



CAMBRIDGE
SYSTEMATICS

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

presented to

**Chicago Area Travel
Model User Group**

presented by

Cemal Ayvalik

(cayvalik@camsys.com)

April 4, 2018

Outline

- Background
- Motivation
- Study Design
- Data and Analysis
- Preliminary Results
- Conclusions
- Recommendations

Background

- **Census Transportation Planning Products (CTPP) – Special tabulations of Census data, now ACS**
 - » Pioneered by AASHTO in partnership with all states to support transportation policy and planning efforts
- **Transportation Tables**
 - » Demographic characteristics of home locations
 - » Characteristics of work locations
 - » Commuting patterns and modal/temporal distributions
- **2006-2010 CTPP features**
 - » First CTPP based on 5-year ACS data for small geographic units such as Census Tracts and Traffic Analysis Zones
 - » 343 tables for over 200,000 geographies.

2006-2010 CTPP

- Released October 31, 2013
- Reduced sample size
- Spanning over five years
- Data quality for small geographies
 - » Need to incorporate uncertainty in the estimates
 - » New disclosure proofed data
- Updated data dissemination software
- Extensive training materials and workshops
- Technical support

Person 1 (continued)

29 a. LAST WEEK, did this person work for pay at a job (or business)?
☐ Yes → SKIP to question 30
☐ No – Did not work (or retired)

b. LAST WEEK, did this person do ANY work for pay, even for as little as one hour?
☐ Yes
☐ No → SKIP to question 35a

30 At what location did this person work LAST WEEK? If this person worked at more than one location, print where he or she worked most last week.
 a. Address (Number and street name)

 If the exact address is not known, give a description of the location such as the building name or the nearest street or intersection.
 b. Name of city, town, or post office

 c. Is the work location inside the limits of that city or town?
☐ Yes
☐ No, outside the city/town limits
 d. Name of county

 e. Name of U.S. state or foreign country

 f. ZIP Code

31 How did this person usually get to work LAST WEEK? If this person usually used more than one method of transportation during the trip, mark (X) the box of the one used for most of the distance.
☐ Car, truck, or van
☐ Motorcycle
☐ Bus or trolley bus
☐ Bicycle
☐ Streetcar or trolley car
☐ Walked
☐ Subway or elevated
☐ Worked at home → SKIP to question 39a
☐ Railroad
☐ Other method
☐ Ferryboat
☐ Taxicab

32 How many people, including this person, usually rode to work in the car, truck, or van LAST WEEK?
 Person(s)

33 What time did this person usually leave home to go to work LAST WEEK?
 Hour _____ Minute _____ ☐ a.m. ☐ p.m.

34 How many minutes did it usually take this person to get from home to work LAST WEEK?
 Minutes _____

35 a. LAST WEEK, was this person on layoff from a job?
☒ Yes → SKIP to question 35c
☐ No
 b. LAST WEEK, was this person TEMPORARILY absent from a job or business?
☐ Yes, on vacation, temporary illness, maternity leave, other family/personal reasons, bad weather, etc. → SKIP to question 38
☐ No → SKIP to question 36
 c. Has this person been informed that he or she will be recalled to work within the next 6 months OR been given a date to return to work?
☐ Yes → SKIP to question 37
☐ No

36 During the LAST 4 WEEKS, has this person been ACTIVELY looking for work?
☐ Yes
☐ No → SKIP to question 38

37 LAST WEEK, could this person have started a job if offered one, or returned to work if recalled?
☐ Yes, could have gone to work
☐ No, because of own temporary illness
☐ No, because of all other reasons (in school, etc.)

38 When did this person last work, even for a few days?
 Within the past 12 months
☐ 1 to 5 years ago → SKIP to L
☐ Over 5 years ago or never worked → SKIP to question 47

39 a. During the PAST 12 MONTHS (52 weeks), did this person work 50 or more weeks? Count paid time off as work.
☐ Yes → SKIP to question 40
☐ No
 b. How many weeks DID this person work, even for a few hours, including paid vacation, paid sick leave, and military service?
☐ 50 to 52 weeks
☐ 48 to 49 weeks
☐ 40 to 47 weeks
☐ 27 to 39 weeks
☐ 14 to 26 weeks
☐ 13 weeks or less

40 During the PAST 12 MONTHS, in the WEEKS WORKED, how many hours did this person usually work each WEEK?
 Usual hours worked each WEEK

Assessing the Utility of 2006-2010 CTPP Data

**Develop a list
of common and
unique applications
of CTPP data**

**Assess common
issues encountered
and remedies
implemented**

**Suggest solutions,
including future
research and/or
resource development**

**Inform
decision-making
for future products**

User Survey and Peer Exchange

The key considerations included

- » Data content
- » Geographic delineation
- » Multiyear data accumulation
- » Margins of error
- » Data perturbation
- » Data dissemination and training
- » Future planning of CTPP data products

Data Content

- Different delineation of workplace data (multiple job holders, more relevant definition of part and full-time)
- More three-way residence and workplace tabulations
- Added-value tabulations such as commute distances
- Concerns with the data quality and timely release
- Unforeseen consequences of scope reduction
 - » ***Smaller CTPP***
 - » ***No TAZ or TAD level tabulations***
 - » ***Less flexible than before***

