Variable Importance

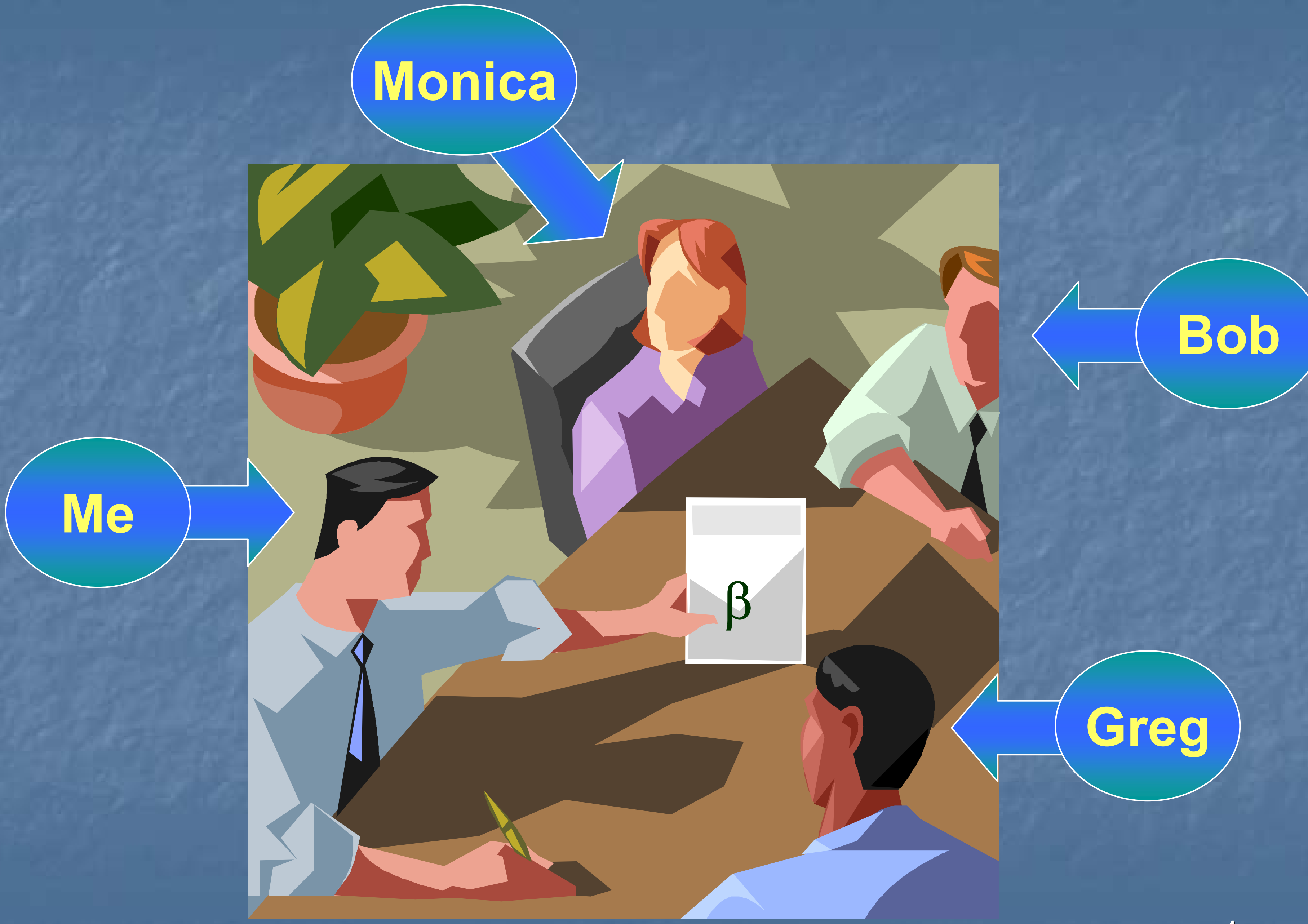
Prof Peter Lenk
Stephen M Ross Business School
University of Michigan
IMSM 2005

Outline

- Whose meaning of "importance"
- Examples
  - Demand analysis via OLS
  - Perceptual Maps via factor analysis
- Bayesian decision theory (BDT)
  - Metric and Non-metric Conjoint
- Importance in BDT
  - Market share simulation

A Vignette:
Why Peter is not allowed to talk with clients.

- Actors:
  - Peter, playing himself
  - Monica, playing the primary client
  - Bob & Greg, supporting cast
- Scene
  - Debriefing on an amazing HB model that Peter just developed. He is very excited.


Monica
Bob
Me
β
Greg


Very nice analysis and presentation. In your opinion, which variable is most important?
θ
Zzzz!
My stock options are in the money.

I’m glad you asked!
As you can see by the posterior means of the coefficients and Bayes factors, ….

“Blah, Blah, Blah.” This is going nowhere

Clueless

What a dork!

Should I trade up to a 5-Series?


Hmm, that's interesting. Now tell me which variable is most important.
Why doesn't she understand? This is so basic.
When is lunch?
I bet he drives a Chevy.

Well, it is not that simple.

As you can see by the posterior means of the coefficients and Bayes factors, ….

Just tell me how I can make my quarterly quota.

Is he still talking?

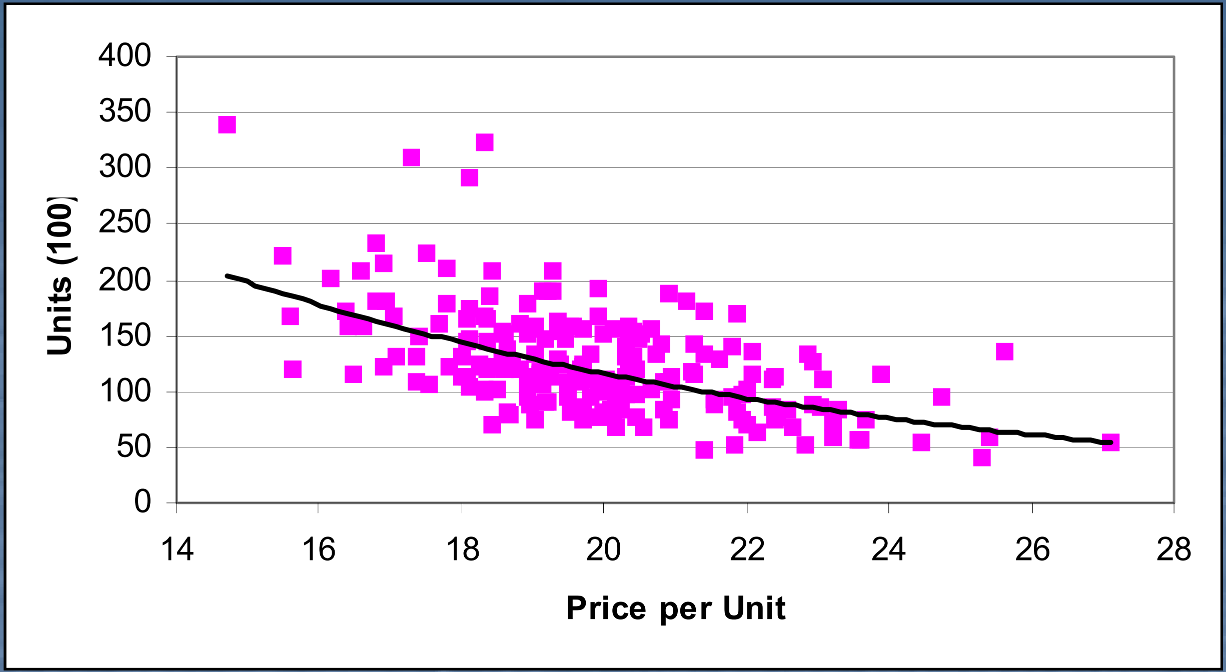
Never again!

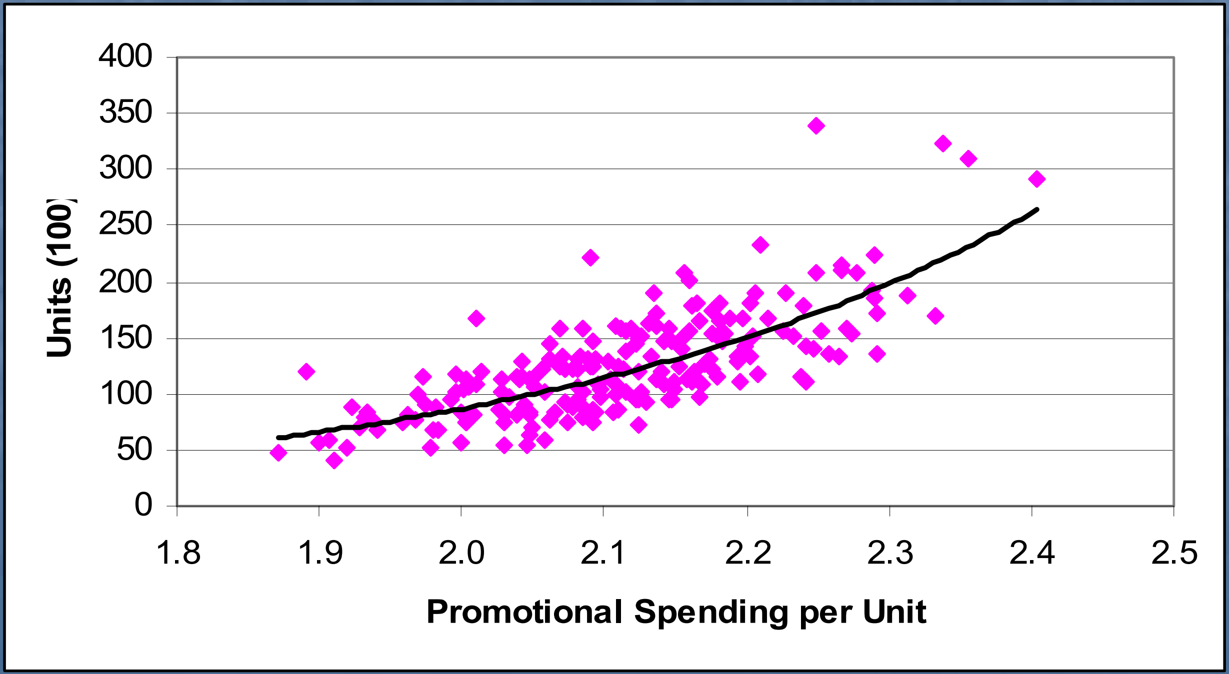
Most Important Variable

- Statistician
  - Statistical significance and power
  - Partial correlation
  - Percent variance explained
- Economist
  - Partial derivatives
- Marketing Manager
  - Impact on revenues, share, profits, brand equity, …

Simulation Fun

- Log-log model for demand
  - Log Units
  - Log Retail Price
  - Log Promotion/Advertising Spending
- Manufacturer sets suggested retail price
  - MSRP is $20 per unit
  - Retailer adjusts price
  - Retailer spends on promotion/advertising

