Total Error in Surveys, Forecasts, and Randomized Social Experiments: Modeling Approaches

Bruce D. Spencer
Northwestern University

Workshop on Total Survey Error
March 17-18, 2005
Washington, D.C.

Key Points

- Total survey error part of larger picture
- Modeling essential
- Total survey error not enough
- Need to identify & model error components
- Past total error modeling
  - Sometimes inaccurate numerical estimates
  - Useful nonetheless
- Work on total error work should expand

Key Points (ctd.)

- Randomized Social Experiments and Clinical Trials
  - Increasingly important
  - Combine aspects of surveys and forecasting
  - Total error modeling appropriate

Total Error
not just total survey error

$$\text{Total Error} = \underbrace{\hat{Y} - Y}_{\text{survey error}} + \underbrace{Y - Y^{*}}_{\text{invalidity error}}$$

$Y–Y^*$ error due to invalidity

Examples –

- Operational definition
  - Census money income -BEA personal income harder in surveys
  - Unemployment measures
  - Achievement tests – w/ loose links to educational goals/teaching
- Imperfect choice of descriptive statistic
  - Use mean where median should be used (test scores?)
  - Rubin (2005) critique of Fisher's ANCOVA for analysis of randomized experiments
  - Model misspecification
- Validity connects statistic to use
- Statisticians should not ignore uses or validity issues

Total (Survey) Error Model

- ". . . a decomposition of the total error into pieces or components that can be estimated or at least bounded.
- The decomposition is an algebraic identity, possibly derived under simplifying assumptions
- and if the means and variances and covariances of the components can be estimated, the mean and variance of the total error can be approximated."

  - Alho and Spencer (2005) *Statistical Demography and Forecasting.* New York: Springer, to appear.

Not Just for Surveys

- Forecasts
- Dual Systems Estimator (DSE)
- Randomized social experiments
- Physical constants

Ways to Estimate Total (Survey) Error

1. Compare to standard
2. Analyze replications
3. Decompose into pieces
   - Estimate component errors
   - Combine

Each involves modeling.

Population Forecasts

Projections: high, medium, low

Forecast interval

$$(P_{low}, \quad P_{high})$$

Forecast is medium projection

Interval from deterministic scenarios

What is interval's coverage probability?

- Compare past forecasts with realizations
- Replications
  - Years
  - Countries
- Probability model for error in forecast
  - Rate of increase (Keyfitz 1981)
  - Short term and long term components (Stoto 1983)

