Socioeconomic Modeling for Activity Based Models

Simon Choi and Cheol-Ho Lee
Southern California Association of Governments

presented to
COG/MPO Mini Conference on Socioeconomic Modeling

July 17, 2009

SOUTHERN CALIFORNIA ASSOCIATION of GOVERNMENTS

Resolving Regional Challenges

Metropolitan Planning Organization

Why Socioeconomic Modeling?

- Consistent and Accurate Socioeconomic data set both
  - At different levels of geography
  - Temporal perspective
- Federal & State requirements:
  - RTP modeling
  - EJ analyses
  - Air quality conformity
  - Regional Housing Needs Allocation (CA)

Expanded State Requirements for RTP Guidelines & SB 375

- California Transportation Commission adopted an "Addendum to the 2007 Regional Transportation Plan (RTP) Guidelines,". Four largest MPOs in California should develop activity-based models (ABMs) within a few years, to improve modeling assessment on key policy options on reducing greenhouse gas (GHG) emissions during the RTP process.
- California Senate Bill No. 375 (SB 375) was enacted. Travel demand models must assess the effects of land use policies, transit service, congestion pricing, and economic incentives on travel.

Timeline of Activity Based Model Implementations in the United States

Region Organization	Started	First Implemented	Planned Implementation	Use in the Latest RTP Analysis
Portland METRO	1996	1998		No
SF County SFCTA	1999	2001		N/A
New York NYMTC	2000	2003		Yes
Columbus MORPC	2001	2005		Yes
Sacramento SACOG	2005	2006		Yes
Atlanta ARC	2002		2009	N/A
Bay Area MTC	2006		2009	N/A
Denver DRCOG	2006		2009	N/A
Seattle PSRC	2008		2011	N/A
San Diego SANDAG	2008		2011	N/A
Los Angeles SCAG	2009		2011	N/A
Tampa Bay for FDOT	2009 (?)		2010	N/A

Other Activity-Based Models

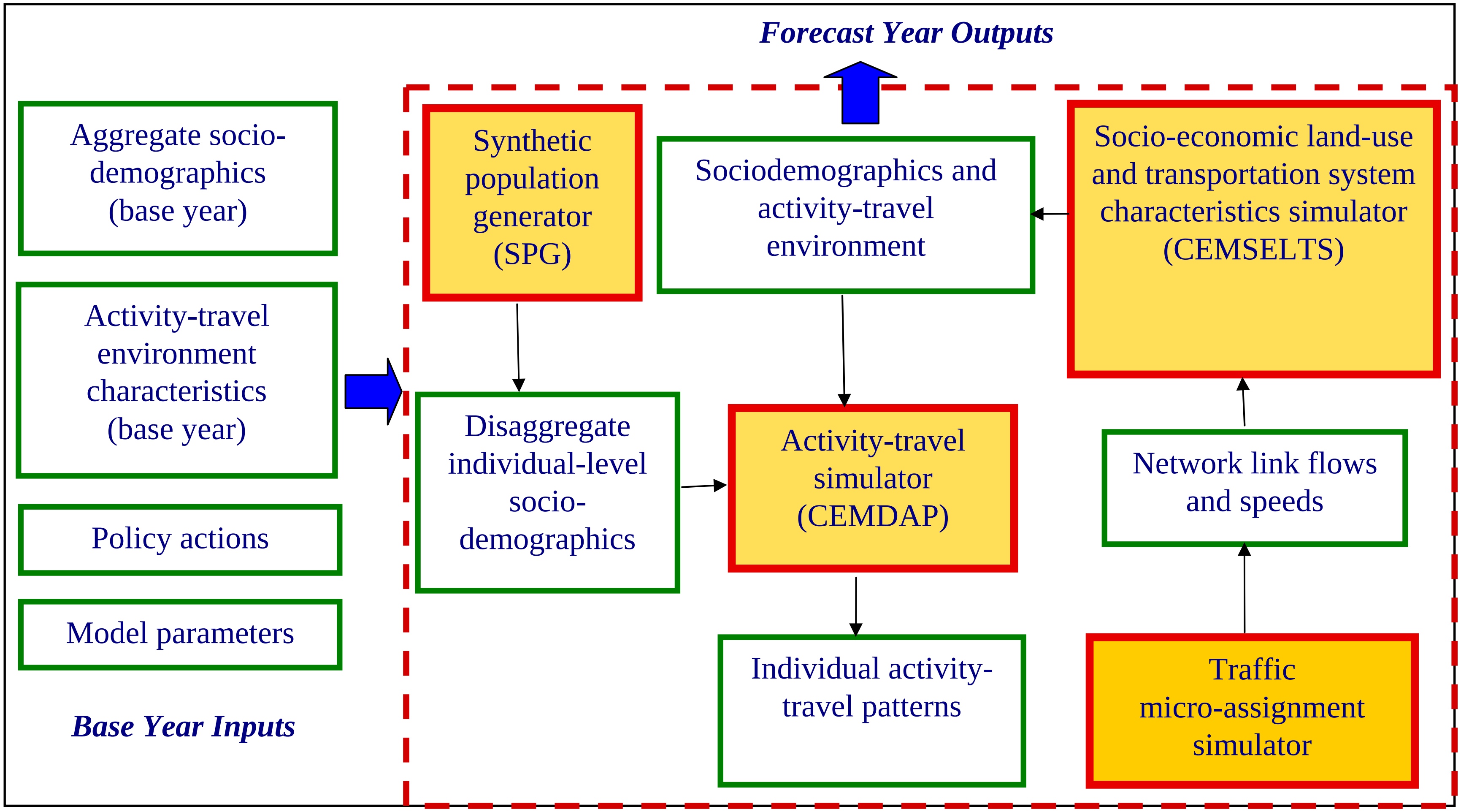
Region	Developer	University	Model Name	Note
Dallas NCTCOG	Bhat et al	Univ. of Texas at Austin	CEMDAP	Developed but not yet implemented and used by planning agency
Miami for FDOT	Pendyala et al	Univ. of South Florida	FAMOS	Developed but not yet implemented and used by planning agency
Portland METRO	Gliebe et al	Portland State Univ.		Under development

