January 2009

Linking highway improvements to changes in land use with quasi-experimental research design: A better forecasting tool for transportation decision-making

R. G. Funderburg
The University of Iowa

Hilary Nixon
San Jose State University, hilary.nixon@sjsu.edu

M. G. Boarnet
University of California, Irvine

M I N E T A   T R A N S P O R T A T I O N   I N S T I T U T E

Linking Highway Improvements to Changes in Land Use with Quasi-Experimental Research Design: A Better Forecasting Tool for Transportation Decision-making

MTI Report 09-02

MINETA
TRANSPORTATION INSTITUTE
Created by Congress in 1991

MINETA TRANSPORTATION INSTITUTE

The Norman Y. Mineta International Institute for Surface Transportation Policy Studies (MTI) was established by Congress as part of the Intermodal Surface Transportation Efficiency Act of 1991. Reauthorized in 1998, MTI was selected by the U.S. Department of Transportation through a competitive process in 2002 as a national "Center of Excellence." The Institute is funded by Congress through the United States Department of Transportation's Research and Innovative Technology Administration, the California Legislature through the Department of Transportation (Caltrans), and by private grants and donations.

The Institute receives oversight from an internationally respected Board of Trustees whose members represent all major surface transportation modes. MTI's focus on policy and management resulted from a Board assessment of the industry's unmet needs and led directly to the choice of the San José State University College of Business as the Institute's home. The Board provides policy direction, assists with needs assessment, and connects the Institute and its programs with the international transportation community.

MTI's transportation policy work is centered on three primary responsibilities:

Research

MTI works to provide policy-oriented research for all levels of government and the private sector to foster the development of optimum surface transportation systems. Research areas include: transportation security; planning and policy development; interrelationships among transportation, land use, and the environment; transportation finance; and collaborative labor-management relations. Certified Research Associates conduct the research. Certification requires an advanced degree, generally a Ph.D., a record of academic publications, and professional references. Research projects culminate in a peer-reviewed publication, available both in hardcopy and on TransWeb, the MTI website (http://transweb.sjsu.edu).

Education

The educational goal of the Institute is to provide graduate-level education to students seeking a career in the development and operation of surface transportation programs. MTI, through San José State University, offers an AACSB-accredited Master of Science in Transportation Management and a graduate Certificate in Transportation Management that serve to prepare the nation's transportation managers for the 21st century. The master's degree is the highest conferred by the California State University system. With the active assistance of the California Department of Transportation, MTI delivers its classes over a state-of-the-art videoconference network throughout the state of California and via webcasting beyond, allowing working transportation professionals to pursue an advanced degree regardless of their location. To meet the needs of employers seeking a diverse workforce, MTI's education program promotes enrollment to under-represented groups.

Information and Technology Transfer

MTI promotes the availability of completed research to professional organizations and journals and works to integrate the research findings into the graduate education program. In addition to publishing the studies, the Institute also sponsors symposia to disseminate research results to transportation professionals and encourages Research Associates to present their findings at conferences. The World in Motion, MTI's quarterly newsletter, covers innovation in the Institute's research and education programs. MTI's extensive collection of transportation-related publications is integrated into San José State University's world-class Martin Luther King, Jr. Library.

MTI Report 09-02

LINKING HIGHWAY IMPROVEMENTS TO CHANGES IN LAND USE WITH QUASI-EXPERIMENTAL RESEARCH DESIGN: A BETTER FORECASTING TOOL FOR TRANSPORTATION DECISION-MAKING

October 2009

Richard G. Funderburg, Ph.D.
Hilary Nixon, Ph.D.
and
Marlon G. Boarnet, Ph.D.

a publication of the
Mineta Transportation Institute
College of Business
San Jose State University
San Jose, CA 95192-0219
Created by Congress in 1991

TECHNICAL REPORT DOCUMENTATION PAGE

<table>
<tr><td>1. Report No.
CA-MTI-09-2703</td><td>2. Government Accession No.</td><td colspan="2">3. Recipients Catalog No.</td></tr>
<tr><td colspan="2" rowspan="2">4. Title and Subtitle
Linking Highway Improvements to Changes in Land Use with Quasi-Experimental Research Design:
A Better Forecasting Tool for Transportation Decision-making</td><td colspan="2">5. Report Date
October 2009</td></tr>
<tr><td colspan="2">6. Performing Organization Code</td></tr>
<tr><td colspan="2">7. Authors
Richard G. Funderburg, Ph.D., Hilary Nixon, Ph.D., and Marlon G. Boarnet, Ph.D.</td><td colspan="2">8. Performing Organization Report No.
MTI Report 09-02</td></tr>
<tr><td colspan="2" rowspan="2">9. Performing Organization Name and Address
Mineta Transportation Institute
College of Business
San José State University
San Jose, CA 95192-0219</td><td colspan="2">10. Work Unit No.</td></tr>
<tr><td colspan="2">11. Contract or Grant No.
65W136</td></tr>
<tr><td colspan="2" rowspan="2">12. Sponsoring Agency Name and Address

California Department of Transportation
Sacramento, CA 94273-0001

U.S. Department of Transportation
Office of Research—MS42
Research & Special Programs Administration
P.O. Box 942873
400 7th Street, SW
Washington, D.C. 20590-0001</td><td colspan="2">13. Type of Report and Period Covered
Final Report</td></tr>
<tr><td colspan="2">14. Sponsoring Agency Code</td></tr>
<tr><td colspan="4">15. Supplementary Notes</td></tr>
<tr><td colspan="4">16. Abstract
An important issue for future improvement and extensions of highways will be the ability of projects to sustain challenges to Environmental Impact Statements based upon forecasts of regional growth. A legal precedent for such challenges was established in 1997 when a U.S. District Court judge ruled that the EIS for a proposed Illinois toll road was deficient because the growth projections were the same in the build and no-build scenarios. This paper incorporates popular regional growth forecasting models into a quasi-experimental research design that directly relates new highway investments in three California counties to changes in population and employment location, while controlling for no-build historical counterfactuals. The authors model simultaneous employment and population growth from 1980 to 2000 in Merced, Orange, and Santa Clara counties, three California counties that received substantive highway improvements during the mid-1990s. The strategy permits a comparison of the before-and-after tests for effects of investments on economic growth and land use in three regions that contrast how increased highway access affects development patterns: (1) for an urban center in Santa Clara County, (2) for an exurban region in Orange County, and (3) for a small town in Merced County.

We find that traditional forecast approaches, which lack explicit control selection, can lead to erroneous conclusions about an impact. Our integrated form of the lagged adjustment model confirms results from a conventional form of the model that includes all cross-sectional units as observations; in both forms of the model we estimate a statistically significant increase in employment development in the exurban region in Orange County where new toll roads were constructed. In the case of Santa Clara County, neither our quasi-experimental integrated approach nor the conventional lagged adjustment approach estimates a significant effect on population or employment growth that can be attributed to the new highways constructed in the urban center. For the small town environment in Merced County, the conventional simultaneous growth regressions produce a materially different estimate than the approach we develop and examine in this paper. Isolating effects to local spatial units where the intervention occurred and their no-build counterfactual produces estimates of a statistically significant decrease in employment growth in the small town near the newly constructed highway bypass.</td></tr>
<tr><td>17. Key Words
Cultural resources;
Geographic information systems;
Information systems;
Decision support systems;</td><td colspan="3">18. Distribution Statement
No restrictions. This document is available to the public through The National Technical Information Service, Springfield, VA 22161</td></tr>
<tr><td>19. Security Classif. (of this report)
Unclassified</td><td>20. Security Classifi. (of this page)
Unclassified</td><td>21. No. of Pages</td><td>22. Price
$15.00</td></tr>
</table>

Form DOT F 1700.7 (8-72)

Library of Congress Catalog Card Number: 2009931015

To oder this publication, please contact the following:
Mineta Transportation Institute
College of Business
San Jose State University
San Jose, CA 95192-0219
Tel (408) 924-7560
Fax (408) 924-7565

E-mail: mti@mti.sjsu.edu
http://transweb.sjsu.edu

ACKNOWLEDGMENTS

We would like to thank the leadership and staff at the Mineta Transportation Institute for their support of all phases of this work, including former-MTI Research Director Trixie Johnson, the current Research Director, Karen Philbrick, Ph.D., Communication Director Donna Maurillo, Research Support Manager Meg Fitts, Student Publication Assistant Sahil Rahimi, Student Webmaster Ruchi Arya, Student Research Support Assistant Chris O'Dell and Student Graphic Artists JP Flores and Vince Alindogan.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

An important issue for future improvement and extensions of highways will be the ability of projects to sustain challenges to Environmental Impact Statements based upon forecasts of regional growth. A legal precedent for such challenges was established in 1997 when a U.S. District Court judge ruled that the EIS for a proposed Illinois toll road was deficient because the growth projections were the same in the build and no-build scenarios. This paper incorporates popular regional growth forecasting models into a quasi-experimental research design that directly relates new highway investments in three California counties to changes in population and employment location, while controlling for no-build historical counterfactuals. The authors model simultaneous employment and population growth from 1980 to 2000 in Merced, Orange, and Santa Clara counties, three California counties that received substantive highway improvements during the mid-1990s. The strategy permits a comparison of the before-and-after tests for effects of investments on economic growth and land use in three regions that contrast how increased highway access affects development patterns: (1) for an urban center in Santa Clara County, (2) for an exurban region in Orange County, and (3) for a small town in Merced County.

