Building a Framework for Evaluating Reasonableness of Travel Time Estimates and Margins of Error

presented at

Applying Census Data for Transportation
50 Years of Transportation Planning Data Progress
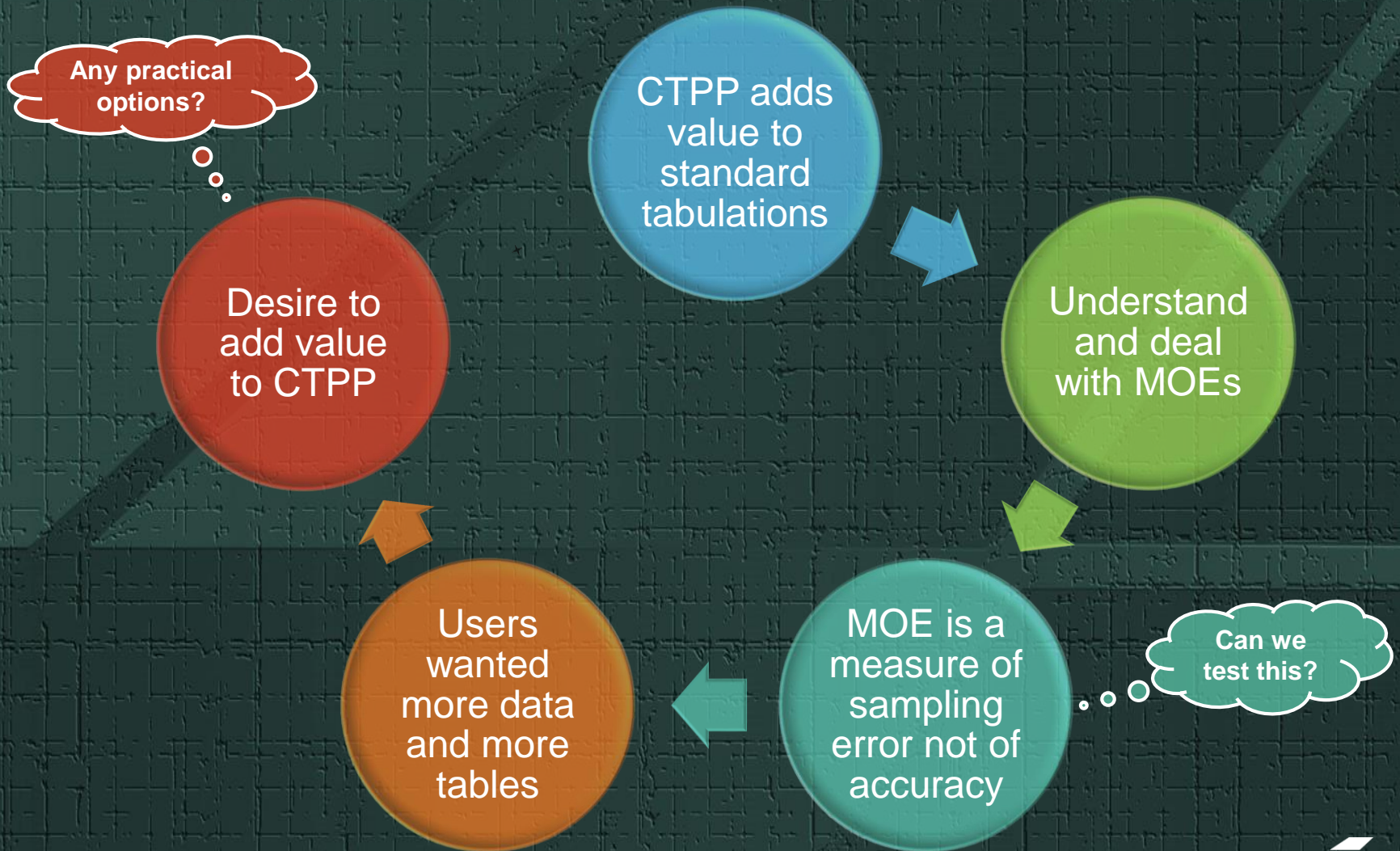**Kansas City, MO**

