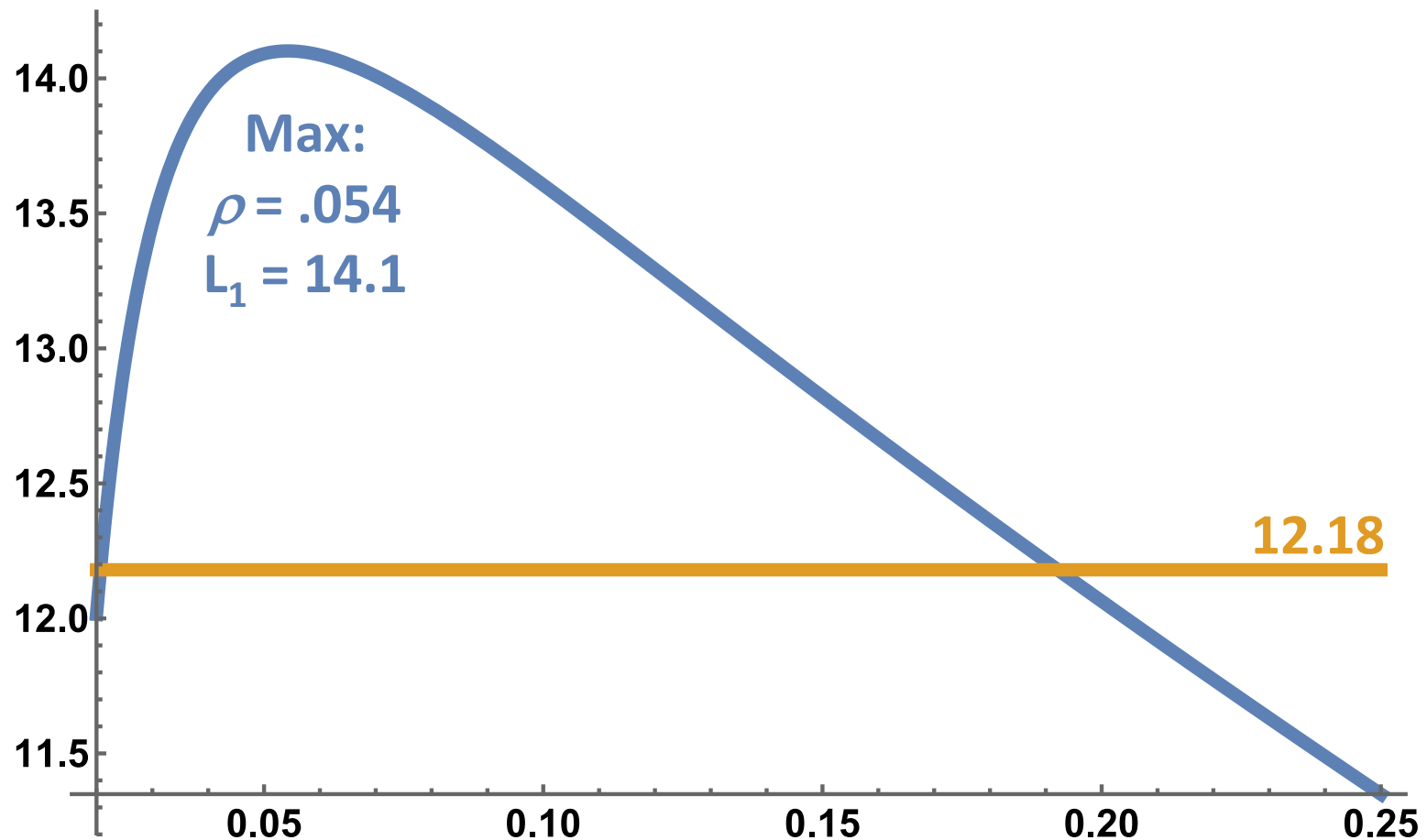
Suppose there are five data records: 1%,2%,3%,4%,5%.