Margins of Error

- 90 percent – understand the concept, but roughly half use the CTPP data without accounting for those
- Experts use margins of error
 - » To evaluate the reasonableness of the estimates qualitatively
 - » To decide which geographic level of detail to use
- Guidance on communicating data with margins of error

Recommendations

Long-Term Census ACS Improvements

- » Second Jobs
- » Better Information on cellphone availability
- » New modes (ridesourcing) and sub travel modes (access/egress to transit)

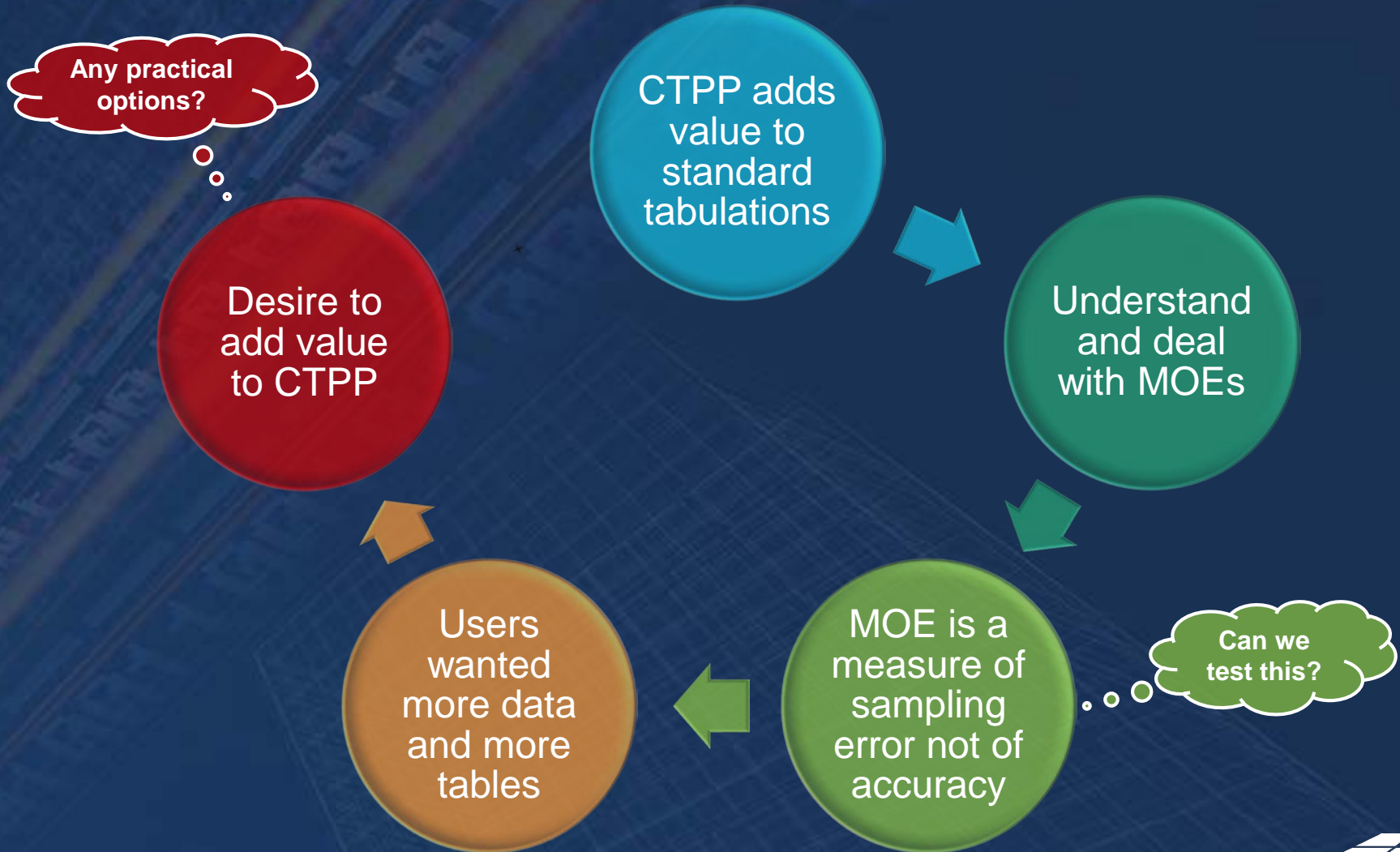
More Multiway SE Tables and Flow Tabulations

- » Age, gender
- » Employment, occupation, earnings
- » School enrollment
- » Internet access/use

Value-Added Enhancements to CTPP

- » Supplement with travel distance data,
- » Help users access multiple datasets,
- » Facilitate data fusion with other sources (such as LODS, NHTS)

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 (PPS)
 - » Collect data at a higher rate for a sample of tract pairs

