3rd Symposium on Activity-Based Modeling

Raitenhaslach, Germany, December 2024

The Potential for Linked Longitudinal Data in Transportation Research

Gregory D. Erhardt
Department of Civil Engineering
Martin School of Public Policy
University of Kentucky

Senior Advisor, Bay Area Metro
Hans Fischer Fellow, Technical University of Munich
greg.erhardt@uky.edu

With:
Pat Mokhtarian
Department of Civil Engineering
Georgia Institute of Technology

Chris Bollinger
Department of Economics
University of Kentucky

ABM Approach

Estimated from cross-sectional data → Implemented in an equibrium-based simulation → Applied to predict change 20+ years in future

Many advantages, but several risks:

- **Possible self-selection:** I move to a walkable neighborhood because I like to walk.
- **Possible endogenity:** Transit operator adds service to popular routes.
- **Missing habits and triggers:** I started teleworking during the pandemic, and never went back.

With possible consequences

Bias in ridership forecasts for major US transit capital projects: 1987-2018.

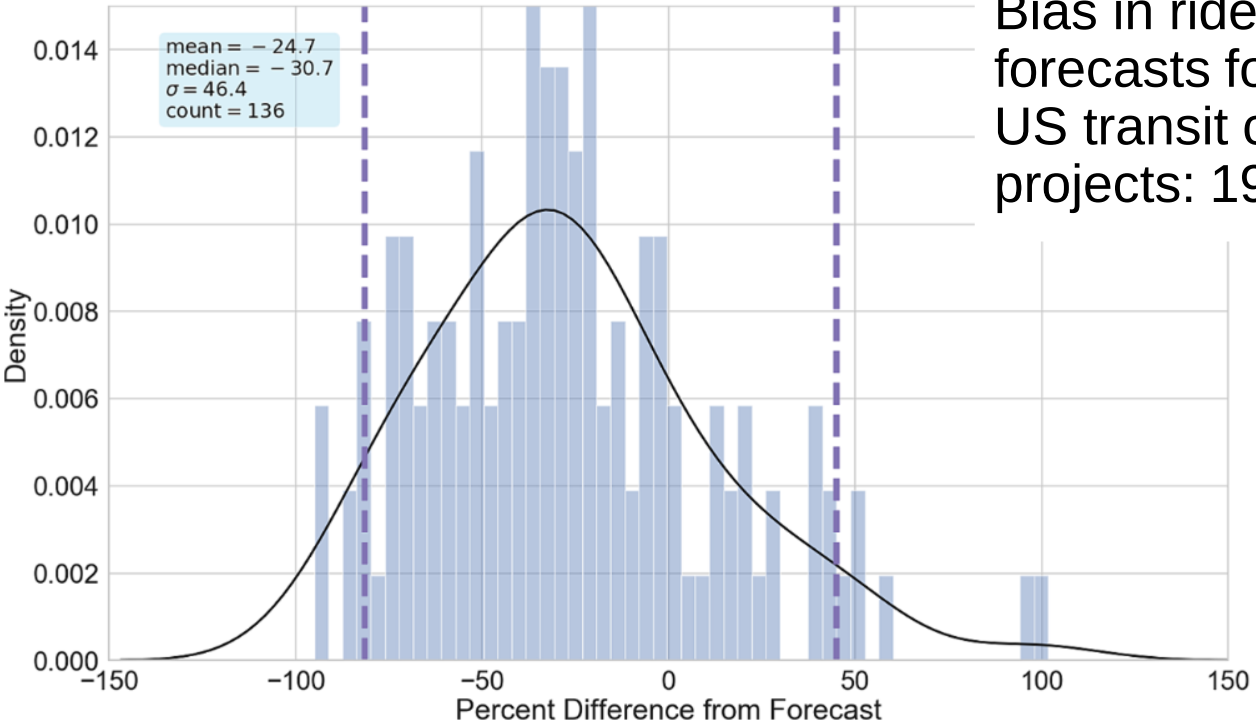


Fig. 2. Percent difference of observed ridership from forecast.

Hoque, et al 2024

We build models to predict change. Let’s start by observing the change we’re trying to predict.

* Actually, there are a lot of reasons we build models (Eptstein 2008), but this one is pretty important.

Panel of convenience

- For repeated cross-sections, there is a non-zero probability that the same unit is sampled in more than one year.
- With a large enough cross-sectional sample, the “panel of convenience” becomes usable.
- Strategy used successfully in the Danish National Survey & Australian Census.

American Community Survey (ACS)

- Annual survey of ~1.5% of US households with >90% response rates from 2005-present.
- Includes home location, work location, usual commute mode, car ownership, etc.
- Restricted-use data includes individual records with a personal identification key (PIK) assigned to >90% of records.