We find that traditional forecast approaches, which lack explicit control selection, can lead to erroneous conclusions about an impact. Our integrated form of the lagged adjustment model confirms results from a conventional form of the model that includes all cross-sectional units as observations; in both forms of the model we estimate a statistically significant increase in employment development in the exurban region in Orange County where new toll roads were constructed. In the case of Santa Clara County, neither our quasi-experimental integrated approach nor the conventional lagged adjustment approach estimates a significant effect on population or employment growth that can be attributed to the new highways constructed in the urban center. For the small town environment in Merced County, the conventional simultaneous growth regressions produce a materially different estimate than the approach we develop and examine in this paper. Isolating effects to local spatial units where the intervention occurred and their no-build counterfactual produces estimates of a statistically significant decrease in employment growth in the small town near the newly constructed highway bypass.

INTRODUCTION

Understanding linkages of new highway construction or capacity expansion to regional growth patterns is crucial for transportation planners and policymakers. Particularly important will be the ability of new projects to avoid or sustain challenges to Environmental Impact Statements (EIS) based upon forecasts of regional growth. A legal precedent for such challenges was established in 1997 when a U.S. District Court judge ruled that the EIS for a proposed Illinois toll road was deficient because the growth projections were the same in the build and no-build scenarios (Sierra Club v. United States DOT, 1997). Despite considerable research on the topic, a fundamental debate in urban and regional planning remains as to whether new highway infrastructure induces growth, shifts growth from previously developed regions to regions that gain access, or the new infrastructure merely follows the path of development to service regions that would have grown with or without the new investment. Our research design gives insight into the question of causality. Do highways cause growth or vice versa? A second key question in this field focuses on spatial shift. If growth occurs along the new highway path, was it at the expense of job loss elsewhere in the metropolitan region, or is it new growth that wouldn't have happened without the new infrastructure? Before the question of spatial shift can be addressed, we must first focus on causality. In this paper, we incorporate popular regional growth forecasting models into a quasi-experimental research design that directly relates new highway investments in Merced, Orange, and Santa Clara Counties to changes in population and employment location, while controlling for no-build historical counterfactuals. We study this mix of urban, small town, and exurban highway projects to examine the possibility of differential effects.

The central finding of the paper is that while improvements in surface transportation infrastructure tend to have large impacts on growth patterns, the nature of the effects is materially dependent on the context of the highway investment. Our models estimate that, on average, 338 to 11,103 new Orange County jobs occurred within a typical census tract in the formerly exurban region after gaining highway access when compared to no-build counterfactuals. This employment gain for tracts near the new highways is statistically and economically significant as it represents upwards of an 18 percent addition from 1990 levels to these tracts on average. Our models predict a starkly different outcome as a result of a highway bypass built outside the small town of Livingston in Merced County. In that instance, we estimate a statistically significant 12 to 83 job losses per square kilometer as a result of the new bypass, which at minimum translates to an 11 percent loss in new jobs for the town that otherwise might be anticipated if the bypass had not been built. We find no significant effects on population or employment growth that can be attributed to the new highway investment near the urban center of Santa Clara County. The differential effects from highway investments in the three contexts illustrate the importance of choosing appropriate comparison groups in forecasts of population and employment growth for build and no-build scenarios.

Recent empirical studies confirm effects from transportation infrastructure improvements on intrametropolitan location choices of people and employers to be both statistically and economically significant. For example, Boarnet and Chalermpong (2002) studied the population and employment growth impact of the early segments of the Orange County toll road system that opened in 1993-1996 and they found statistically significant and economically meaningful employment growth impacts. Similarly, Chalermpong (2004) examined the population and employment growth impact of the I-105 Freeway in Los Angeles and again found large impacts, and Holl (2004) studied firm location effects in the first decade after improvements to Portugal's highway network and found impacts. In a study of 24 freeway lane expansion projects in California, Cervero (2003) was able to demonstrate that infrastructure projects induce growth through increased building activity within a 4-mile project corridor. More generally, Baum-Snow's (2007) examination of the suburbanization effects from the U.S. Interstate Highway System estimates that building the first new highway through a central city reduces central city population growth by 17 percent, while increasing suburban population growth. Nationally, Baum-Snow (2007) estimates that building the Interstate Highway System resulted in central city population growth that was 8 percent lower than what would have otherwise occurred, again shifting growth to the suburbs. Baum-Snow's (2007) findings present a strong rebuttal to earlier opinions that interstate highways were not on net associated with the decentralization of metropolitan areas. Furthermore, Baum-Snow's findings provide strong reasons to believe that highway infrastructure is associated with urban and regional growth patterns. In this paper, we examine such effects at a finer scale to address language and nuances posed by the National Environmental Protection Act regarding projections under "build" and "no build" scenarios and judicial decisions thereof. In short, given the recent evidence that the Interstate Highway System contributed to the decentralization of U.S. metropolitan areas, what is the land use / growth impact of specific highway projects?

Alongside the question of whether a highway investment causes economic development and changes in land use, an unbiased forecast model would respect the endogenous relationship between population and employment location. The simultaneous spatial interaction between employment and population is grounded in modern location theory (Muth, 1971; Steinnes, 1978; Greenwood, 1985). Although the debate continues over whether jobs follow people or people follow jobs, it is well established that people often do not choose to live where they choose to work and vice versa and thus choose to commute over some distances to work.

In this paper, we build on the considerable amount of research that has been conducted on regional growth forecasting models to explore how transportation infrastructure improvements have led to changes in population and employment location. Our work avails from insights gained from a long tradition of simultaneous population and employment location models (Bradford and Kelejian, 1974; Steinnes and Fisher, 1974; Carlino and Mills, 1987). Boarnet (1992 and 1994) adapts these earlier models to include explicit spatial relationships within an urban area so that population growth is not just a function of the immediate spatial unit's indigenous characteristics, but also a function of the job opportunities in the regional labor market comprising other nearby areal units of observation. Similarly, the expansion of a local job base is in part a function of the regional labor supply to which residents of other nearby census tracts (or other spatial units) contribute. The spatial econometric adjustment model has since been adapted to examine impacts on growth from location of highways and access to rail transit and the link between urban and rural development (Boarnet, 1996; Bollinger and Ihlanfeldt, 1997; Henry, Barkley, and Bao, 1997; Henry, Schmitt, Kristensen, Barkley, and Bao, 1999; Schmitt and Henry, 2000). Building on the recent specification of the endogenous growth model in Boarnet, Chalermpong, and Geho (2005), our contribution is to directly incorporate a selection of controls into the system of simultaneous equations so as to devise natural experiments for each of the three study counties.

The selection of controls is designed to incorporate the appropriate no-build counterfactuals into the forecast model. Propensity score matching is the quasi-experimental technique used to select, as controls, regions similar in every respect to those receiving (or in proximity to) transportation improvements, except that the controls lacked any similar sort of intervention. Quasi-experimental techniques have been used in a variety of settings to find and match the cases among the set of potential controls that are most similar in every respect to the treatment group, except that the control group did not experience the intervention, thus preserving the intent behind random assignment in experimental design (Cook and Campbell, 1979; Rosenbaum and Rubin, 1983; Dehejia and Wahba, 1999 and 2002; Holzer, Quigley, and Raphael, 2003; O'Keefe, 2004; Smith and Todd, 2005). In a natural experiment such as the improvement or extension of a highway, "treatment" cases experiencing the intervention have been predetermined by project siting and policy, and not by random assignment. Lacking the ability to randomly assign cases and restrict which group receives the treatment, the researcher can use quasi-experimental techniques to mimic the research design of controlled experiments, allowing the possibility of easy-to-understand inferences about the impact of the intervention (Boarnet, 2001).

Our basic approach involves identifying the "experimental" or treatment group, identifying the superset of geographic units from which to select the "control" group, implementing alternative matching methods, analyzing changes in population and employment growth as difference in differences (treatment from matched control and over time), and incorporating the selection of matched controls into models of simultaneous population and employment growth to examine temporal changes in growth before and after the investments, while controlling for the counterfactual no-build scenario—what would have happened to population and employment in the regions that gained transportation access if the projects had never been developed.

The remainder of the paper is structured as follows. The next section describes the three study counties and the new highway investments constructed in each county during the mid-1990s that constitute the policy interventions. Section 3 describes the data and the propensity score matching methods we employ to select a control for each spatial unit that later receives access to new highway infrastructure in the mid-1990s. The third section also details the lagged adjustment model and the difference-in-differences estimators we employ to measure impacts when compared to no-build counterfactuals. In Section 4 we provide our findings and we conclude the paper with implications and suggestions for further research in the last section.