Study Design

Develop and Test Sample Hypotheses

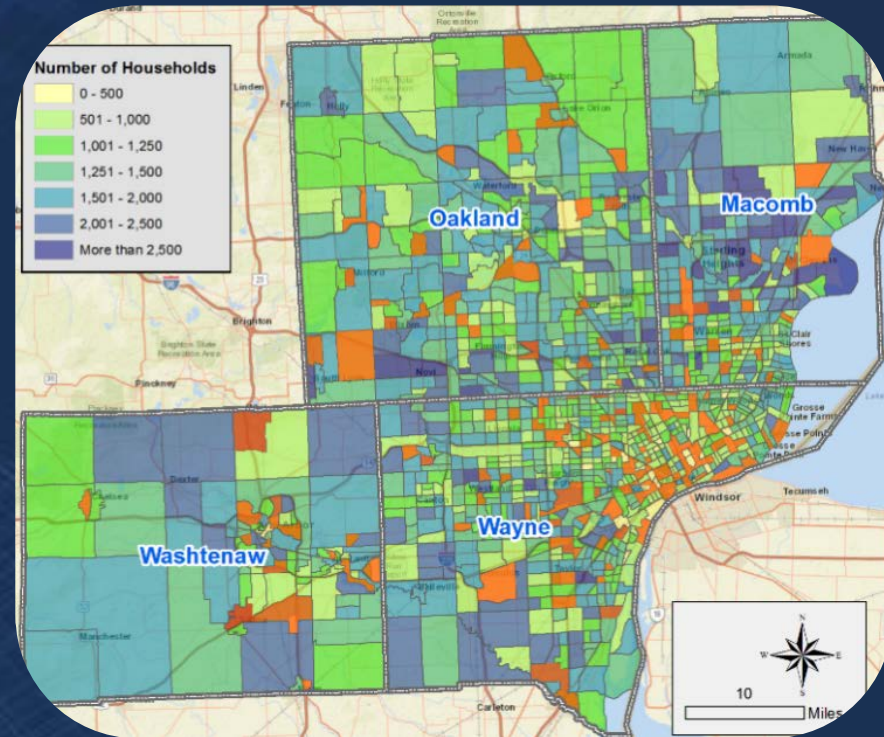
1. CTPP Mean Travel Times are Equivalent to Google Estimates by Strata
2. Accuracy of Mean Travel Times is Independent of MOE (Sampling Error)
3. Accuracy of Mean Travel Times is Independent of the Strata
4. CTPP and Google MOE are Equivalent across the Strata

Data Development and Analysis

Data Development

- Study Area: Part of the Detroit Metropolitan Area
 - » Population¹: 4.23 M
 - » Employment²: 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 –
Stratified selection
of a set of tract
pairs

Synthesize Commutes –
A set of probable O-D
pairs for each of the
selected tract pairs

Data Development – CTPP Sample

- Download CTPP Tables
 - » A112100; A110106; A202100; B306201; B302106
- Stratified sample to allow testing effects of select characteristics
 - **TRACT SIZE – Place of Residence**
 - **WORKER DENSITY – Workplace**
 - **AERIAL DISTANCE**
- 10% MOE with 90% confidence (n=70)

			AERIAL DISTANCE					
			(in Miles)					
			SIZE	4 or Less	4 to 7	7 to 11	11 to 17	More
			(Area in Sq.Miles)					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						
		Greater than 1.5						
	Greater than 2,500	Less than 0.6						
		0.6 to 1.5						
		Greater than 1.5						

Data Development – CTPP Sample

➤ Probability-Proportional-to-Size Sampling

- » $P(\text{selection}) = f(\text{size})$
- » 45 strata with 70 pairs w/o replacement
- » Used worker flows as the size variable
- » 3,150 O-D pairs were selected



Data Development – Test Sample

- Build point level O-D locations
 - » SEMCOG's Building Footprints
 - » Establishment locations (Info USA)
- PPS with replacement to select
 - » HHs from sampled residence (RES) tracts
 - » Establishments from sampled Place of Work (POW) tracts
- For each sampled tract pair, randomly match RES and POW points.
- 137,100 O-Ds in the test sample pool.

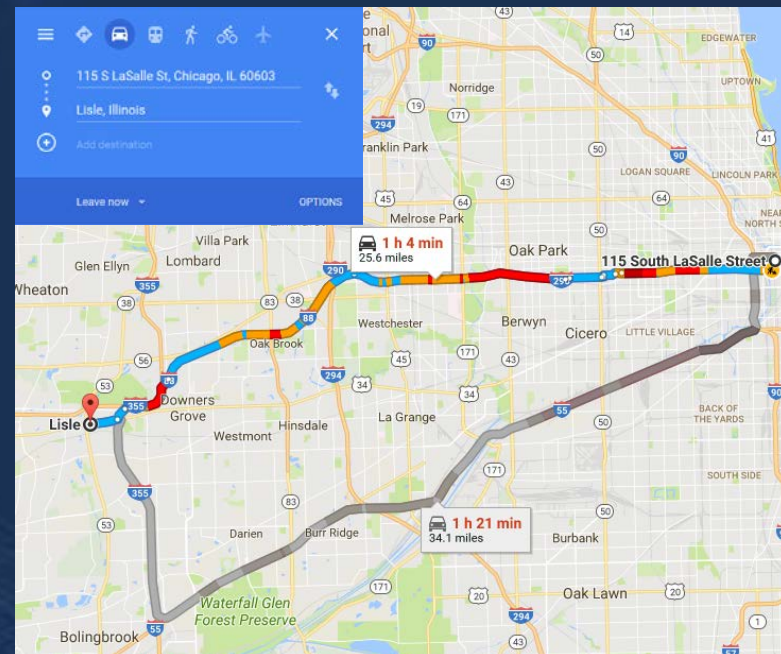


Data Development – Test Sample

- A custom built Google Maps API
 - » Lat/Lon pairs to highway travel times
 - » Collects “Directions” data at desired times and frequency

