Measurement Error in the CE: Monitoring the Quality of the Estimates

Roger Tourangeau, Westat
Lucilla Tan, Brian Meekins
and Brandon Kopp, Bureau of Labor Statistics

International Total Survey Error Workshop
October 2, 2014

BLS

Westat

Introduction

- Based on our review of the state of knowledge about the error in the Consumer Expenditure Survey (CE), we concluded that less was known than was desirable
- In addition, we should be able to track how well the CE is doing over time
- As a practical matter, most of the measures proposed track overall error in the CE, not just measurement error

MMMI approach

- Many methods have been used to assess error in the CE, each with their strengths and weaknesses
- We recommend an multi-method-multi-indicators (MMMI) approach that consists of three main categories:
  - Internal indicators (based solely on CE data or paradata)
  - External indicators (compare estimates from the CE to an external data source)
  - Comparison of CE production estimates with "gold standard" interviews

MMMI approach—II

- Precisely because no one approach is perfect, we think coming at this from several angles will provide a much more comprehensive picture of the CE quality
- It is time to move away from reliance on the PCE estimates as the main basis for evaluating the CE

Criteria for External Indicators

- Comparability: Is the external estimate comparable to the CE
- Consistency: Do the estimates show a similar magnitude difference from CE estimates over repeated survey administrations?
- Ease of producing the estimate: How difficult are the benchmark estimates to produce? Can they be produced in a timely manner without undue burden on the CE Survey staff?
- Timeliness and periodicity of the benchmark estimate: What is the elapsed time between data collection and publication of the benchmark estimate? How frequently are the data collected?
- Comprehensiveness: Taken together do the various benchmarks provide an overall picture of error in CE estimates (across multiple sections, waves, and time periods)?

External Indicators

- Comparison to external data sources
- Two main external sources
  - Personal Consumption Expenditures from NIPA (National Income and Product Accounts)
  - Compare CE estimates with other surveys (e.g., MEPS, PSID, RECS)
- Weakness—Although PCE covers many categories and a lot of work has gone into establishing "concordance" of PCE/CE categories, errors in PCE are not well established; not clear external benchmarks are really more accurate than the CE

Some Candidate Indicators

CE Category	MEPS	PSID	NHEA
Physician Services	X		X
Dental Services	X		X
Eyecare services	X		
Nonphysician services (excluding dental and eyecare)	X		X
Hospital-inpatient	X	X	X
Prescription drugs	X	X	X
Vision aids	X		

Some Candidate Indicators--2

CE Category	ACS	AHS	RECS	PSID	PCE
Electricity	X	X	X	X	X
Natural Gas	X	X	X		X
Total Other Fuels	X	X	X		
Fuel Oil		X	X		
Propane/LPG			X		
Kerosene			X		
Other Fuels (Wood, Coal, etc.)		X			
Water/Sewer	X	X		X	X
Garbage		X			X
Primary Mortgage	X	X		X	
Rent	X	X		X	
Homeowner's Insurance	X	X		X	X
Property Tax	X	X		X	

External Indicators

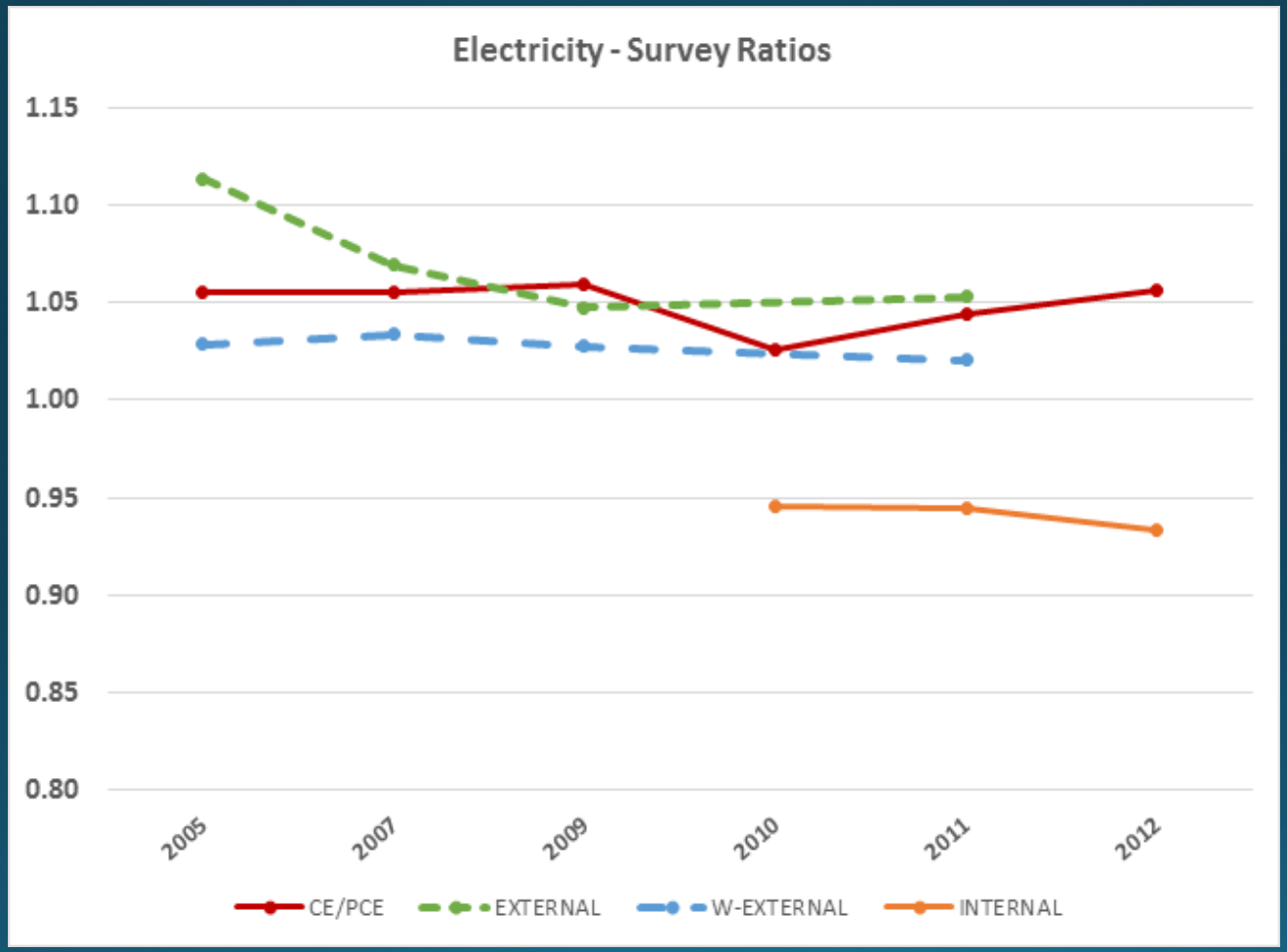
Expenditure Share for our 3 illustrative examples

CE Category	Expenditure Share (2011)
Electricity	2.9%
Rent	6.1%
Prescription Drugs	1.0%
Average Annual Spending	$49,705