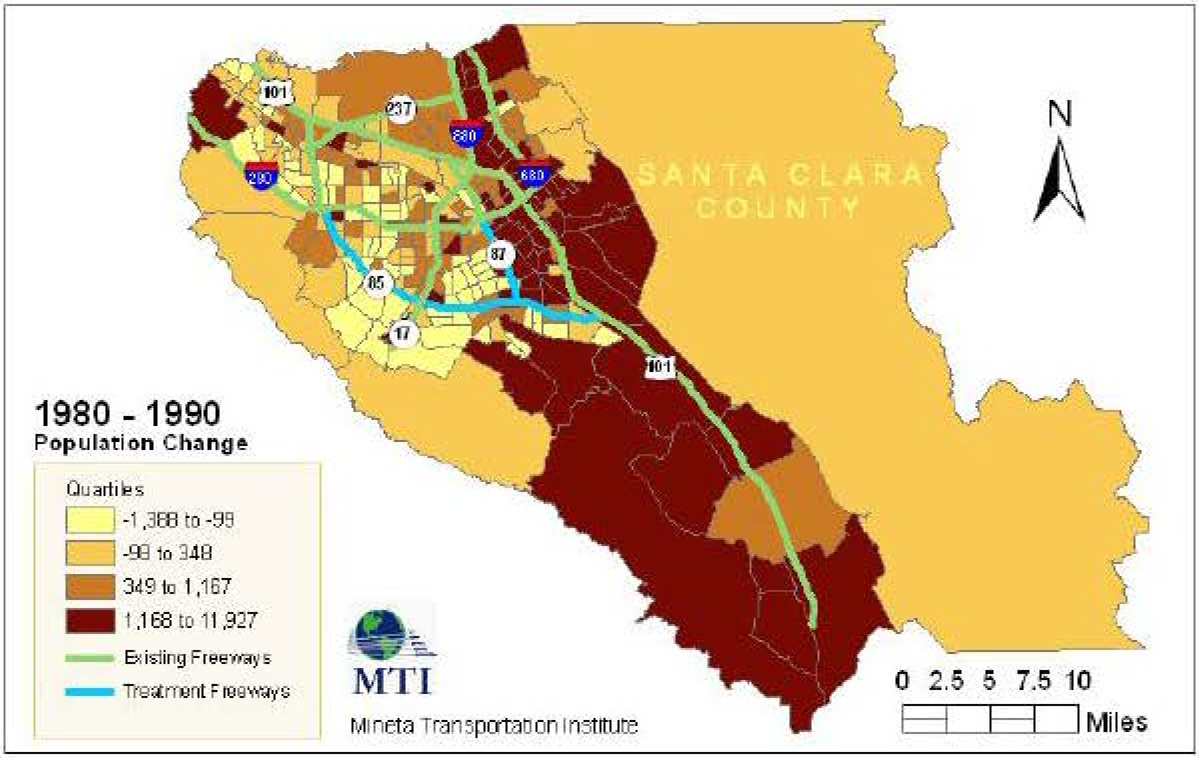
STUDY REGIONS AND HIGHWAYS

To examine the causal linkage between highway infrastructure improvements and population and employment growth near access points to these improvements, we selected four substantive highway projects constructed in three California counties during the mid-1990s. Two projects are located in Santa Clara County, a predominantly urbanized region. The third project is a new limited-access extension of highway that bypasses the small town of Livingston in Merced County, a rural region of the state. The fourth highway project is the new system of toll roads constructed in Orange County. On the whole, Orange County is a metropolitan region; although the new highways service its formerly exurban reaches.

These cases were selected as they represent major highway infrastructure investments constructed and opened during the 1990s and data was readily available on population, employment, and other variables of interest for the decade preceding opening of the new highway project (i.e. 1980-1990) as well as for a comparable post-construction analysis (i.e. 2000). Although, in some cases, plans for these projects were developed decades prior to construction, particularly in the case of Santa Clara County, evidence suggests that during our pre-intervention period it is a reasonable assumption that individuals and businesses were not anticipating the construction of the highway project. As a result, we can be fairly confident that our results represent population and employment change attributed to the infrastructure project.

SANTA CLARA COUNTY

Santa Clara County is an urban county on the southern end of the San Francisco Bay Area's Silicon Valley. The county grew rapidly both in population and in jobs between 1980 and 1990 and then growth slowed somewhat in 2000. For the 259 census tracts that averaged 12.98 km^2 in size and collectively exhausted the county in 1980, population in the average tract grew at an annualized rate of 1.46 percent from 5,000 people in 1980 to 5,780 in 1990. The number of jobs expanded even more vigorously than population in the 1980s at a compound annualized growth rate of 2.74 percent, with a typical tract in 1990 having 3,369 jobs, up from 2,570 a decade earlier. By the year 2000, the growth rates cooled to 1.17 percent for population and 0.82 percent for employment, increasing 2000 levels to 6,495 people and 3,656 jobs per census tract.



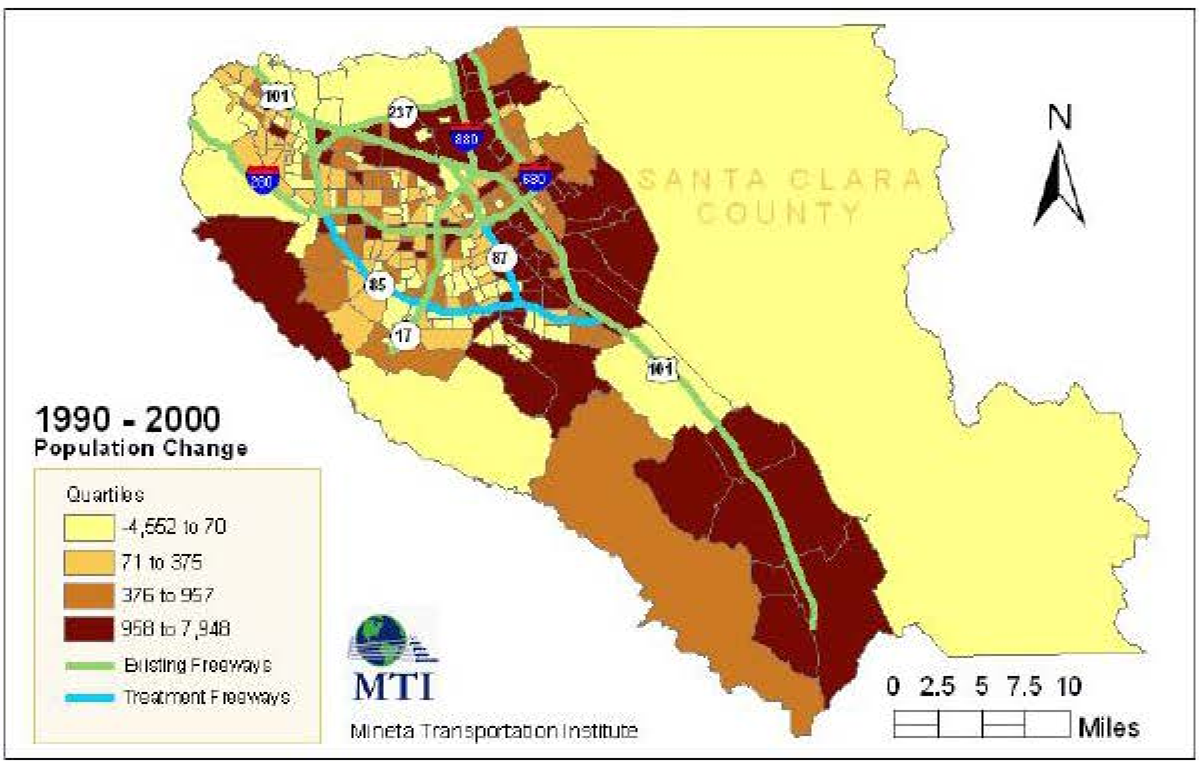


Figure 1 Population Growth in Santa Clara County Census Tracts Before and After Opening of State Route 85 and State Route 87 Highway Extensions

Our study area in Santa Clara County consists of two highway extension projects, both completed during the 1990s. Figures 1 and 2 provide an overview of population and employment growth in the county before and after the projects, the highway network as it existed before 1990, and the new sections of highway that are the interventions of interest for the natural experiments. The first project is the extension of State Route 87, also called the Guadalupe Freeway, from Interstate 280 near downtown San José to State Route 85, or the West Valley Freeway. Our second project is the extension of State Route 85, which links I-280 in Cupertino to the north and joins Highway 101 in the southern part of San José.