Source: Mark Bradley & John Bowman, SCAG PROJECT 09-012 Strategy for Activity-Based Travel Demand Model Development with Travel Survey, Technical Memorandum 2, Activity-Based Models for Major MPOs for Southern California Association of Governments, January 22, 2009.

SCAG ABM: Two-Phase Model Development Approach

- Phase 1: April 2009-January 2010
  - Apply an existing activity-based model that has been implemented by a planning agency in the U.S. and incorporate it into SCAG's existing trip-based model.
  - Use the current zonal socioeconomic data, and would use externally determined model coefficients.
  - The activity-based model software and the SCAG trip-based model system with TransCAD software would be adapted as needed so that the entire modeling system would converge satisfactorily and run within a reasonable amount of time.
- Phase 2: January 2010-June 2011
  - Develop a full-function activity-based model.

CEMDAP II



Source: Abdul Pinjari, Naveen Eluru, Rachel Copperman, Ipek N. Sener, Jessica Y. Guo, Sivaramakrishnan Srinivasan, Chandra R. Bhat., FHWA/TX-07/0-4080-8 Activity-Based Travel-Demand Analysis for Metropolitan Areas in Texas: CEMDAP Models, Framework, Software Architecture and Application Results for Texas Department of Transportation, October 2006.

Socioeconomic Data Needs for Activity-Based Model

- TAZ level socioeconomic data for base year and any forecast year
  - Employment (retail, service, and other), population, households, median household income
- Aggregate distribution of household-level & person-level variables for SPG
  - Person-level variables: gender, race, and age
  - Household-level variables: household type, family or non-family household, age of household head, household size, presence of children
- Additional Demographic Variables for Population for CEMSELTS
  - Person-level variables: education and employment characteristics
  - Household-level variables: residential tenure, dwelling unit type, and auto ownership at the household level.