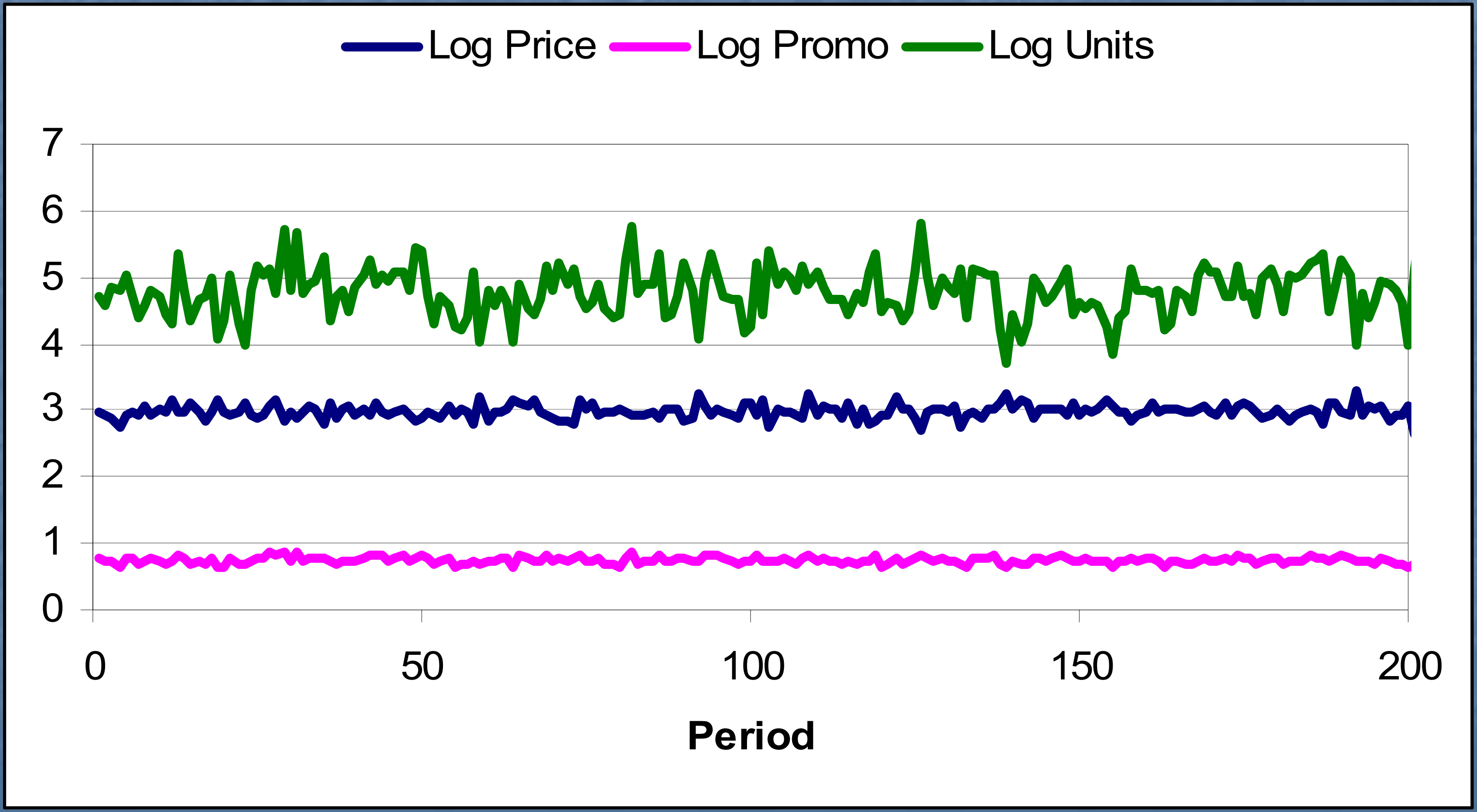

400
350
300
250
200
150
100
50
0
Units (100)
14
16
18
20
22
24
26
28
Price per Unit


400
350
300
250
200
150
100
50
0
Units (100)
1.8
1.9
2.0
2.1
2.2
2.3
2.4
2.5
Promotional Spending per Unit

Simulated Data

- Log(Units) =
  $7 - 2*\log(Price) + 5*\log(Promo) + error$

Very Stable Market



Which Variable is More Important?

Regression Statistics	
Multiple R	0.933
R Square	0.870
Adjusted R Square	0.869
Standard Error	0.134
Observations	200

	Coefficients	*STD Coef*	*Standard Error*	*t Stat*	*P-value*
Intercept	6.613	0.000	0.321	20.596	0.000
Log Price	-2.014	-0.559	0.093	-21.715	0.000
Log Promo	5.584	0.719	0.200	27.944	0.000

Statistician's Importance

- Variable selection
- Log(Promo) has the larger standardized coefficient and t-stat
- Partial Correlations
  - tstat / sqrt(tstat^2 + df_Error)
  - Price: -0.840
  - Promo: 0.894

Economist’s Importance

- Specify objective function
  - Revenue, profits, market share, etc
  - Assume continuity
- Find gradients
- Move in direction to optimize objective function
- Bigger steps are better
- Easy with log-log demand

Do it With Math

$$E\left[Y\right] = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$

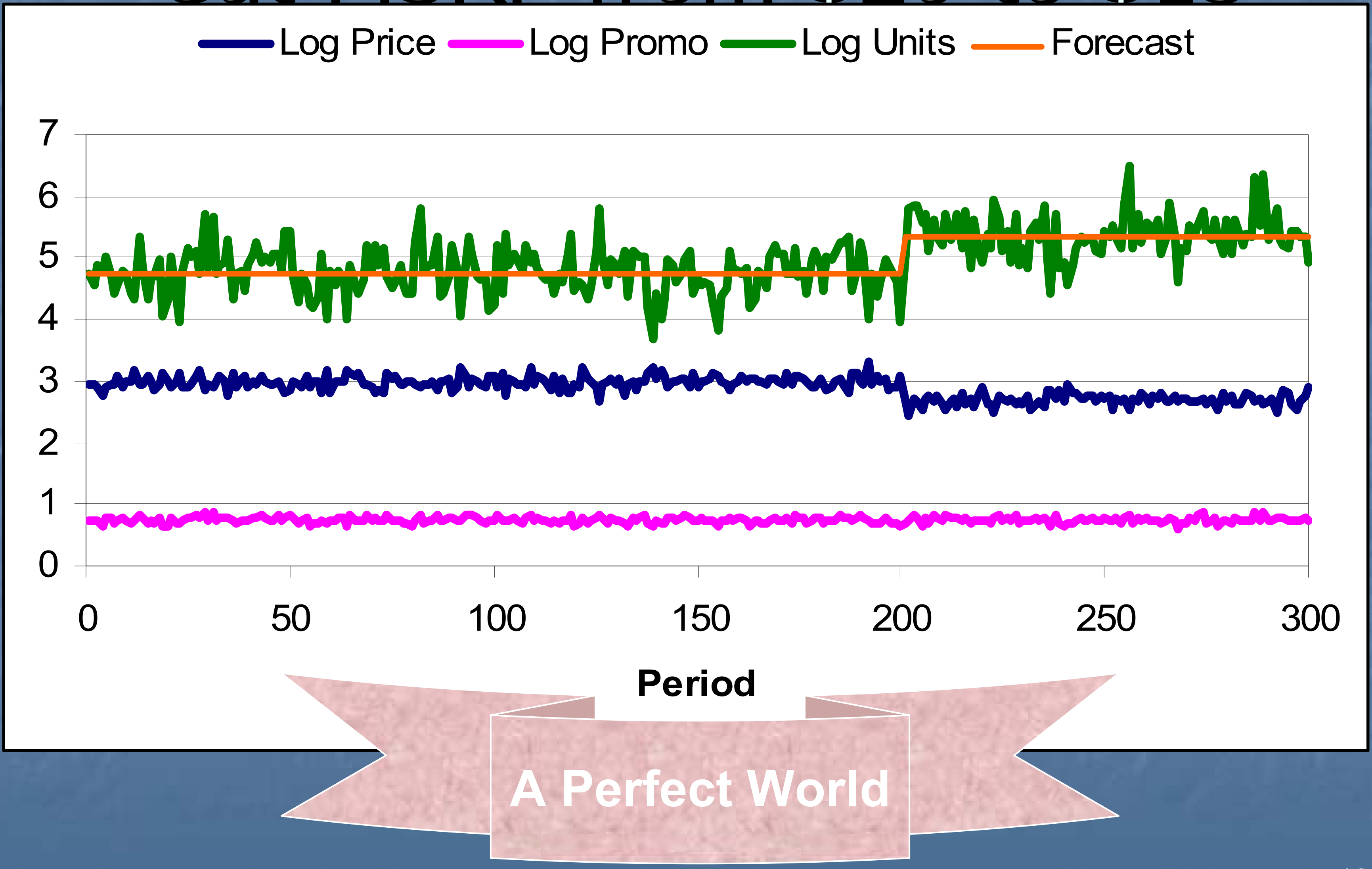
$$E\left[\frac{\partial Y}{\partial X_1}\right] = \beta_1$$

$$E\left[\Delta Y\right] = \beta_1 \Delta X_1$$

Estimates

- log(Units) = 6.163–2.015 log(Price)+5.584 log(Promo)
- Change MSRP from \$20 to \$15 or change log(Price) by -0.3
- Expected change in log(Units) is (2.015)(0.3) = 0.605

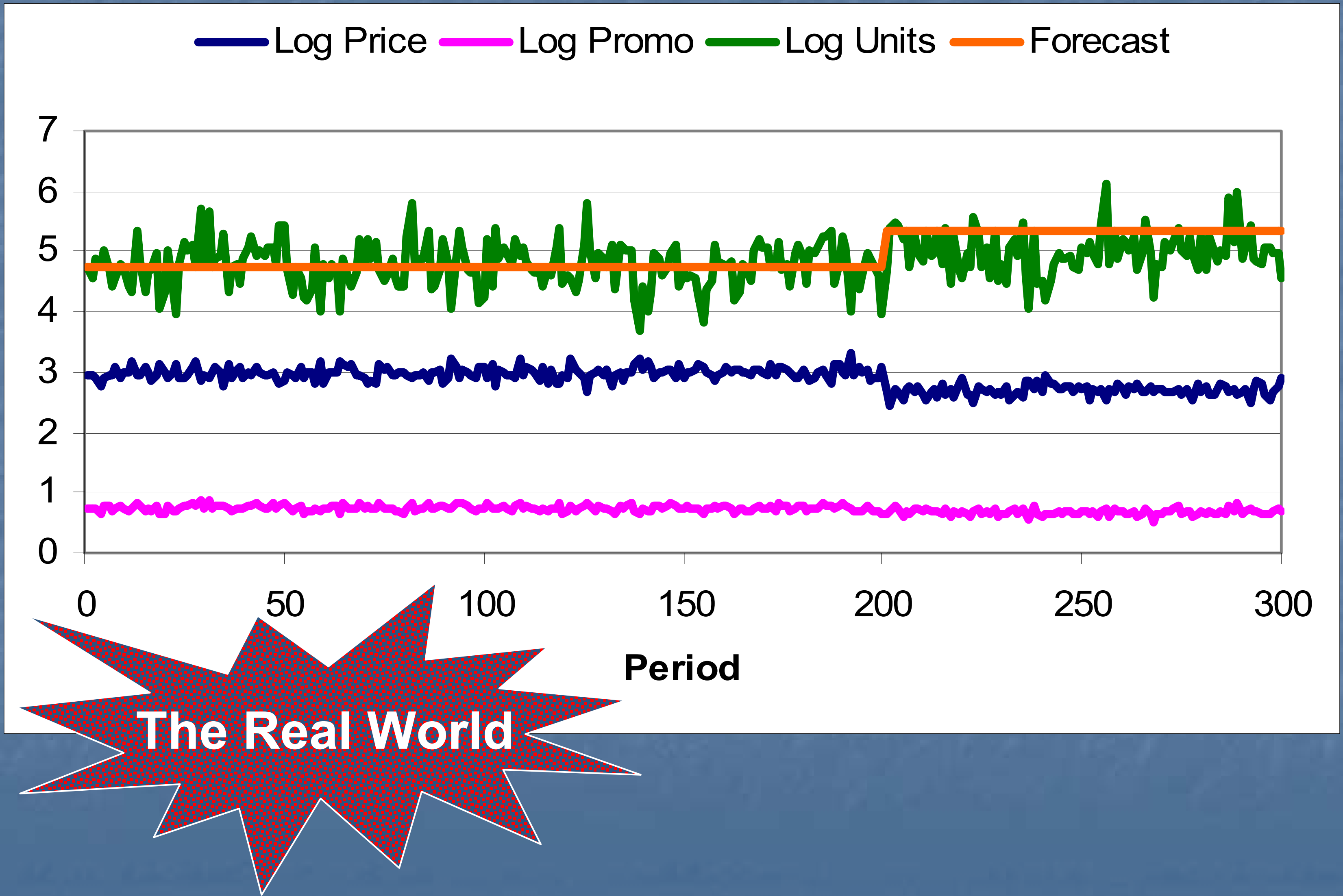
Cut MSRP from $20 to $15



Can Client Change X?