November 14-16, 2017

presented by

Cemal Ayvalık

Kimon Proussaloglou

Outline

➤ Motivation

➤ Study Design

➤ Data and Analysis

➤ Preliminary Results

➤ Conclusions

➤ Recommendations

CAMBRIDGE SYSTEMATICS

Motivation

Study Design

➤ Compare part of the CTPP flow data to an external data source
  » Auto travel times (shortest path) via Google Maps

➤ Synthesize ACS sampling
  » Two-step probability-proportional-to-size sampling
  » Collect data at a higher rate for a sample of tract pairs

➤ Develop and Test Sample Hypotheses
  » CTPP Mean Travel Times are Equivalent to Google by Strata
  » Accuracy of Mean Travel Times is Independent of MOE
  » Accuracy of Mean Travel Times is Independent of the Strata
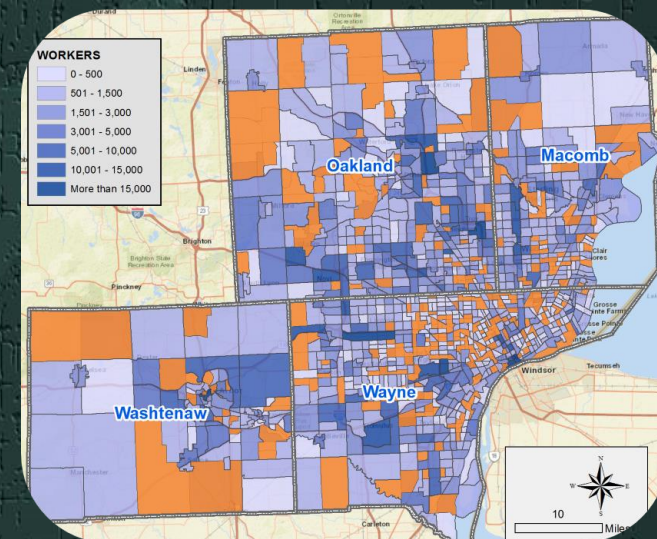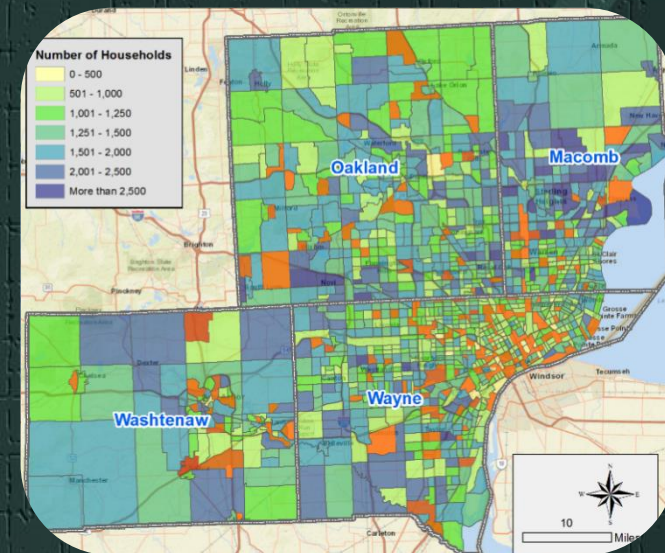 » CTPP and Google MOE are Equivalent across the Strata

CAMBRIDGE SYSTEMATICS

Data Development and Analysis

Data Development

➤ Study Area: Part of the Detroit Metropolitan Area

» Population[1]:   4.23 M
  Employment[2]:   1.95 M

» 2006 – 2010 CTPP: 1.75 M flows among 82,452 tract pairs

1. 2016 Census Bureau Population Estimates
2. 2016 Quarterly Census of Employment and Wages