Beyond *total* population

Subgroups

Age-dependency ratio

Fiscal forecasts

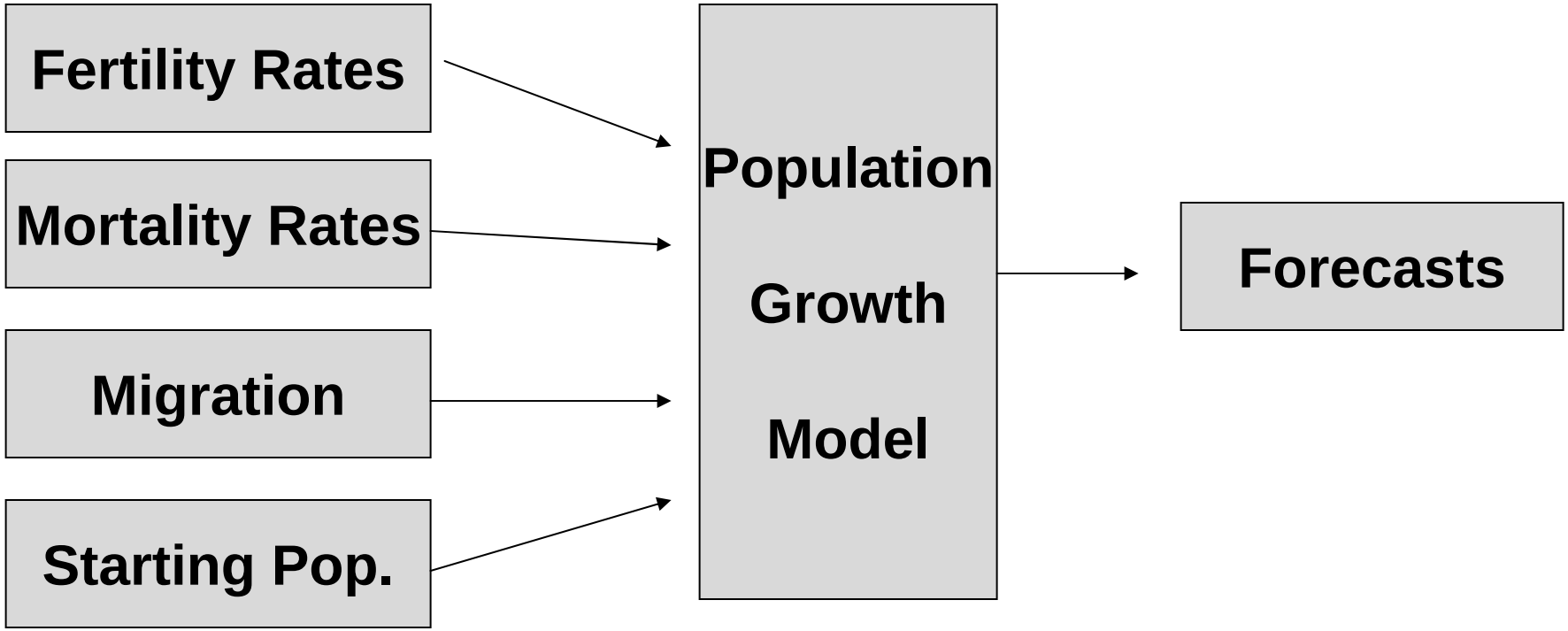
- Social Security
- other

Forecast-Error Components

Hoem (1973):

- data error
- estimation error (in past vital rates)
- erroneous trends in the mean vital rates
  - model misspecification
  - imperfect expert judgment
- errors due to random fluctuations

Forecast Model: Error Propagation



Error Propagation in Demographic Forecasts

- Stochastic forecasts – probability distribution for errors in inputs and hence outputs
- Stochastic forecasts for population vector
  - Alho and Spencer (1985)
  - Lee and Tuljapurkar (1994)
- Initially, analytic approximations
- Currently, simulation based
  - generate sample paths
  - software: PEP (Alho), S^4 (Tuljapurkar et al.)

Parameterize Covariances

Need model to relate past volatility to future

More covariances than data points

- need model
  - Alho and Spencer (1997, 2005)