- Driver analysis may not mean much if you cannot change the “most important” X
- Assume manufacturer only has control of MSRP, but not actual price or promotional expenditures
- Retailer plays with price and promotional expenditures

Cut MSRP from $20 to $15



What Could Happen?

- Retailer sets promotional spending as a fraction of MSRP
  - Mean of log(Promo) = C*log(MSRP)
- Price reduction is offset by reduction in promotion spending

Do it With Math

Assume $X_2 = CX_1$

May not be zero

$$\frac{dY}{dX_1} = \frac{\partial Y}{\partial X_1} + \frac{\partial Y}{\partial X_2}\frac{\partial X_2}{\partial X_1}$$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 (C)(X_1)$$

$$\frac{dY}{dX_1} = \beta_1 + \beta_2 (C)$$

$$\Delta Y = (\beta_1 + \beta_2 C)\Delta X_1$$

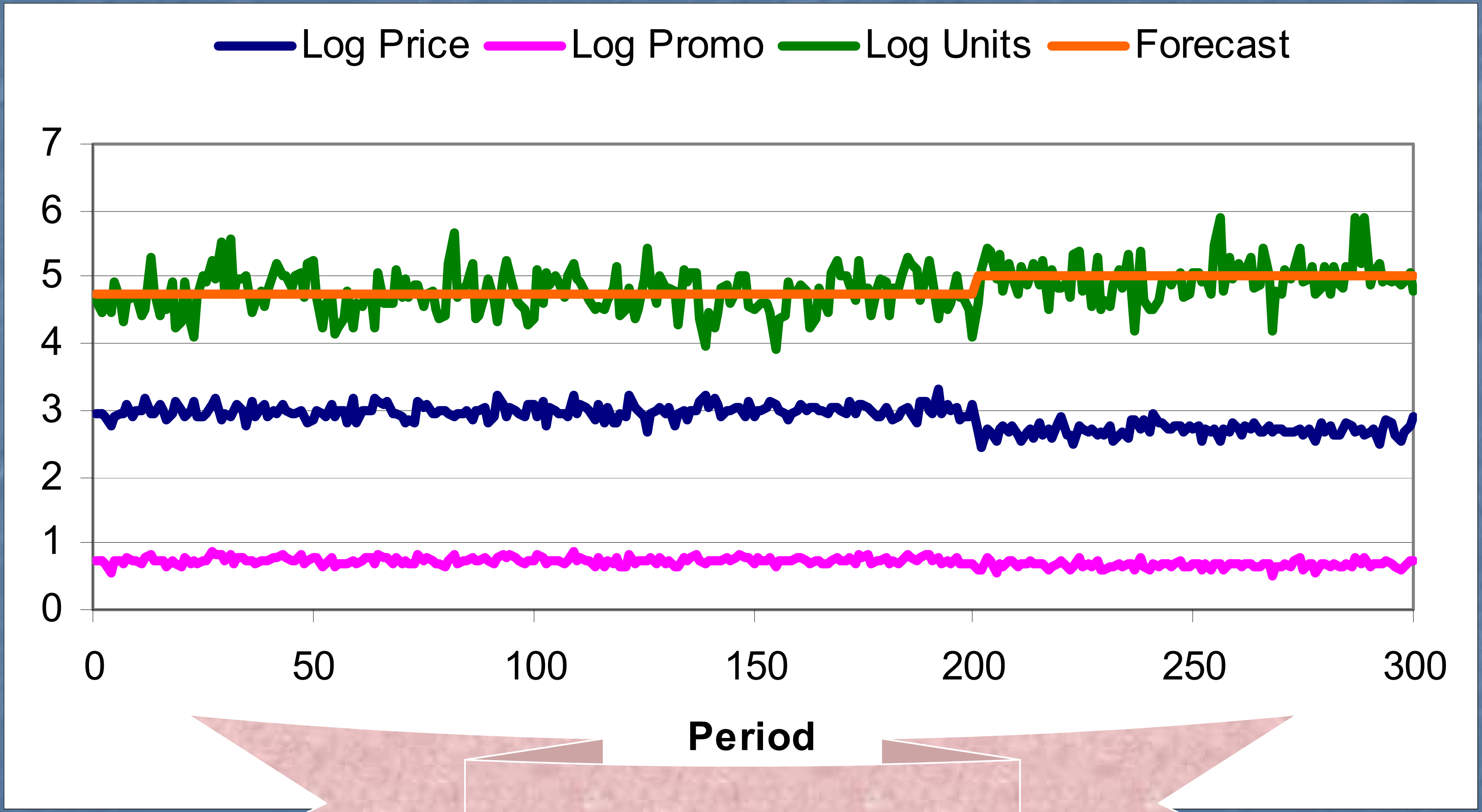
Structural Model

- $\text{Log(Promo)} = \alpha_0+\alpha_1\log(\text{Price})+\varepsilon_1$
- $\text{Log(Units)} = \beta_0+\beta_1\log(\text{Price})+\beta_2\log(\text{Promo})+\varepsilon_2$
- Expected change in Log(Units) for a Δ change in price is $(\beta_1 + \beta_2\alpha_1)\Delta$

Estimates

- log(Promo) = 0.061 + 0.226 log(Price)
- log(Units) = 6.163–2.015 log(Price)+5.584 log(Promo)
- Change MSRP 20 to 15, or change in log Price is -0.3
- Expected change in log Units is only [-2.015 +(5.584)(0.226)](-0.3) = **0.223**
- Compare to (2.015)(0.3) = **0.605**

Better Forecasts



Paradise Reclaimed

Endogenous Variables

- Two equations:
  - log(Promo) is a function of log(Price)
  - log(Units) is a function of log(Promo) and log(Price)
- Log(Promo) is endogenous if errors are correlated between equations.
  - Results in inconsistent estimates

Example

- Retailer plans promotional spending in anticipation of demand
  - Expectation of low demand results in higher promotional spending
  - Expectation of high demand results in lower promotional spending

Most Important Variable?

- Simulated model:
  - Log(Units) = 7 – 2*log(Price) + **5***log(Promo) + e
  - Corr(e,log(Promo)) = -0.86

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	9.906	0.256	38.702	0.000
Log Price	-2.012	0.074	-27.221	0.000
Log Promo	1.178	0.159	7.393	0.000

Next Steps

- Durbin-Wu-Hausman Test
- Some fixes
  - Instrumental variables
  - Two Stage Least Squares
- Good News
  - If you bill by the hour
- Bad News
  - If you have a fixed contract

Marketing Manager’s Importance

- “Econometric niceties may interest you, but they do not reflect the world I live in.”
- Competitive response
- Perceptions and attitudes
- Considerations outside scope of study
  - Organizational constraints
  - Institutional inertia
  - Time horizons

Competitive Response

- Huge academic literature
- "Strategic variables" instead of "Drivers"
- Fragile Models
  - Theoretical and empirical results are sensitive to model assumptions and initial conditions
- Overly simple models
  - Two competitors, one product, and homogeneous customers, rational actors

Attitudes and Perceptions

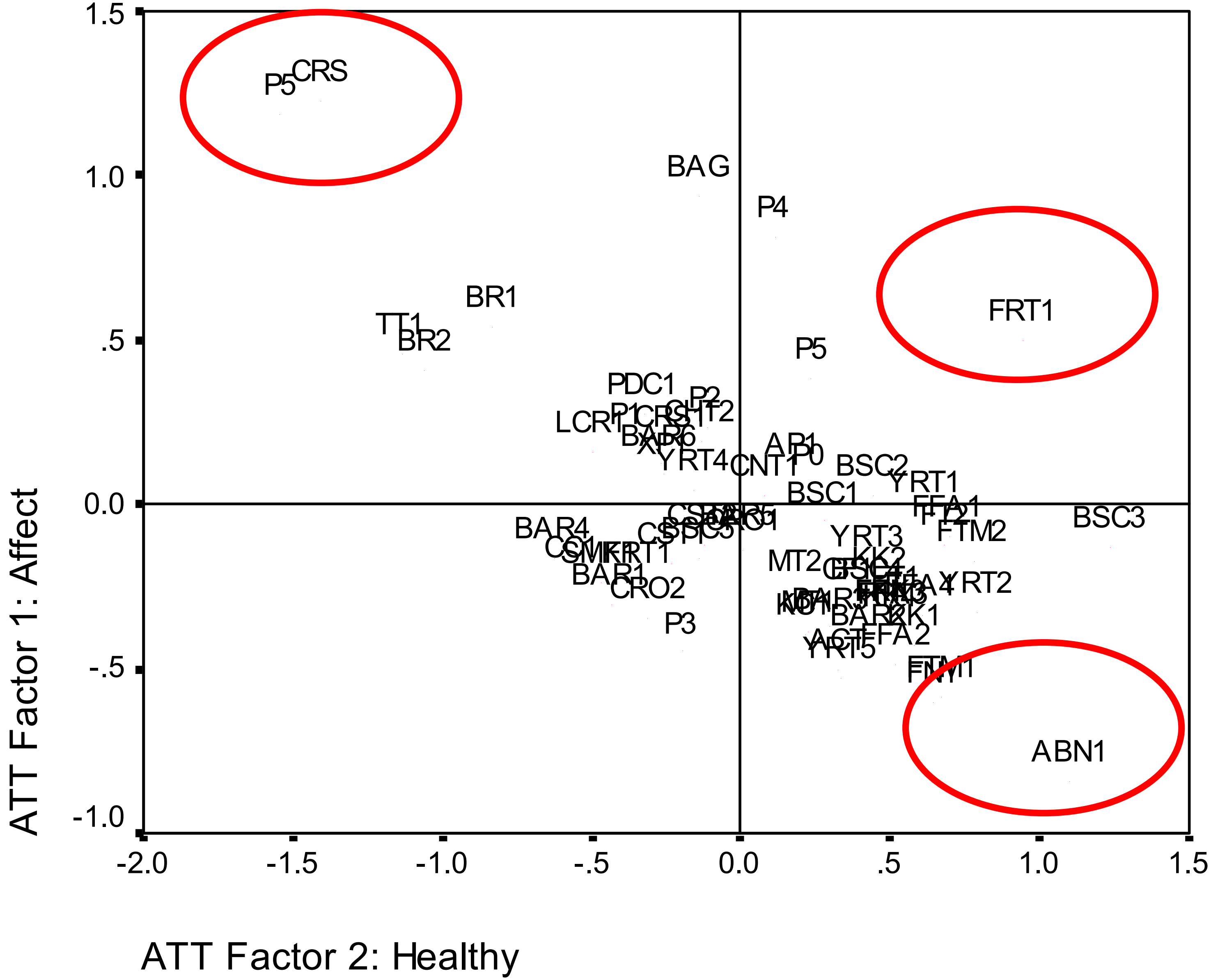
- 6000 subjects evaluated more than 100 attitudinal or perceptual items for a familiar product concepts in a meal category
- More than 50 product concepts
  - Branded FCPG
  - Generic foods (apples, bread, …)
- Each subject evaluated only one concept
- Stack data by subject

Perceptual Maps or Market Structure

- Discriminate analysis
- MDS
- Cluster analysis
- Factor analysis
- etc

Factor Analysis

- Four Factors
  - Affect
  - Health
  - Easy
  - Yummy
- Factor scores sum to 0 across products
- Mean factor scores within product do not sum to zero


P5CRS
BAG
P4
BR1
TT1
BR2
FRT1
P5
PDC1
LCRT
YRT4
CNT1
BSC2
YRT1
BSC1
BSC3
BAR4
FTM2
YRT3
MT2
YRT2
BAR1
CRO2
P3
FFA2
YRT5
ABN1
1.5
1.0
.5
0.0
-.5
-1.0
-2.0
-1.5
-1.0
-.5
0.0
.5
1.0
1.5
ATT Factor 1: Affect
ATT Factor 2: Healthy

Importance?