Data Development – CTPP Sample

➤ Download CTPP Tables

» A112100; A110106; A202100; B306201; B302106

➤ Stratified sample to allow testing effects of select characteristics

▪ SIZE

▪ WORKER DENSITY

▪ DISTANCE

➤ 10% MOE with 90% confidence (n=70)

			AERIAL DISTANCE (in Miles)				
		SIZE (Area in Sq.Miles)	4 or Less	4 to 7	7 to 11	11 to 17	More than 17
WORKER DENSITY (Workers per Sq.Mile)	1,000 or Less	Less than 0.6					
		0.6 to 1.5					
		Greater than 1.5					
	1,000 to 2,500	Less than 0.6					
		0.6 to 1.5		(3x3x5)			
		Greater than 1.5					
	Greater than 2,500	Less than 0.6					
		0.6 to 1.5					
		Greater than 1.5					

CAMBRIDGE SYSTEMATICS

Data Development – CTPP Sample

➔ Probability-Proportional-to-Size Sampling

» P(selected) = f(size)

» 45 strata with 70 pairs w/o replacement

» Worker flows as the size variable

» 3,150 O-D pairs were selected

CAMBRIDGE SYSTEMATICS

Data Development – Test Sample

➤ Build point level O-D locations
  » SEMCOG Data Portal
    ▪ Building Footprints
  » Establishment locations (Info USA)

➤ PPS with replacement to select
  » Twice the households from sampled residence (RES) tracts
  » Twice the establishments from sampled Place of Work (POW) tracts

➤ For each tract pair in the sample, randomly match RES and POW points.

➤ 137,100 O-Ds in the test sample pool

CAMBRIDGE SYSTEMATICS

Data Development – Test Sample

➔ A custom built Google Maps API

» Lat/Lon pairs to highway travel times

» Collects "Directions" data at desired times and frequency

➔ Data collected

» Over a one-month period (August - October) Mondays thru Thursdays

» Between 7:00 AM – 8:30 AM @ 30-min intervals

➔ By approximating ACS sampling rates, a test sample was developed using 11,235 O-D pairs

CAMBRIDGE SYSTEMATICS
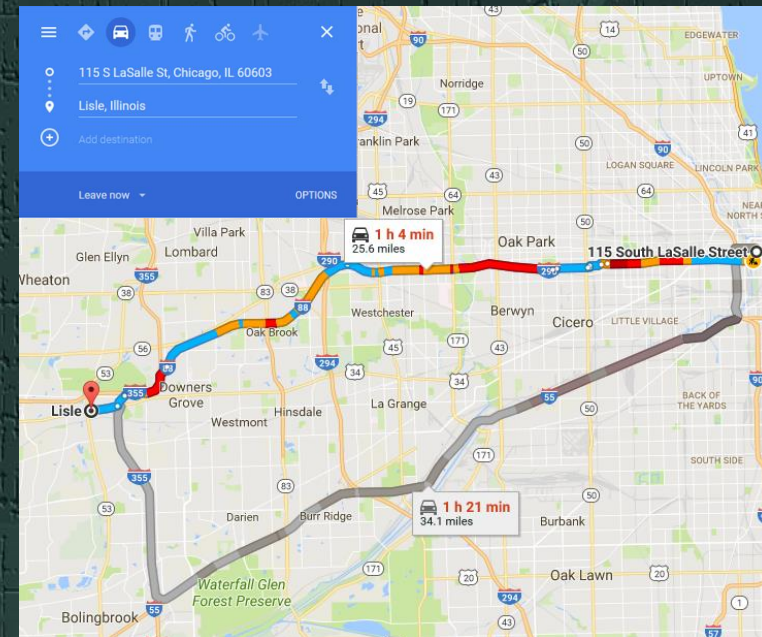
Analysis Approach

Differences in Mean Travel Time Estimates and Sampling Errors

Differences in Travel Time Bin Distributions

CTPP Sampling Error and Relative Error Relationship

Analysis of Variance (ANOVA)