- Data collected
 - » One-month period (Late August and early October) Mondays thru Thursdays
 - » 7:00 AM to 8:30 AM @30-min intervals

- Test sample of 11,235 O-D pairs to scale to CTPP sample



Analysis Approach

Differences in
Mean Travel Time
Estimates and
Sampling Errors

Analysis of
Variance
(ANOVA)

Differences in
Travel Time Bin
Distributions

Cochran-Mantel-
Haenszel (CMH)
Statistics

CTPP Sampling
Error and Accuracy
Relationship

Correlation
Analysis

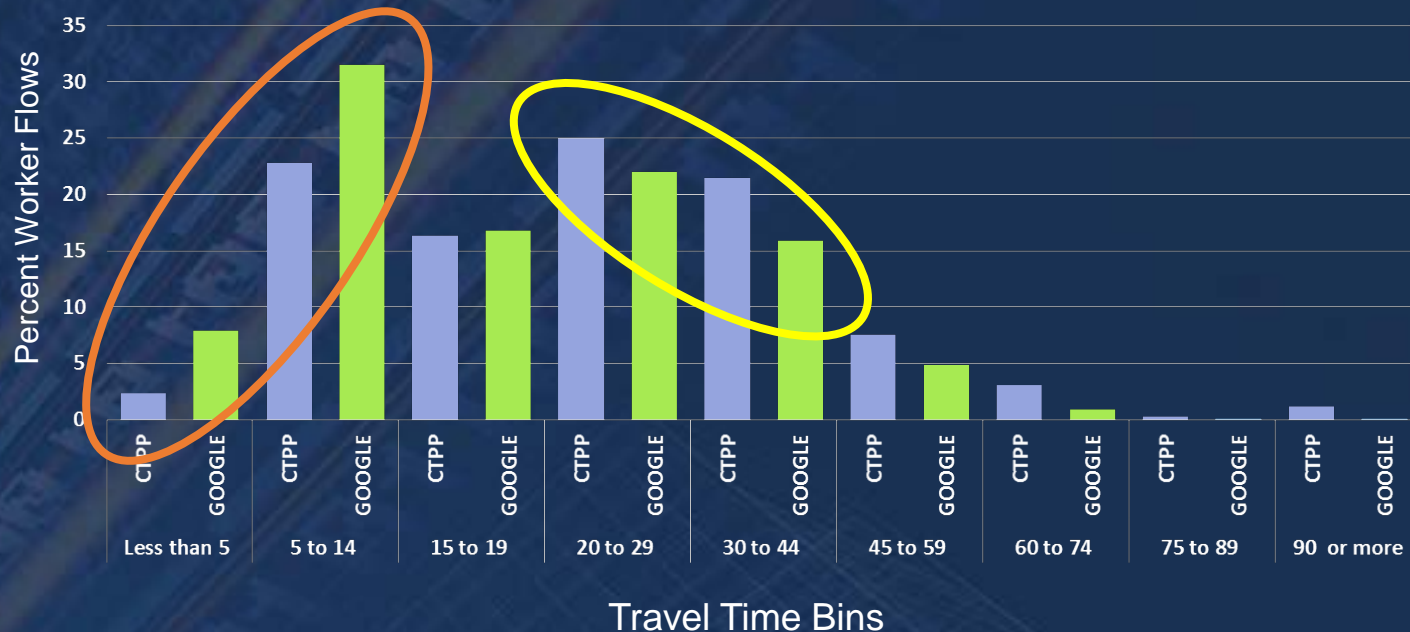
Preliminary Results

H1A: Mean Travel Times

- CTPP vs. Google Maps (Main Effect)
 - » 26.3 vs. 23.7 Minutes
 - » Small but significant (N>3,000)
- Differences Across the Strata (Interactions)
 - » Minor differences in **Tract Size** and **Worker Density**
 - » Greater variance across **Distance** categories