- Plethora of “internal measures”
  - Chi-squared statistics
  - % variance or eignevalues
  - Cornbach’s alpha
  - Stress measures

Client's Perspective

- Understanding market structure is good
- What can Monica do to achieve her objectives?
- Maps lack dependent variable

Include Behavioral Variables

- Behavioral items
  - Intention to buy
  - Future frequency of purchases
- Fancy models
  - Structural equations
  - PLS models
- Rough and ready
  - CFA with regression of factor scores

Regress ITB on Attitude Factors

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.656[a]	.430	.429	.75535

a. Predictors: (Constant), ATT Factor 4: Yummy, ATT Factor 3: Easy, ATT Factor 2: Healthy, ATT Factor 1: Affect

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	3.161E-15	.010		.000	1.000
	ATT Factor 1: Affect	.592	.010	.592	60.237	.000
	ATT Factor 2: Healthy	.235	.010	.235	23.929	.000
	ATT Factor 3: Easy	.111	.010	.111	11.315	.000
	ATT Factor 4: Yummy	-.106	.010	-.106	-10.787	.000

a. Dependent Variable: ITB Factor

The Experiment not Performed

- Study explores cross sectional correlations of items among subjects
- Long & tenuous causality chain
  - Change in X produces change in Affect that creates change in ITB that leads to more sales.
- Study did not manipulate X
  - Grumpy Gus
  - Perky Pat

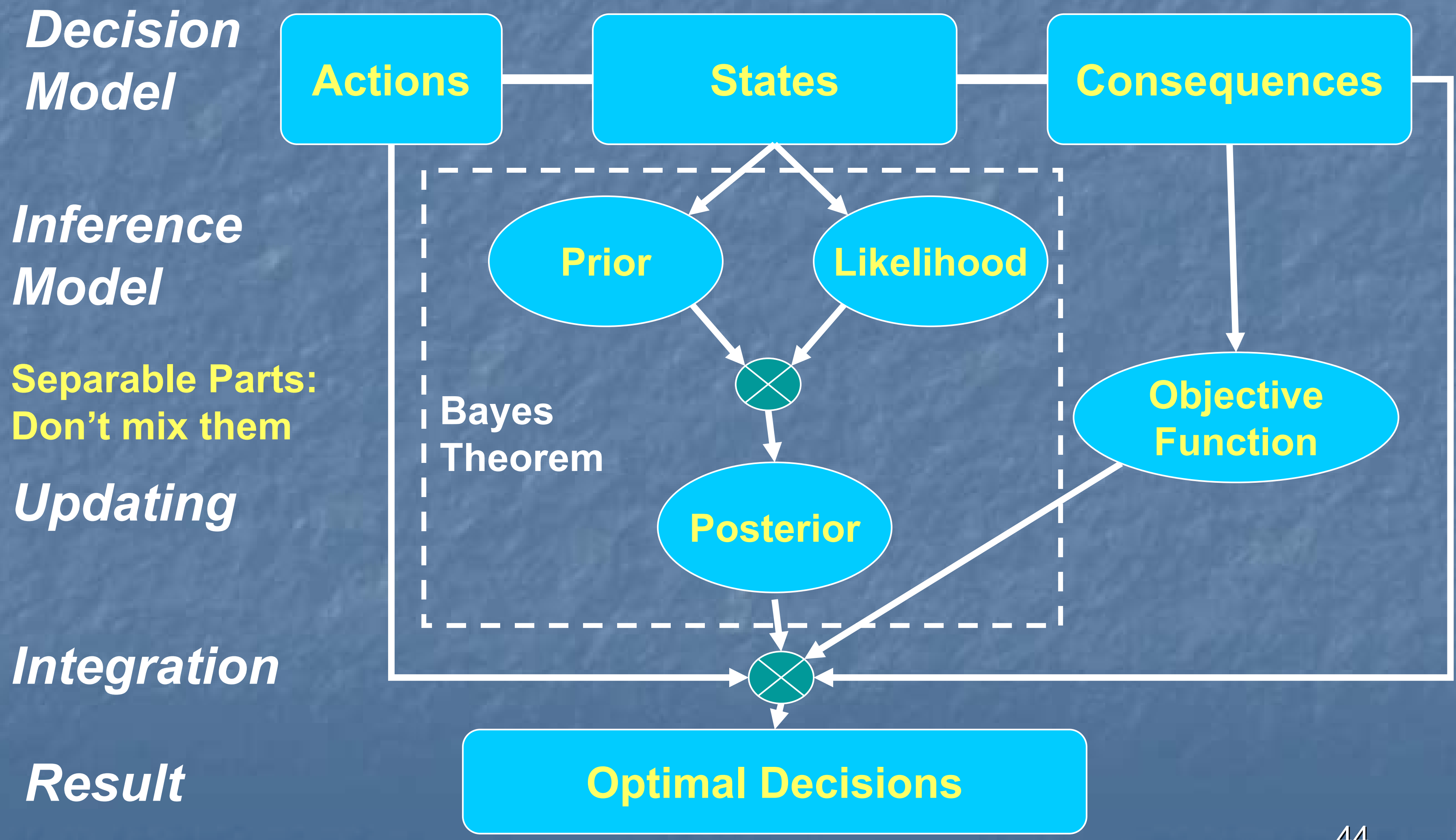
Tracking Studies

- Correlate scores to marketing activity
- May be infeasible to use a panel
- Need to connect study outcome – satisfaction, loyalty, ITB, brand image, … - to business goals
- Adjust for econometric anomalies

Unified Framework

- Bayesian Decision Theory – the ***Real BDT***
- Bayes models can combine statistical estimation with decision making
- Merges statistical and managerial importance

Bayesian Decision Model



Simply Bayes:
Estimating a Mean

- $Y_i = \mu + \varepsilon_i$
- Error terms $\{\varepsilon_i\}$ are iid normal
  - Mean is zero
  - Standard deviation of error terms is σ.
  - *Assume that σ is known*
- Prior distribution for μ is normal
  - Prior mean is $\mathbf{m_0}$
  - Prior variance is $\mathbf{v_0}^2$

Posterior Distribution

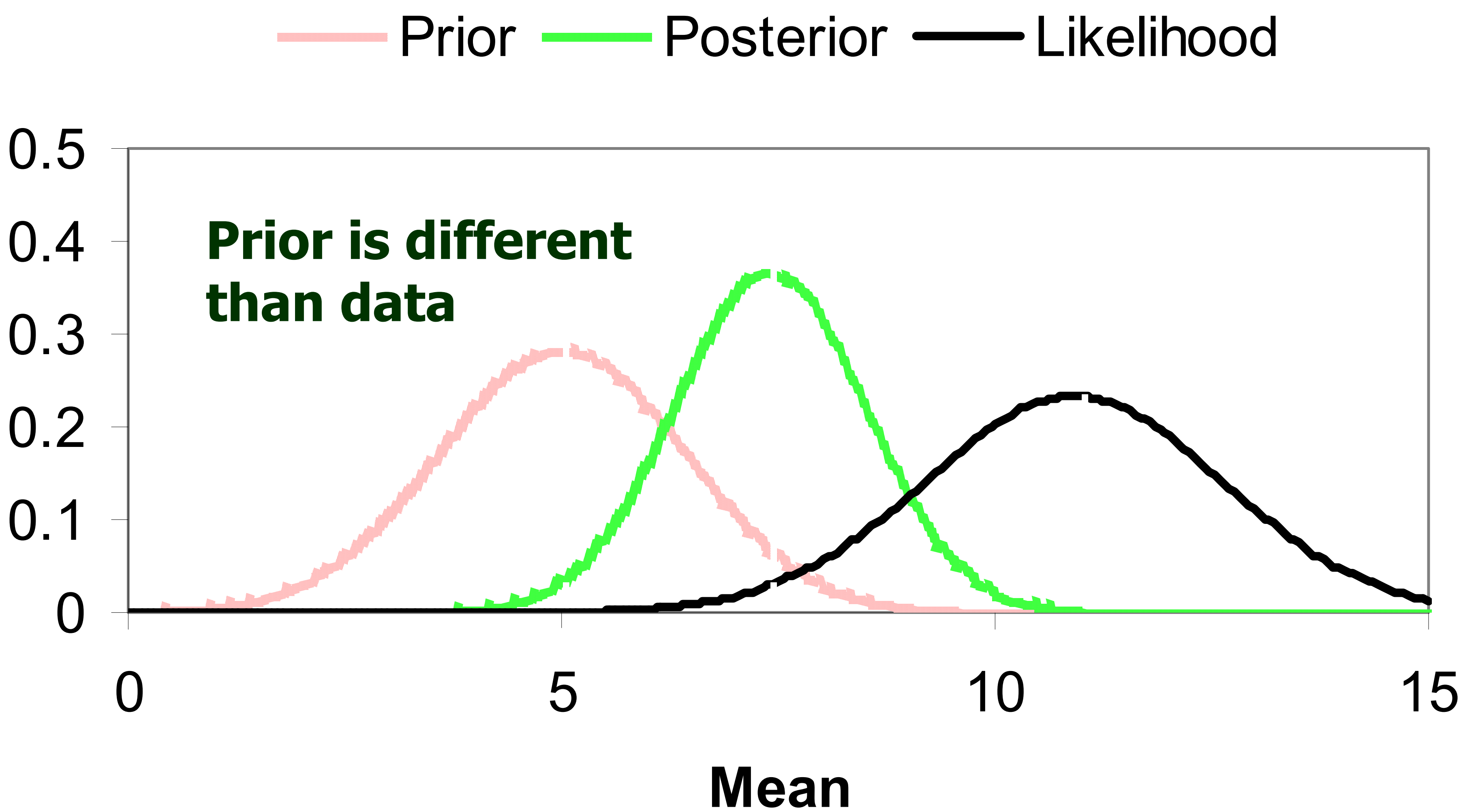
- Observe n data points
- Posterior distribution is normal
  - Mean is m_n
  - Variance is v_n^2
- Posterior mean shrinks sample mean towards prior mean

$$m_n = w\bar{y} + (1-w)m_0$$

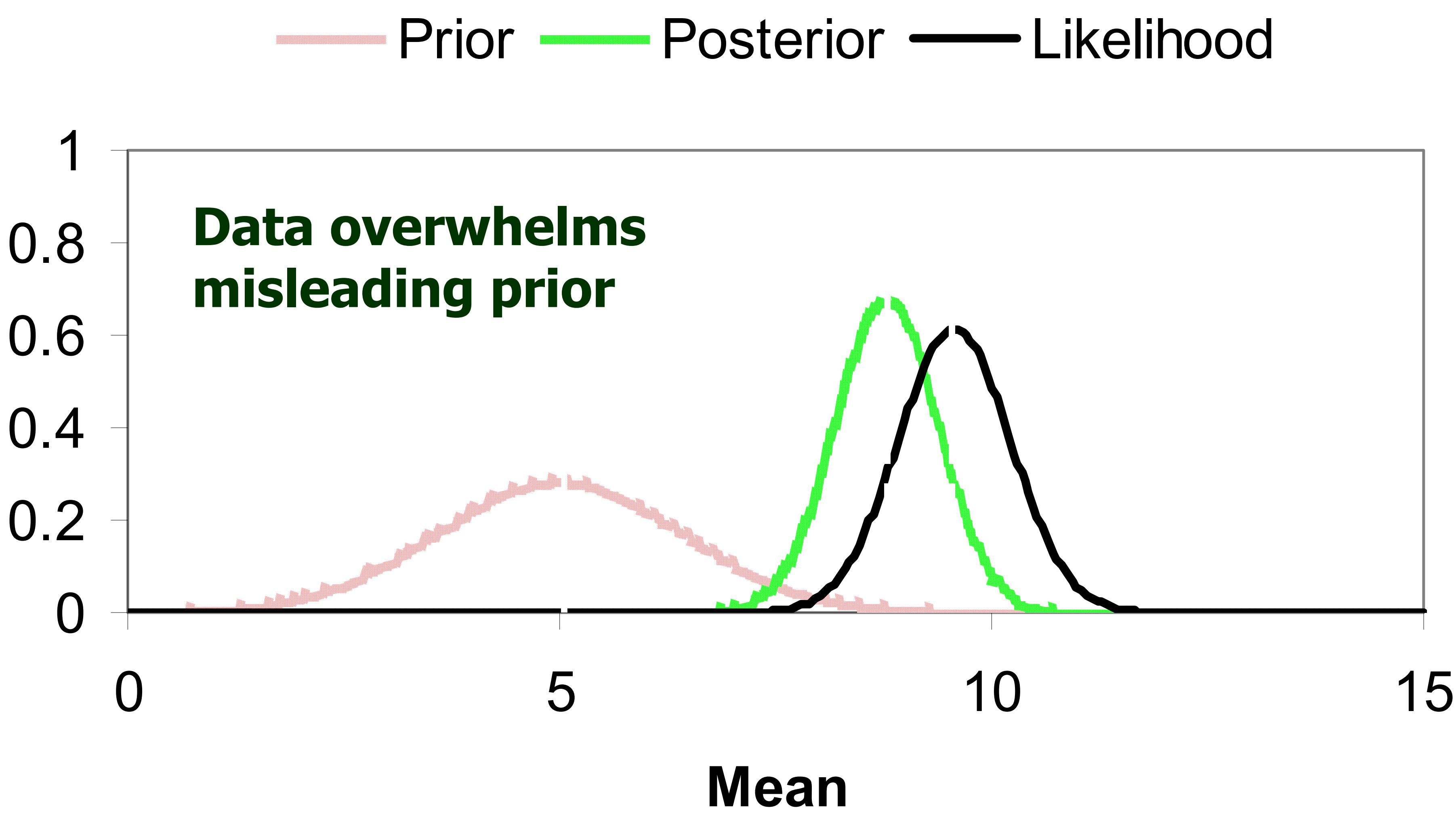
$$w = \frac{\frac{n}{\sigma^2}}{\frac{n}{\sigma^2} + \frac{1}{v_0^2}} \text{ and } 0 < w < 1$$