Individual-Level Control Variables for the Base/Forecast Year

Variable Name	Value	Value Description
P_RACE	0	White alone
	1	African-American alone
	2	American Indian and Alaska Native alone
	3	Asian alone
	4	Native Hawaiian and other Pacific Islander alone
	5	Some other race alone
	6	Two or more races
P_GENDER	0	Male
	1	Female
P_AGE	0	Under 5 years
	1	5 to 14 years
	2	15 to 24 years
	3	25 to 34 years
	4	35 to 44 years
	5	45 to 54 years
	6	55 to 64 years
	7	65 to 74 years
	8	75 to 84 years
	9	85 and more

Household-Level Control Variables for the Base/Forecast Year

Variable Name	Value	Value Description
HH_FAM	0	Family
	1	Non-family
HH_TYPE	1	Family: married couple
	2	Family: male householder, no wife
	3	Family: female householder, no husband
	4	Non-family: householder alone
	5	Non-family: householder not alone
HH_CHILDREN	0	No own children under 18
	1	Own children under 18
HHLDER_AGE	0	15-64
	1	65 and over
HH_SIZE	0	1 person
	1	2 persons
	2	3 persons
	3	4 persons
	4	5 persons
	5	6 persons
	6	7 or more persons

Developing Socioeconomic Data (SED) for ABM Model Phase 1

- County SED: PROFAMY Extended Cohort-Component Method for Households (Household and Consumption Forecasting, Inc., Household Projections for Southern California Six Counties and Region, 2000-2050 for Southern California Association of Governments, June 2009)

- TAZ SED: Locally Weighted Scatterplot Smoothing (LOESS) (Heonsoo Park, Advanced Programming Support for Developing Traffic Analysis Zone/Grid Cell Socioeconomic Data and Assessing Selected Small Area Allocation Models for Southern California Association of Governments, June 2009)

County SED:PROFAMY Extended Cohort-Component Method for Households

- Zeng, Vaupel, and Wang (1997; 1998) developed a two-sex dynamic macro model for projections of households and living arrangement. Zeng, Land, Wang, and Gu (2006) further extended the initial model.
- A multi-state accounting model & Use age-sex-status-specific schedules of demographic rates and summary parameters thereof to specify projected demographic rates in the future years.
- The model groups all individuals of the population and projects forward the groups' status changes by cohort and by age, sex, race (optional), marital/union status (including cohabitation), parity, number of co-residing children and parents, rural or urban (optional), and whether living in a private or institutional household.

County SED:PROFAMY Extended Cohort-Component Method for Households

- The basic mechanism of this dynamic household projection model is that forecasts are made about the components (marriage/union, fertility, leaving parental home, mortality, and migration) of changes in demographic parameters that produce household distributions in the future years. This is analogous to, and a substantive extension of, the cohort component population projection model

- While most other macro-simulation models, which require stringent data on transition probabilities of household type statuses, this model requires as input only conventional demographic data (vital statistics, censuses, surveys), as listed in Table 1, to compute the individual groups' status changes by cohort and age. These data can be obtained from vital statistics, censuses, and routinely conducted surveys.

Table 1. Data and the data sources for household projections at the national, state, and sub-state Levels