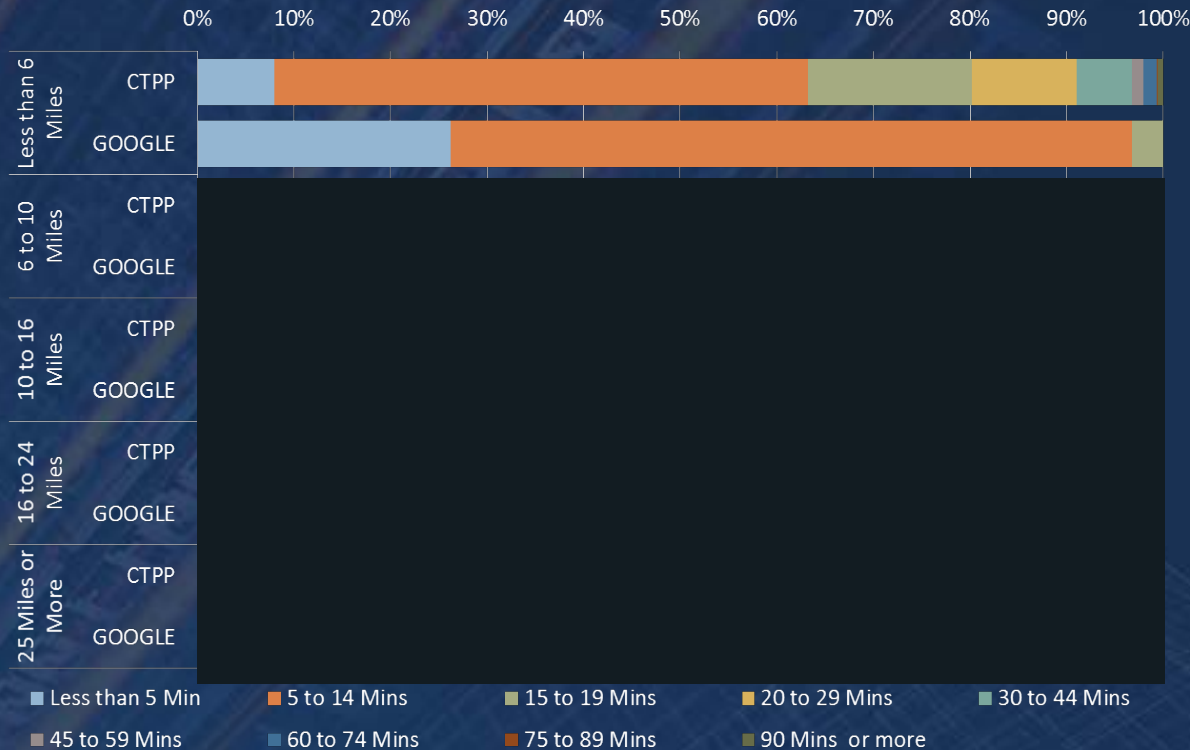


H1B: Travel Time Distributions



- Minor variations across **Tract Size** and **Worker Density** categories, greater variations in **Distance** factor.

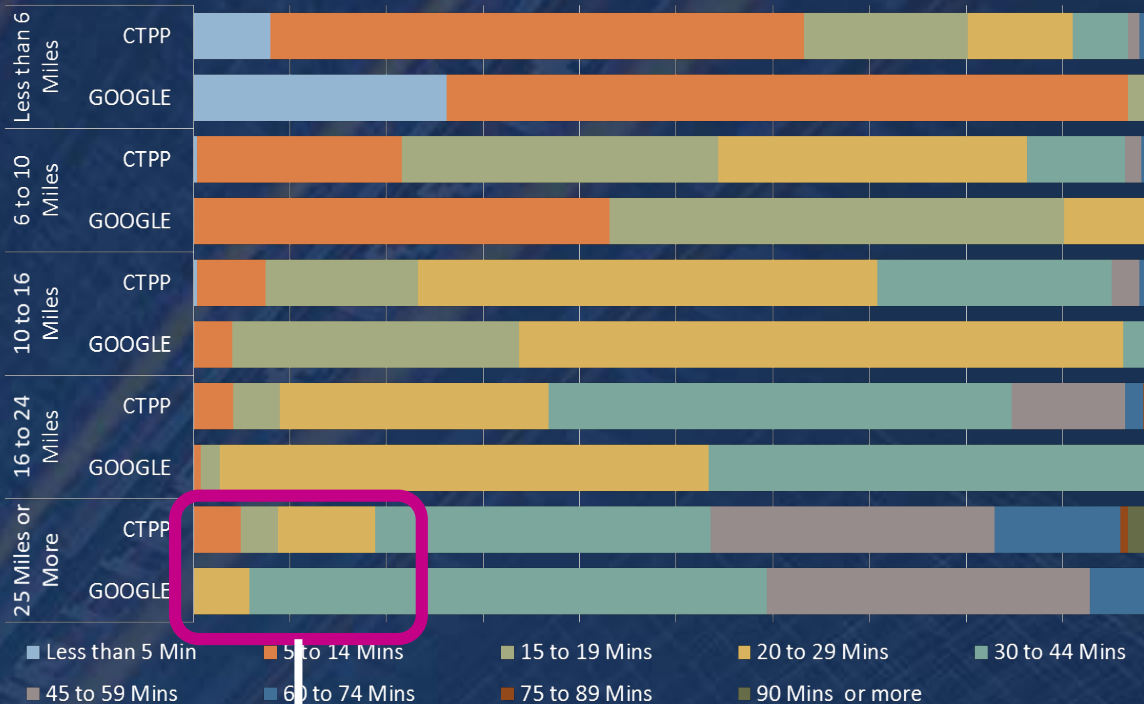
H1B: Travel Time Distributions



- CTPP & Google times statistically different by **Distance**
- CTPP data show “more noise” in reported travel times
- Google has higher share of shorter trips