$$v_n^2 = \frac{1}{\frac{n}{\sigma^2} + \frac{1}{v_0^2}}$$

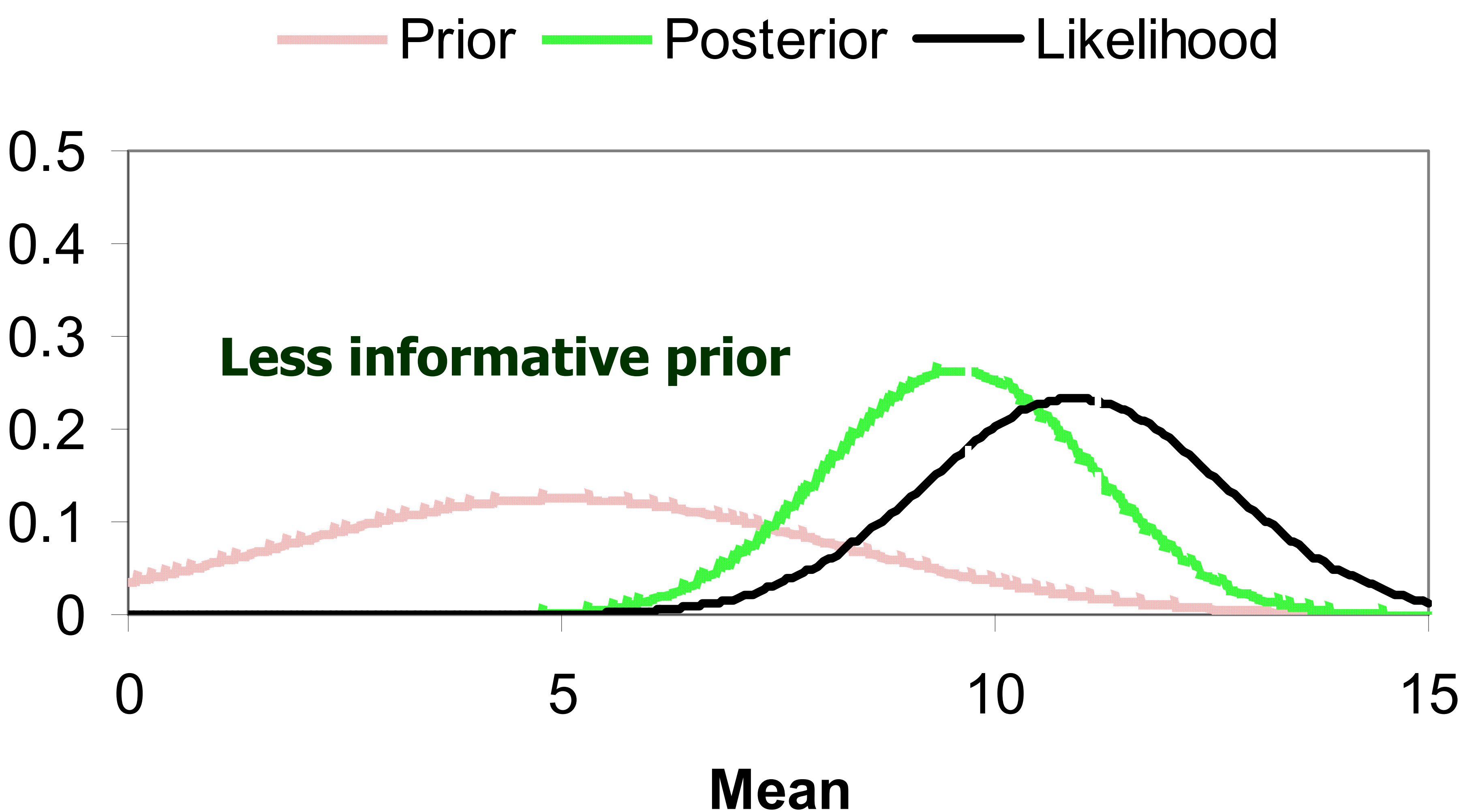
Prior & Posterior n=5



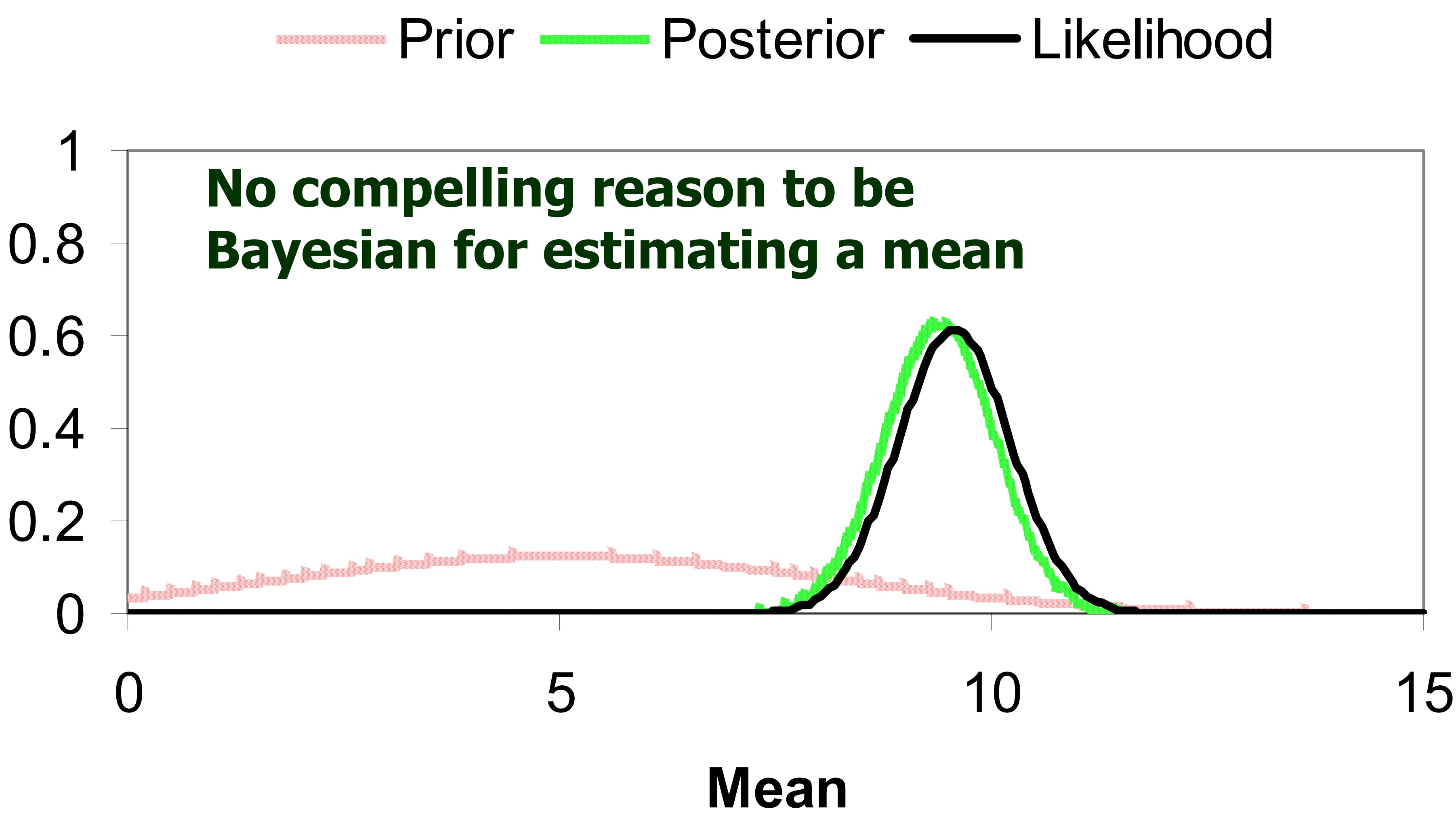
Prior & Posterior n=50



Prior & Posterior n=5



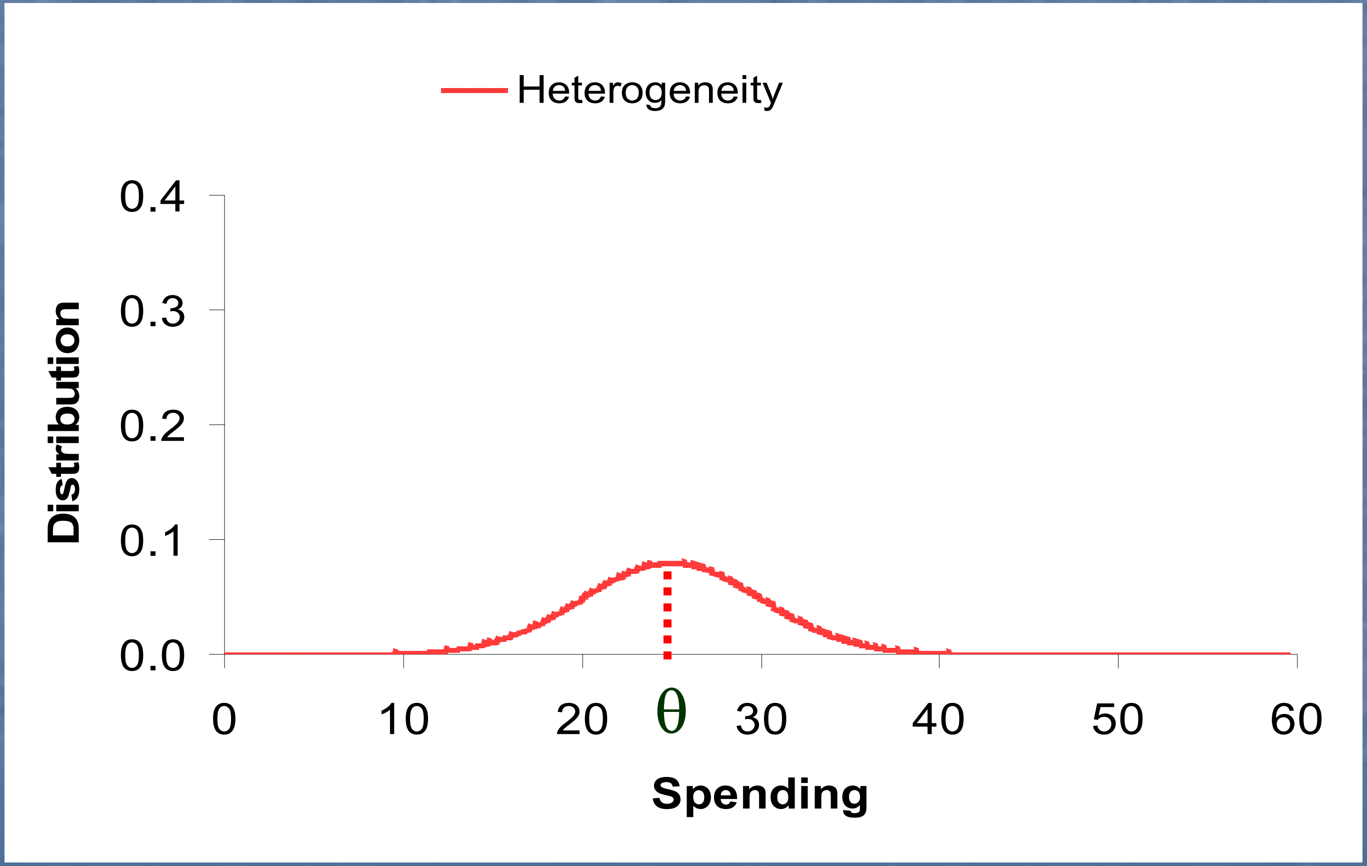
Prior & Posterior n=50



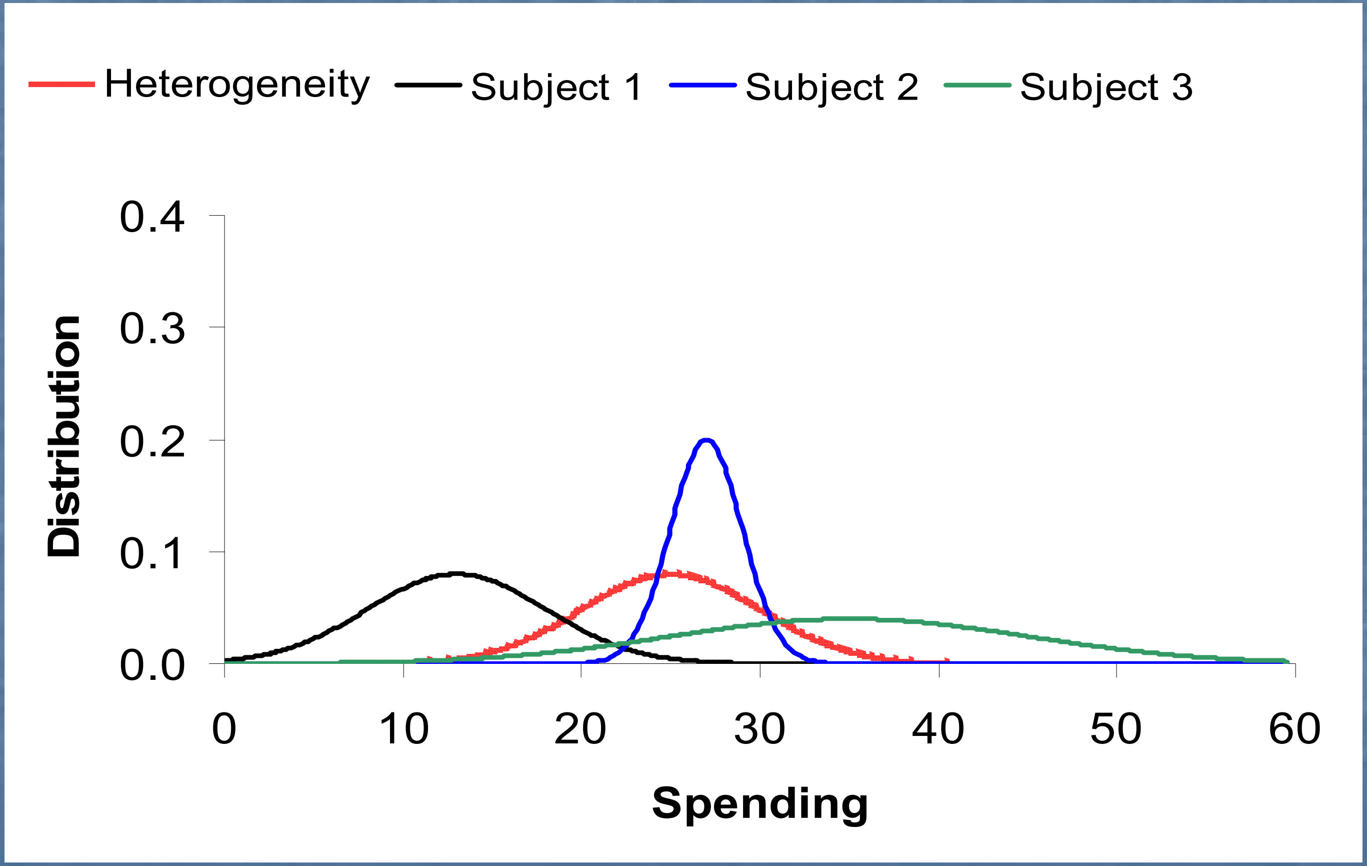
HB Model for Weekly Spending

- Within-subjects or subject-level model