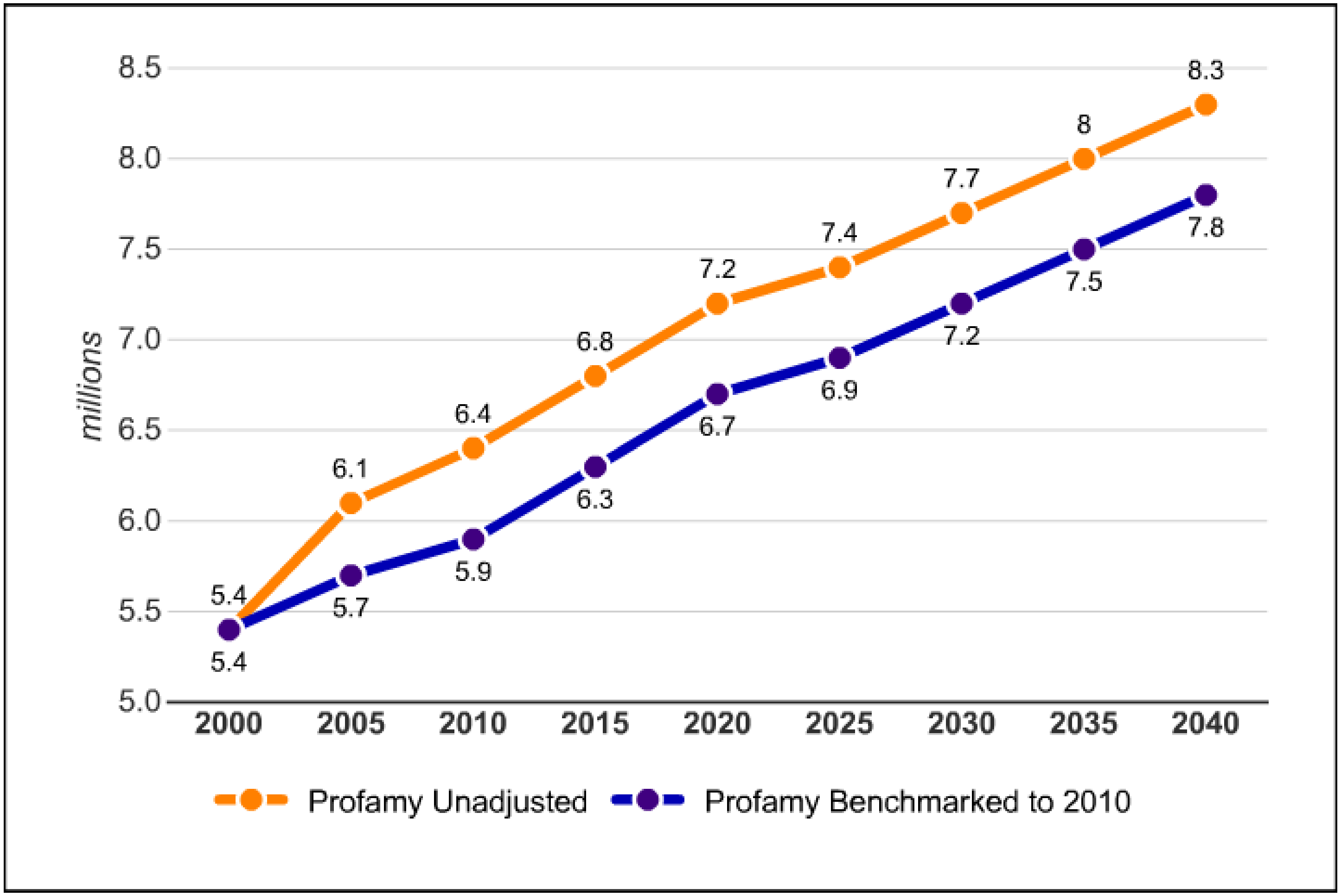
Contents of Data	Sources
(1) Base Population for the nation, states and sub-state areas.	Census PUMS 5%, American Community Survey
(2) Model standard schedules at the national level (not necessary for the states and sub-state areas) For example, For example, race-sex-age -specific o/e rates of marriage/union formation and Dissolution , race-sex-age specific net rates of leaving the parental home, estimated based on two adjacent census micro data files and the intra cohort iterative method (Coale1984; 1985; Stupp 1988; Zeng, Coale et al., 1994).	Pooled NSFH, NSFG, CPS, SIPP data sets,
(3) Demographic summary measures for the nation and states (not necessary for sub-state areas). For example, Race-specific Total Fertility Rates (TFR) by parity	Based on estimates released by the Census Bureau and the National Center for Health Statistics