Cochran-Mantel-Haenszel (CMH) Statistics

Correlation Analysis

CAMBRIDGE SYSTEMATICS

Preliminary Results

Hypothesis 1A: Mean Travel Times

➤ CTPP vs. Google Maps (Main Effect)

» 26.3 vs. 23.7 Minutes – Statistics → Significant (N>3,000)

➤ Differences Across the Strata (Interactions)

» Minor Differences in Size (Low, 4 minutes) and Worker Density (High and Mid; 3 minutes)

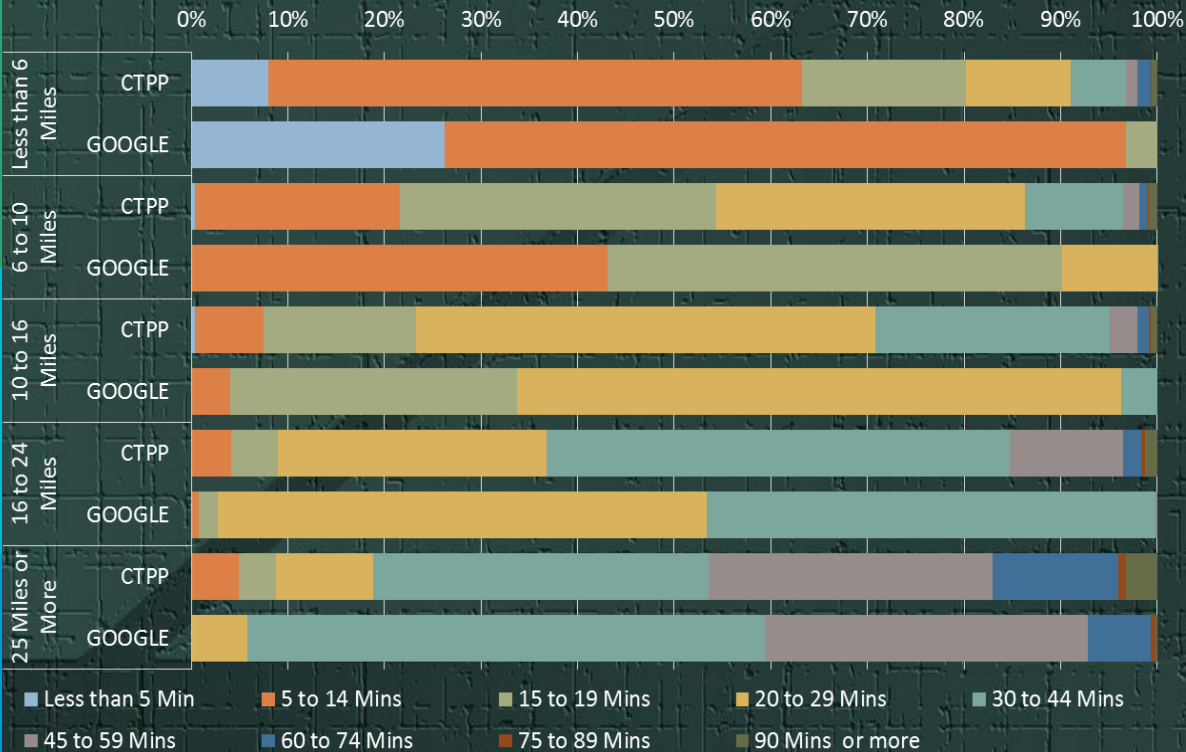» Varying results across Distance categories