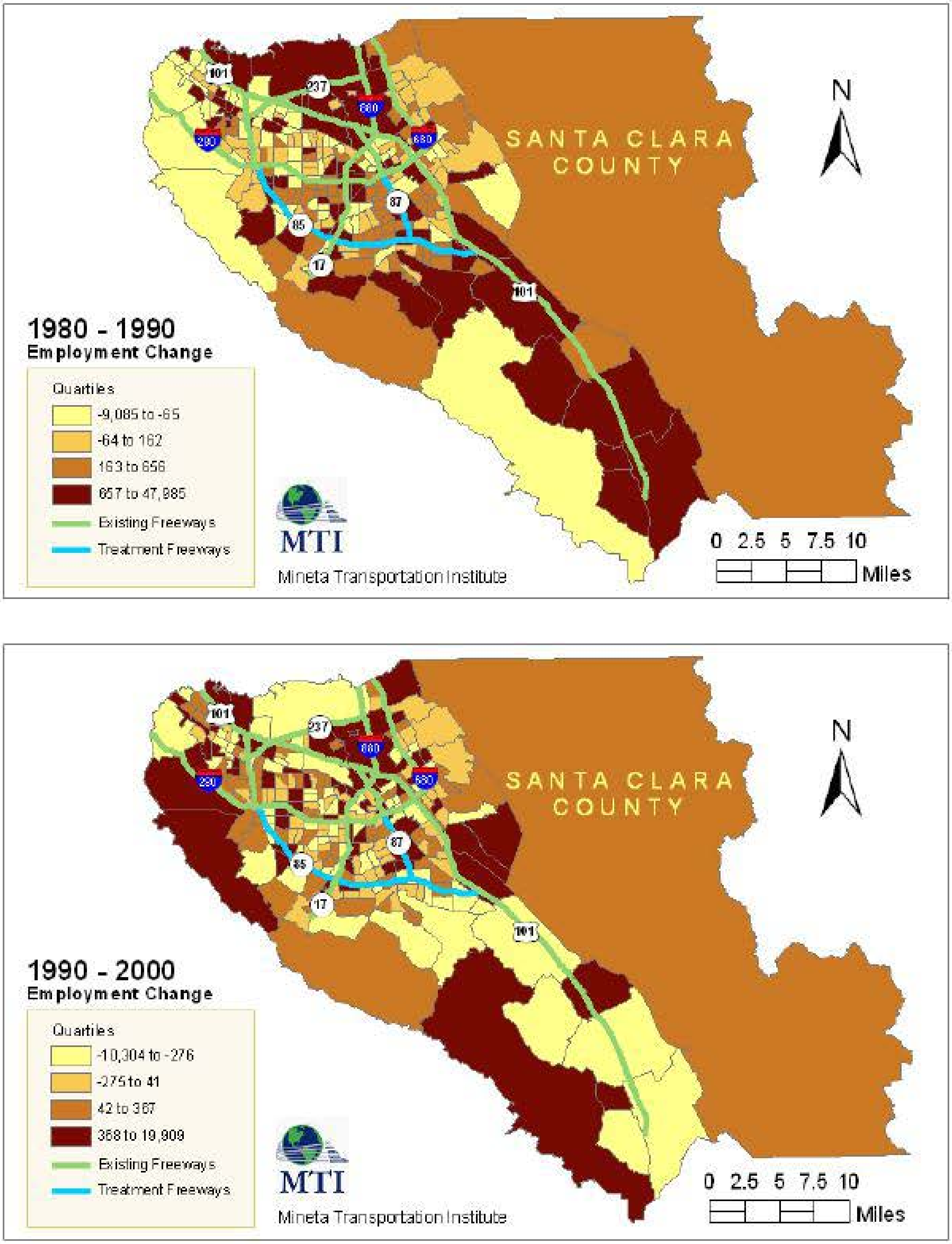


Figure 2 Employment Growth in Santa Clara County Census Tracts Before and After Opening of State Route 85 and State Route 87 Highway Extensions

State Route 87

As early as the 1960s, the Guadalupe Parkway provided a connection for drivers from Highway 101 in the north to downtown San José. Construction to convert the parkway to a grade-separated, controlled-access freeway began in the 1970s. The first stage involved the construction of a multi-level interchange at I-280, just south of downtown San José. Work on the northern section of the freeway continued during the 1970s and 1980s, although one traffic light remained in the northern section near Hedding Avenue until 2004.

The intervention modeled in this paper includes five new centerline miles that comprise the highway's southern extension. Construction of the southern portion of Highway 87, which extended the freeway from I-280 to Highway 85 in south San José, was completed in 1993. Completion of this section had been delayed approximately two years as a result of lawsuits regarding alignment, overpass design, and environmental concerns associated with the removal of serpentine rock containing asbestos along the construction route (Wyman, 1990). The construction delay impacted several residential developers planning to build in the southern region of the city.

State Route 85

The original plans to construct Highway 85 stem back to 1957 when the California Division of Highways signed an agreement with cities in the South Bay to build Highway 85 (Richards, 1994). The first seven miles of the route, from Cupertino to Mountain View (I-280 north to Highway 101), opened in 1971. No new construction occurred along the designated route for the next decade and a half. During the mid-1970s, Governor Jerry Brown and Caltrans Director Adriana Gianturco focused attention away from highway construction and toward rail transit. As a result, some building construction was allowed within the Highway 85 right-of-way, structures that later had to be demolished to make way for the freeway.

During the early 1980s, congestion along arterial streets and existing freeways in the county led County Supervisor Zoe Lofgren and Silicon Valley Manufacturing Group President Peter Giles to spearhead Measure A, the first local sales tax in California for transportation. Funds raised from the half-cent sales tax over a 10-year period were used to widen Highway 101, upgrade Highway 237 in the northern part of the County, and build Highway 85 (Richards, 1994).

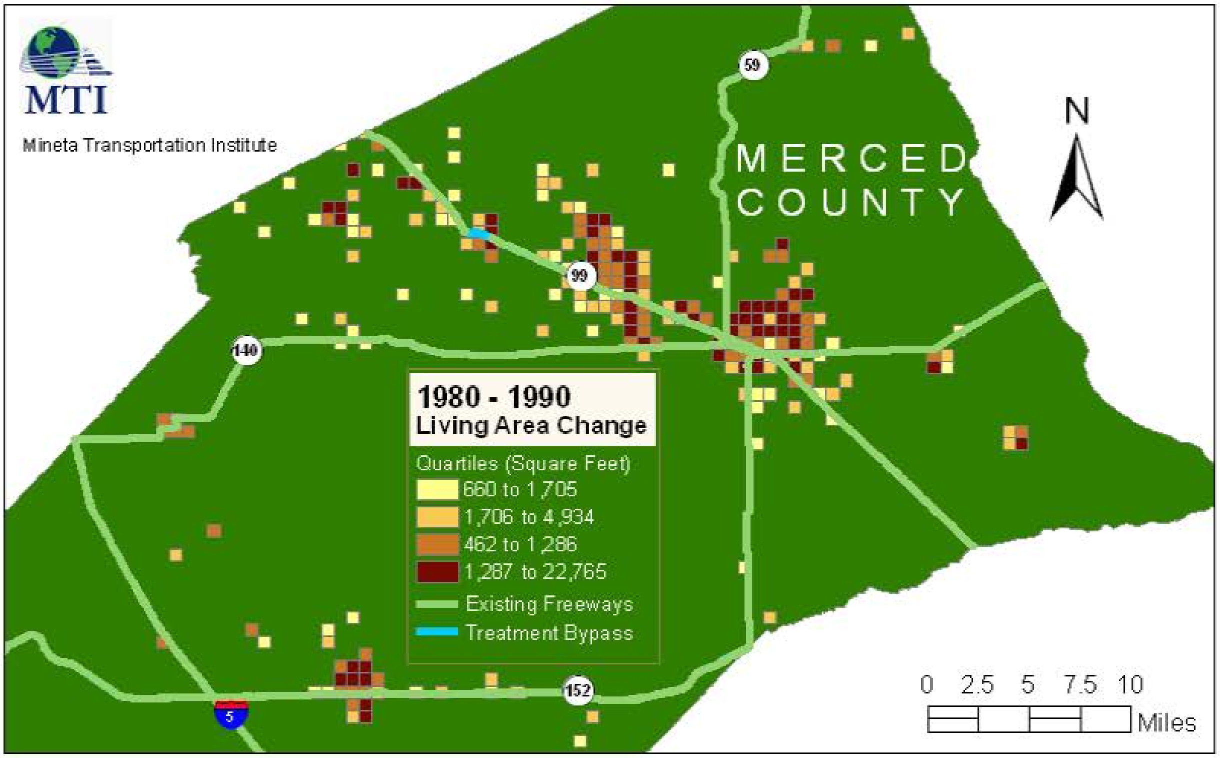
The intervention modeled in this paper includes all 19 new centerline miles of Highway 85 that connect Interstate 280 South to Highway 101. Construction to extend Highway 85 from I-280 south to Highway 101 began in the late 1980s and the first two-mile section, from Cottle Road to Santa Teresa Boulevard, opened in 1991. A year later, three additional miles opened, from Santa Teresa Boulevard to Almaden Expressway. The last 12.5 miles, completing the route, opened October 19, 1994 (Richards, 1994). For residents living in the Almaden Valley portion of San José, Highway 85 was designed to reduce commute time, which had been shown to be 14 minutes longer than anywhere else in the region (Richards, 1994).