Propagation of Error in Fiscal Forecasts

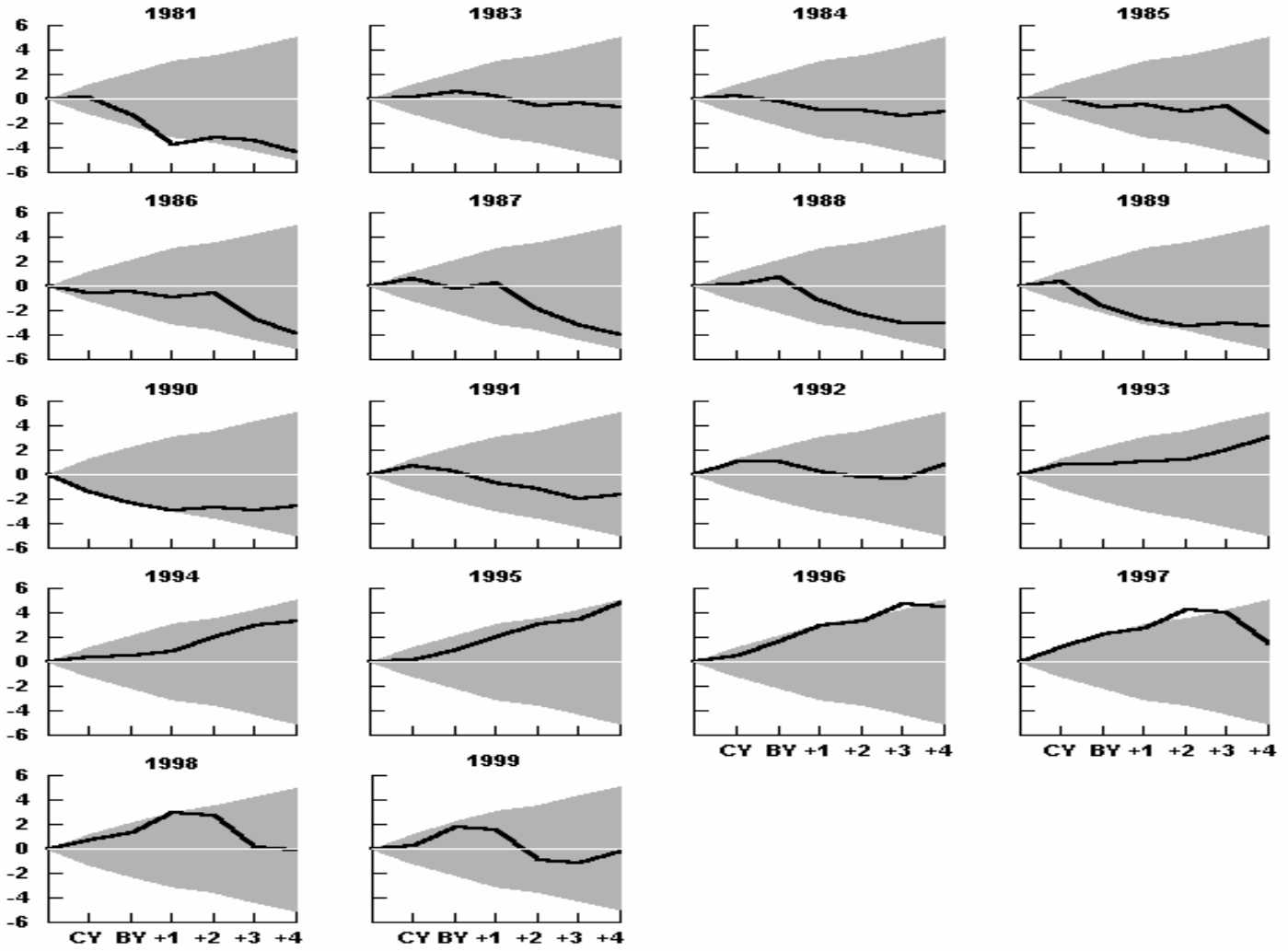
Stochastic Forecasts

- Social Security Trust Fund balance
  - Lee and Tuljapurkar (1998)
  - U.S. Office of the Actuary, SSA
    - Uses alternative models for prediction uncertainty
- Budget deficit or surplus
  - CBO (2005) *The Uncertainty of Budget Projections: Discussion of Data and Methods*

