Linking Goodness of Fit to Economic Gains

What's the dollar benefit of more accurate predictions?

Bob Stine
Department of Statistics, Wharton

slides on web page after conference

Wharton
Department of Statistics

Common Class Project

Challenge students taking regression to build a model used to

- predict sales
- forecast demand ...

Success shown by

- Student communicates business implications

- Higher R^2

How does fit convert to $?

- Students recognize that higher R^2 means a "better fit", but what is the economic value?
  - No matter which is used, how to you convert improvements of the model into dollars?
- For example
  - R^2 of a model increases from 50% to 70%
  - Can you convert that to money?
- Generic scenario?

News-vendor Problem

Context

- Unknown demand
- Perishable good
- How much to make?

electronics

foods fashion

cars

Uncertainty in demand implies wasted opportunity

- Make too much: leftover product
- Make too little: unmet demand

Aside

- Another chance to address Type I/Type II error

Solution to News-vendor

Notation

Demand D is random variable

Cost to manufacture   cost

Selling price   price

How much to supply?

Expected marginal value of m^{th} item is

price P(D > m) – cost

Continue making so long as positive marginal value

$$\frac{\text{price-cost}}{\text{price}} = P(D \le m)$$

$$\pi = \%\text{profit} \quad = \quad P(\text{demand met})$$

Implication of Uncertainty

Add assumption

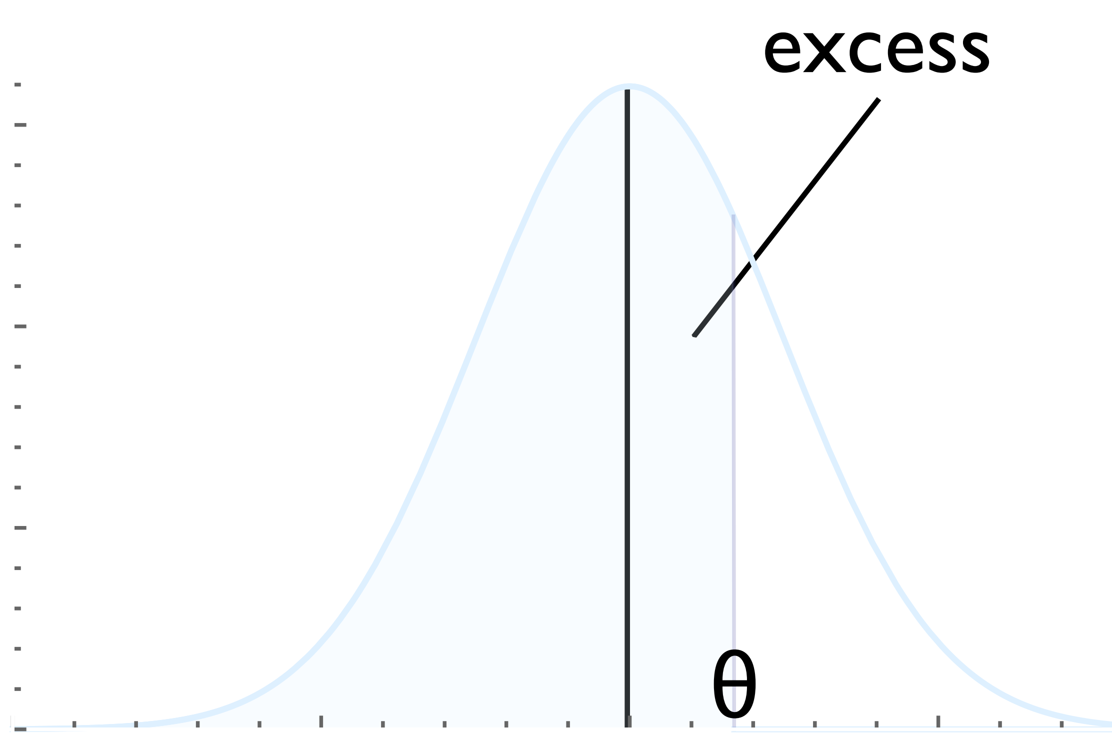
$$D \sim N(10000, \sigma^2)$$

Label target to make $\theta = \mu + \sigma z_{\pi}$

%profit=0.75

σ=2500

$\theta \approx$ 11,700



Wharton
Department of Statistics

Implication of Uncertainty

Add assumption

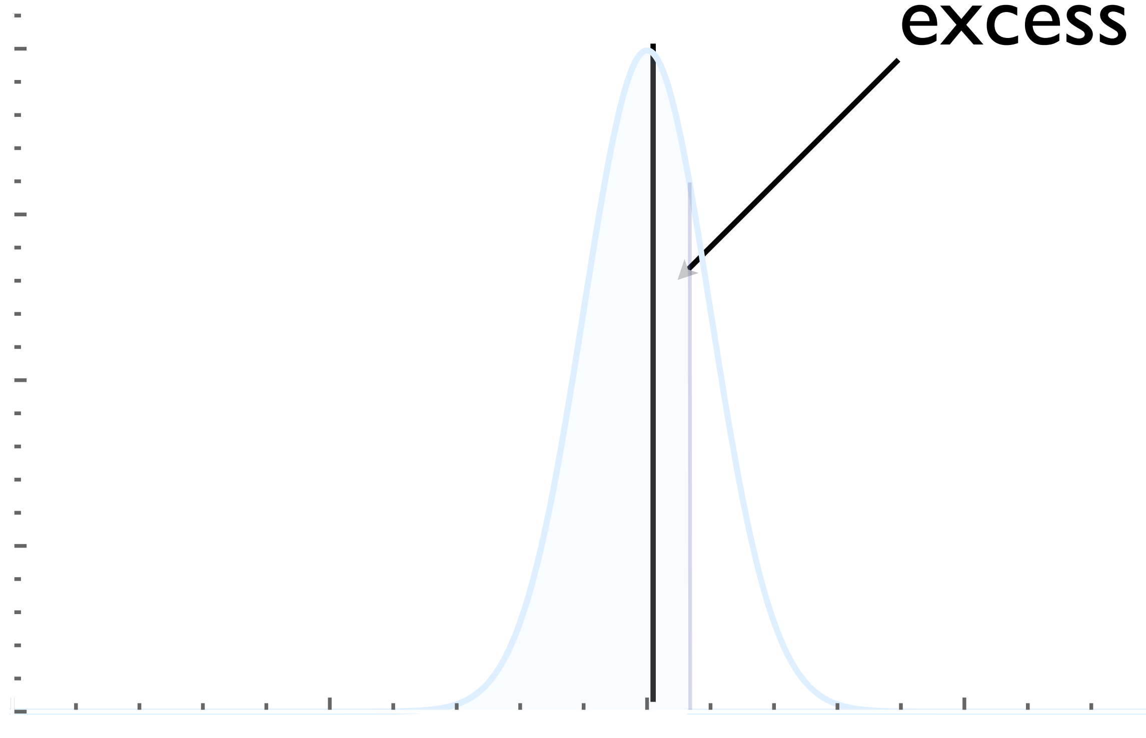
$$D \sim N(10000, \sigma^2)$$