CAMBRIDGE SYSTEMATICS
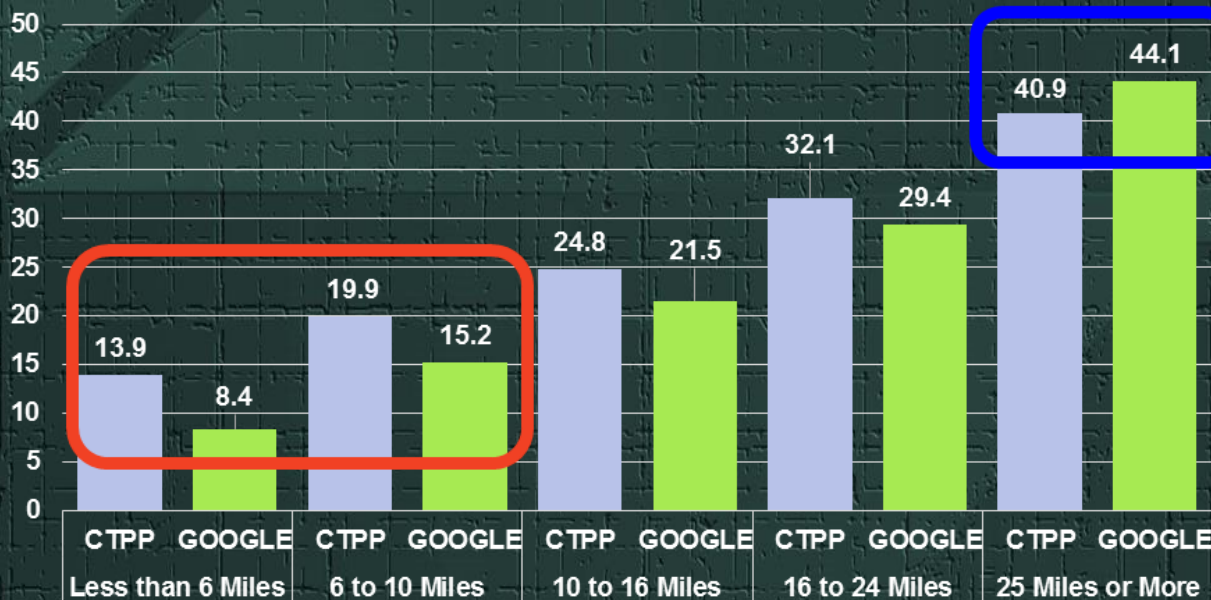
Hypothesis 1B: Travel Time Distributions
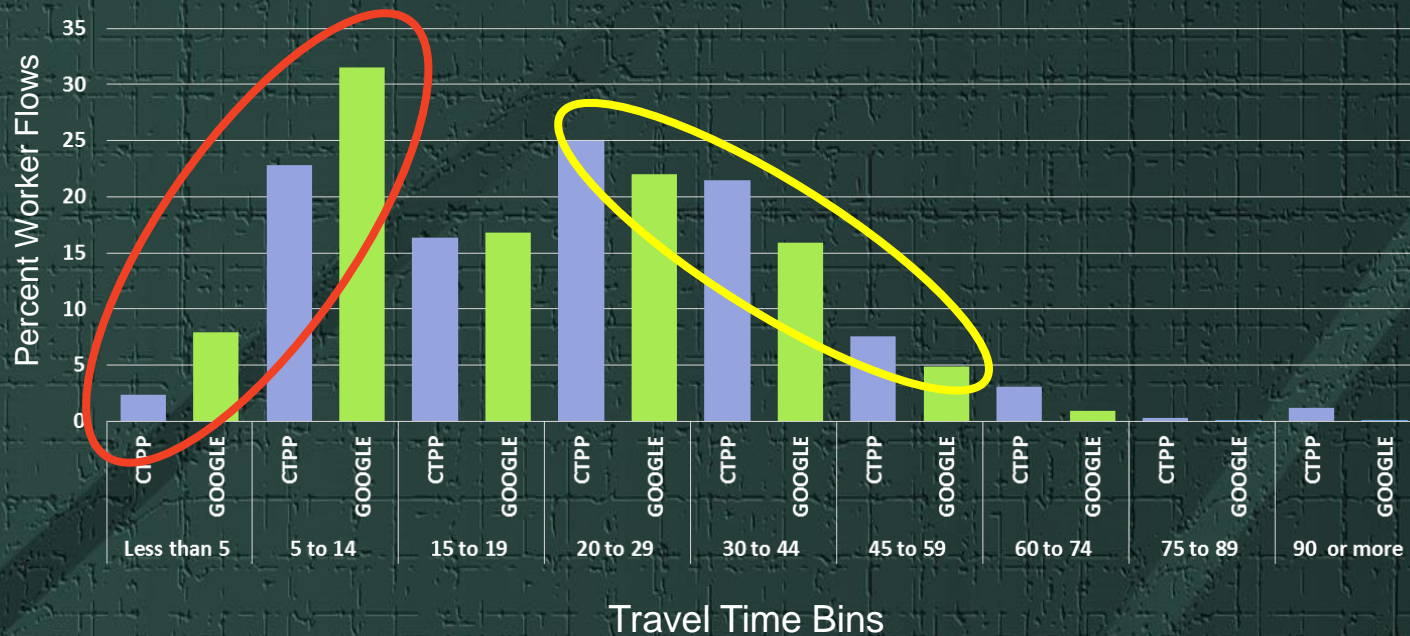
➤ Similar effects across the Size categories.
  » Greater differences between small and large tracts in first 3 bins.