H1B: Travel Time Distributions

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%



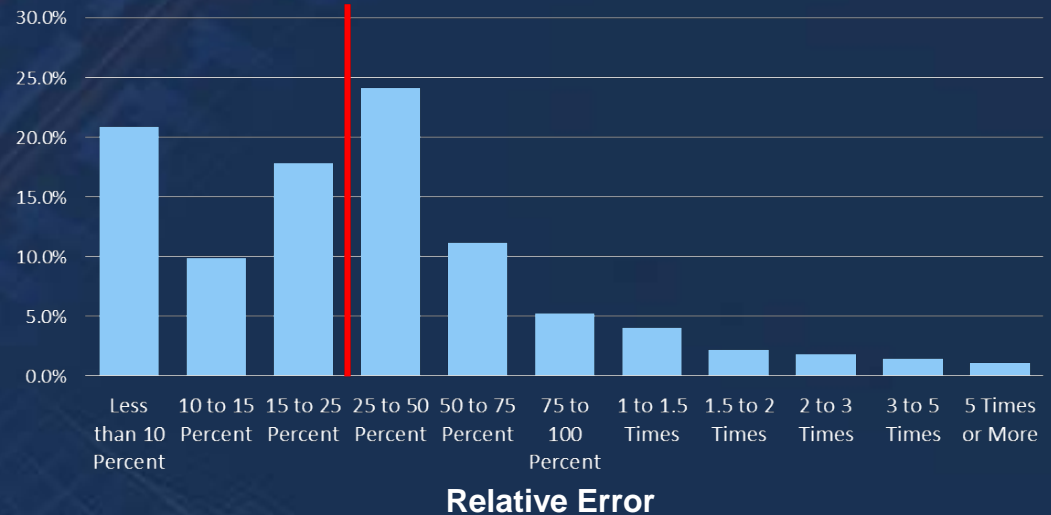
Impact of noise in CTPP data on travel time estimates for long distance commutes.



H2: Accuracy vs. Sampling Error

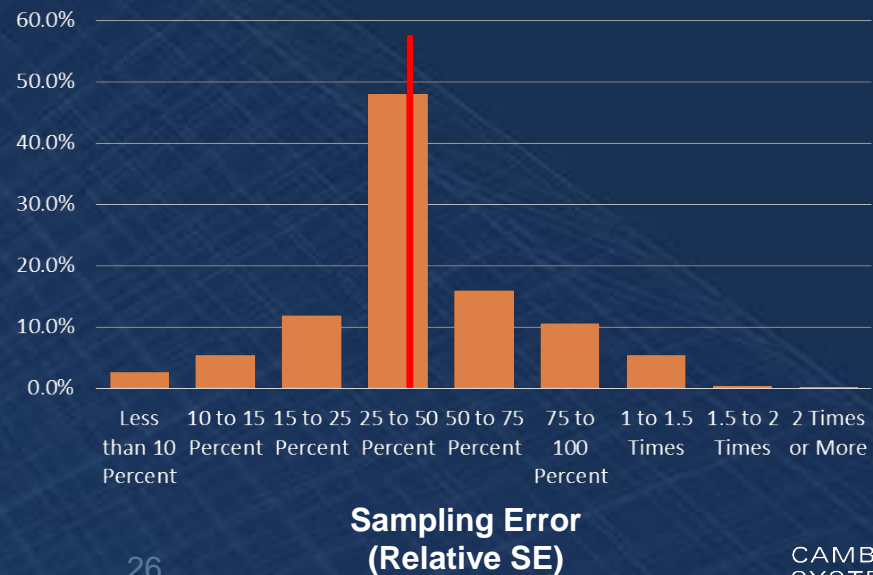
Relative Error

$$\frac{ABS(EST_{CTPP} - EST_{GOOGLE})}{EST_{GOOGLE}} \times 100$$



Sampling Error (Relative SE)

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



H2: Accuracy vs. Sampling Error

Relative Error

$$\frac{ABS(EST_{CTPP} - EST_{GOOGLE})}{EST_{GOOGLE}} \times 100$$

Sampling Error (Relative SE)