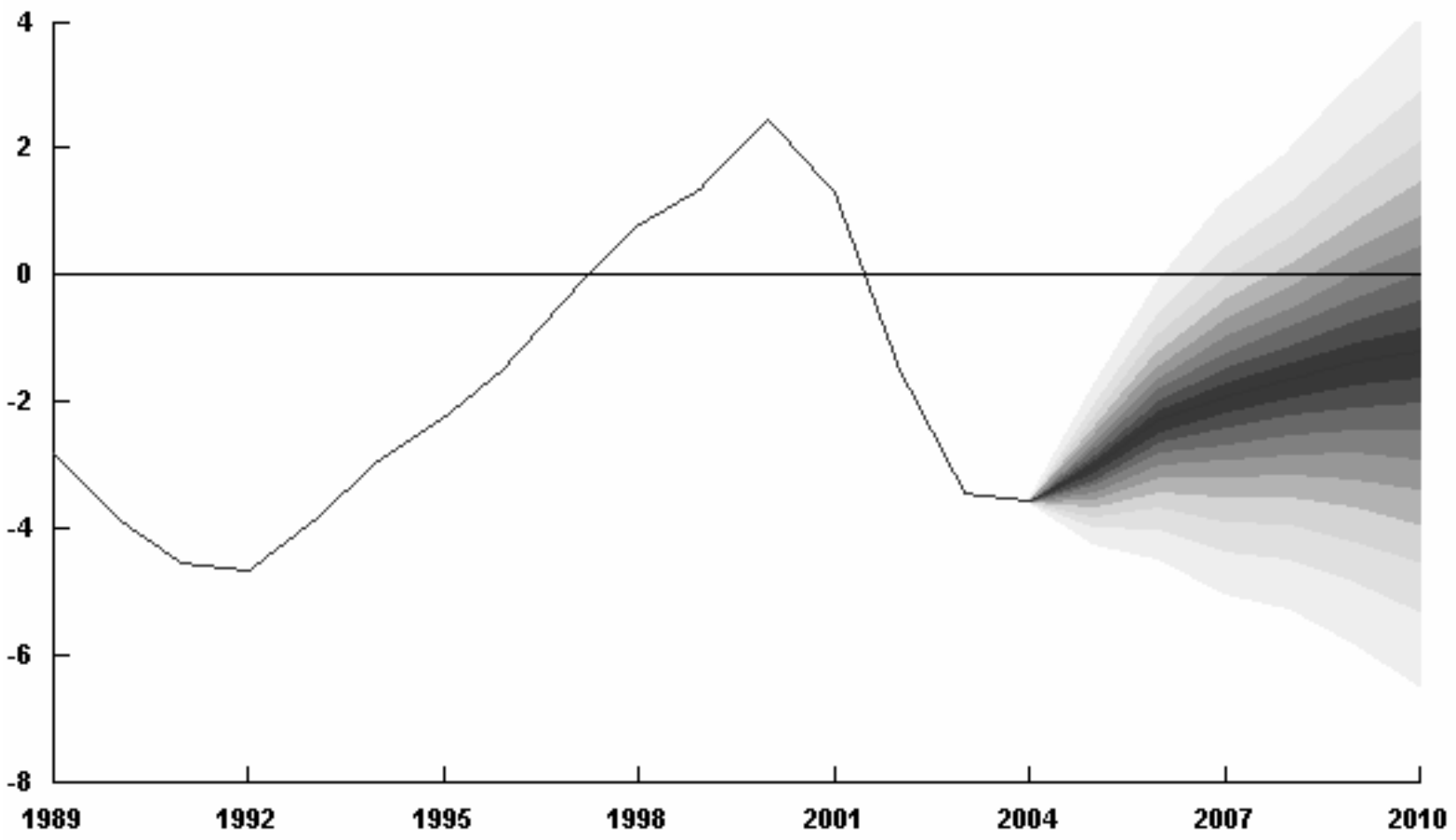
Forecast Budget Surplus/Deficit

- Want probability distribution for error in forecast *t* years ahead
- CBO (2005) forecasts
  - Assume no changes in law or taxes
  - Short term and long term error components
  - Analyze past volatility, predict future volatility

Track Record (scale is percentage of GDP)



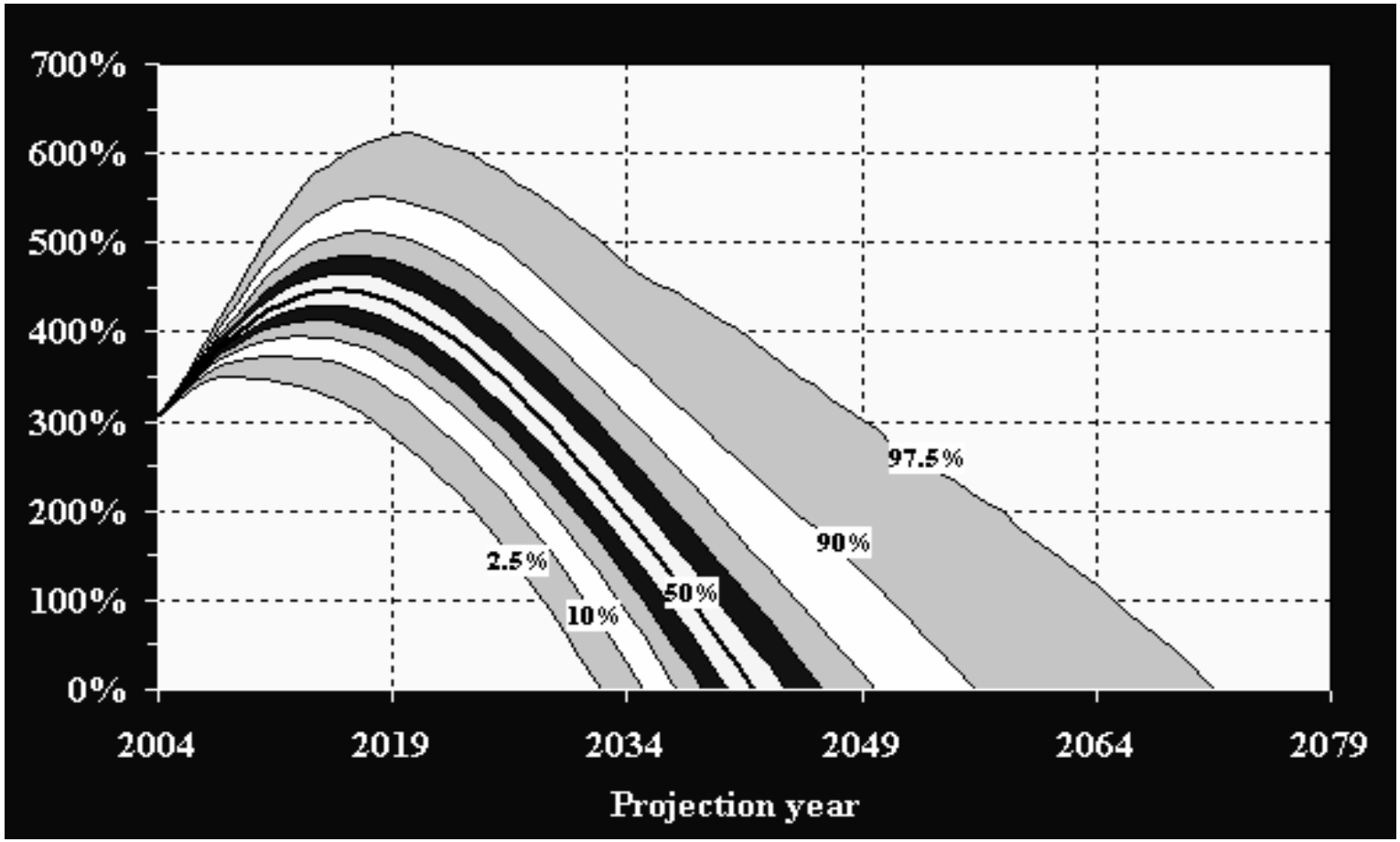
90% Interval for CBO Projection of Budget Surplus (as % of GDP)