➤ Similar effects across the Worker Density categories.

CAMBRIDGE SYSTEMATICS

Hypothesis 1B: Travel Time Distributions

- Statistical differences across Distance categories.

- CTPP shows a higher level of variance in reported travel times.

- Google data showed higher shares of lower travel time bins per Distance category.

CAMBRIDGE SYSTEMATICS

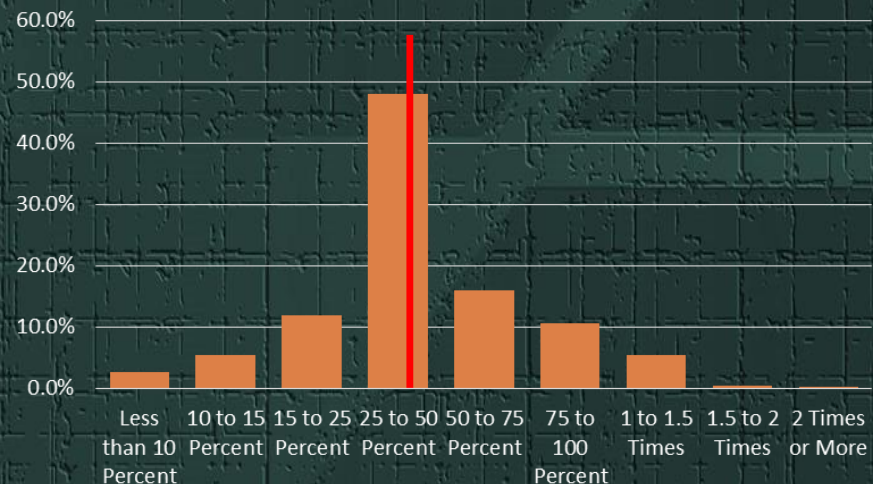# Hypothesis 2: Accuracy vs. Sampling Error

Relative Error

$$\frac{ABS(EST_{CTPP} - EST_{GOOGLE})}{EST_{GOOGLE}} X\ 100$$
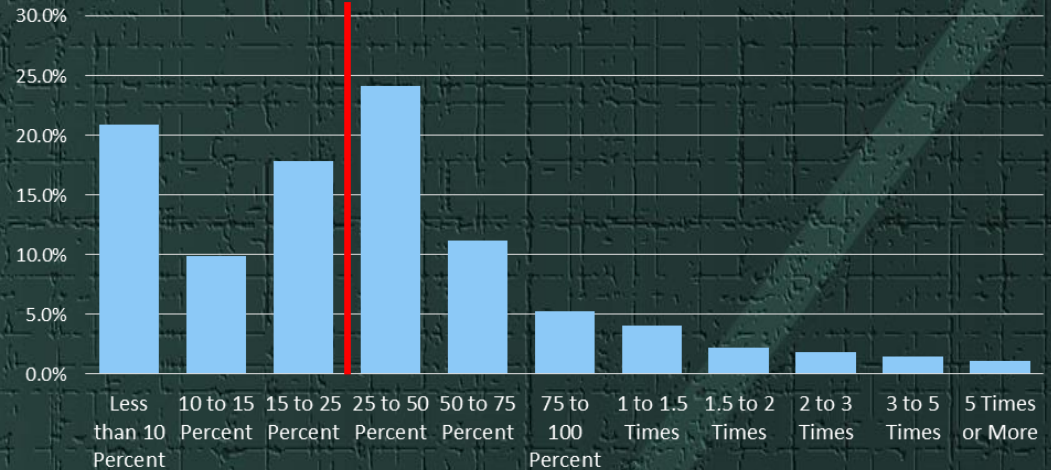
Sampling Error (Relative SE)