Expected addresses resampled

Year	Sub-Frame	Final Interviews in ACS	Sampling Rate within Subframe	Expected Addresses Resampled from:			
				5 Years Prior	10 Years Prior	15 Years Prior	Any Year
2005	1	1,924,527	7.70%	0	0	0	0
2006	2	1,968,362	7.70%	0	0	0	0
2007	3	1,937,659	7.50%	0	0	0	0
2008	4	1,931,955	7.40%	0	0	0	0
2009	5	1,917,748	7.30%	0	0	0	0
2010	1	1,917,799	7.30%	140,490	0	0	140,490
2011	2	2,128,104	8.00%	157,469	0	0	157,469
2012	3	2,375,715	8.90%	172,452	0	0	172,452
2013	4	2,208,513	8.30%	160,352	0	0	160,352
2014	5	2,322,722	8.60%	164,926	0	0	164,926
2015	1	2,305,707	8.50%	163,013	163,585	0	326,598
2016	2	2,229,872	8.20%	174,505	161,406	0	335,910
2017	3	2,145,639	7.80%	185,306	151,137	0	336,443
2018	4	2,143,000	7.70%	170,056	148,761	0	318,816
2019	5	2,059,945	7.40%	171,881	141,913	0	313,795
2020	1	1,406,935	5.00%	115,285	95,890	96,226	307,402
2021	2	1,950,832	6.90%	153,861	146,839	135,817	436,517
2022	3	1,980,550	6.91%	148,343	164,250	133,964	446,558
Total		36,855,584		2,077,940	1,173,781	366,007	3,617,728

Result: Longitudinal commute data that are:

- Cheap!
- Large sample
- Disaggregate
- Nationally representative…sort of.

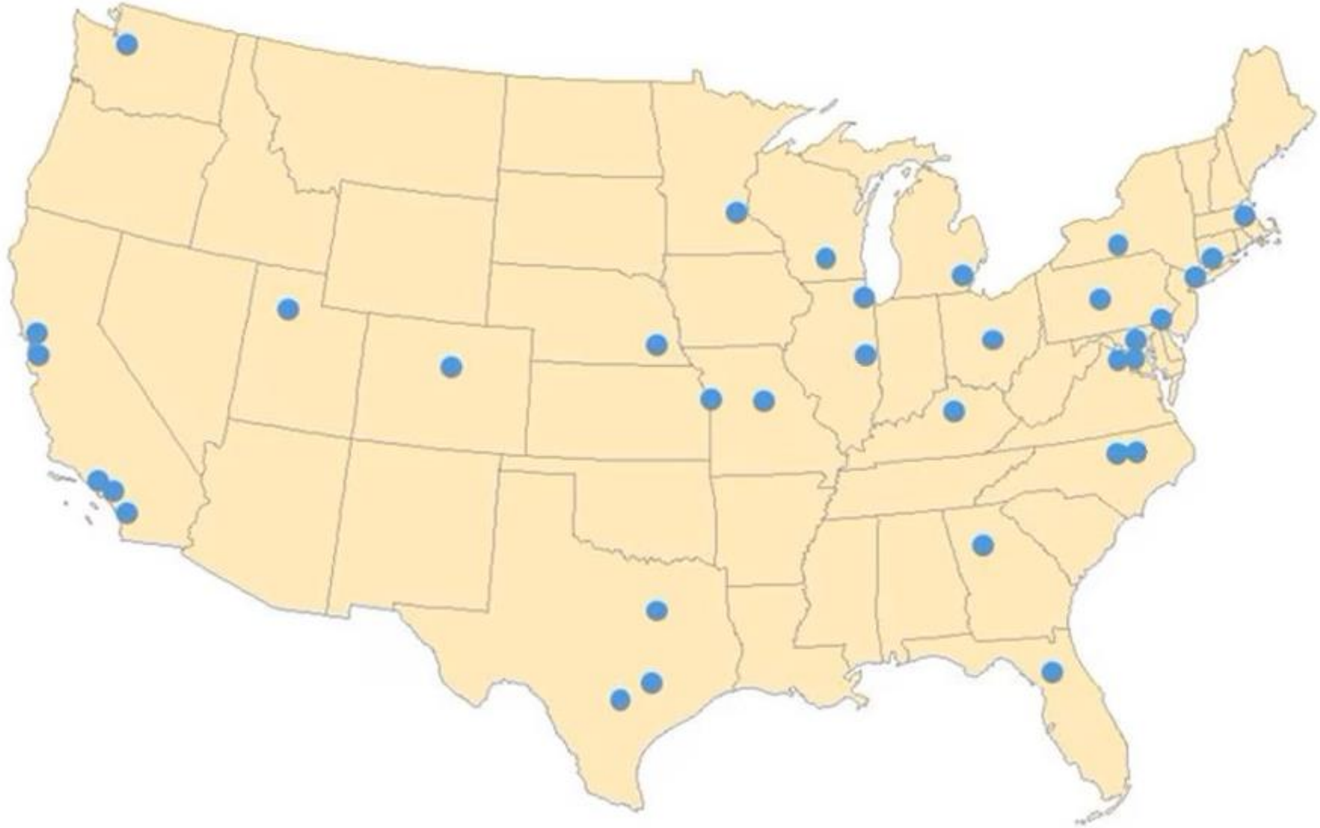
Inclusive of people living continously in the US over the panel period