Combining External Indicators

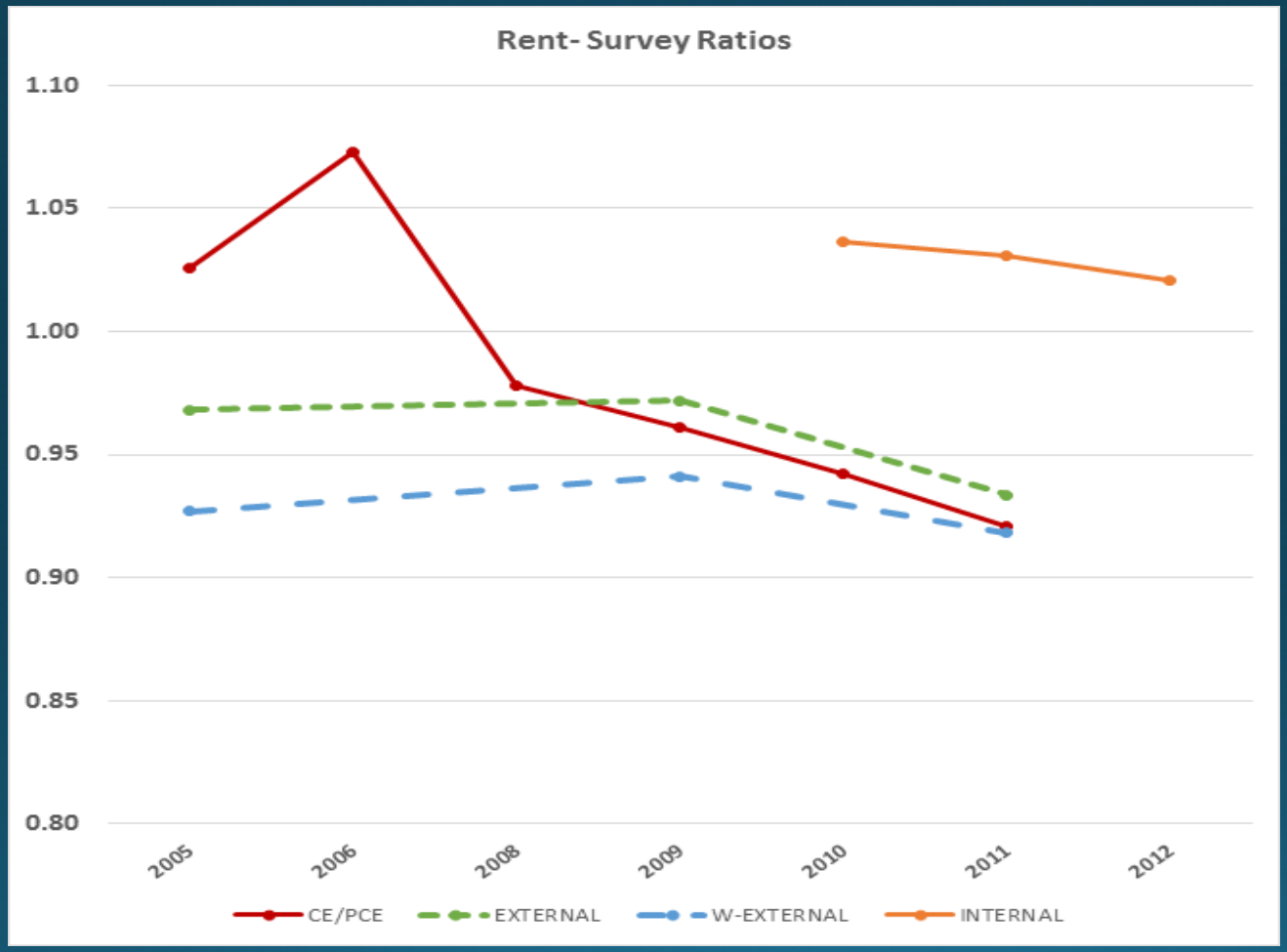
- To increase the robustness of the external comparisons, we recommend taking weighted averages of the external estimates for the commodity
- The weights would reflect the reliability of the ratio of the CE estimate to the external indicator over time and would downweight estimates that show large fluctuations over time
- For a given commodity category, a ratio $r_{j,t}$ is constructed for each selected implementation from source j (t), dividing the CE estimate for the commodity with the external source estimate.
- An average of the ratios is taken, where $\bar{r}_j = (\sum^T r_{j,t})/T$, where T is the total number of time points from external source j. Next a variance is computed, $s_j^2 = \sum^T (r_{j,t} - \bar{r}_j)^2$ for each external source j. The estimates from all external sources are then combined using a weighted average at each time, t.
- Where the weighted average is given by, $WR_t = \frac{\sum_{j=1}^{J} (r_{j,t} * 1/s_j^2)}{\sum_{j=1}^{J} 1/s_j^2}$, where J is the total number of external sources

An Example



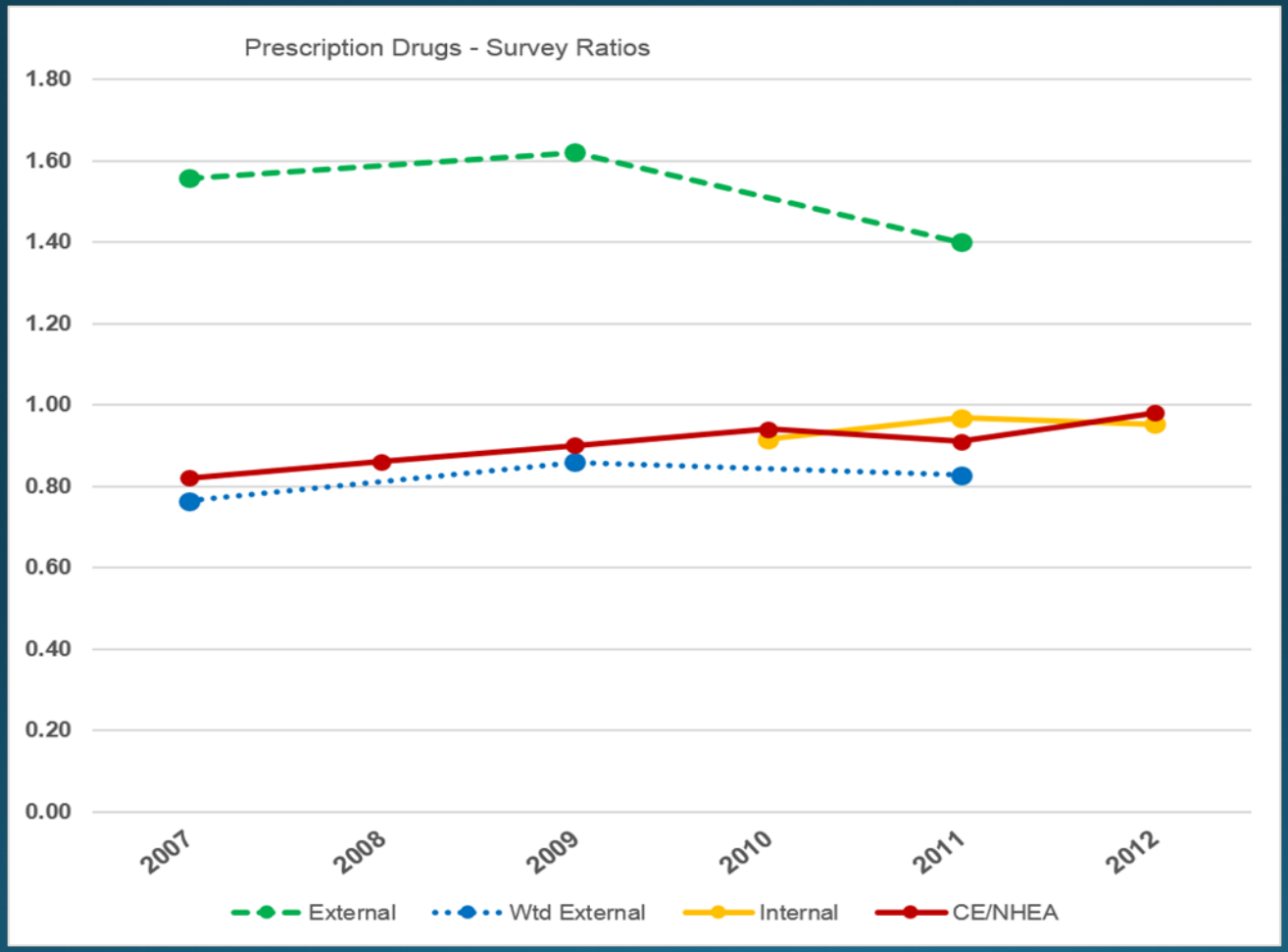
- The external sources for electricity are PCE, ACS, AHS, RECS, and PSID.