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

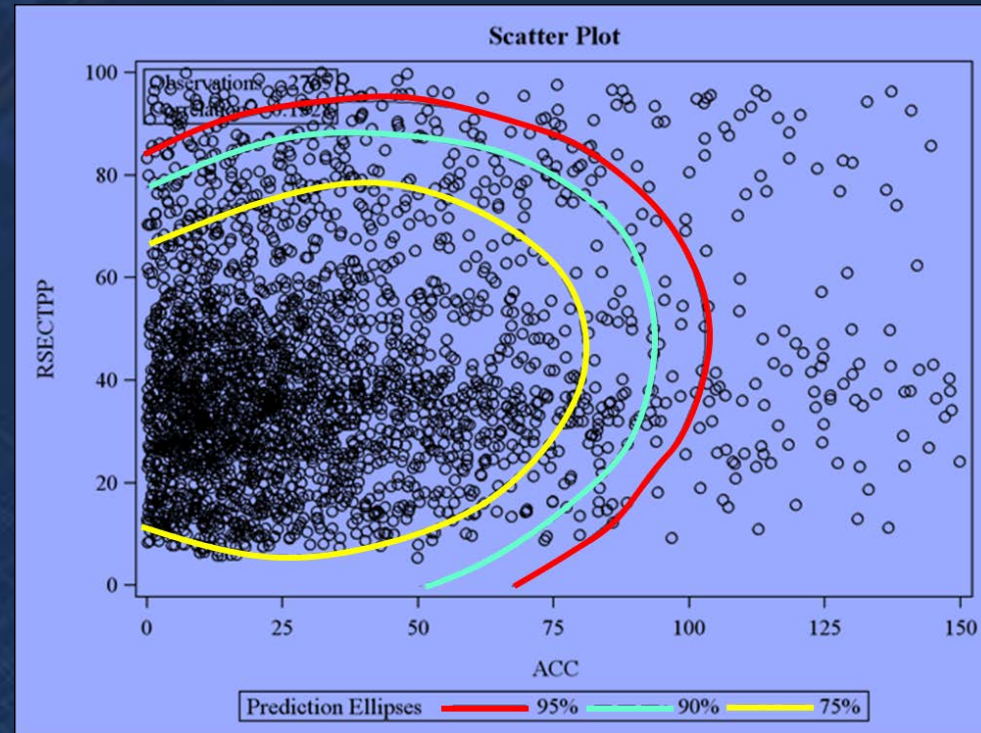
Relative Error	Relative Standard Error								
	Less than 10 Percent	10 to 15 Percent	15 to 25 Percent	25 to 50 Percent	50 to 75 Percent	75 to 100 Percent	1 to 1.5 Times	1.5 to 2 Times	2 Times or More
Less than 10 Percent	23	50	94	323	101	44	23	1	
10 to 15 Percent	13	25	39	146	58	22	9		
15 to 25 Percent	22	33	66	288	82	40	31	1	
25 to 50 Percent	14	36	87	359	129	90	42	4	
50 to 75 Percent	5	14	41	166	61	42	22	3	
75 to 100 Percent	2	4	12	72	26	31	14	2	2
1 to 1.5 Times		3	14	56	10	29	15		
1.5 to 2 Times	2	1	4	31	12	14	6		
2 to 3 Times			4	26	11	10	6		
3 to 5 Times		2	5	24	8	4	3		
5 Times or More		1	5	23	3	4			

H2: Accuracy vs. Sampling Error

Correlation = 0.133

95% CI = (0.096 – 0.169)

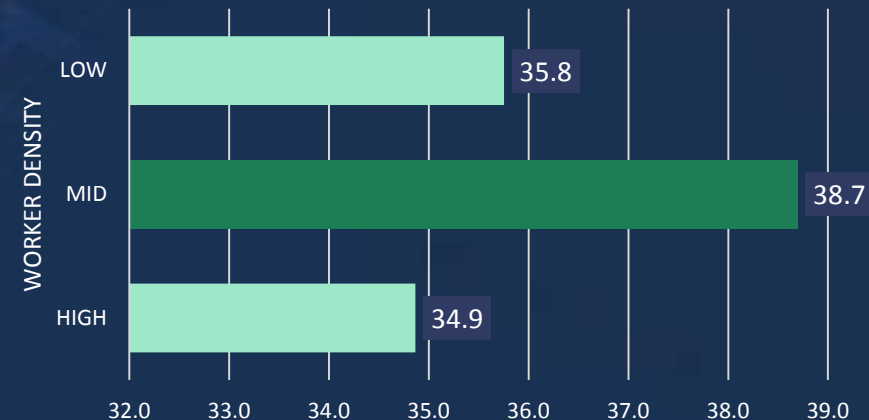
FACTOR LEVELS	SIZE	WORKER DENSITY	AERIAL DISTANCE
LOW	0.127	0.108	0.070
LOW MID			0.255
MID	0.128	0.136	0.199
MID HIGH			0.167
HIGH	0.148	0.178	0.141



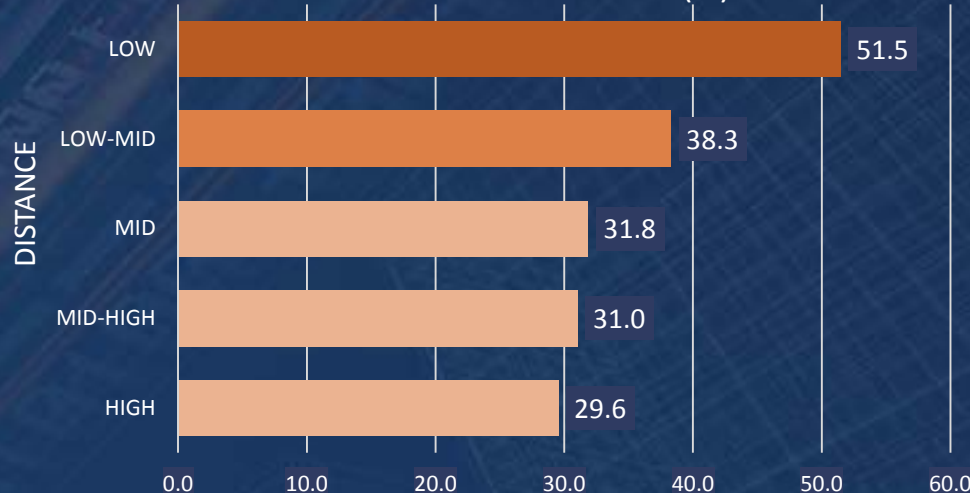
H3: Relative Errors by Strata (Accuracy)