  - $Y_{i,j} = \mu_i + \varepsilon_{i,j}$ for subject i and week j
  - Mean for household i is μ_i & $\varepsilon_{i,j}$ is error
- Between-subjects or heterogeneity in household means
  - $\mu_i = \theta + \delta_i$
  - θ is population mean and δ_i is random error

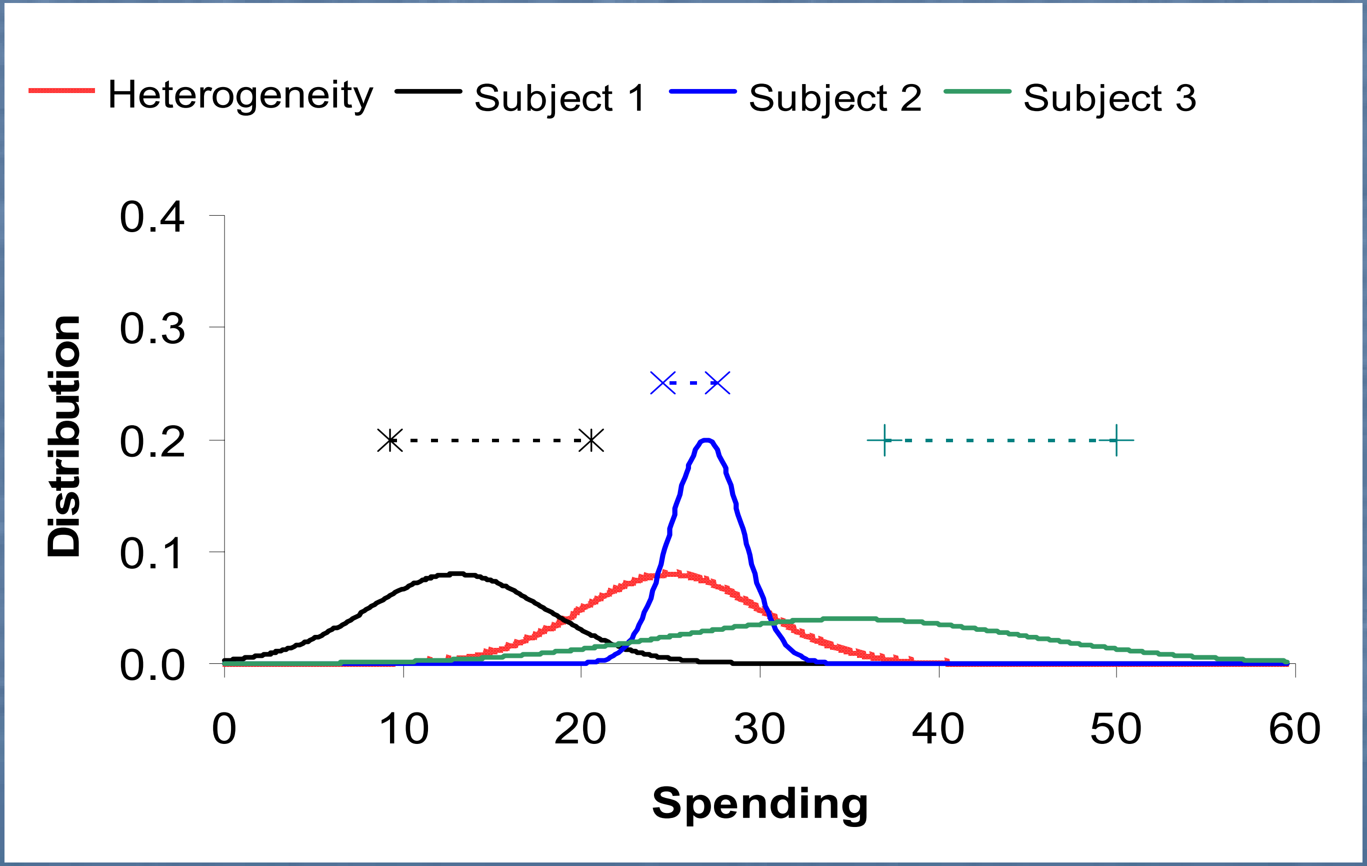
Between-Subject Heterogeneity in Mean Household Spending