Another Example



- The external sources for rent are PCE, ACS, and AHS.

One More


Prescription Drugs - Survey Ratios
1.80
1.60
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
2007
2008
2009
2010
2011
2012
External
Wtd External
Internal
CE/NHEA

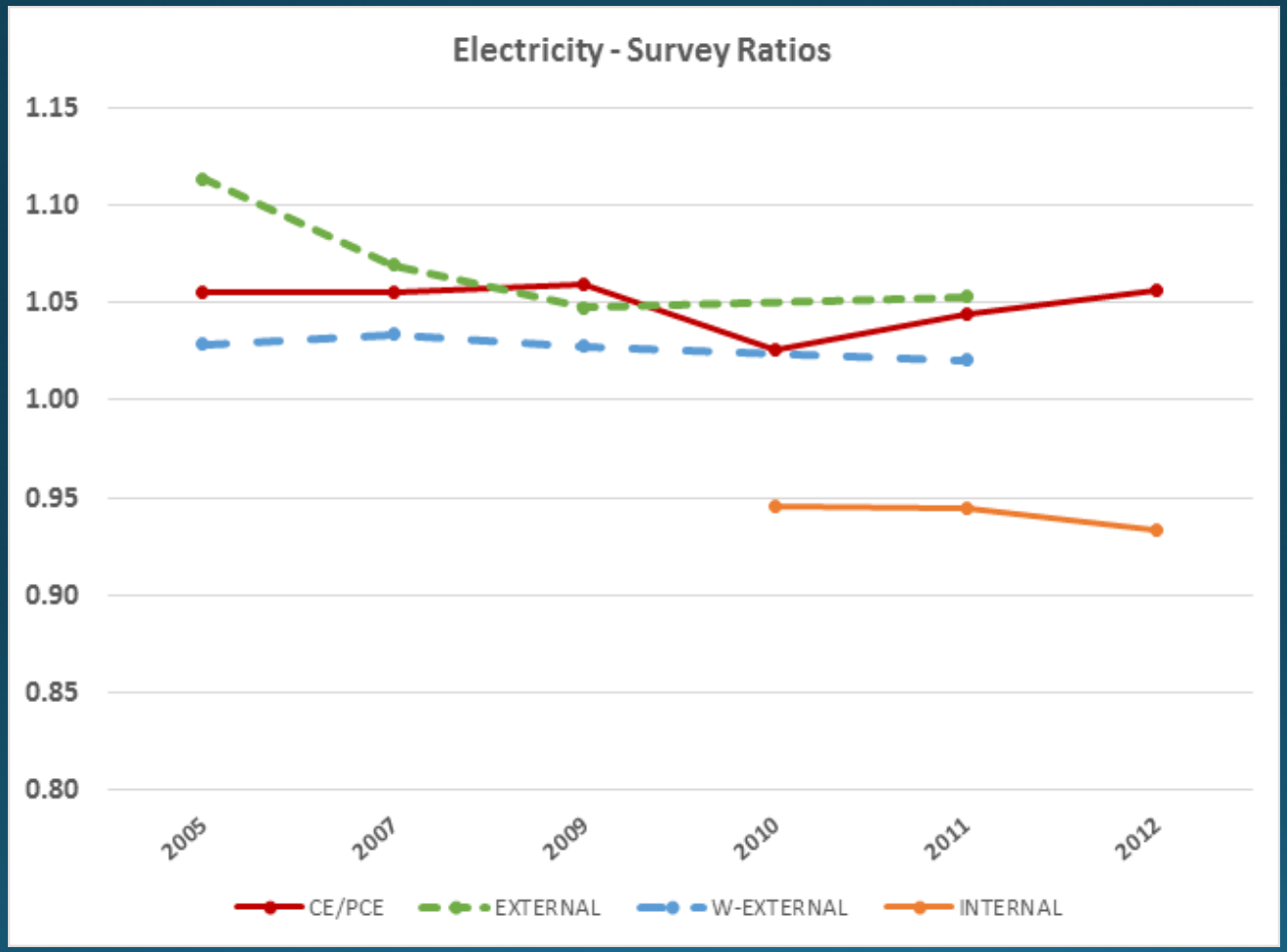
Internal Indicators

- Internal indicators should be robust, easy to interpret, and based on similar metric to the external indicators.
- Candidate indicators include both commodity- or section-specific indicators and interview-level indicators.
  - Section specific indicators:
    - record use
    - section interview time
  - General indicators
    - Willingness to provide income data
    - The number of attempts required to complete and interview
  - The indicators are then evaluated by examining their relationship with the reported expenditure of the commodity category. Those showing no relationship or a weak relationship with expenditure are discarded.