MERCED COUNTY

Merced County is a rural county in the middle of California's Central Valley. The county grew rapidly in population and jobs during both decades that we study. For the 24 census tracts that averaged 212.85 km² in size and collectively exhausted the county in 1980, population in the average tract grew at an annualized rate of 2.85 percent from 5,607 people in 1980 to 7,429 in 1990. The number of jobs expanded faster than population in the 1980s at a compound annualized growth rate of 3.80 percent, with the average tract in 1990 having 1,703 jobs, up from 1,173 a decade earlier. Rapid growth continued through the year 2000, and over the decade population increased 1.68 percent per year to 8,773 people per tract and tract employment increased 4.15 percent annually to 2,558 jobs per tract in 2000. As indicated in Figures 3 and 4, the small town of Livingston in the region surrounding the intervention experienced little population and employment growth between the before-intervention (1980-1990) and the after-intervention (1990-2000) periods.

State Route 99

Stretching across California's Central Valley from its junction with Interstate 5 south of Bakersfield to the town of Red Bluff far up north, Highway 99 has a long and storied history. Originally designated in 1928 as part of US 99, this stretch of road provided key access through the central part of the state for "dust bowl refugees" as depicted in John Steinbeck's The Grapes of Wrath (Reese, 2004). In 1964, the highway was designated as State Route 99. Currently, Highway 99 serves as an alternative to the heavily traveled I-5, and there are plans to upgrade the route south of Stockton to Interstate Highway standards (California Department of Transportation, 2004).



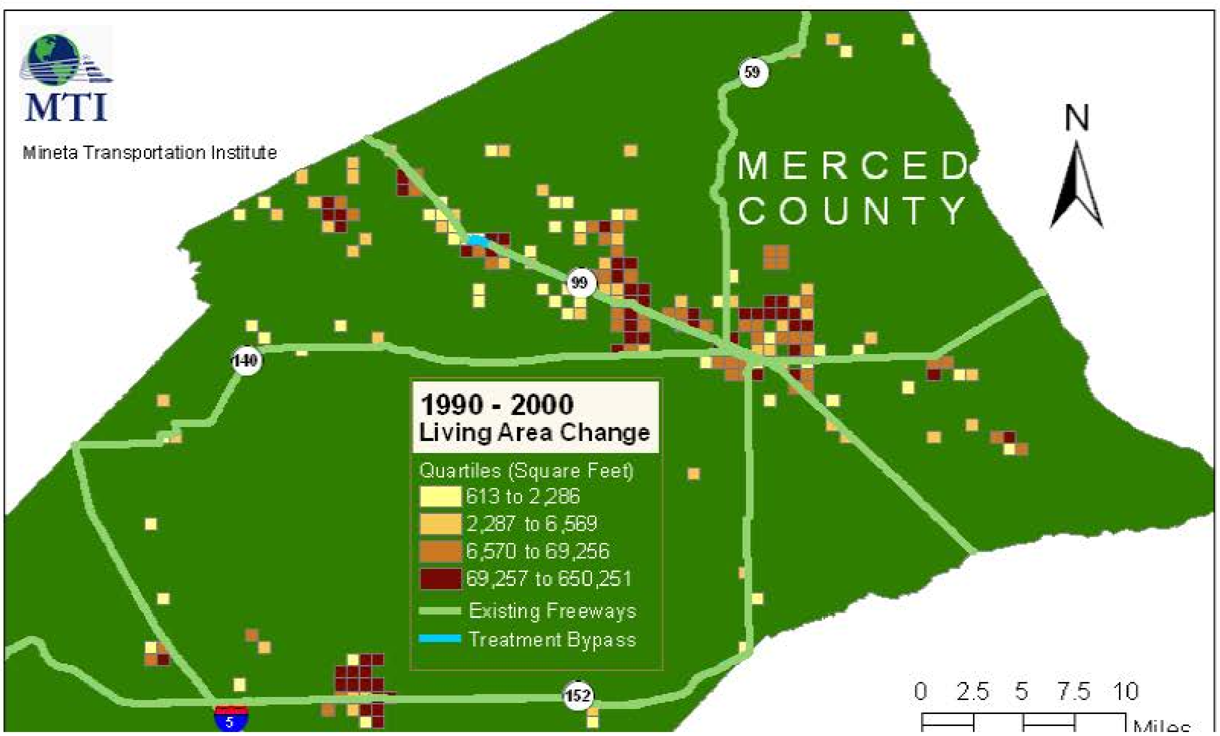
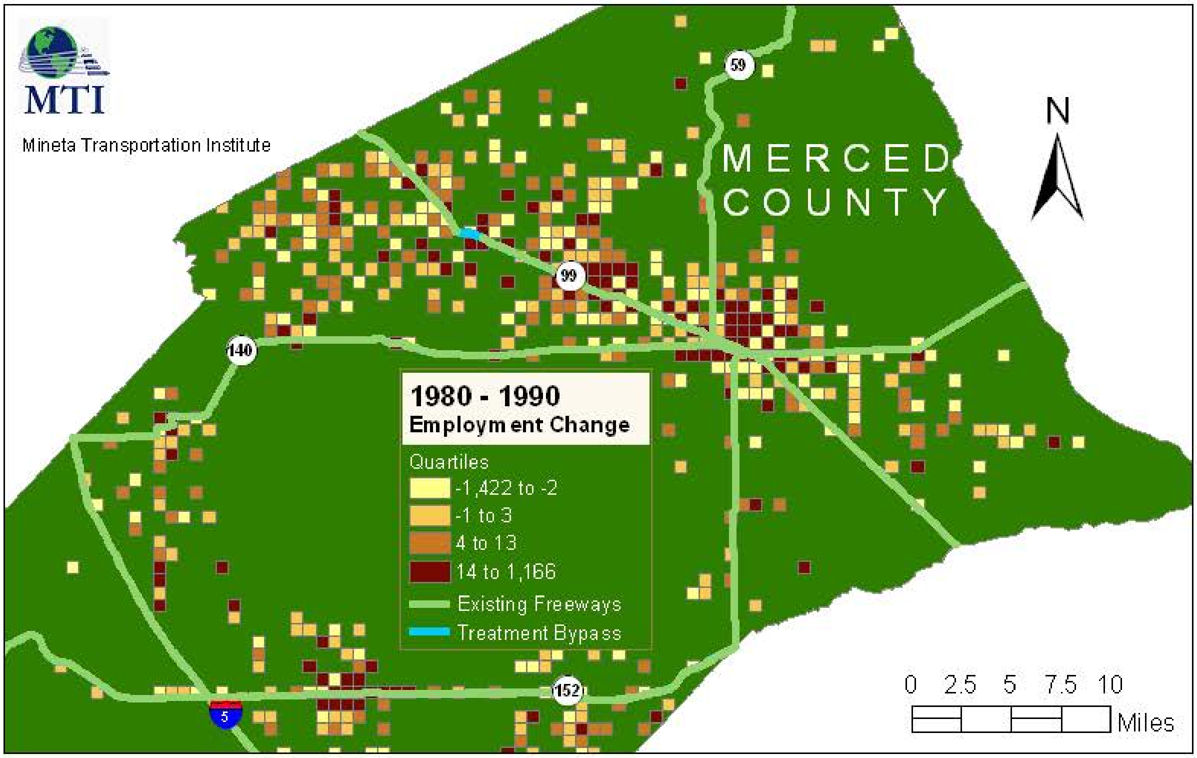


Figure 3 Residential Building Square Footage Growth in Merced County 1 km^2 Grid Cells Before and After Opening of State Route 99 Livingston Bypass