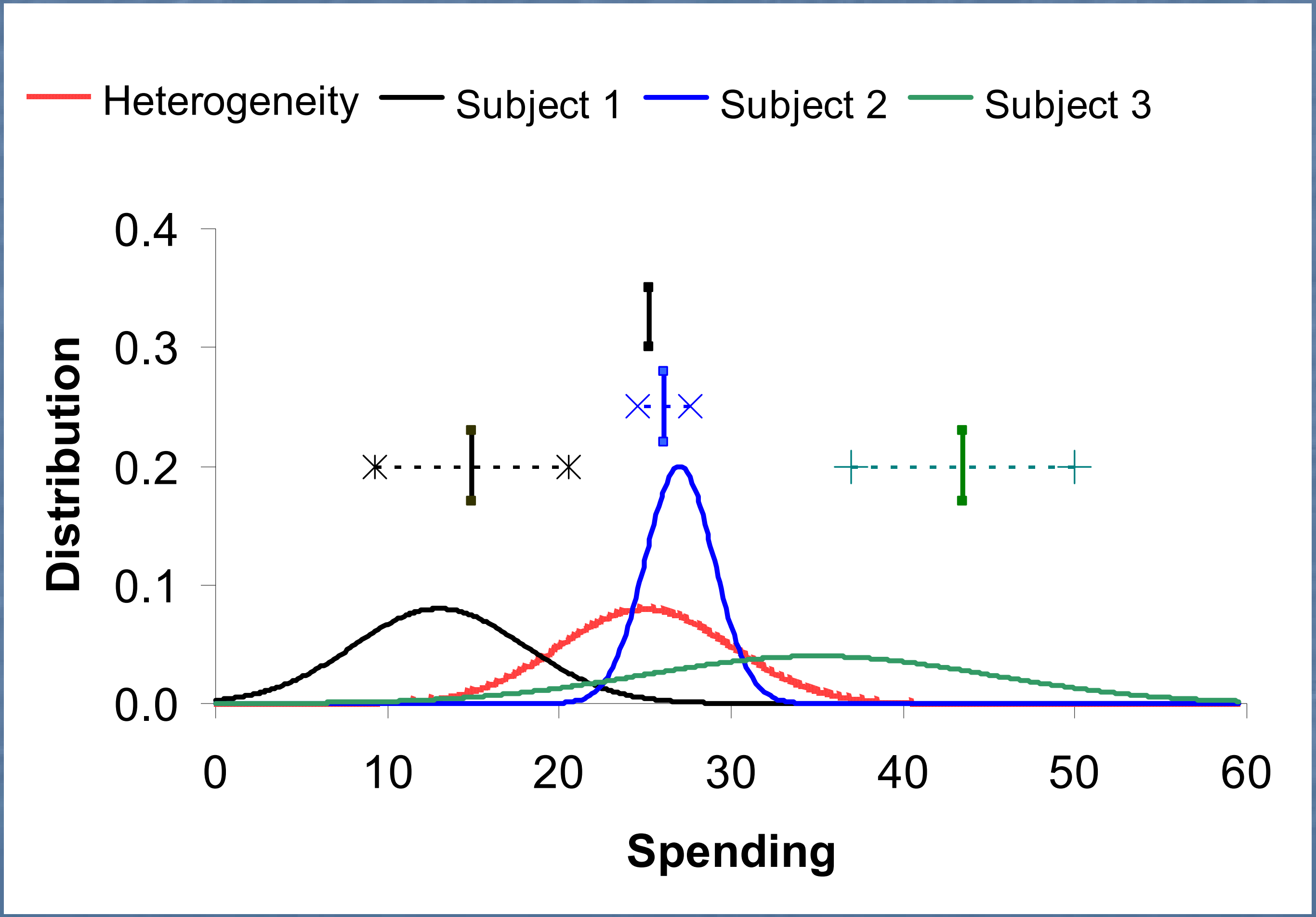
Between & Within Subjects Distributions



2 Observations per Subject



Pooled Estimate of Mean



HB Shrinkage Estimator

- Combines individual average and pooled average

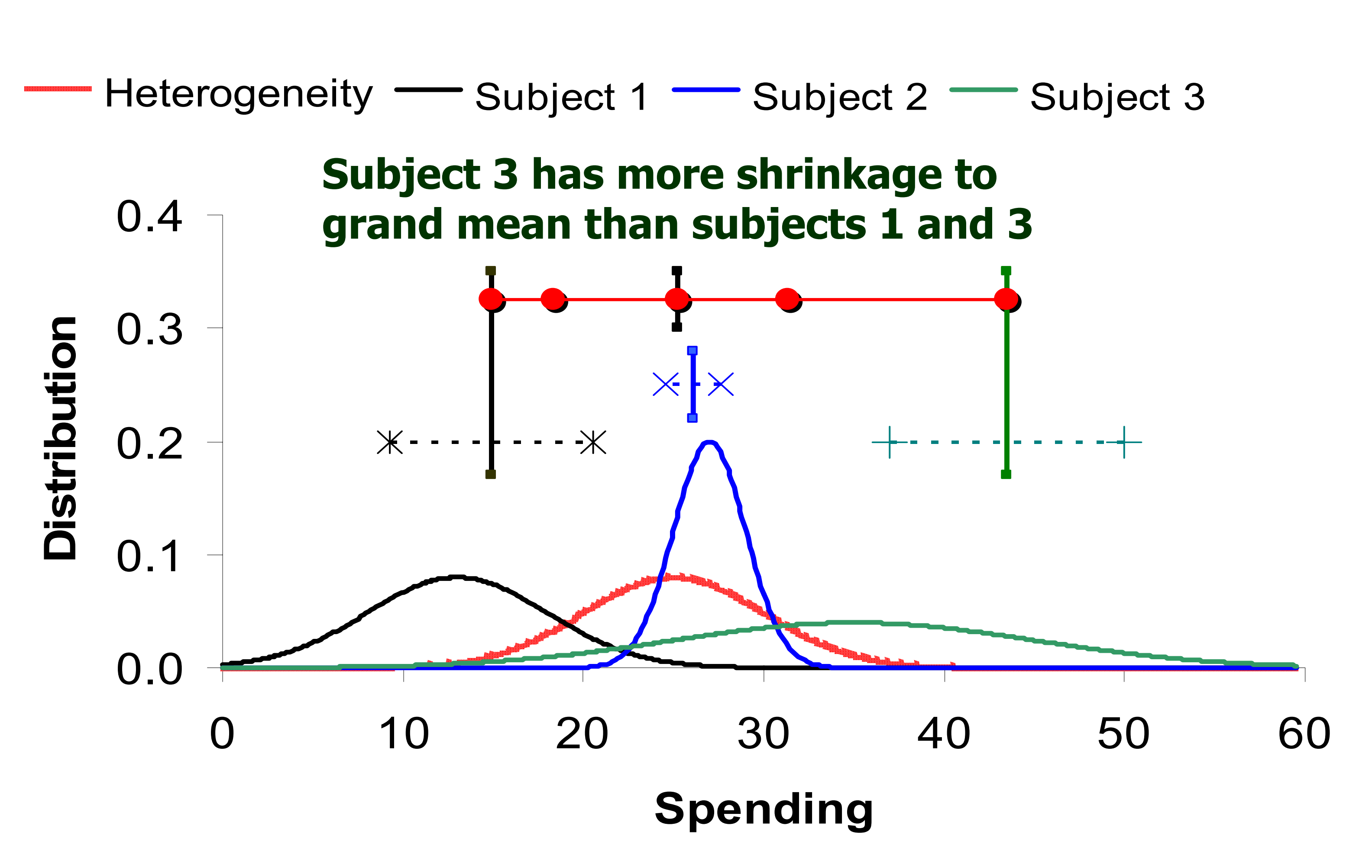
$$w_i \overline{Y}_i + (1 - w_i)\overline{\overline{Y}}$$

- HB automatically gives optimal weights based on
  - Prior variance of μ_i
  - Number of observations for subject i
  - Variance of past spending for subject i
  - Number of subjects
  - Amount of heterogeneity in household means

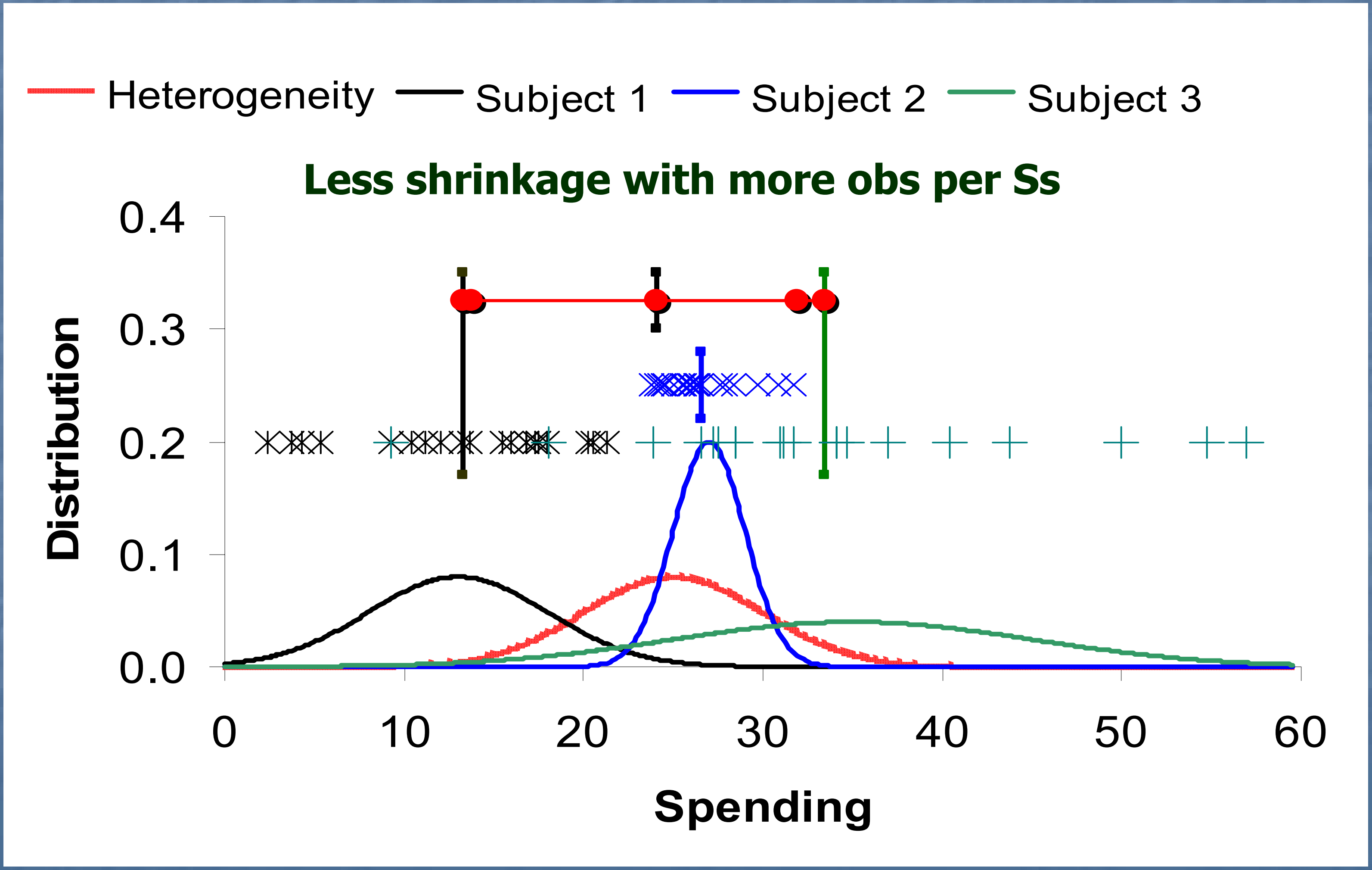
Real benefit if data are broad and shallow

(Many Ss & few obs/Ss)

Shrinkage Estimates
Two observations per subject



Shrinkage Estimates
20 Observations per Subject



Bayes & Shrinkage Estimates

- Automatically determine optimal shrinkage
- Minimizes MSE
- Borrows strength from all subjects
- Tradeoff some bias for variance reduction

Posterior Expectations by Monte Carlo

- Compute posterior mean of function T(θ).

$$E[T(\theta) \mid y] = \int T(\theta) p(\theta \mid y) d\theta$$

- Generate random draws θ_1, θ_2, …, θ_m from posterior distribution using a random number generator.

$$E[T(\theta) \mid y] \approx \frac{1}{m} \sum_{j=1}^{m} T(\theta_j)$$

Good & Bad News

- If your computer has a random number generator for the posterior distribution, Monte Carlo is a snap to do.
- Your computer almost never has the correct random number generator.
- Markov chain Monte Carlo (MCMC) and Metropolis algorithms get the job done
  - *if your client can wait*

Example

- Metric (ratings) conjoint experiment
- 179 subjects
- 16 personal computer profiles
- 13 binary attributes
- 7 subject-level covariates
- Y = likelihood of purchasing computer described by profile on 0 to 10 scale.
  - Lenk, DeSarbo, Green, and Young (1996)

Model

- Within subject i
- $$Y_i = X_i\beta_i + \varepsilon_i$$
- Between subjects: parameter heterogeneity
- $$\beta_i = \Theta z_i + \delta_i$$

Estimated Θ

	Constant	FEMALE	YEARS	OWN	NERD	APPLY	EXPERT
Constant	3.719		-0.115				0.175
Hot Line		0.233					
RAM	0.514				0.164	0.046	-0.064
Big Screen							
Fast CPU							0.060
Hard Disk		-0.160					
Multimedia	0.580						
Cache							0.047
Black	0.302						
Retail			0.021				-0.030
Warranty		0.147	0.024				
Software	0.322		-0.032				
Guarantee			0.024				
High Price	-1.522	0.386					

Displayed posterior means are bigger than two posterior standard deviations

Important Variables

- So, screen size is not an important factor?
- Not so fast, this is an HB model.
  - Θ is the mean of the heterogeneity in partworths.
  - A zero θ only means that the distribution of heterogeneity is at zero
  - Some people like big screens, and others don't.
  - You need $\theta=0$ **and** var(δ) very small.