Latent Class Models for Combining Internal Indicators

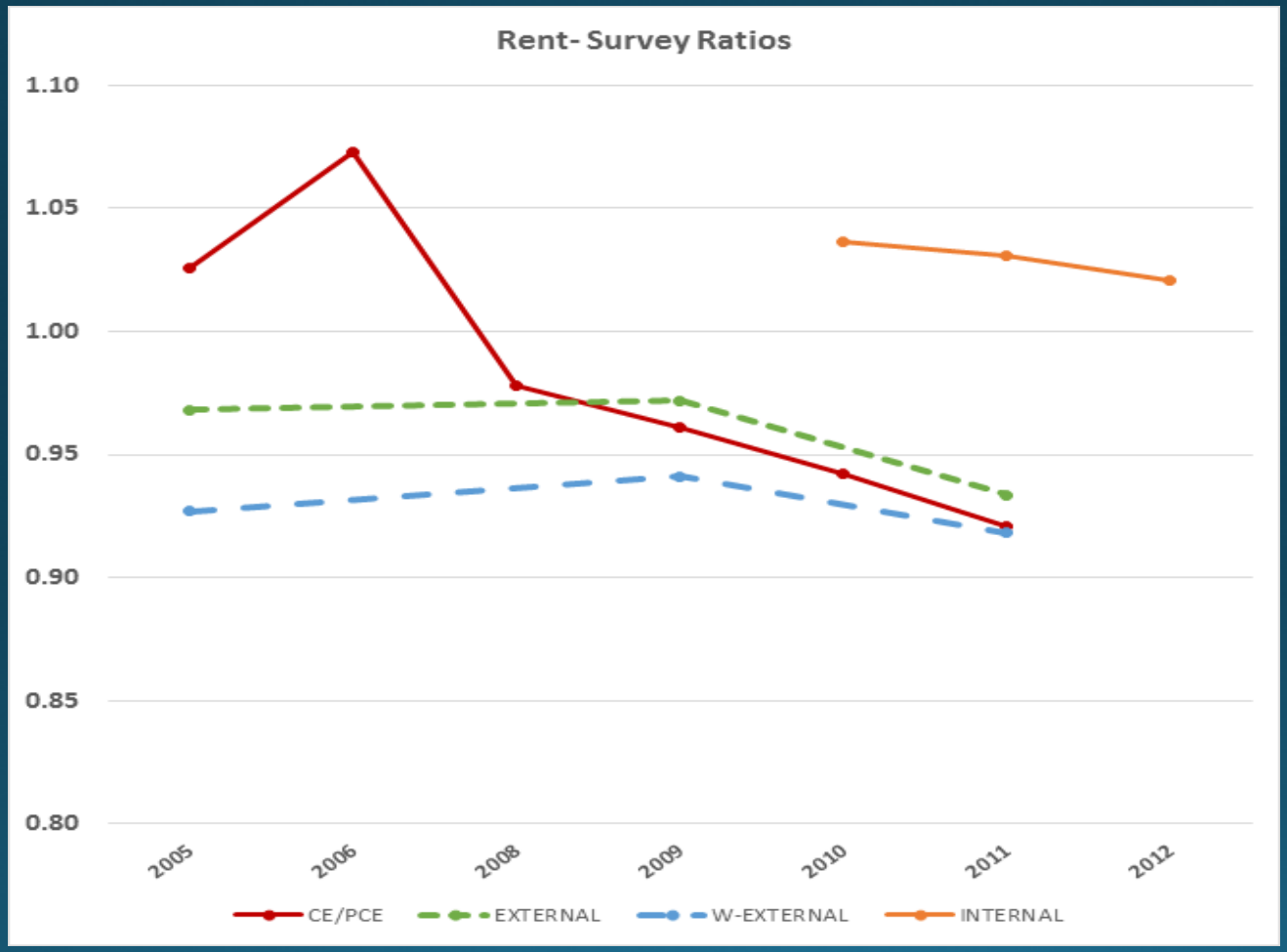
- Four variables seem to have strong relation to reporting across a number of commodity categories:
  - Use of the infobook (+);
  - Whether the interview is done by telephone (-);
  - Whether R used records (+);
  - Commodity-specific time (+)
- Classify respondents into one of two latent classes based on these variables
- Construct ratios of mean expenditures reported by all reporters over "good reporters"

An Example



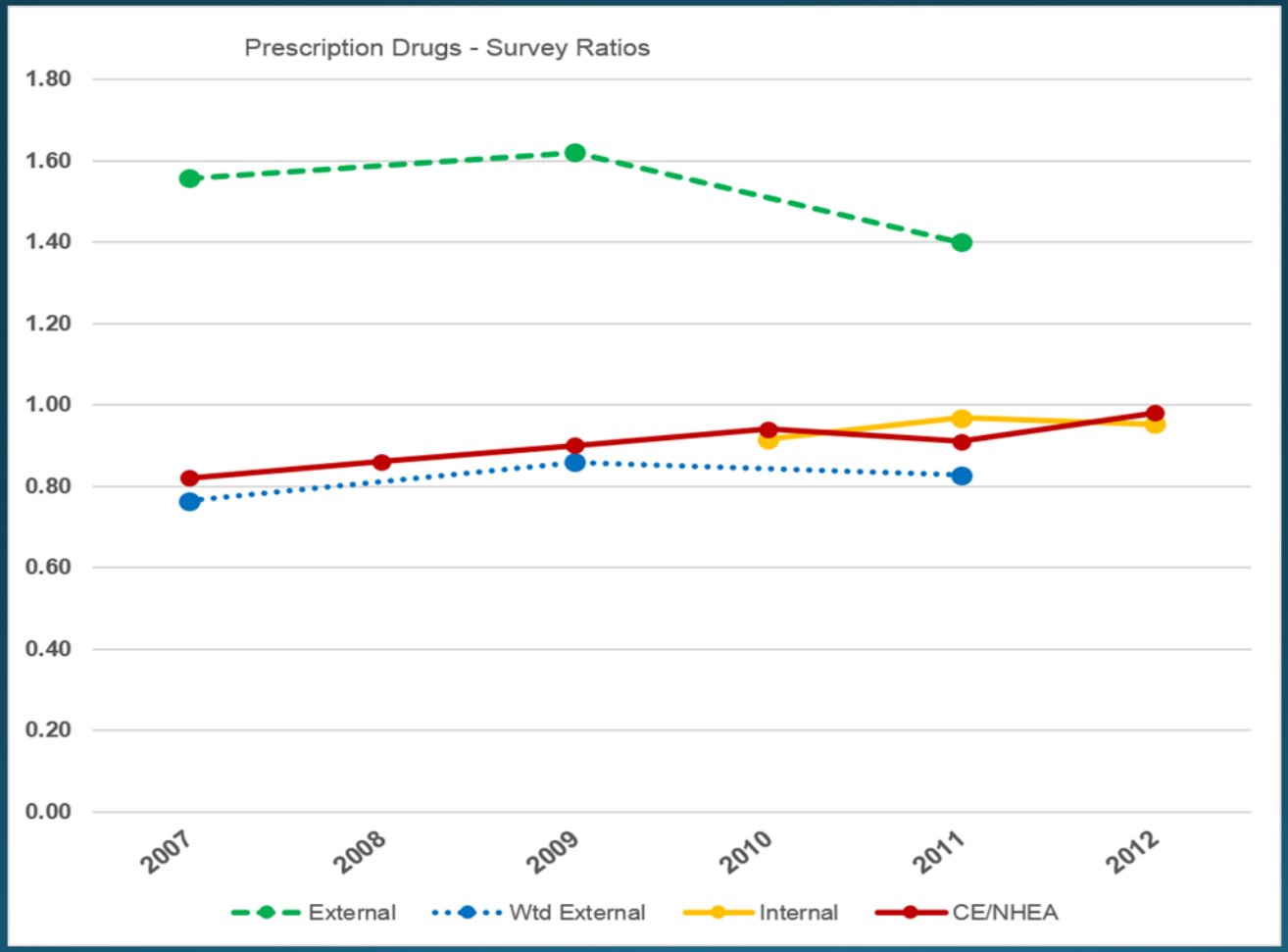
- The external sources for electricity are PCE, ACS, AHS, RECS, and PSID.

Another Example



- The external sources for rent are PCE, ACS, and AHS.