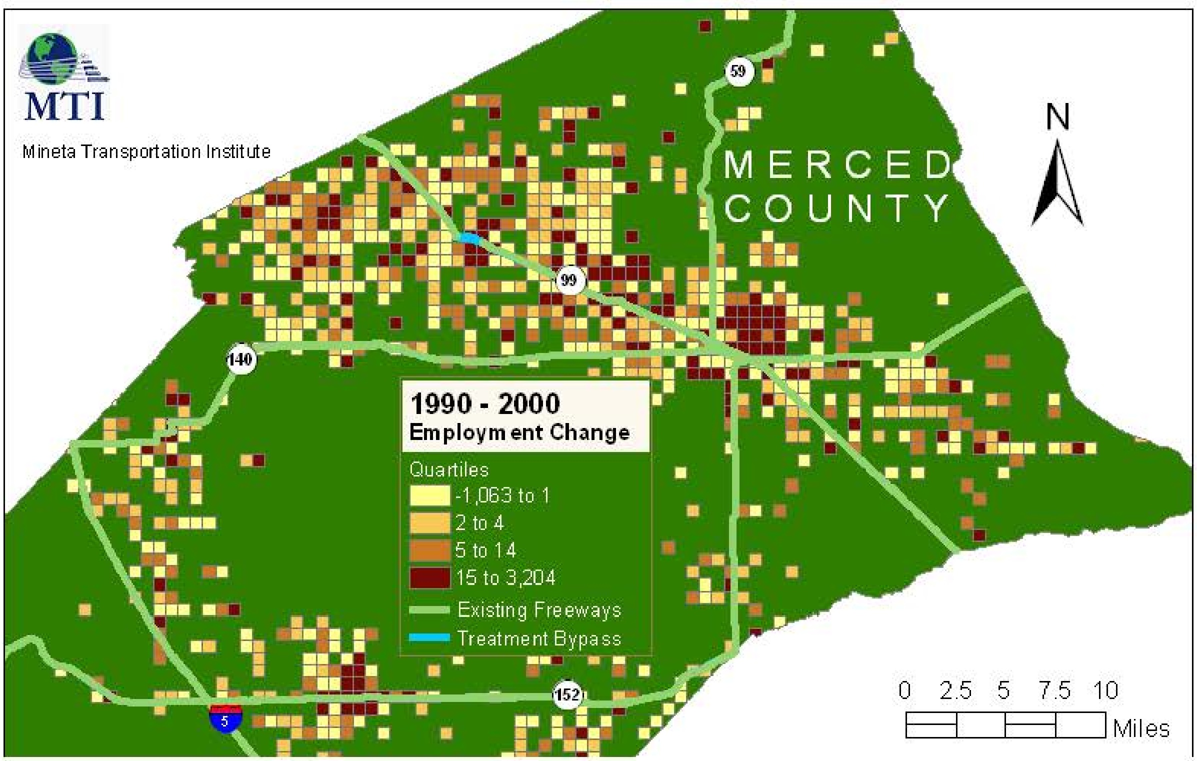


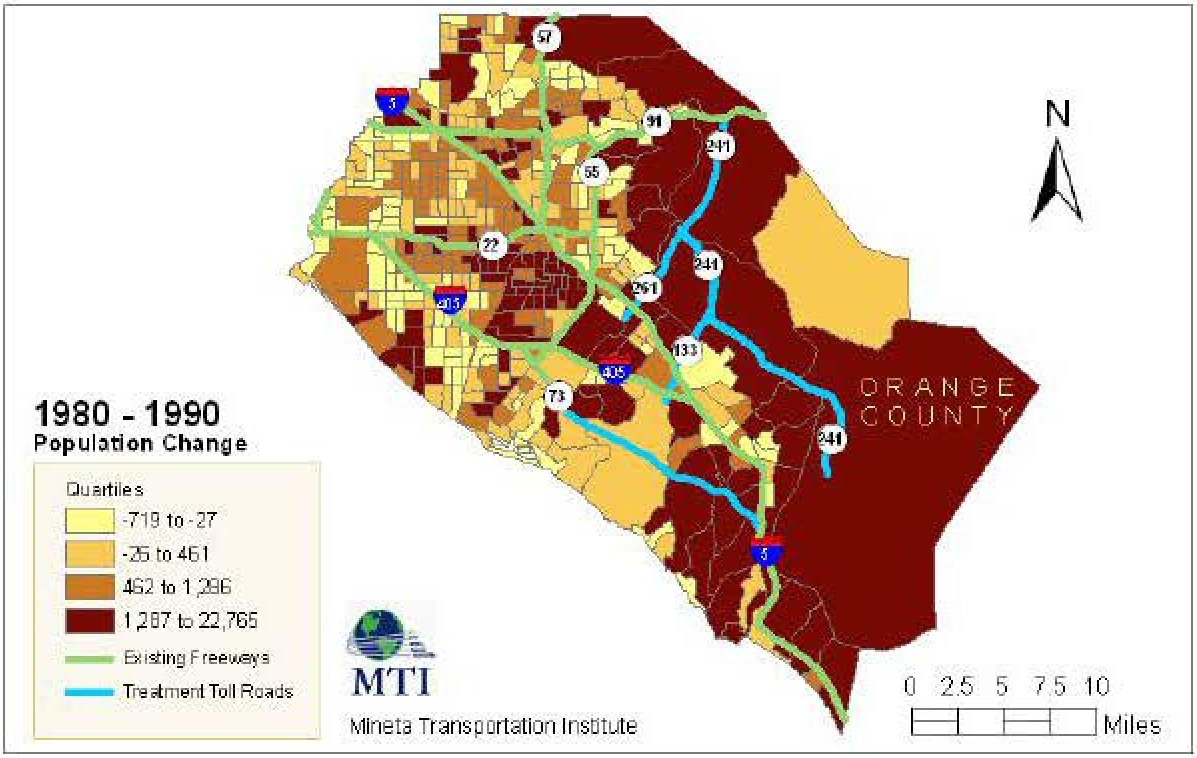
Figure 4 Employment Growth in Merced County 1 km² Grid Cells Before and After Opening of State Route 99 Livingston Bypass

Located in Merced County, the Livingston Bypass is the extension of Highway 99 modeled as an intervention in the natural experiment; the five new centerline miles of Highway 99 freeway that now bypass the 1.5 mile strip of highway that ran through the small town of Livingston are illustrated in Figures 3 and 4. Historically, Highway 99 had multiple stoplights in the business districts of small towns along the route. Gradually, most of these lights were removed and the road was converted to a controlled-access freeway. As early as 1958, there had been discussions to remove the lights at Hammat Avenue and Winton Parkway in Livingston, yet no progress was made for almost four decades (Fimrite, 1996). During the late 1980s, lights were removed in nearby towns, but the lights in Livingston remained as the only stoplights along the 400-mile route (Fimrite, 1996; Thome, 1990).

Labeled “blood alley,” the intersection in Livingston was the site of 46 vehicular deaths between 1976 to 1990 (Houston, 1990). Construction on the bypass began in 1994 and was initially scheduled to be finished a year later. Delays postponed completion a year with southbound lanes opening in September 1996 and northbound lanes opening in December (Fimrite, 1996). Anecdotal reports suggested widespread support for the project, although some business owners expressed concern that there would be a negative impact on their livelihood (Fimrite, 1996). A total of 61 properties, both businesses and homes, were purchased by the California Department of Transportation to construct the project (Thome, 1990).

ORANGE COUNTY

Orange County is a metropolitan county in southern California, nested between Los Angeles County to the north, Riverside County to the east, and San Diego County to the south. The county grew rapidly in population between 1980 and 1990, but experienced little job growth during that decade. Population continued to grow steadily in 2000 and employment growth also gained steam between 1990 and 2000. For the 418 census tracts that averaged 4.94 km^2 in size and collectively exhausted the county in 1980, population in the average tract grew at an annualized rate of 2.23 percent from 4,624 people in 1980 to 5,765 in 1990. The number of jobs remained virtually the same in the 1980s; with a compound annualized growth rate of 0.29 percent, the average tract in 1990 had 2,254 jobs, up from 2,190 a decade earlier. In the 1990s, the population expanded at a slower rate of 1.68 percent and employment growth accelerated to 3.67 percent and in 2000, average tract levels adjusted to 6,808 people and 3,231 jobs.