Label target to make $\theta = \mu + \sigma\, z_\pi$

%profit=0.75

σ=1000

$\theta \approx$10,700



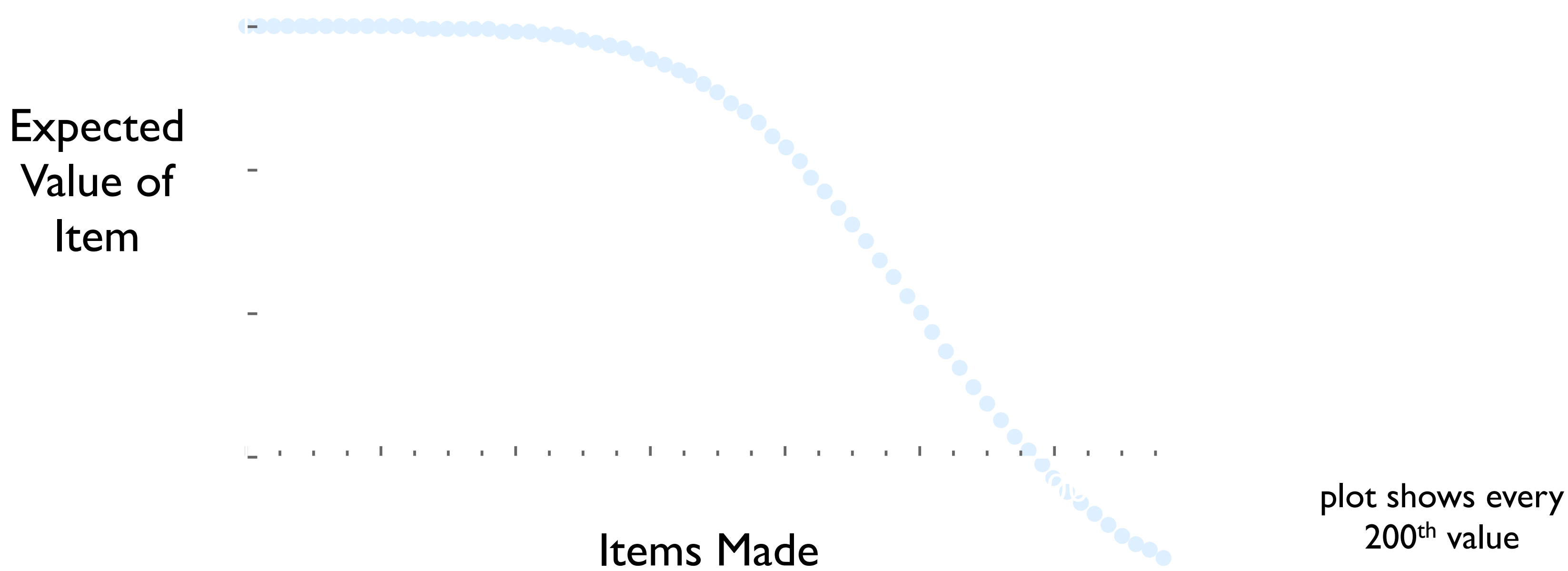
Expected Profit

What's the expected profit?

- Easy to find the optimal amount to produce
- Harder to find associated expected profit

'Modern' approach

- Just do it: calculate the expected profit of each item
- Add them up



Expected Profit

What's the expected profit?

- Discrete calculation is exact
- Hard to see forest for all the trees!

Approximation is simple in normal case

- Assume choose optimal production amount
- Expression for profit is linear in σ
- Maximum 'gain' for small σ if cost = 50% of price.

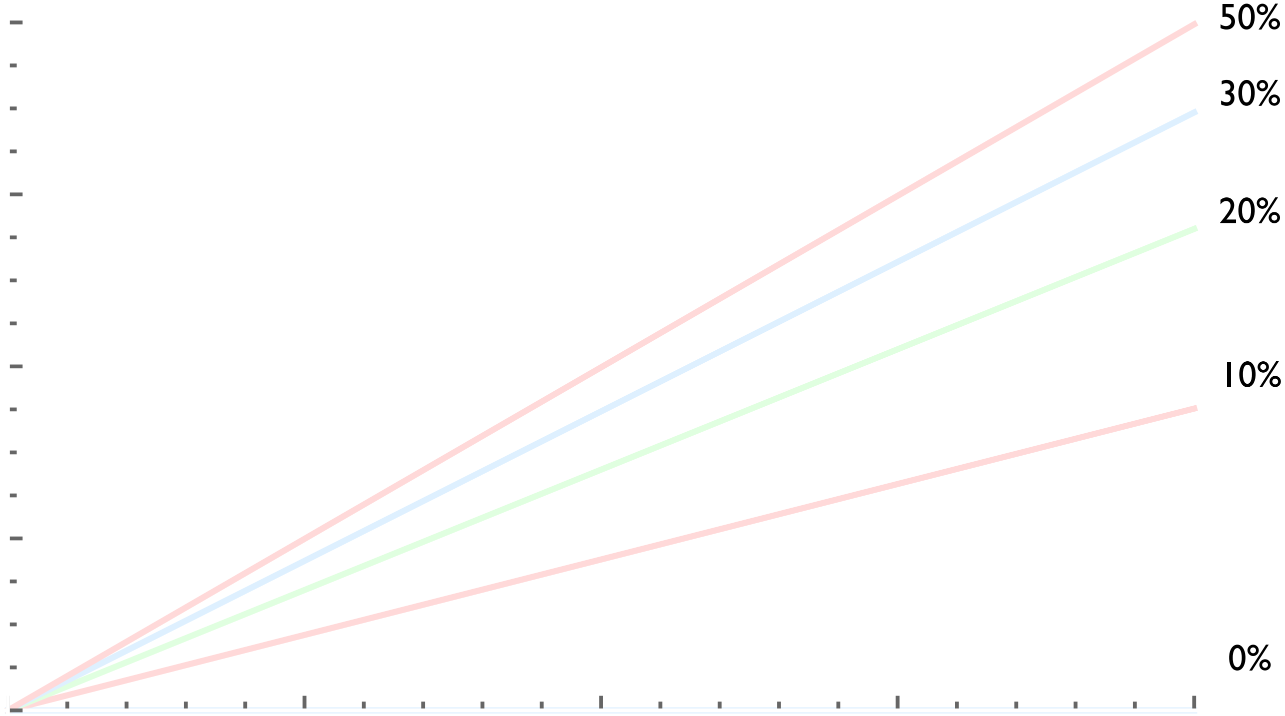
$$E(\text{profit}) \approx \underbrace{\mu\,(\text{price} - \text{cost})}_{\text{profit if no uncertainty}} - \underbrace{\sigma\ \text{price}\ \varphi(z_\pi)}_{\text{cost of uncertainty}}$$