A Comparison between ProFamy Projected and ACS 2005-2007 Estimation: SCAG Region

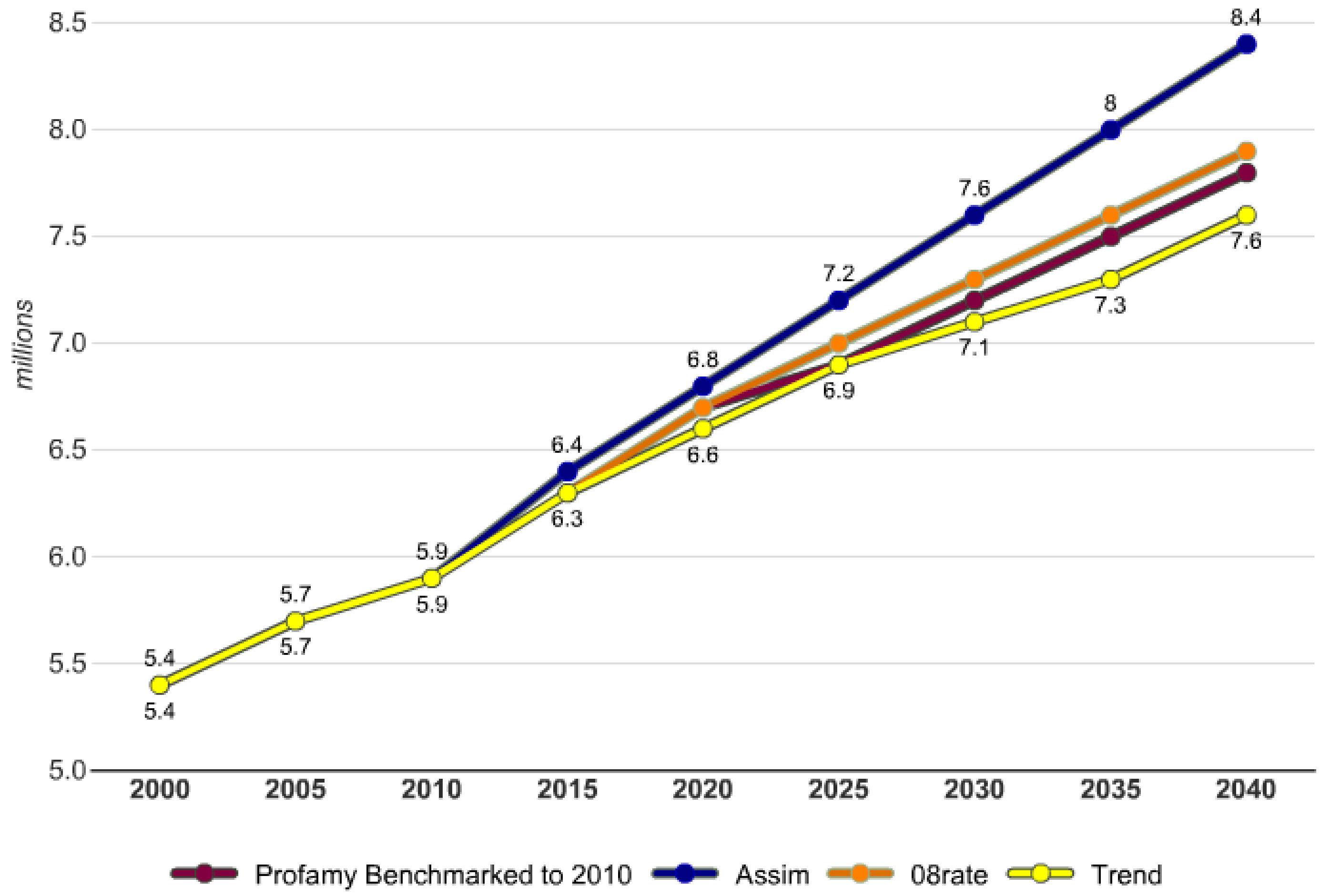
	ACS	Profamy	% Difference
Total Number of Households	5,677,465	6,107,566	7.6
Average Household Size	3.1	3.0	-3.2
% 1 Person Household	23.8	22.8	-4.2
% 2-3 Person Household	43.8	44.5	1.6
% 4+ Person Household	32.4	32.7	0.9
% Couple Household	55.4	55.8	0.7

Note: (1) ACS：ACS 2005-2007 estimation; (2) PROF: ProFamy projection in year 2006; (3)%Difference = (ProFamy-ACS)/ACS*100