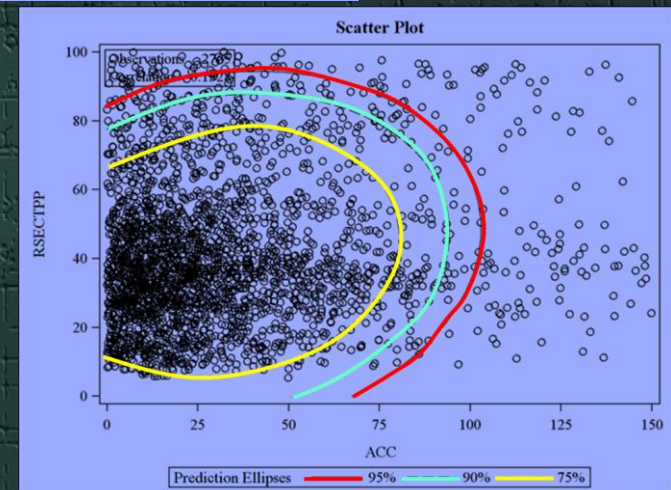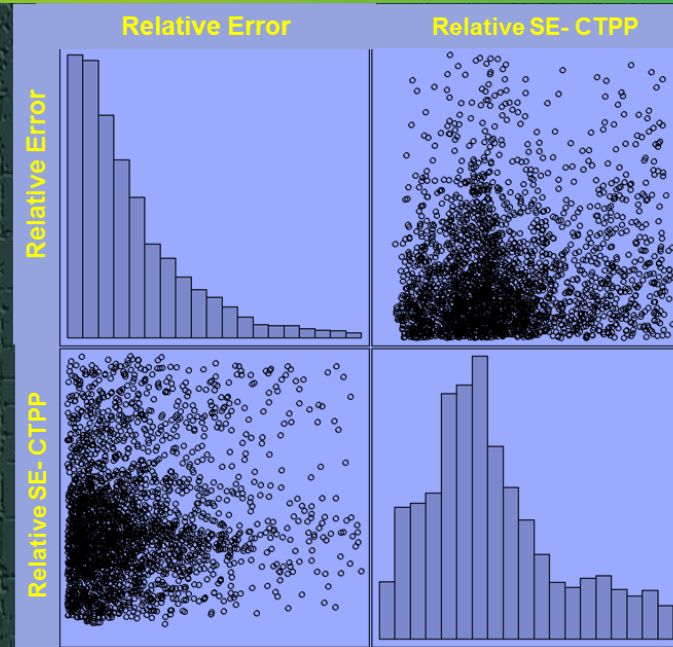$$\frac{SE_{CTPP}}{EST_{CTPP}} X\ 100$$

EST = Mean Travel Time

Relative Error

Relative Standard Error

CAMBRIDGE SYSTEMATICS

Hypothesis 2: Accuracy vs. Sampling Error

Correlation = 0.133

95% CI = (0.096 – 0.169)

FACTOR LEVELS	SIZE	WORKER DENSITY	AERIAL DISTANCE
LOW	0.127	0.108	0.070
LOW MID			0.255
MID	0.128	0.136	0.199
MID HIGH			0.167
HIGH	0.148	0.178	0.141