One More


Prescription Drugs - Survey Ratios
1.80
1.60
1.40
1.20
1.00
0.80
0.60
0.40
0.20
0.00
2007
2008
2009
2010
2011
2012
External
Wtd External
Internal
CE/NHEA

Gold Standard Interview

- Key to assessing validity of internal and external indicators—Are the ratios in the internal and external indicators for a given commodity category similar (e.g., <1) to those from the gold standard interview (that is, GS estimate/production estimate)?
- Also, key to establishing level and direction of errors
- Borrows many features of the proposed new CE interview
- Five key features:
  - Initial bounding interview
  - Short reference period
  - Reliance on aided recall (records, diaries); prospective collection of records
  - Reduced burden
  - Contingent incentives

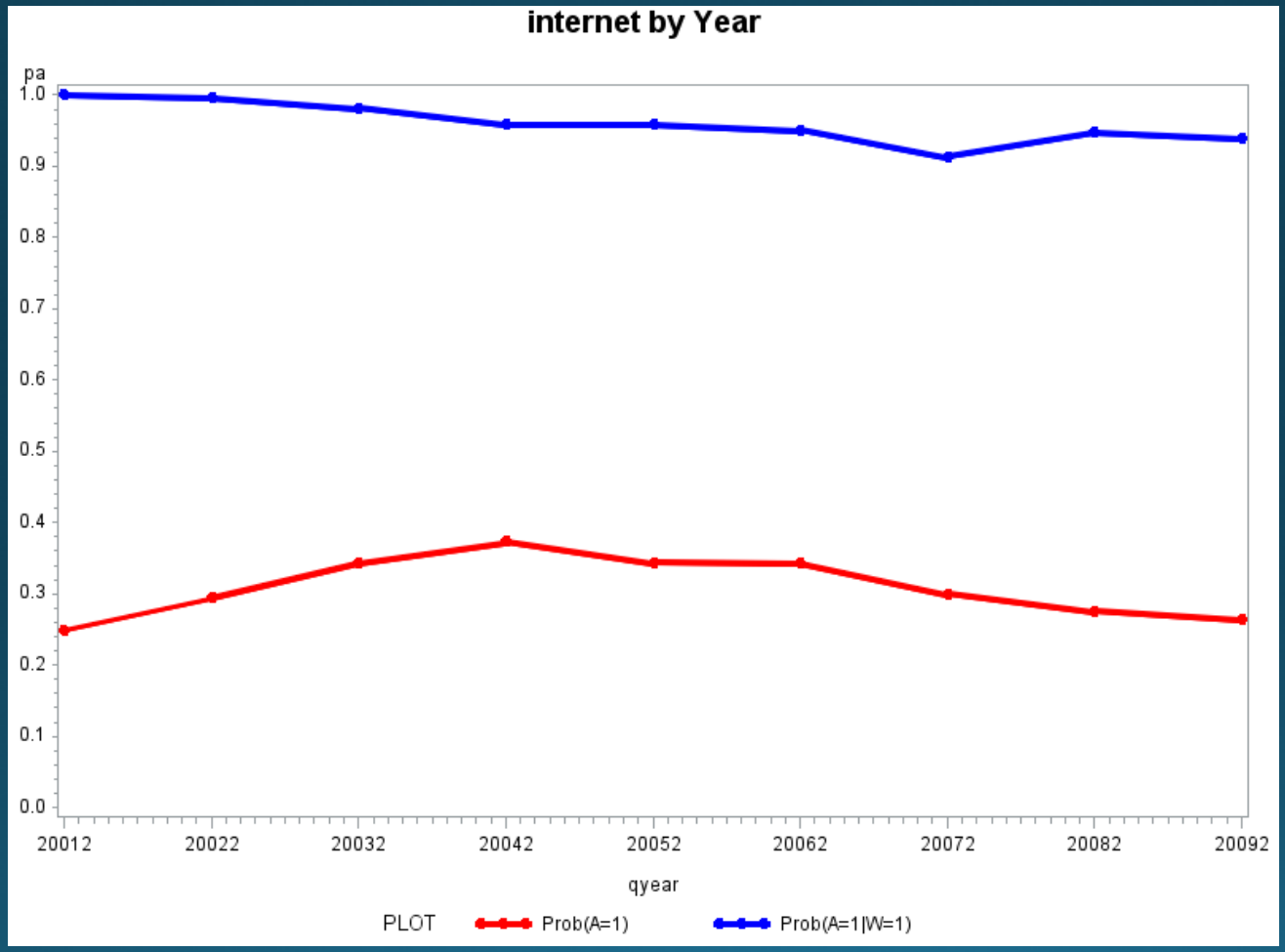
Some Topics for Research

- We see at least four factors as critical for successful gold standard interview
  - Incentives for records collection or diary keeping
  - Other inducements for encouraging record keeping
  - Length of reference period (burden versus stability of estimates)
  - Selection of commodity categories