Profamy Household Projections Benchmarked to 2010: SCAG Region



Profamy Household Projections and Alternatives: SCAG Region



Profamy Percent of Households by Type: SCAG Region

	2000	2010	2020	2030	2040
All households					
One person only	22.7%	22.6%	21.4%	20.3%	19.5%
One person and other	5.4%	6.5%	6.6%	6.6%	6.8%
Married couple, no co-residing kids	17.6%	18.9%	19.7%	19.1%	19.0%
Cohabiting couple, no co-residing kids	3.0%	4.5%	4.3%	4.1%	4.0%
Married couple, with co-residing kids	30.2%	29.7%	29.1%	29.5%	29.7%
Cohabiting couple, with co-residing kids	3.5%	3.1%	3.4%	3.5%	3.6%
Lone mother, with co-residing kids	12.7%	9.8%	10.6%	11.6%	12.1%
Lone father, with co-residing kids	4.8%	5.0%	5.0%	5.2%	5.4%

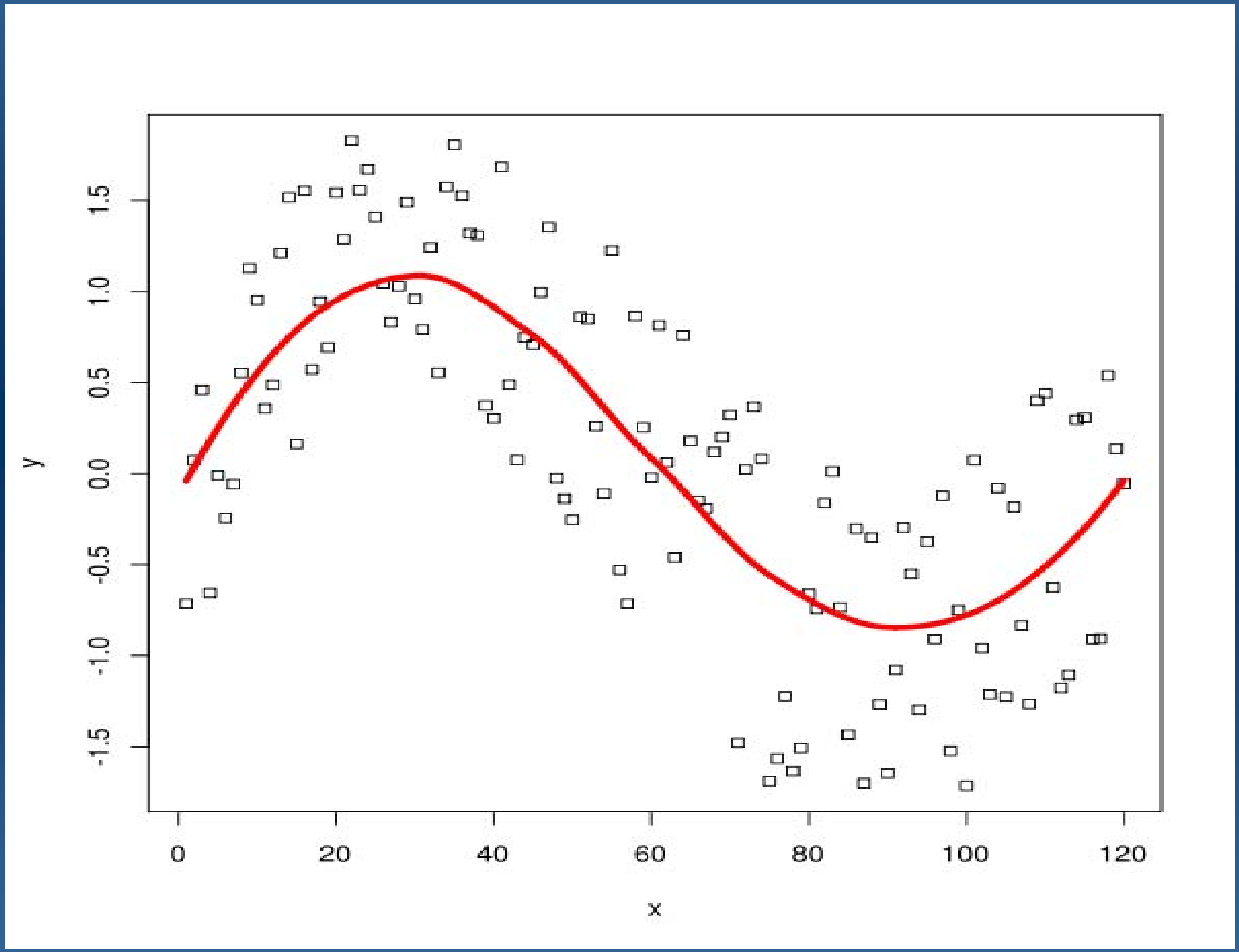
Profamy Percent of Households by Size: SCAG Region