Variable Selection in OLS

- OLS is fragile
- Need to be circumspect when adding a variable because bad things can happen
  - Degrees of freedom
  - Lack of model fit
  - Outliers
  - Multicollinearity
  - Endogeniety

Bayes Model Selection

- Pick model to maximize utility – same as generic discriminate analysis
  - Posterior probability of model given data if miss-classification costs are equal
  - Bayes factors if prior probabilities of models are equal
- Ad hoc procedures
  - Posterior Means/ Posterior STD DEV

Bayes World is Different

- Variable selection is not as important as in OLS
- Jimmy Savage said, "Use models as big as an elephant."
- Prior for coefficients helps to ameliorate adverse affects of adding non-significant variables
  - Set prior mean to zero
  - Set prior std dev to reflect problem
  - Shrinkage estimator (ridge regression)

Better Yet

- Work with the marketing manager to answer her real question
  - Objective function f
  - f = f(Ω,X) where Ω are model parameters
  - Example: f is choice share, and X are product features, and Ω are partworth heterogeneity
- Pick $X^{\#}$ to optimize f

Bad: Plug-in Estimators

- Estimate Ω with Ω^
- Plug Ω^ into objective function: $f(X, \Omega\hat{})$
- Find $X^{\#}$ to maximize $f(X, \Omega\hat{})$

  $X^{\#} = \arg\max f(X, \Omega\hat{})$
- Works if f is linear or nearly linear
- Does not account for the uncertainty in Ω.
- Results are too "sharp"

Bayes it Up

- In-line optimization
- Generate $\Omega_1, \ldots, \Omega_M$ from the posterior distribution (via MCMC?)
- Find $X_m^\#$ to maximize $f(X,\Omega_m)$ for each of the simulated values
- Explore posterior distribution of $X^\#$ by means of $X_1^\#, \ldots, X_M^\#$
  - Means std devs, histograms, …

BDT

- MCMC Function Optimizer Optima
- Ω_1 $f(X,\Omega_1)$ $X_1^{\#}$ $f(X_1^{\#},\Omega_1)$
- Ω_2 $f(X,\Omega_2)$ $X_2^{\#}$ $f(X_2^{\#},\Omega_2)$
- Ω_3 $f(X,\Omega_3)$ $X_3^{\#}$ $f(X_3^{\#},\Omega_3)$
- …
- Ω_M $f(X,\Omega_M)$ $X_M^{\#}$ $f(X_M^{\#},\Omega_M)$
- Posterior distributions

Example: Choice Based Conjoint

- Random Utility Model
  Subject's i utility for Brand j is
  $U_{i,j} = \beta_{i,0} + \beta_{i,1}x_{1,j} + \ldots + \beta_{i,p}x_{p,j} + \varepsilon_{i,j}$
- Error term is multivariate normal (probit)
- X's are product attributes
- Pick brand j if $U_{i,j}$ is maximum
- Heterogeneity: $\beta_i = \Theta z_i + \delta_i$

Example

- Data provided by Sawtooth Software
- Joint work with Robert Zeithammer
- 326 IT purchasing manager
- 5 brands of personal computers
- 8 choice tasks per subject
- 4 alternatives per choice task
  - 3 brands and "None"

Utility Covariance Matrix

	BrandA	BrandB	BrandC	BrandD	BrandE
BrandA	1.02	0.01	-0.12	-0.16	-0.41
BrandB	0.01	0.95	0.08	-0.13	-0.45
BrandC	-0.12	0.08	1.18	-0.44	-0.53
BrandD	-0.16	-0.13	-0.44	1.21	-0.08
BrandE	-0.41	-0.45	-0.53	-0.08	1.00

If subject likes Brand A more than expected, he or she will like Brands D and E less than expected

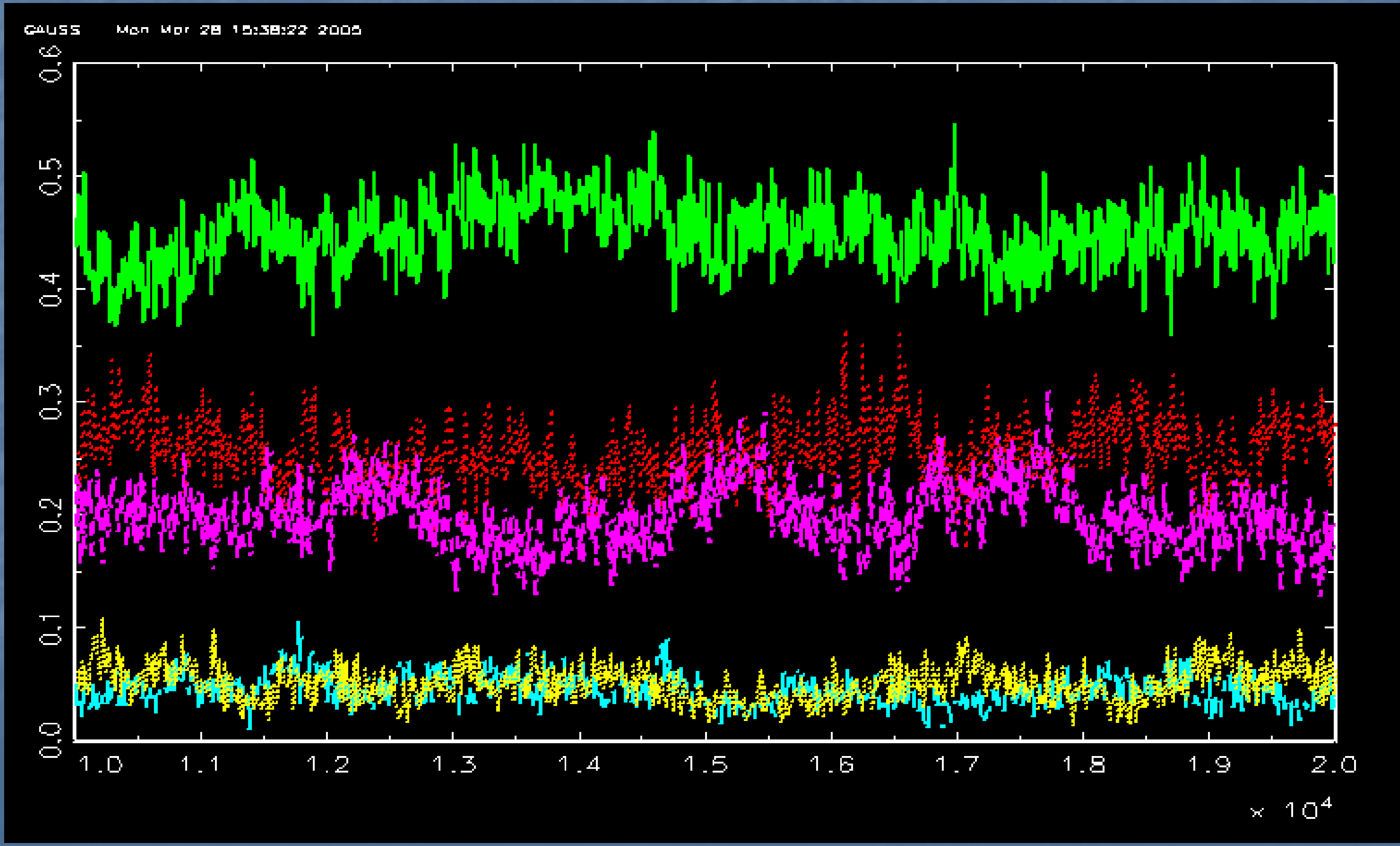
Posterior Mean of Θ

	CNST	ExPayLow	ExPayHig	Expert	Female	SmallCo	LargeCo
BrandA	0.768				-0.421		0.302
BrandB	0.882		-0.382		-0.406		
BrandC	0.459	0.455	-0.458		-0.471		
BrandD	0.400		-0.584		-0.544		
BrandE			-0.597	-0.354	-0.691		
LowPerfo	-1.574						-0.326
HighPerf	0.566		0.267			0.371	
TeleBuy	-0.192	0.231					
SiteBuy		0.328					
ShortWar							
LongWar	0.401						
MFGFix	-0.679	-0.399					
SiteFix	0.342						
Price2	0.315			-0.291			-0.291
Price3	-0.723	-0.296					
Price4	-0.977	-0.661			0.287		

Simulated Market Share

- Fix 5 product specifications
- During each iteration
  - Generate subject's latent utility for each product
  - Pick the product with maximum utility
  - Compute market share
- Distribution of market shares

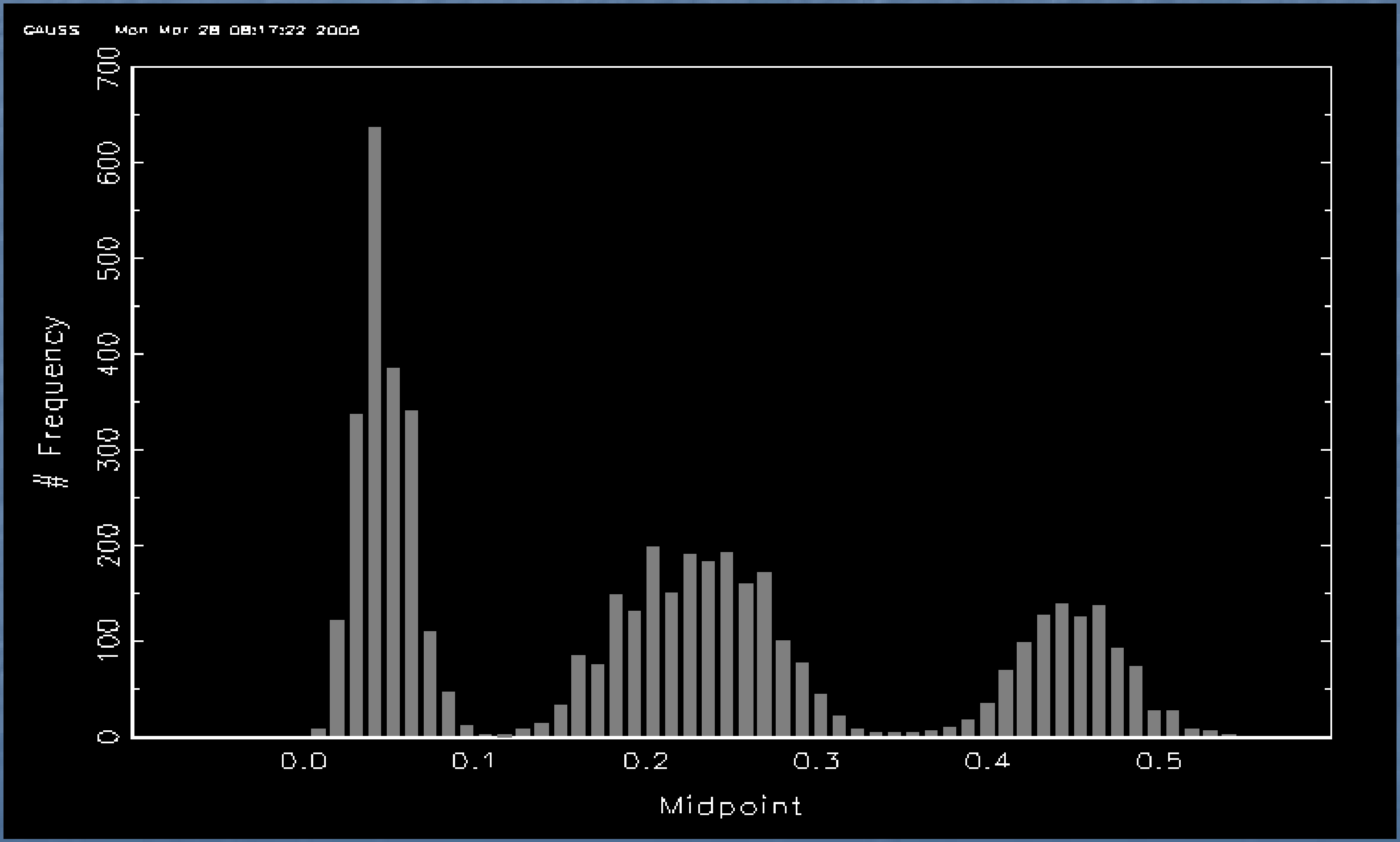
Iteration Plots of Market Shares



Posterior Means and STD DEV

Brand	Mean	STD DEV
A	0.45	0.030
B	0.04	0.014
C	0.26	0.028
D	0.20	0.029
E	0.05	0.015

Histogram of Posterior Distribution



Important Variables: Sensitivity Analysis

- If client is currently at X_0
  - $f(X_0,\Omega_m)$ for $m = 1, \ldots, M$
- Change components of X_0
  - $f(X_0+\Delta X,\Omega_m)$ for $m = 1, \ldots, M$
- Base importance on
  - $\Delta f = f(X_0+\Delta X,\Omega_m) - f(X_0,\Omega_m)$ for $m = 1, \ldots, M$

Conclusion

- Managerial importance is different from statistical importance
- Bayesian decision theory provides a unified framework to account for statistical uncertainty in managerial meaning of "importance"