Social Security

- Asset : Expenditure Ratio
  - Jan 1. assets to
  - Yearly expenditure
- Forecast the ratio for each year
- Forecast which year ratio will hit zero
- Probability distribution for error in forecast
  - relies on total error model

Prediction Intervals, Social Security Trust Fund Ratios of Assets to Expenditures



from SSA OASDI *2004 Trustees Report to Congress*
http://www.socialsecurity.gov/OACT/TR/TR04/VI_stochastic.html

Median curve crosses axis in 2042

"By 2042, when workers in their mid-20s begin to retire, the [Social Security] system will be bankrupt – unless we act now to save it."

http://www.whitehouse.gov/infocus/social-security/ 3/17/05

note added 3/23/05 – Trustees Report for 2005 has median cross axis at 2041

Total Error for Surveys

Examples

Population Census

Dual Systems Estimator (DSE)

Total Error in Survey
Estimates of Population

- 1990 PES and 2000 A.C.E.
- P sample – estimate census omissions
- E sample – estimate erroneous enumerations
- Evaluation studies to estimate component errors

Major Error Components (1990)

- Measurement error
  - Sampling error
  - Reporting error
  - Matching error
  - Imputation error (missing data)
- Model error
  - Correlation bias
  - Synthetic estimation error
- Data processing error

-(Mulry and Spencer 1991, 1993)

Some Component Errors in Estimate of 1990 US Undercount Rate (est = 2.11)

Source	Bias	Std Dev
P-sample matching	+.21	.05
Model (correlation) bias	–.29	.09
Missing data	0.00	.08
Sampling	+.11	.19
Data error*	~.65	~.20
All	+.49	.23

Each estimate shown assumes other errors are 0

* Data error was approximated after the analysis in Mulry and Spencer (1993)

Some Error Estimates Required
Modeling and Judgment

- Correlation bias – from Demog. Analysis
- Imputation bias – assessed by sensitivity analysis
- Bias and variance of error components for numerous poststrata – hierarchical models
- Synthetic est. error – sensitivity analysis

March 2001 Adjustment Decision

Should the census be adjusted for use in redistricting?

March 2001 Adjustment Decision

- Data error components not available
- March 2001 Total Error Model (TEM)
  - used 1990 data error component model
  - diagnostic information not incorporated into TEM raised doubts
  - flawed but transparent
  - supported use of adjustment in March 2001
- Census Bureau disregarded TEM

March 2001 Adjustment Decision

- Census Bureau made the right decision
- TEM
  - Based on inadequate estimates for data error
  - Misleading but *transparent* in assumptions

- Mulry and Spencer (2003)

A.C.E. Rev. II

- Adjusted for
  - measurement error
  - coding error
  - correlation bias
- Estimates of residual bias not available
- New estimates of duplicates
  - record matching between A.C.E. and census
  - evaluations of duplicates available for TEM

Total Error Model, A.C.E. Rev. II

- No time to assess residual biases
- TEM: adjustment would improve accuracy
- Census Bureau
  - Ignored TEM
  - Decided not to adjust census with Rev. II
- NAS Panel affirmed the Census Bureau decision

Numerical Results from TEM Sometimes Incorrect

- TEM critically dependent on component error evaluations
- Some evaluation data not easily incorporated into TEM at this time
  - Sensitivity analysis
  - Want variance of estimate of total bias