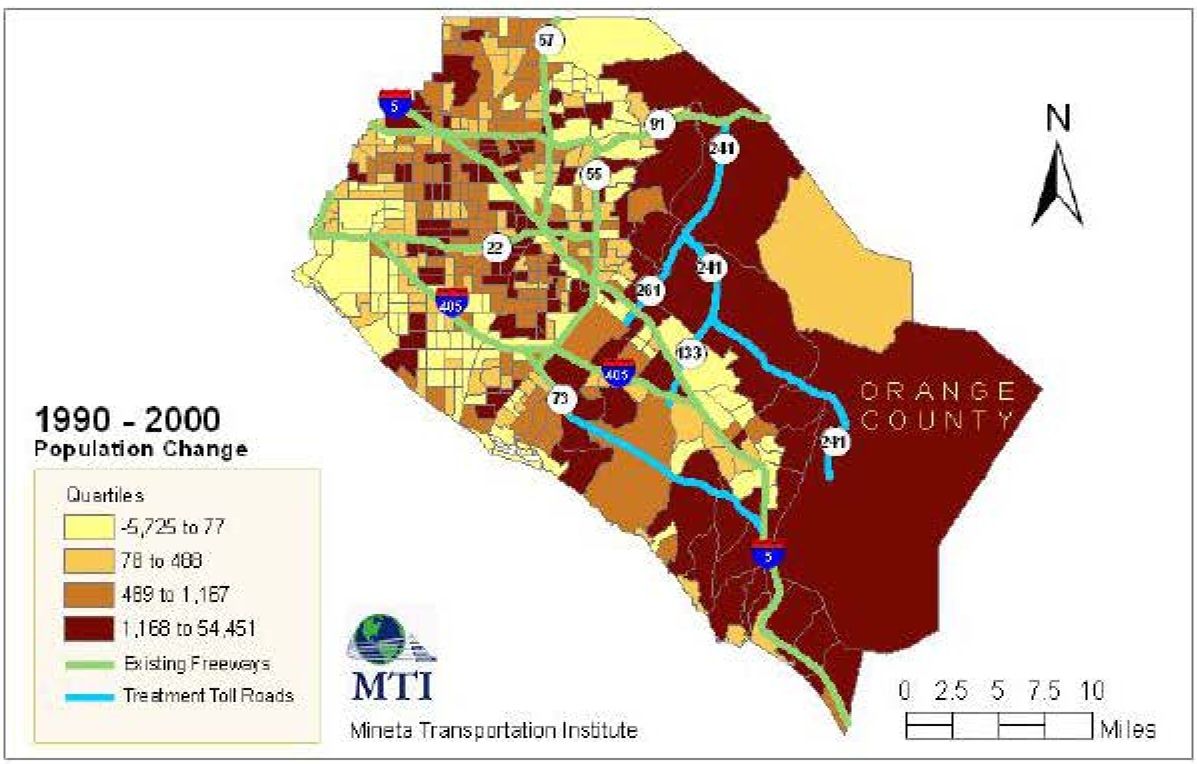
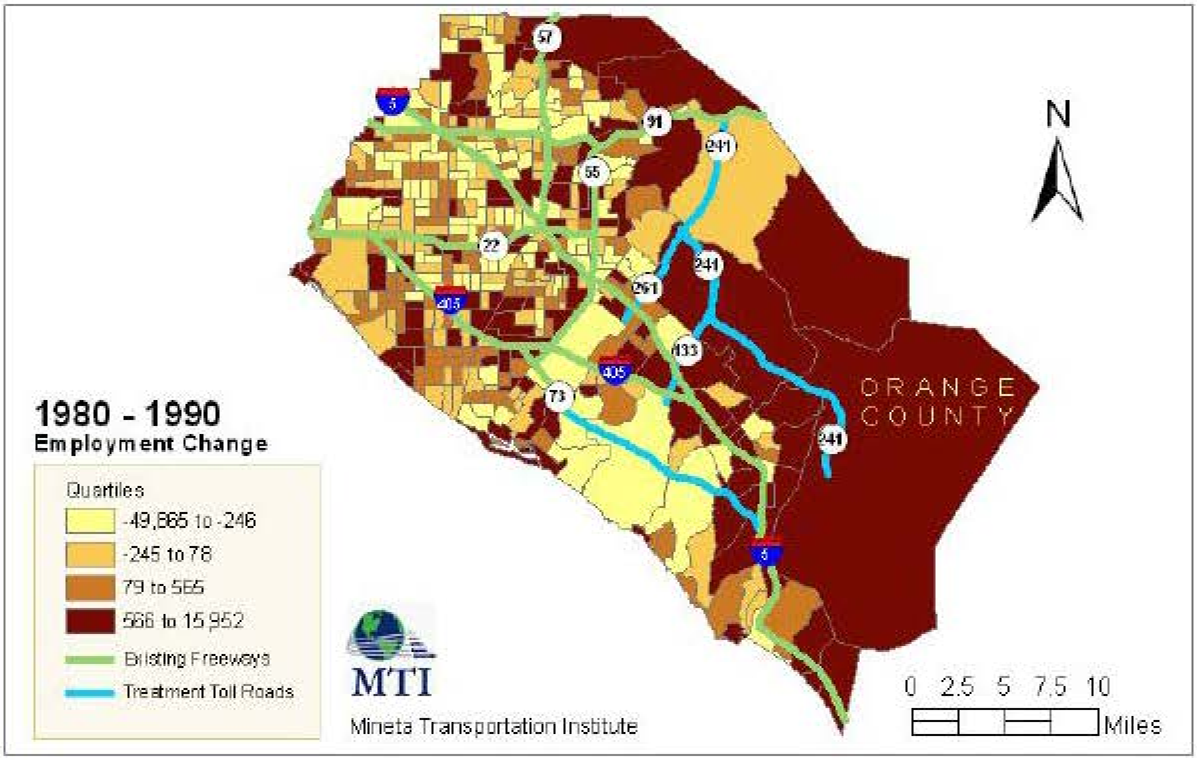


Figure 5 Population Growth in Orange County Census Tracts Before and After Opening of the San Joaquin Hills, Eastern, and Foothill Corridors of the new toll road network (State Routes 73, 241, 261, and portions of 133)

Since 1993, 51 new centerline miles of toll road have opened in Orange County. Collectively, the roads extend the county's relatively dense highway network into the rapidly growing southern, exurban region of the county. Figures 5 and 6 depict the highway and toll road network in the county and growth in its 418 census tracts before and after the toll roads opened. The San Joaquin Hills, Eastern, and Foothill Corridors (California State Routes 73, 241, 261, and portions of 133) were part of Orange County's Master Plan for Arterial Highways since the 1970s, but planning for the toll roads began in earnest when the Transportation Corridor Agencies (TCAs) were created in 1986. In 1987, the TCAs determined that state and federal funds would not be sufficient to finance the roads, and state legislation signed that year allowed the roads to be built as toll facilities. Although in concept the three corridors were imagined during the early 1970s, a case study finds it unlikely that land development would have anticipated the roads before the early 1990s when construction began on the toll road network (Boarnet, DiMento, and Macey, 2002). All 51 centerline miles of new toll road are modeled as intervention in our study.



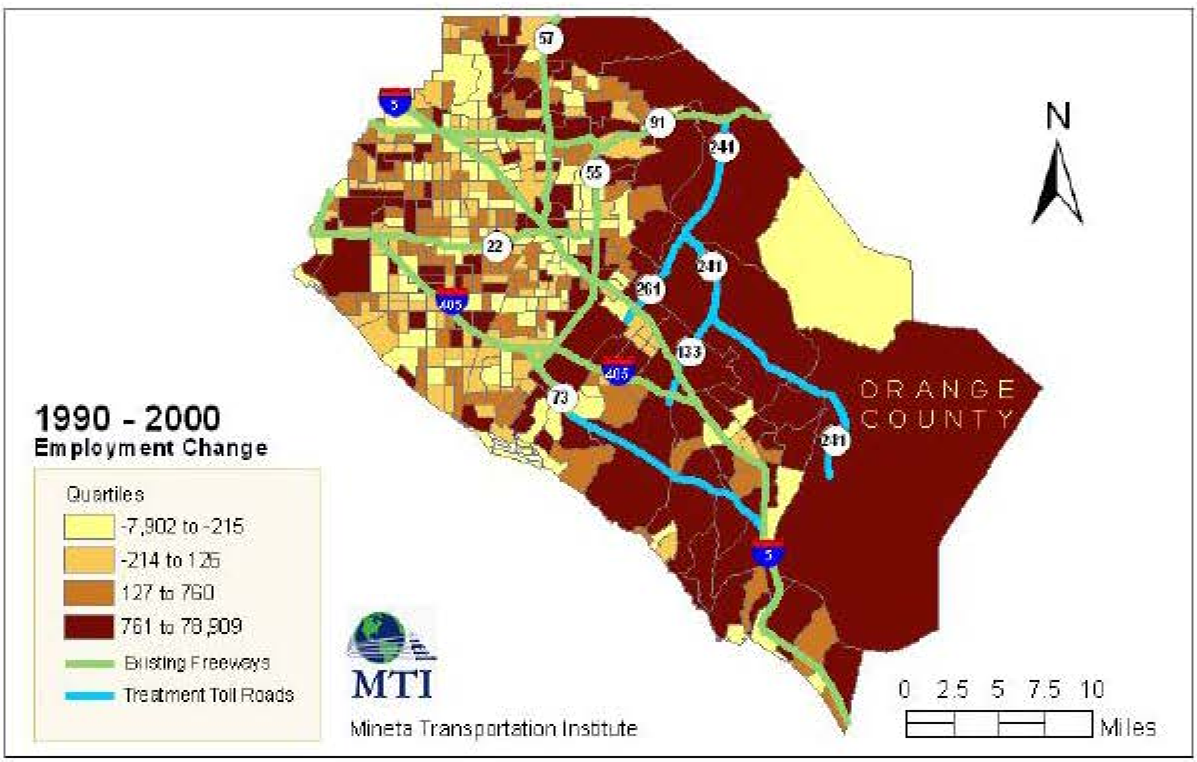


Figure 6 Employment Growth in Orange County Census Tracts Before and After Opening of the San Joaquin Hills, Eastern, and Foothill Corridors of the new toll road network (State Routes 73, 241, 261, and portions of 133)

San Joaquin Hills Corridor

In 1986, Orange County and the cities of Costa Mesa, Irvine, Newport Beach, San Clemente, and San Juan Capistrano entered into a Joint Exercise of Powers Agreement that established and authorized the San Joaquin Hills Transportation Corridor Agency (SJHTCA) to plan, design, and construct the project. The San Joaquin Hills Corridor is a tolled extension of State Route 73 that spans 15 centerline miles north-south, joining Irvine and San Juan Capistrano. Although the SJHTCA had planned to open the stretch by July 1, 1992, the first segment of the San Joaquin Hills Corridor opened to the public on July 24, 1996, and the completed 15-mile corridor opened later that year on November 21.