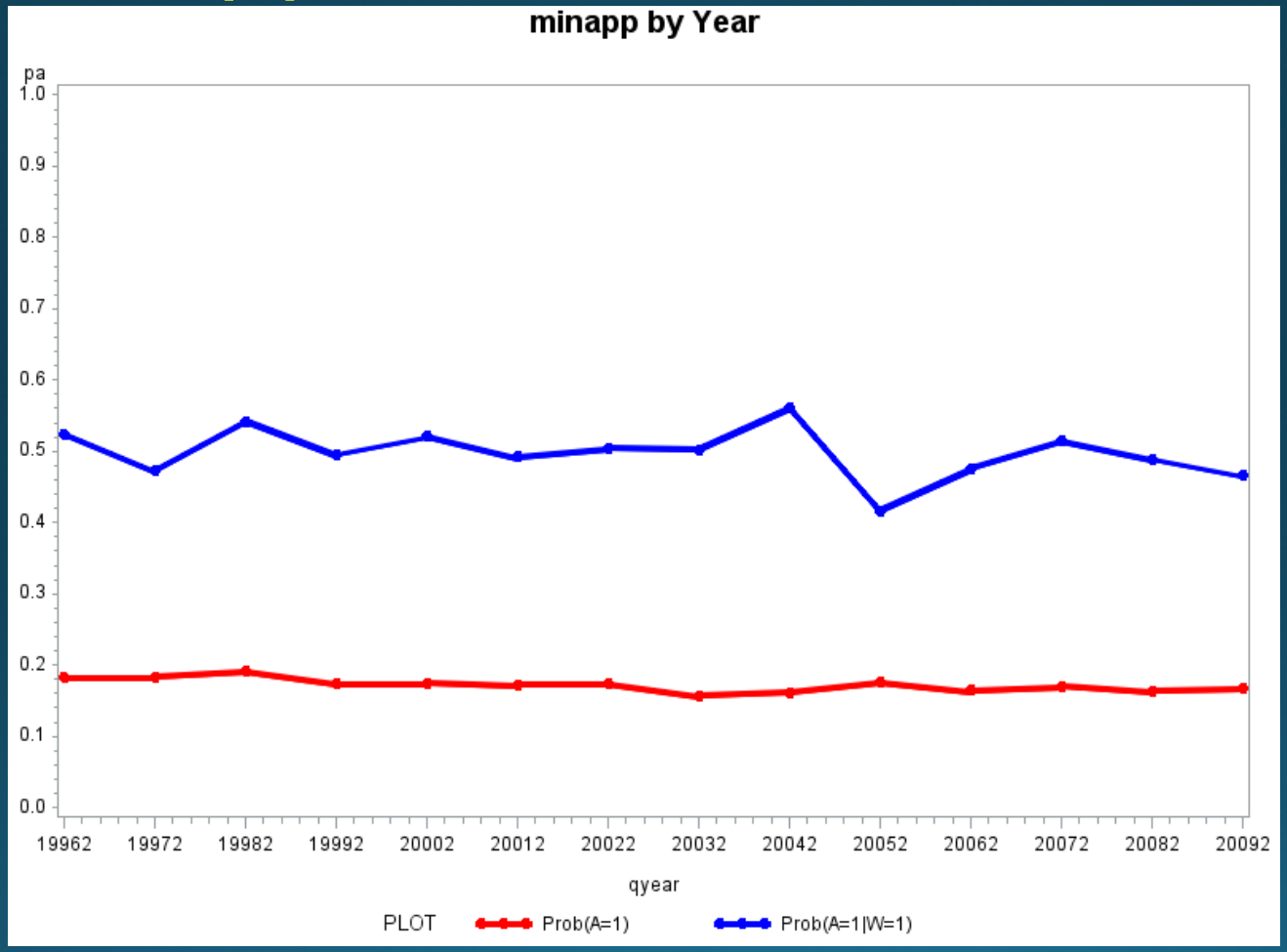
Markov Latent Class Analysis

- Uses repeated measurements from panel survey data to estimate classification error
- Does not require external validation data; estimates of error directly from panel data
- LCA used to study measurement or response error (VandePol and deLeeuw 1986; Tucker 1992; Van de Pol and Langeheine 1997; Bassi et al. 2000; Biemer and Bushery 2000; Tucker, et al. 2002, 2003, 2004, 2005, 2006, and 2008); Meekins et al. (2011)
- Accuracy rates by subgroup - percent of true purchasers that reported purchasing that commodity
- Could easily produce a ratio of reported expenditure/reported expenditure*1/accuracy

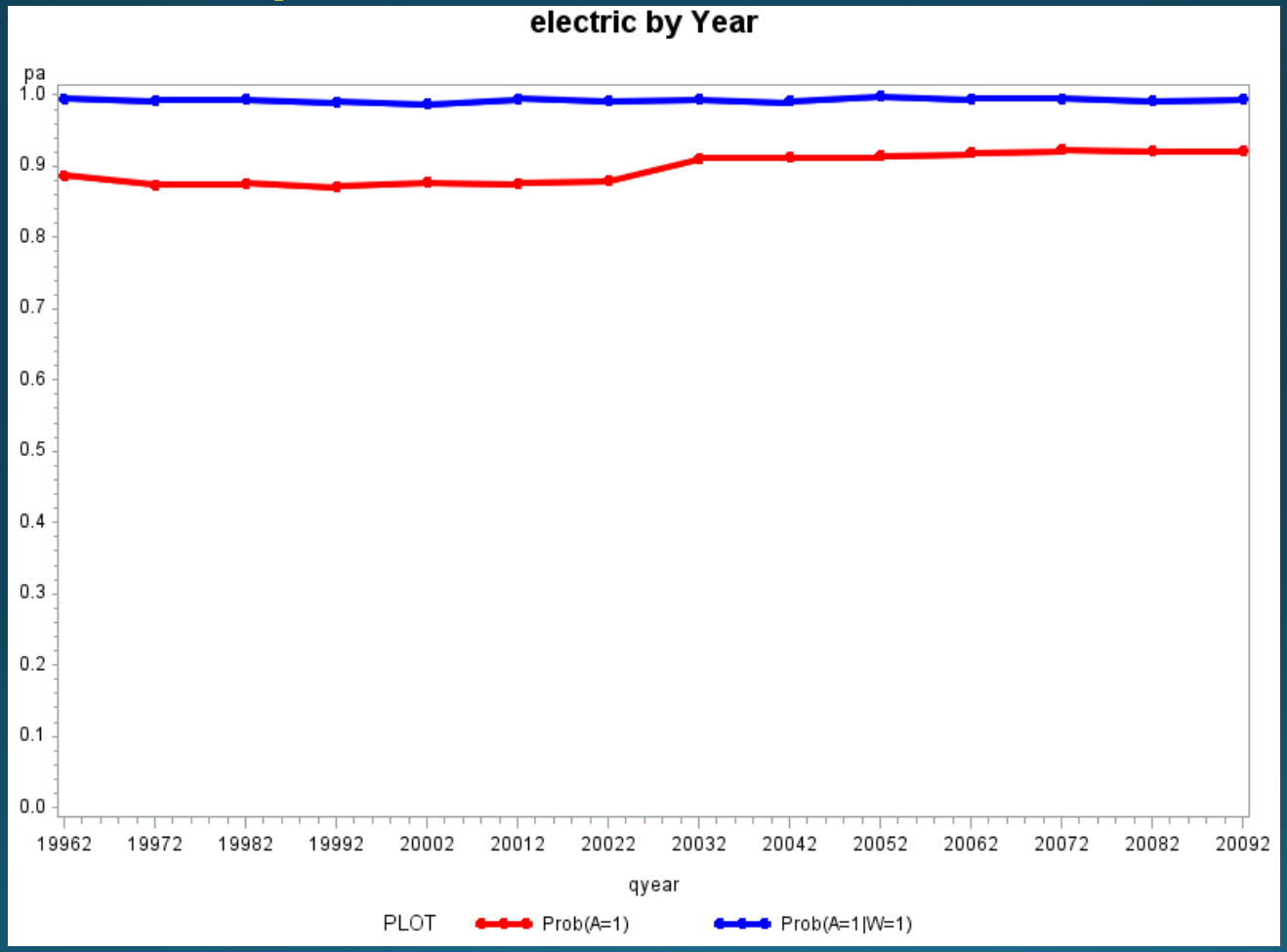
Internet


internet by Year
pa
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
20012
20022
20032
20042
20052
20062
20072
20082
20092
qyear
PLOT
Prob(A=1)
Prob(A=1|W=1)