Total error modeling is a process

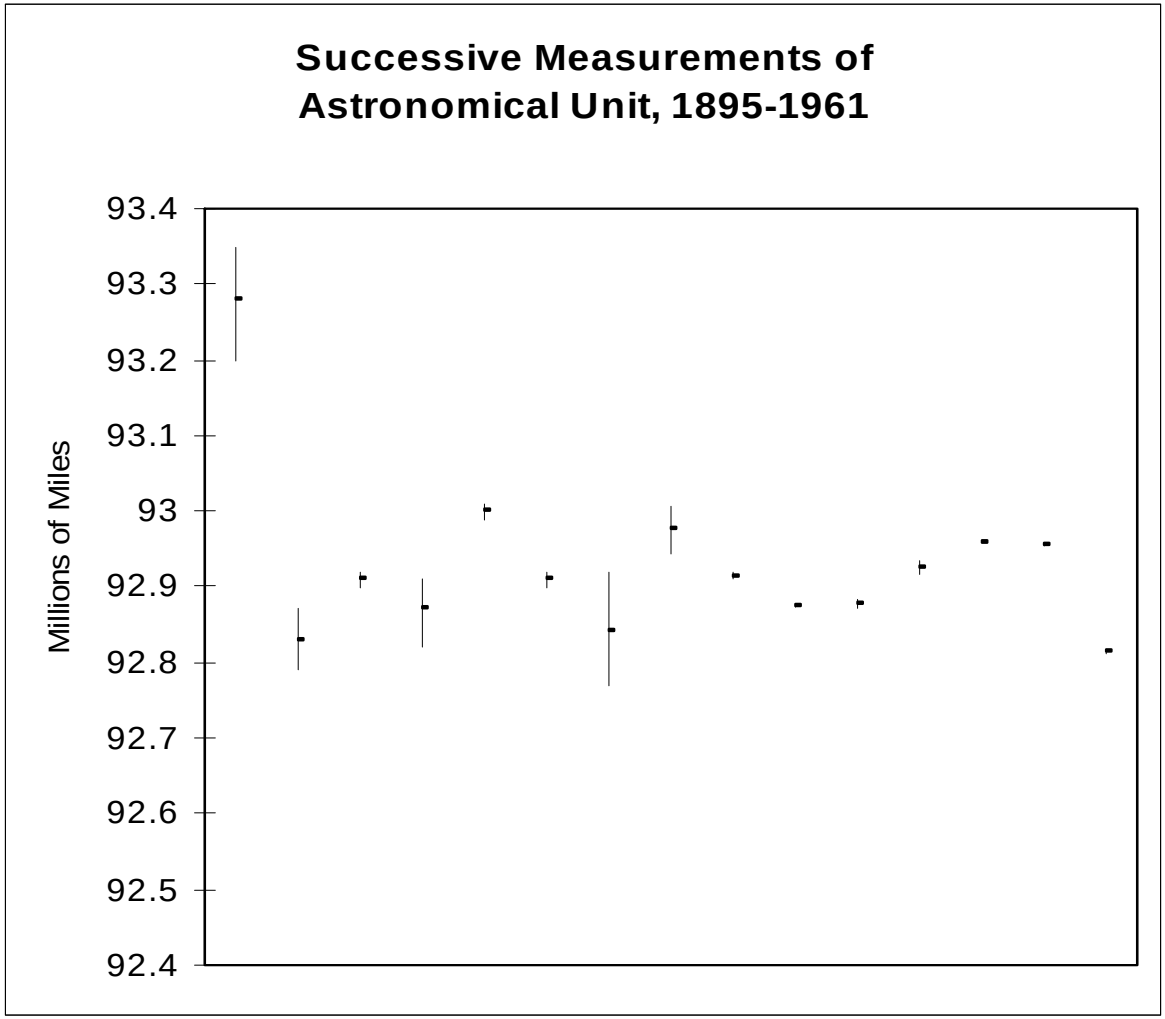
- As additional information emerges, update TEM
- Observe effect of alternative error specifications
- No guarantee that either
  - All errors are identified
  - Identified errors are estimated accurately
- Framework for understanding net error
  - Some errors cancel

Total error modeling is a process (ctd.)

- Original TEM for 1990 PES did not detect large processing error when wrong computer program used to edit clerical match codes
- Limitations of TEM should be state clearly
  - TEM for March 2001 undercount did not have current models for data error, and was misleading
  - TEM for A.C.E. Rev II. was incomplete
- Sensitivity analysis can be useful for understanding effect of limitations

Have physical sciences done better ?

(A past look)



Each successive value lies outside the previous error range. (Youden 1972)