- Test differences in mean travel time by strata
- Flows with higher levels of error:
 - » Smaller residential tracts
 - » Mid level worker density tracts
 - » Shorter distance commutes

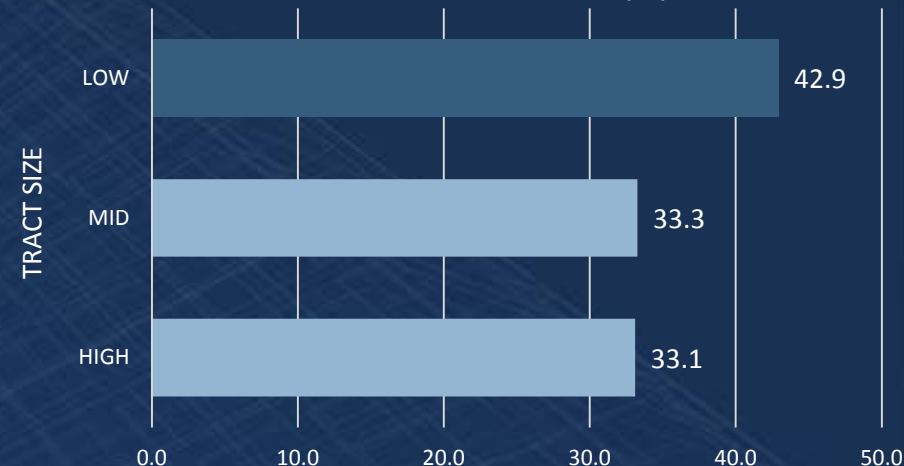
LSMEAN - Relative Errors (%)



LSMEAN - Relative Errors (%)



LSMEAN - Relative Errors (%)



H4: MOE by Strata (Sampling Errors)

- Compared sampling errors in CTPP vs. Google
 - » Google SEs are much lower
- CTPP errors are similar across strata
- Google errors did not vary across **Tract 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”

Conclusions and Recommendations

Conclusions

- **A first step: Comparing CTPP to an external source**
 - » Synthetic approach to pair point-level O-D
 - » Examine differences across market segments
- **CTPP and Google mean travel times similar at the regional level**
 - » Differences for short and long distance commutes
 - » CTPP showed greater noise in travel time distributions
- **Little correlations between sampling error and accuracy**
- **Early comparisons of sampling errors between CTPP & Google**
- **Promise of data fusion with traditional data sources**

Recommendations

Procedural Improvements

- Better ACS process synthesis in sample building
- Access/egress consideration
- Test new factors

Added Value

- Quality Control
- Additional data for users
- Validation of published SEs

Research

- Impute demographics to add more dimensions to CTPP
- Add travel time and demographics to LEHD

New CTPP is Coming



2012 - 2016 CTPP in early 2019

Web: <http://ctpp.transportation.org>

Listserv: <http://www.chrispy.net/mailman/listinfo/ctpp-news>

Technical Support: CTPPSupport@camsys.com

Small Geography Policy Change:

<http://ctpp.transportation.org/Pages/Policy-Change-on-Small-Geography.aspx>

Q&A

Q: Accuracy vs. Sampling Error

Low correlations between accuracy and MOEs imply presence of both good estimates with large MOEs and poor estimates with low MOEs.

A cursory analysis on the right shows a relatively favorable picture for the good estimates with large MOEs (805 vs. 517 tract pairs) for the data used in the study.

Relative Error

$$\frac{ABS(EST_{CTPP} - EST_{GOOGLE})}{EST_{GOOGLE}} \times 100$$

Sampling Error (Relative SE)

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

Relative Error	Relative Standard Error								
	Less than 10 Percent	10 to 15 Percent	15 to 25 Percent	25 to 50 Percent	50 to 75 Percent	75 to 100 Percent	1 to 1.5 Times	1.5 to 2 Times	2 Times or More
Less than 10 Percent	23	50	94	323	101	44	23	1	
10 to 15 Percent	13	25	39	146	58	22	9		
15 to 25 Percent	22	33	66	288	82	40	31	1	
25 to 50 Percent	14	36	87	359	129	90	42	4	
50 to 75 Percent	5	14	41	166	61	42	22	3	
75 to 100 Percent	2	4	12	72	26	31	14	2	2
1 to 1.5 Times		3	14	56	10	29	15		
1.5 to 2 Times	2	1	4	31	12	14			
2 to 3 Times			4	26	11	10	6		
3 to 5 Times		2	5	24	8	4	3		
5 Times or More		1	5	23	3	4			