# Hypothesis 3: Relative Errors (Google – CTPP)

➔ ANOVA to test group differences
   » Mean Sample Relative Error = 36%
   » All main effects and interaction between Size and Worker Density were significant

➔ Flows with higher levels of error:
   » From smaller (Size) residence tracts (43 vs. 33%)
   » To mid level Worker Density tracts (39% vs. 35%)
   » Shorter Distance commutes with OD tracts that are within 10 miles (51% and 38% vs. 31%)

➔ Interactions
   » Mid and High Worker Density tracts showed interaction effects with Size.

CAMBRIDGE SYSTEMATICS
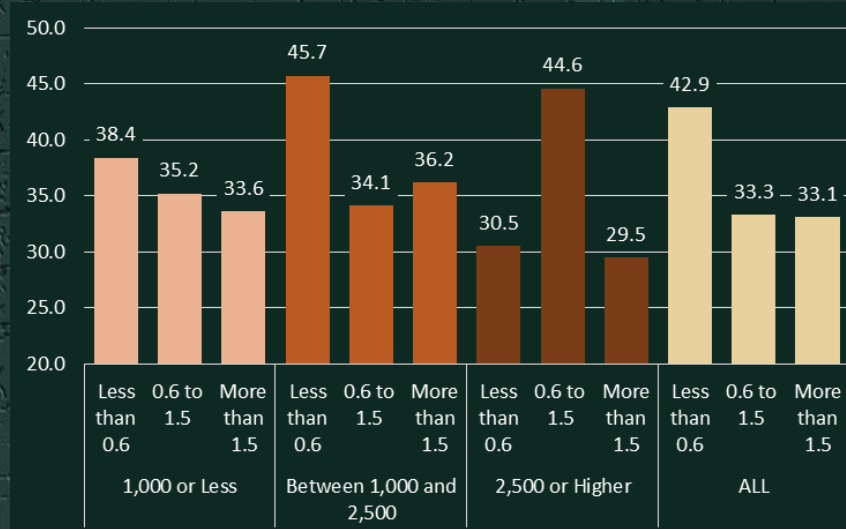
Hypothesis 4: Sampling Errors

➔ Compare sampling errors between CTPP and Google Estimates

» Google SE are much lower than those in CTPP

➔ CTPP errors slightly (± 5 percent) varied within every factor

➔ Google errors did not vary across Size and Worker Density categories

➔ Google errors were inversely related to distance

» 10 percent for 6 Miles or less

» 2 Percent for 25 Miles or more

CAMBRIDGE SYSTEMATICS

Conclusions and Recommendations

Conclusions

➤ A first step for a framework comparing CTPP to an external source.

» Synthetic approach to link point-level O-Ds to approximate tract level flows.

» Statistical tests for differences across market segments.

➤ CTPP and Google Maps average travel times showed agreement at the regional level.

» Differences were observed for short and long distance commutes

» CTPP showed greater variance in travel time distributions within distance levels.

➤ Correlations between sampling error and accuracy are very weak across strata.

➤ Findings on comparing sampling errors between the CTPP and procedure is preliminary.

CAMBRIDGE SYSTEMATICS

Recommendations

Procedural Improvements

- Better ACS process synthesis in sample building

- Testing new factors

- Simulations of synthetic sample draws

Added Value

- Quality Control in data production

- Additional data for users

- Validation of published SEs

Research

- Variance estimation (re-estimation, aggregation)

- Extensions for other sources (LEHD, Activity Based Models)

CAMBRIDGE SYSTEMATICS

Thank You

Cemal Ayvalık
Principal | Regional Manager
Travel Demand Forecasting
Cambridge Systematics, Inc.
115 S. LaSalle Street, Suite 2200
Chicago, IL 60603
t 312 665 0209

www.camsys.com

CAMBRIDGE SYSTEMATICS