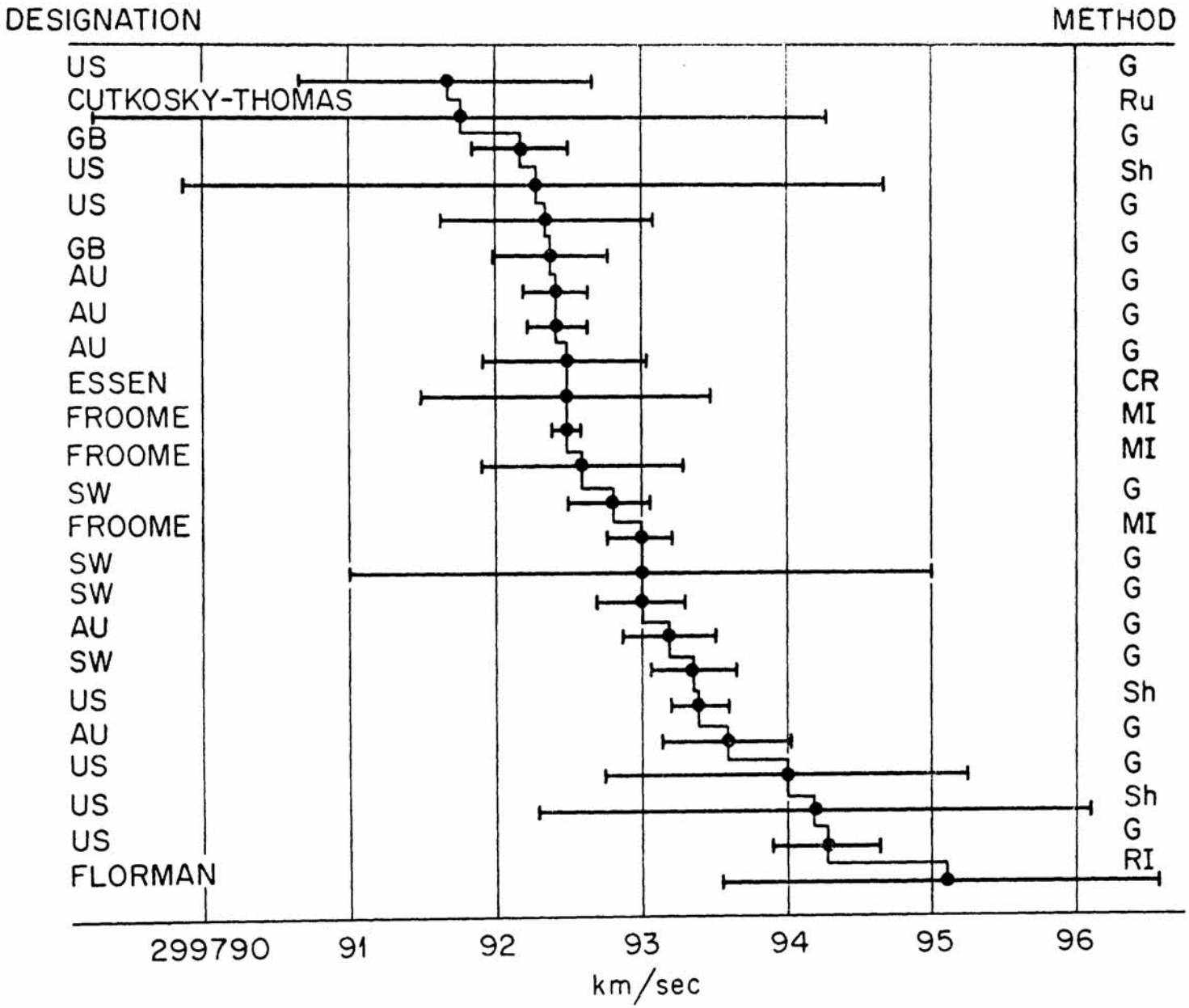


FIGURE 1—Measurements of the speed of light with the reported errors.

“In spite of the difficulties that arise in estimating the error in a constant, most scientists agree that the effort should be made.”

- W. J. Youden (1962)

NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/ Jan. 26, 2005.

- "Uncertainty analysis" [TEM] is part of "Measurement Process Characterization"

- Some errors evaluated not from statistical analysis of data, but subjectively:
  - Reference standards calibrated by another laboratory
  - Physical constants used to calculate reported value
  - Environmental effects that cannot be sampled
  - Others

Randomized Social Experiments

Treatment effect

- definable as difference between means of different variables in same population
- peer effects $\Rightarrow$

  definition of treatment effect is design dependent
- estimated from random sample

Peer Effects

- Peer effects may be present in some kinds of experiments
- Inoculation against a contagious disease
  - Treatment assigned to one person may benefit neighbors
- Educational programs
  - Performance of an individual in a class or school is affected by interactions with classmates, both directly and indirectly

Head Start

- National program providing comprehensive childhood development services to low-income pre-school age children and their families
- Services are provided in centers
- *Building Futures: The Head Start Impact Study Interim Report*, September 2003. Report to Congress. Washington, D.C.: Administration for Children and Families

Head Start Impact Study

- Sample of 378 centers (2002-03) that
  - met performance standards
  - had enough applicants that some would need to be rejected
- At sampled Head Start centers, list of enrollees was extended to allow for assignment of ~11 to control group
- Samples were selected of ~ 27 of the newly entering 3- and 4- year old applicants to each sampled center
- Sample children were randomly assigned to treatment or control (~11 control, ~16 treatment, per center)
- 2829 treatment children
- 1921 control children

Control Group

- The control group are not admitted to Head Start program.
- They choose which services to (try to) get
  - Head-Start-like services (e.g., state funded)
  - Or other services
  - Or no non-parental services