- If the records are independent draws,
 - then the PDF of the data set is the product of five Vasicek PDFs.
- Suppose that $PD = 0.03$.
- Which values of ρ would not be rejected?

The next slide plots the value of L_1 . It also has a horizontal line 1.92 below the peak.

- $1.92 = 3.84 / 2$. Any Null Hypothesis that produces likelihood above the horizontal line is not rejected.
- Big enough likelihoods stem from $0.021 < \rho < 0.192$.
 - The data reject that $\rho = 0$. (BTW, what would the PDF look like if $\rho = 0$?)
 - Otherwise, just about anything is OK.
- This is one way to find a confidence interval for ρ .

Log[L₁] of the 5-record sample



The p-value

There would be no rejection of a Null Hypothesis that ρ is a number on the interval 0.021 to 0.192.

Suppose a Null hypothesis that names a value, ρ_0 .

- Under this Null hypothesis, we can simulate**
 - drawing data samples from the population, and**
 - finding the associated MLEs of ρ .**
- Some fraction of the simulated estimates of ρ are further from ρ_0 than the MLE based on the real data.**
 - This fraction is the p-value of the test.**
 - If it is low, then the MLE is not likely to have been produced by the Null.**

The usual standard is to reject the Null if p-value < 0.05 .

What the p-value is not

The p-value is not the probability that the Null is true.

Listen: Of course the Null is not true.

Models are not reality.

A map is not a territory.

All models are wrong

From Wikipedia, the free encyclopedia

All models are wrong is a common aphorism in statistics; it is often expanded as “All models are wrong, but some are useful”.

Hypothesis testing summary

Define two “nested” hypotheses.

- With certain parameter values, the Alternative equals the Null.

Estimate the parameters of each hypothesis using MLE.

Find $D = 2 \log [L_1 / L_0]$.

If $D >$ the 95th percentile of the χ^2 distribution with k degrees of freedom, then reject the null hypothesis.

- k is the number of extra parameters in the alternative.

If the Alternative does “better enough” on the calibration data, then the Null hypothesis is judged to be not a good guide to future data.

More perspectives

The purpose of a hypothesis test is to improve forecasts.

- We want a model that fits the data we haven't seen yet.

A wrong, simple model can be useful.

- Example: Newtons' inverse square law of gravity.

Even though the data say that people die younger if they carry a cigarette lighter, people should quit smoking.

- A rejection can tell you that something is wrong with the Null, but it doesn't tell you what is wrong or how to fix it.

Progress does not occur by rejecting the same model that has been rejected countless times before.

- If we know that the Null is wrong, why test it?
 - “This is how you get your work published,” is not a good enough answer.
- Instead, find a Null Hypothesis that is difficult to reject.

Questions? Comments?

There are two kinds of model:

**Those that have been rejected by the data,
and those that have not been rejected, yet.**

How *search* defeats *significance*

How search defeats significance

**This section shows how something modelers do
(search among many possible model specifications)
invalidates something modelers want
(a valid measure of statistical significance).**

**When a modeler performs ad hoc model search, the
model is chosen based on the data.**

- **It is not valid to say that the data support the model
when the model has been chosen to fit the data.**
 - **Model search is often called “building the model.”**
 - **That sounds better than “shameless data fitting.”**

Incidentally

Sometime when you aren't listening to me, put "XKCD" into a browser. Insight!