φ is normal density

Penalty for Uncertainty

Explicit role for error variation

Effect of profit/price ratio



$\frac{\text{price-cost}}{\text{cost}}$

graph shows values for price=1

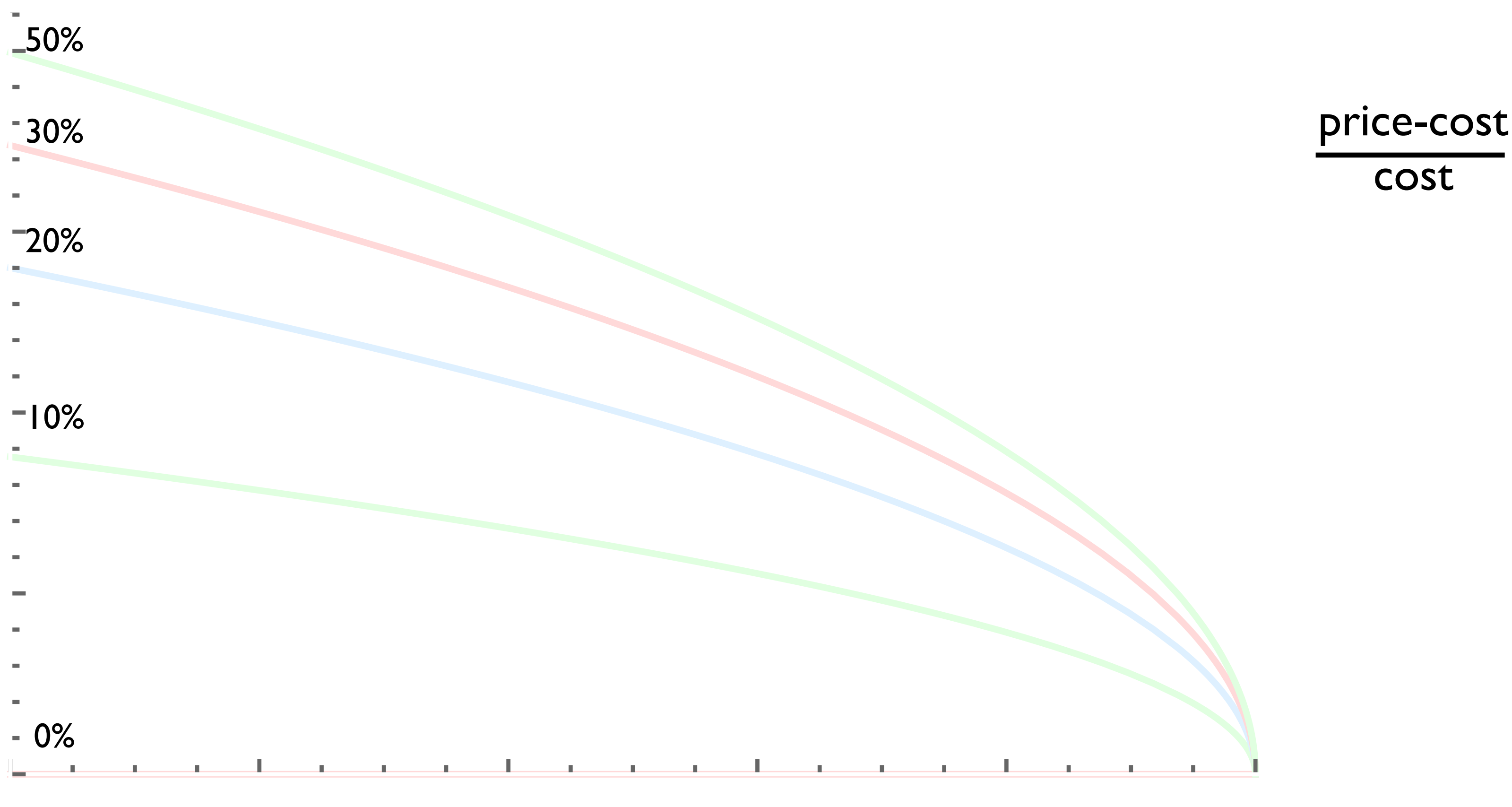
Wharton
Department of Statistics

$$E(\text{profit}) \approx \text{best} - \sigma \text{ price } \varphi(z_\pi)$$

Penalty for Uncertainty

If you really do like R^2 over σ then...