Assignment to Treatment or Control Group

- Stratify new applicants by age (age 3 or 4)
- Center decides which new applicants it would admit, under normal procedures
- Relax the threshold to accommodate as if additional applicants would be admitted (~11, to allow enough children to assign to control)
- Sample ~27 and randomly assign to treatment (~16) or control (~11)
- The rest are admitted

Treatment Effects when Peer Effects Are Present

- Assume that outcome of interest for children within a center depends on who else is in that center, but not other centers
- Treat continuing students as part of makeup of center
- Define treatment effect for individual as average effect with respect to all possible configurations of new applicants (within design context)

Sources of Survey Error

- Sampling error
- Non-response / incomplete frame
- Measurement errors
  - the usual ones, plus
  - non-compliance (individuals do not follow treatment assignment

Additional Errors: Prediction

Future treatment effect in non-experimental settings (Manski 1995)

1. Sample is not from target pop.
2. Treatment to be received by non-random parts of target pop.
3. Future treatment(s) may differ from tested treatment
4. Control may differ $\Rightarrow$ treatment effect changes

TEM should include those prediction errors

We should: Measure what is needed for policy guidance, even if it can only be measured poorly.

- J. W. Tukey (1979)

Concluding Remarks

Total error modeling

- requires modeling and judgment
- illuminates interplay of component errors
  - relative importance
  - cancellation and interactions of errors
  - allows for cost effectiveness analysis
- necessary for cost-benefit analysis

Concluding Remarks

TEM estimates of size of error

- can improve meta-analysis (w/ weighting of studies)
- can sharpen policy debates:
  - Is disagreement about
    - status quo
    - predictions of policy outcomes
    - values?
- help us learn if better information is needed.

References

- ACF (2003) *Building futures: the Head Start Impact Study interim report*, September 2003. Report to Congress. Washington, D.C.: Administration for Children and Families.

- Alho J.M. and Spencer B.D. (1985) Uncertain population forecasting. *Journal of the American Statistical Association* 80, 306-314.

- Alho J.M. and Spencer B.D. (1997) The practical specification of the expected error of population forecasts. *Journal of Official Statistics* 13, 203-225.

- Alho J.M. and Spencer B.D. (2005) *Statistical demography and forecasting*. New York: Springer (to appear).

References

- CBO (2005) *The uncertainty of budget projections: a discussion of data and methods.* Washington, D.C.: Congressional Budget Office.

- Hoem J. (1973) Levels of error in population forecasts. *Artikler 61.* Oslo: Central Bureau of Statistics.

- Keyfitz N. (1981) The limits of population forecasting. *Population and Development Review* 7, 579-593.

- Lee R.D. and Tuljapurkar S. (1994) Stochastic population forecasts for the United States: beyond high, medium, and low. *Journal of the American Statistical Association* 89, 1175-1189.

- Lee R.D. and Tuljapurkar, S. (1998). Uncertain economic futures and social security finances. American Economic Review, May, 237-241.

References

- Manski C. (1995) *Identification problems in the social sciences*. Cambridge: Harvard University Press.

- Mulry M.H. and Spencer B.D. (1991) Total error in PES estimates of population: the dress rehearsal census of 1988. *Journal of the American Statistical Association* 86, 839-854 with discussion 855-863.

- Mulry M.H. and Spencer B.D. (1993) Accuracy of the 1990 census and undercount adjustments. Journal of the American Statistical Association 88, 1080-1091.

References

- Mulry M.H. and Spencer B.D. (2001) *Accuracy and coverage evaluation: overview of total error modeling and loss function analysis.* DSSD Census 2000 Procedures and Operations Memorandum Series B-19*. Washington, D.C.: U.S. Census Bureau. http://www.census.gov/dmd/www/pdf/Fr19.pdf

- Rubin D.B. (2005) Causal inference using potential outcomes: design, modeling, decisions. *Journal of the American Statistical Association* 100, 322-331.

- Social Security Administration (2004) *The 2004 annual report of the board of trustees of the Federal Old-Age and Survivors Insurance and the Federal Disability Insurance Trust Funds.* Washington, D.C.: U.S. Government Printing Office.

References

- Stoto M. (1983) Accuracy of population projections. *Journal of the American Statistical Association*, 78, 13-20.

- Tukey J.W. (1979) Methodology, and the statistician's responsibility for BOTH accuracy AND relevance. *Journal of the American Statistical Association* 74, 786-793.

- Youden W.J. (1962) Systematic errors in physical constants. *Technometrics* 4, 111-123.

- Youden W.J. (1972) Enduring values. *Technometrics* 14, 1-11.