	2000	2010	2020	2030	2040
1 person	22.7%	22.6%	21.4%	20.3%	19.5%
2 person	27.4%	29.6%	30.7%	30.5%	30.5%
3 person	16.0%	16.5%	17.4%	18.4%	19.2%
4 person	15.4%	13.7%	13.5%	14.2%	14.5%
5 person	9.0%	8.6%	8.3%	8.3%	8.1%
6 person	4.6%	4.6%	4.5%	4.4%	4.3%
7 person	2.5%	2.3%	2.2%	2.1%	2.0%
8 person	1.3%	1.1%	1.1%	1.0%	1.0%
9+ person	1.1%	1.0%	0.9%	0.9%	0.8%

TAZ SED Modeling: LOESS

- One of many "modern" modeling methods that build on "classical" methods, such as linear and nonlinear least squares regression.
- Originally proposed by Cleveland (1979) and further developed by Cleveland and Devlin (1988), specifically denotes a method that is (somewhat) more descriptively known as locally weighted polynomial regression. Cleveland and Devlin (1988) approximate the general nonparametric function with a simple linear regression, with more weight given to observations that are closer to the target zone (e.g., census tract).
- A procedure for fitting a regression surface to data through multivariate smoothing: The dependent variable is smoothed as a function of the independent variables in a moving fashion analogous to how a moving average is computed for a time series (Cleveland and Devlin, 1988, p.596).
- http://www.itl.nist.gov/div898/handbook/pmd/section1/pmd144.htm. Urban applications are found in Fu and Somerville (2001), McMillen and McDonald (1997), and Meese and Wallace (1991).

TAZ SED Modeling: LOESS



http://en.wikipedia.org/wiki/File:Loess_curve.svg

Table 1 Estimation results of Household type (Los Angeles county)

Variable	Family: married couple	Family: male householder, no wife	Family: female householder, no husband	Non-family: householder alone	Non-family: householder not alone
Const	-1.542	3.348	-1.085	-1.653	1.781
	(0.326)	(4.015)	(0.467)	(0.444)	(1.058)
AGE0517	0.264	0.025	0.082	-0.278	-0.081
	(31.088)	(16.808)	(19.561)	(41.606)	(26.746)
AGE1824	-0.103	-0.011	0.032	0.054	0.009
	(6.889)	(4.032)	(4.366)	(4.579)	(1.704)
AGE1664	0.085	0.013	-0.030	-0.056	0.001
	(10.715)	(9.374)	(7.624)	(9.048)	(0.253)
AGE65	0.457	0.037	-0.054	-0.296	-0.027
	(14.679)	(6.813)	(3.513)	(12.103)	(2.447)
HHR1824	-0.394	-1.225	-5.908	4.843	2.479
	(0.901)	(15.884)	(27.498)	(14.069)	(15.933)
HHR2544	-0.711	-1.196	-5.792	5.259	2.068
	(1.654)	(15.799)	(27.458)	(15.561)	(13.544)
HHR4564	-0.284	-1.190	-5.547	4.894	1.821
	(0.667)	(15.834)	(26.492)	(14.588)	(12.010)
HHR65	-0.940	-1.243	-5.700	4.974	1.888
	(2.232)	(16.761)	(27.576)	(15.019)	(12.614)
INC25k	0.440	1.210	5.816	-4.381	-1.838
	(1.050)	(16.391)	(28.274)	(13.292)	(12.345)
INC2550	0.588	1.229	5.817	-4.453	-1.837
	(1.399)	(16.612)	(28.205)	(13.476)	(12.302)
INC5000	0.762	1.176	5.778	-4.604	-1.831
	(1.778)	(15.584)	(27.476)	(13.665)	(12.028)
INC100	0.877	1.166	5.673	-4.543	-1.777
	(2.065)	(15.584)	(27.217)	(13.603)	(11.774)
R^2	0.931	0.883	0.856	0.94	0.913
adjusted R^2	0.931	0.882	0.856	0.94	0.912
RMSE (OLS)	93.271	16.429	45.785	73.359	33.15
RMSE(LOESS)	2.772	0.926	1.942	2.232	1.266
window size	17	17	17	17	17