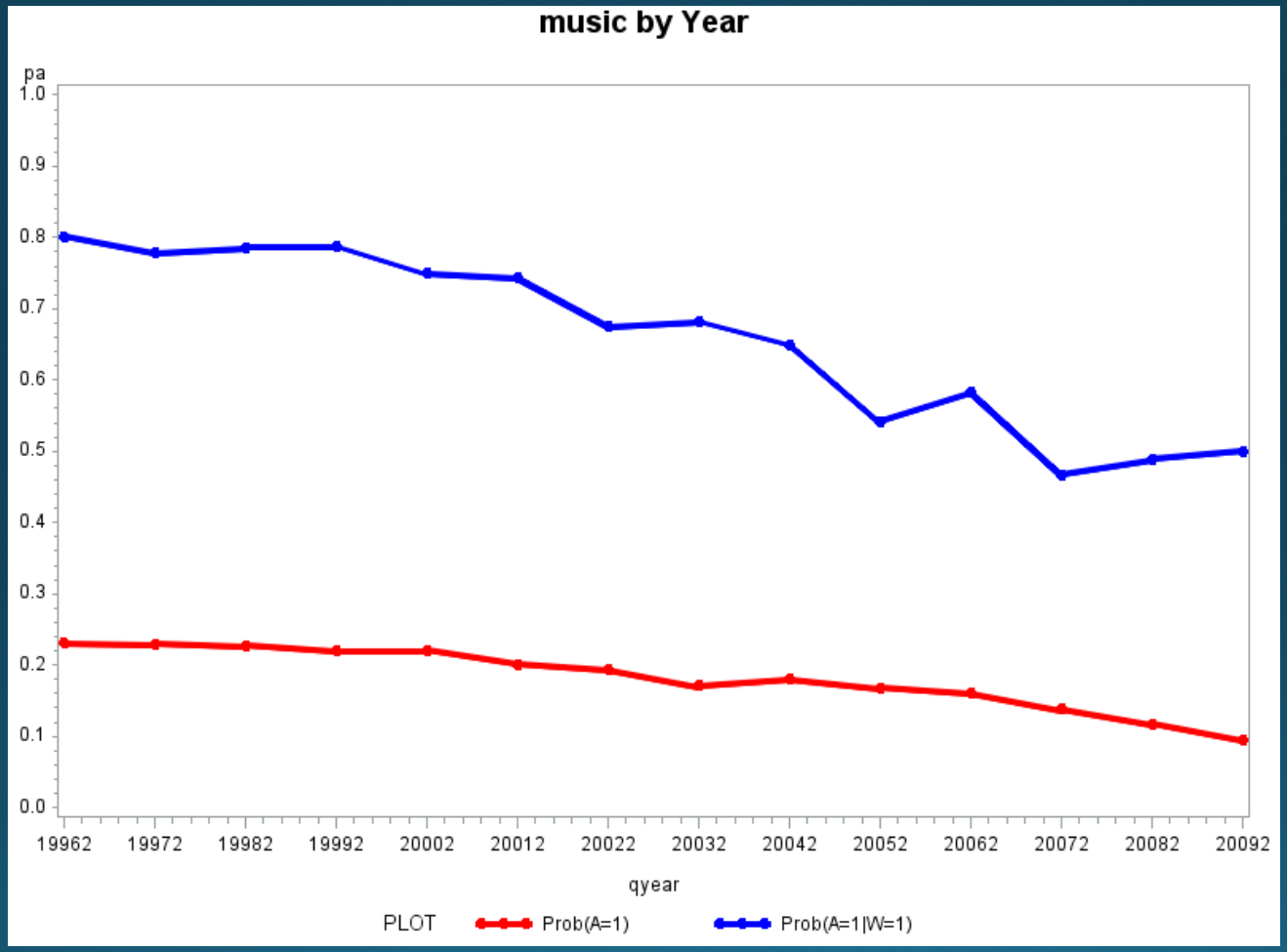
Minor Appliances



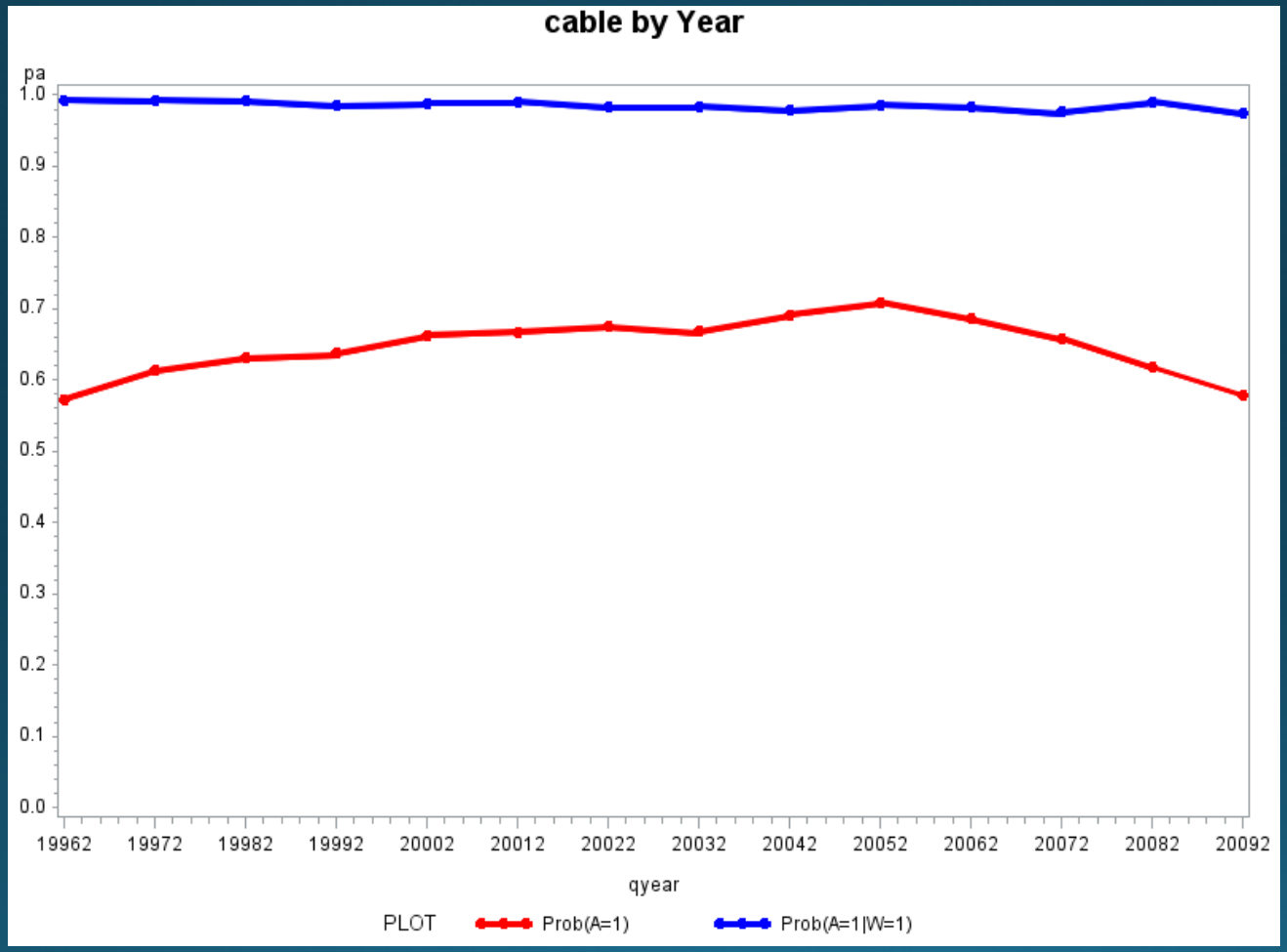
Electricity



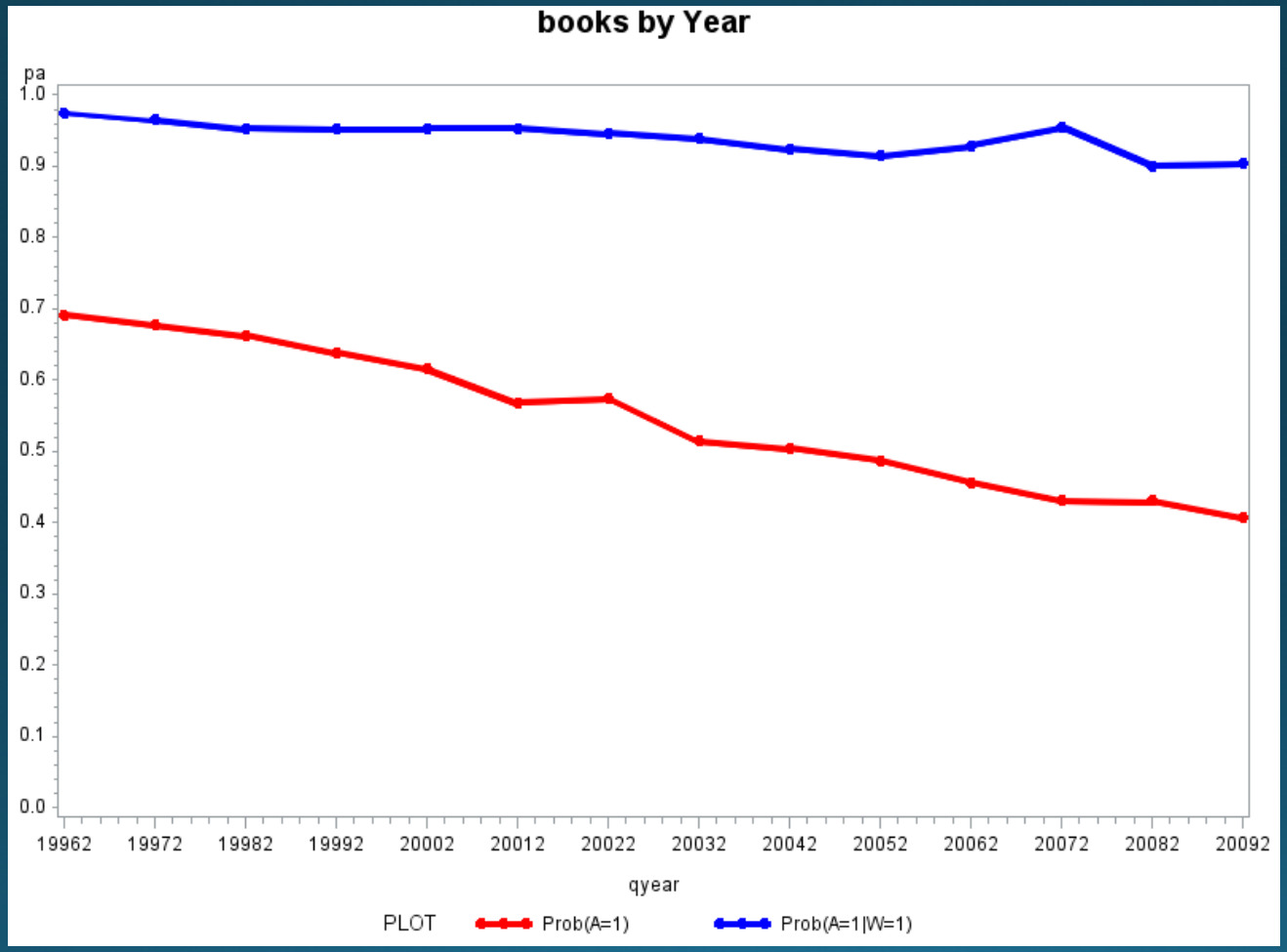
Music


music by Year
pa
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
19962
19972
19982
19992
20002
20012
20022
20032
20042
20052
20062
20072
20082
20092
qyear
PLOT
Prob(A=1)
Prob(A=1|W=1)

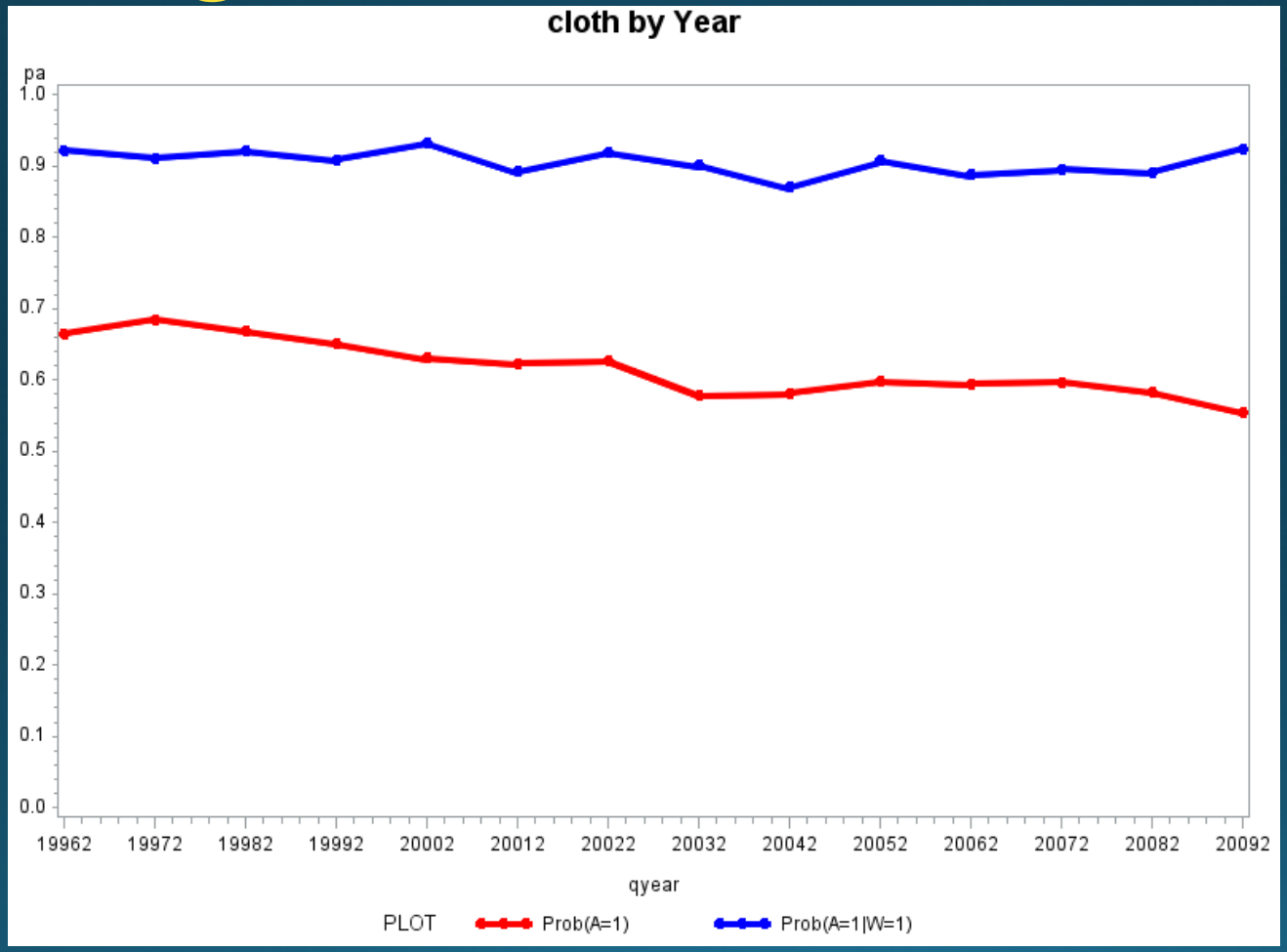
Cable TV


cable by Year
pa
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
19962
19972
19982
19992
20002
20012
20022
20032
20042
20052
20062
20072
20082
20092
qyear
PLOT
Prob(A=1)
Prob(A=1|W=1)

Books


books by Year
pa
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
19962
19972
19982
19992
20002
20012
20022
20032
20042
20052
20062
20072
20082
20092
qyear
PLOT
Prob(A=1)
Prob(A=1|W=1)

Clothing


cloth by Year
pa
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
19962
19972
19982
19992
20002
20012
20022
20032
20042
20052
20062
20072
20082
20092
qyear
PLOT
Prob(A=1)
Prob(A=1|W=1)

Conclusions

- No one approach is perfect
- We recommend building on past efforts
- Develop a time series with multiple indicators
  - Internal indicators
  - External indicators
  - Possibly MLCA models
  - These are all inexpensive
  - Still, given the flaws, they should be supplemented with periodic (but regular) gold standard interview studies
  - Have overlapping expenditure categories to assess convergence across methods

Contact

Brian Meekins
Bureau of Labor Statistics
meekins.brian@bls.gov