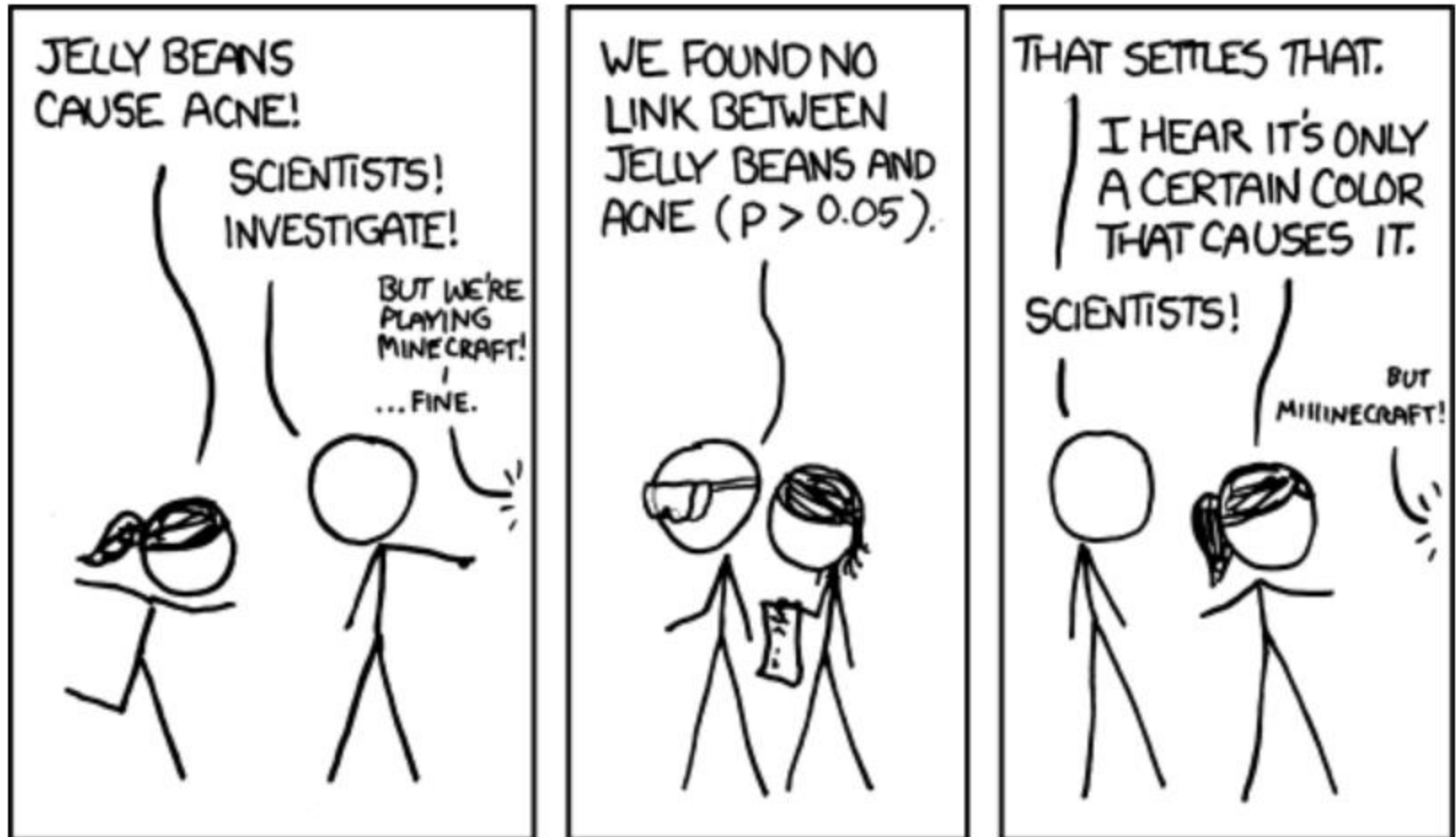
ADVANCED TECHNIQUES

|< < PREV RANDOM NEXT > |



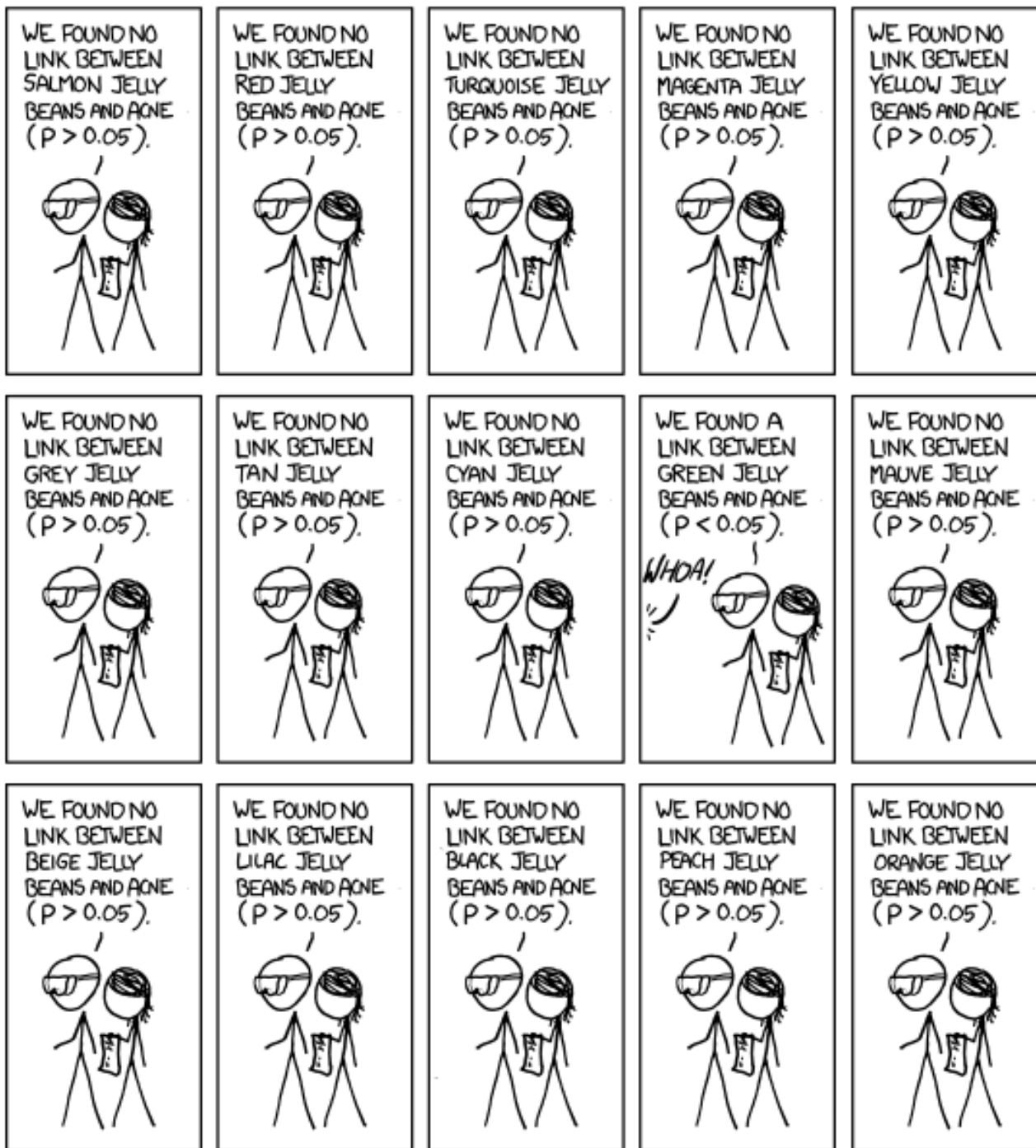
ALL ADVANCED MATH TECHNIQUES

Specification search, part 1

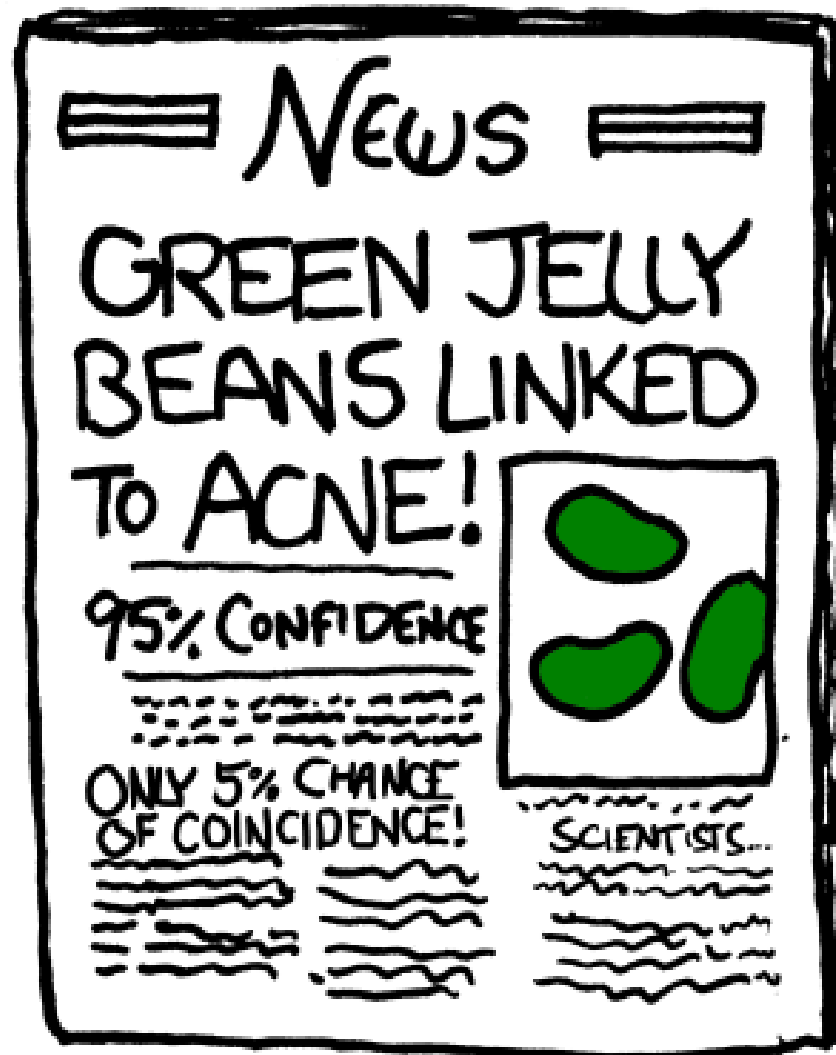


Spec

part 2



Specification search, part 3



How search defeats significance

Note that the scientists made 20 different studies, one for each color, and the summary reported about green ones.

- Of course one of the 20 colors will show the greatest association with acne. You can sort anything!
- Of course one color is likely to show "significance."
 - A well-designed test should show significance in 5% of data samples when the variables have no relation whatever, and 20 things were tried.

Outsiders might not know that 20 things were tried.

- If they did, they could compute that this is an expected result.
 - But a bad scientist is careful to conceal ad hoc model search.

Hey, let's do big-time research on corporate mergers!

- What's the effect of location, CEO academic degrees, and mentions in top newspapers, when these factors interact!

Bad forecasts

The green jellybean model makes bad forecasts.

- It predicts that green jellybeans cause acne.
- It predicts that only green jelly beans cause acne.
 - The search among colors biases the results.

This prediction can shape behaviors.

- If people stop eating green jellybeans, what will they eat?
 - Will they eat other colors of jellybean?
 - Will they eat grilled salmon and broccoli?
 - Will they avoid all jellybeans and eat hot fudge sundaes?

The point: A bad risk model causes risky behavior.

- If a risk model understates the risk of a certain trade, that trade will be done too frequently.

Questions? Comments?

Common practices reduce model validity:

- **Model “specification search”**
- **Model “building”**
- **Preliminary data regressions**
- **“Correlograms” to specify serial dependence**

Don't forget

Homework 3 is due next Thursday at 6 Chicago time.

Lisheng's TA session will be online Sunday at 6.