Note: t values are in parentheses.

Comparison of OLS and LOESS forecast Performance using RMSE

RMSE using OLS = 93.271
RMSE using LOESS = 2.772 (window size= 17.000)

		Estimate		Residuals	
1990	2000	OLS	LOESS	OLS	LOESS
500	748	685	749	63	-1
675	714	629	717	85	-3
676	871	708	871	163	0
469	472	392	477	80	-5
750	918	829	909	89	9
118	738	713	730	25	8
70	585	689	587	-104	-2
612	1,342	1,293	1,342	49	0
12	571	504	571	67	0
677	830	768	827	62	3
740	974	1,024	975	-50	-1
559	845	813	845	32	0
540	1,146	1,119	1,146	27	0
369	519	636	519	-117	0
2	516	601	516	-85	0
425	663	722	663	-59	0
504	601	669	602	-68	-1
3	71	117	71	-46	0
484	570	679	570	-109	0
717	902	1,044	902	-142	0
286	346	455	346	-109	0
652	665	769	665	-104	0
537	561	666	560	-105	1

Table 1 Estimation results of presence of children (Los Angeles county)

Variable	No own children under 18	Own children under 18 years
Const	-37.145005	-3.026997
	(8.7417)	(0.7126)
AGE0517	0.040434	0.421346
	(9.1420)	(95.2903)
AGE1664	-0.004965	0.001762
	(1.5021)	(0.5333)
AGE65	0.438734	-0.019868
	(59.4752)	(2.6939)
WORKER	0.130238	0.087485
	(31.1072)	(20.9010)
INCOME	0.000833	-0.000005
	(14.7252)	(0.0814)
R^2	0.922	0.942
adjusted R^2	0.922	0.941
RMSE (OLS)	67.572	67.555
RMSE(LOESS)	7.049	6.156
window size	10	10

Note: t-values are in parentheses

Comparison of OLS and LOESS forecast performance using RMSE

RMSE using OLS = 67.572
RMSE using LOESS = 7.049 (window size= 10.000)

		Estimate		Residuals	
1990	2000	OLS	LOESS	OLS	LOESS
374	513	458	500	55	13
362	488	384	470	104	18
773	486	460	494	26	-8
445	299	271	290	28	9
385	655	559	653	96	2
80	471	465	490	6	-19
53	459	451	459	8	0
336	996	827	995	169	1
8	336	312	356	24	-20
480	515	504	509	11	6
435	675	676	681	-1	-6
275	612	530	612	82	0
361	779	794	778	-15	1
346	265	330	270	-65	-5
2	324	349	316	-25	8
512	345	347	341	-2	4
589	364	394	365	-30	-1
2	62	59	63	3	-1
505	306	357	306	-51	0
722	457	609	457	-152	0
322	197	225	208	-28	-11
727	377	454	384	-77	-7
522	353	348	347	5	6

What's Next?

- Coordination of Integrated Land Use Model (PECAS) and Activity based model
- Parcel based synthetic SED
  - Region → County → City → CT/TAZ → Grid Cell → Parcel
- Advanced behavioral model vs. simple allocation model.

Thank You

SOUTHERN CALIFORNIA ASSOCIATION of GOVERNMENTS

Resolving Regional Challenges