Better fit means smaller penalty



σ=1000

Wharton
Department of Statistics

$$E(\text{profit}) \approx \text{best} - \sigma \sqrt{(1-R^2)}\ \text{price}\ \varphi(z_\pi)$$

Details: Less Variance

Marginal

μ=10,000 with σ=2500

Optimal production θ = 11,686 units

Expected profit

750,000 – 79,444 ≈ $671,000

cost=$25
profit=$75

Regression R^2 = 0.60

μ =10,000 with σ=1000

θ = 10,675 units

Expected profit

750,000 – 31,778≈ $720,000

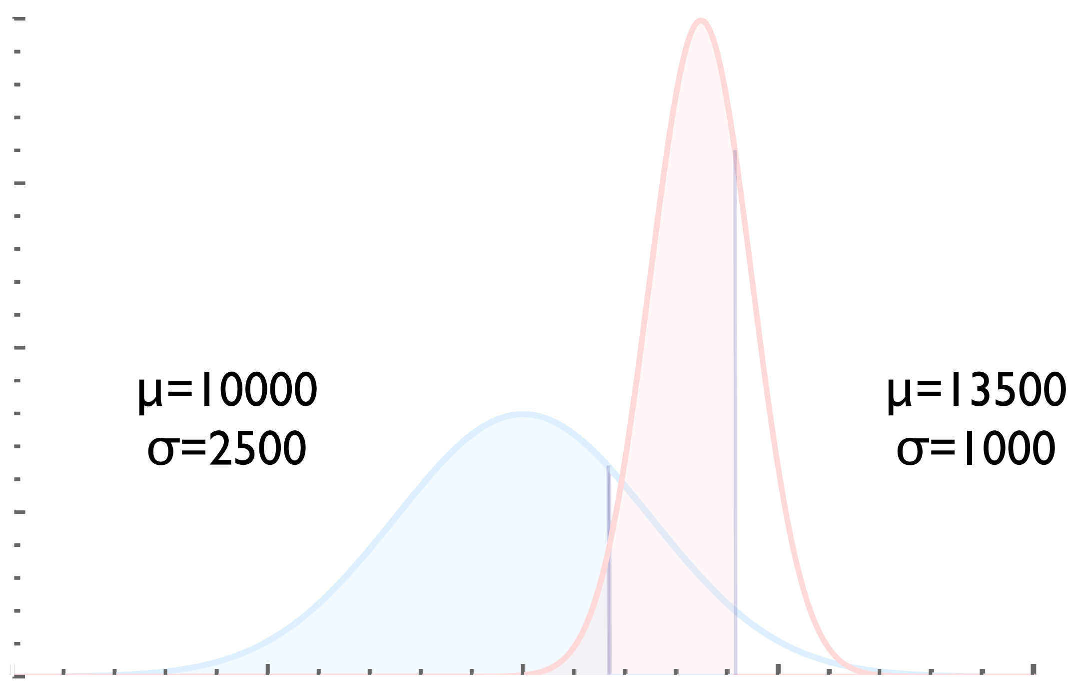
Wharton
Department of Statistics

What if mean changes too?

Marginal mean ignores possible covariate information that affects mean response

Suppose mean grows as SD falls...

Better fitting models imply that expected response varies with covariate



Example: Wrong Mean

Setup

Marginal: μ=10,000 with σ=2500

Conditional: μ=13,500 with σ=1000

cost=$25
profit=$75

Marginal results

Sets θ = 11,686 units, expecting $675,000 profit

Expectation* is really $875,000. Great!

Regression

μ =13,500 with σ=1000 implies θ= 14,175

Expected profit is $980,722

Just left $100,000 on the table!

Wharton
Department of Statistics

*NB: This calculation requires a different expression when not at the optimal production threshold.

Closing Remarks

Opportunity to link to other classes

- Supply chain, operations
- Economics
- Strategic management

Link to upper level classes

- Random variables, quantiles
- Where did that magic formula come from?

Advantage of calculus over discrete sums

- The discrete calculation in Excel hides the elegant role of σ and cost/price ratio

Thanks!