		Previous Interview Year																
		2021	2020	2019	2018	2017	2016	2015	2014	2013	2012	2011	2010	2009	2008	2007	2006	2005
Panel	Sub-Frame	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1
2010	1																	
2011	2																	
2012	3																	
2013	4																	
2014	5																	
2015	1																	
2016	2																	
2017	3																	
2018	4																	
2019	5																	
2020	1																	
2021	2																	
2022	3																	

Additional linked data

- Aggregate ACS measures: 2005-present
- Access Across America data: 2014-present
- EPA Smart Location Database: 2011, 2013, 2021
- Census Master Address File
- Social Security Administration Detailed Earnings Records
- Longitudinal Employer–Household Dynamics (LEHD)
- Future: other travel or health surveys

Access available in 34 Federal Statistical Research Data Centers with Special Sworn Status



What can we learn from these data?

Q1: By how much can changes to the residential built environment reduce car commuting?

- Measure the change in commute VMT as a function of the change in the built environment.
- Use sample selection model to account for residential self-selection
- Compare to cross-sectional estimates

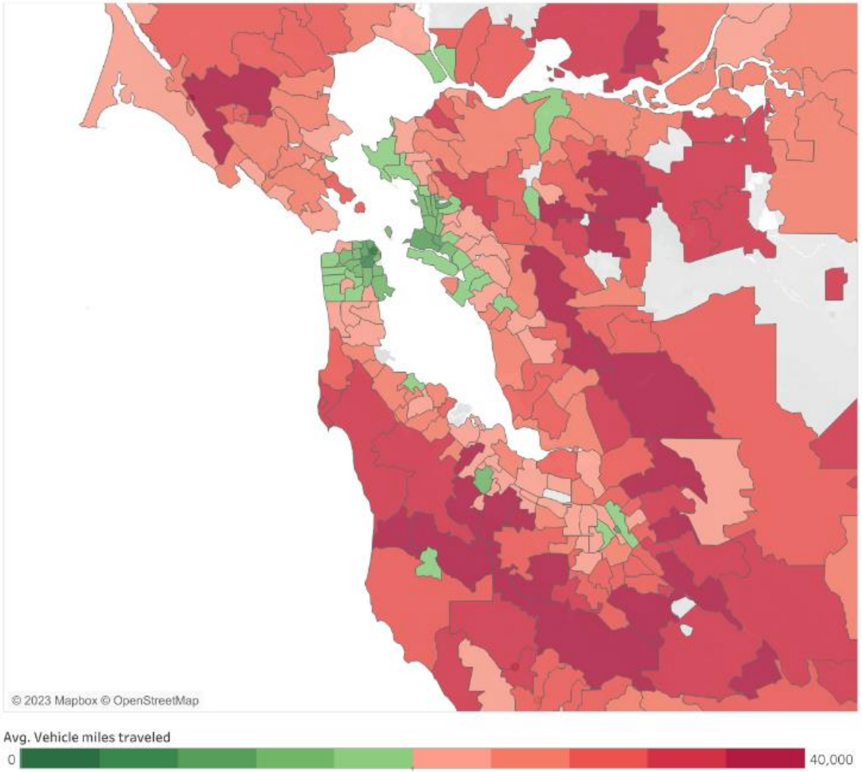


Figure 1 Annual Vehicle Miles Traveled per Household (45)

Q2: Does teleworking induce moves to lower-accessibility locations?

Possible Telework Effects on Residential Relocation

	Resampling Year	
Base Year	Not Teleworking (NT)	Teleworking (T)
Not Teleworking (NT)	Control	New opportunity to move to lower-accessibility location
Teleworking (T)	Not Relevant	Existing opportunity to move to lower-accessibility location

Q3: How does teleworking affect wage growth?

Expected Drivers of Teleworking Effect on Wage Growth

	Resampling Year	
Base Year	Not Teleworking	Teleworking
Not Teleworking	Control	Increased Job Selection
Teleworking	Decreased Networking	Increased Job Selection + Decreased Networking

Q4: Who is leaving transit-rich areas, who is replacing them, and how do their commutes change?

Expected Drivers of Teleworking Effect on Wage Growth

	Ends in:	
Starts in:	Transit-Rich Area	Transit-Poor Area
Transit-Rich Area	Control Group 1	Potentially "Displaced" Workers
Transit-Poor Area	Potential "Gentrifiers"	Control Group 2

What else should we be asking?

Who else can use these data?

Georgia Tech

University of Kentucky

National Science Foundation