Foothill and Eastern Corridors

The Foothill Toll Road and the Eastern Toll Road are governed under the jurisdiction of the Foothill/Eastern Transportation Corridor Agency (F/ETCA), which was also established in 1986 by a joint powers agreement between Orange County and the cities of Anaheim, Irvine, Orange, San Clemente, San Juan Capistrano, Santa Ana, Tustin, and Yorba Linda. Although all of State Route 241 is part of the toll road network, its northern half is part of the Eastern Toll Road while its 12 centerline miles in its southern half comprises the Foothill Toll Road. The Foothill Transportation Corridor joins the Eastern Toll Road outside of Irvine with Oso Parkway, running parallel to I-5. The Eastern Transportation Corridor comprises the entire length of State Route 261, the northern part of State Route 133, and the remaining northern half of State Route 241. The east end of Route 241 connects to State Route 91, the Riverside Freeway, to the north and runs south until it splits into two legs at Santiago Canyon Road outside of Tustin. The western leg of the Eastern Transportation Corridor is Route 261, which connects the west side of Irvine and terminates at Walnut Avenue and Jamboree Road. The eastern leg remains Route 241 to connect the east side of Irvine; then south of the Foothill Toll Road, it becomes State Route 133, which continues as the toll road until it intersects I-5. The last segment of the 24 centerline miles comprising the Eastern Toll Road opened to traffic in 1999.

METHODS

Our approach to the natural experiments in each of the three study counties involves six steps. First, we identify the set of spatial units of observation (census tracts or grid cells) that comprise the intervention, or experimental, group in the particular county. Next, we identify the superset of geographic units from which to select the best possible control group. In the third step, we implement alternative matching methods based on different objective functions, a local versus a global optimum. We then analyze changes in the dependent variable by way of a standard quasi-experiment based on the traditional difference-in-differences estimator. In step five, we estimate the conventional simultaneous growth regressions, which implicitly assume that all county observations outside the treatment corridor represent valid counterfactuals. Last, we integrate the quasi-experimental selection of controls into the evaluation equations and compare results to determine the extent that failure to carefully develop appropriate comparison groups biases estimators of the impact of new highway access on regional growth.

The experimental group is comprised of the geographic units defined to have received a highway intervention described in the previous section. Following guidance from previous research demonstrating that effects from new highway access tend to be localized, we include as experimental groups areal units that intersect a new segment of highway, which we treat as a distance of zero miles, and additional tracts or cells with a centroid within one or two miles of the new highway segment (Boarnet and Chalermpong, 2001, 2002; Chalermpong, 2004). In this manner the 0-1 mile treatment group $I'(\text{Treatment} \leq 1\,\text{mi.})$ is a subset of the 0-2 mile treatment group $I'(\text{Treatment} \leq 2\,\text{mi.})$.

As with the experimental group, we employ alternative criteria for identifying the superset of potential controls. When the treatment is defined as receiving new highway access within a crow's-fly (Euclidean) distance of 0-1 miles, one set of controls is drawn from a superset of all other tracts or cells within the same county that are further than one mile to a new segment of highway and another set is drawn from a superset that includes only those geographic units that are within a 1-2 mile distance of a new highway intervention. Similarly, for treatment observations identified by location within a 0-2 mile distance of a new highway centerline, the two sets of controls are drawn from a superset comprised of spatial units that are greater than two miles from the intervention without a maximum distance restriction and one that is within a 2-4 mile corridor of the intervening highway. There are tradeoffs between using the locationally restricted and less restrictive definitions for the superset of potential controls. While locationally similar to the geographic units receiving the treatment, controls drawn from the locationally restricted superset may be less similar in economic and demographic characteristics to the treatment cases than other observations located beyond the distance limit.

Our analytical method is to examine temporal changes in population and employment growth before and after a substantive highway investment in each county and we use quasi-experimental research techniques to understand what would have happened to population and employment if the projects had not been approved. In classical research design, the basic approach to evaluate an impact of an intervention on an outcome variable Y is to randomly assign cases into two groups, an intervention group of observations that experiences a purposeful change in regimen ($I' = 1$) and a control group that does not experience the intervention ($I'=0$). Random assignment of observations to the two groups assures an unbiased experiment because any observed or unobserved characteristics among cases in the study other than differences attributable to experiencing the intervention under question have an equal probability of falling into either group. Another key aspect of experimental research is measuring the outcome before and after the intervention. In an experiment with two time periods, $A=0$ indicates a period before the treatment group receives the intervention and $A=1$ indicates an after-intervention time period. In this study, every spatial unit has two observations, a before-intervention observation of 1980-1990 growth and an after-intervention observation of 1990-2000 growth, and geographic units are indexed by $i = 1, 2, 3, ..., n$. In classical research design, an outcome Y_i such as population or employment growth is typically modeled as

$$Y_i = \alpha + \beta I'_i + \gamma A_i + \delta(I'_i \cdot A_i) + \varepsilon_i$$

where α is an intercept term, β is a permanent experimental group effect accounting for time-invariant differences between the experimental and the control group; γ is a temporal effect common to both the experimental and the control group; δ is the true impact of the new access on the growth outcome variable of interest; and ε_i is a random, i.i.d. error term. The goal of impact evaluation and forecasting in general is to find the best estimate of δ with the available data.

DATA

Insofar as feasible, the data supporting the models comes from publicly available Bureau of the Census sources. The extent that readily available and free public data can be used in these forecast models will broaden the usefulness and application of the tool by practitioners with a need to predict growth under build and no-build scenarios. In the case of rural counties such as Merced County, U.S. Census data were not reported at a fine enough scale in 1980 to allow for complete consistency among the data sources across the three study counties. In 1980, the Census Bureau recorded demographic and economic activity for only 24 designated census tracts in rural Merced County. In contrast, the urban counties of Santa Clara and Orange had 260 and 420 census tracts in 1980, respectively.

The limited utility of census tract-level data for rural regions, often the smallest geography for reporting estimates from Census long-form surveys, is not surprising; however, the 1980 Census data poses an additional challenge for rural counties that affects short-form data as well. A comprehensive accounting of the Census topology that nests census blocks within block groups and block groups within tracts was not completed until the 1992 TIGER/Line. Estimates from the short-form survey on summary tape file 1 for Merced County are recorded for block groups in incorporated towns and Census designated places only. Although companies such as GeoLytics sell proprietary data that adjusts 1980, 1990, and 2000 Census data into alternative boundary definitions for three Censuses, their use in finer-than-tract-scale analysis is not valid if the object of the study is the human settlement pattern as it is in our research. The only basis available to apportion data reported in the 1990 Census back to a geographical unit smaller than a 1980 census tract requires the absurd assumption that census blocks existed in the early decade exactly as they did in the later decade, and no new blocks were added in the later decade. Roads are often the key delimiters of census blocks and as humans inhabit new landscapes, new subdivisions, and new roads; thus, new blocks are created as a result. Therefore, one should exercise caution before purchasing proprietary data reporting to back adjust census data to 1980 and 1990 boundaries smaller than census tracts if the purpose is to predict fine-scale development or land use change.

For the two urban California counties, Orange and Santa Clara, data for our population variables are collected from the 1980, 1990, and 2000 Censuses of Population. The historical tabular data and their GIS shapefiles were obtained from the National Historic Geographic Information System (Minnesota Population Center, 2004). All data conform to 1980 census tract boundaries; in cases where 1990 or 2000 boundaries were not coterminous with 1980 boundaries, we used the locations of 1990 and 2000 blocks and Census short-form population count data to assist with backward apportionment.

Data for our 1990 and 2000 employment variables in the two urban counties come from the employment-by-place-of-work tables in the Census Transportation Planning Package (CTPP) for each of the two decades and were collected from the Bureau of Transportation Statistics. These data, which are reported by traffic analysis zones (TAZes) smaller than 1980 census tracts, are backward adjusted to 1980 tract boundaries by assigning all TAZ employment to the 1980 tract that shares the greatest overlapping area with the TAZ in question. Unlike population, which bears a relationship to the size of a census tract or TAZ, there is no relationship between the size of a TAZ and number of jobs within its boundary. In fact, employment may be more likely to be concentrated in larger geographic units with metropolitan counties due to Euclidean zoning practices that cause some heavily industrialized geographic units to lack population. Therefore, we choose to assign all the jobs in a TAZ to the 1980 census tract with its greatest overlapping area and we do this consistently for each decade of data.

For 1980 employment data in the two urban counties, we use the employment-by-place-of-work-tables in the Urban Transportation Planning Package (UTPP), which was the predecessor to the CTPP. The tremendous benefit of the UTPP is that it provides a consistent employment series to the CTPP that dates back to 1970 in some cases. Its major drawback is that it was only produced for the country's largest metropolitan planning organizations that purchased it. For our needs we were fortunate to be able to obtain the 1980 UTPP data for Orange County from the Southern California Association of Governments and the 1980 UTPP data for Santa Clara County from the Metropolitan Transportation Commission in the San Francisco Bay Area.

The use of the place of work Census data to measure employment location has been criticized because the locations of jobs are reported by workers responding to the Censuses of Population rather than the employing establishments. To the extent that transportation investments shift the location of new employment, however, those concerns are unfounded because such shifts would also alter these commute patterns to employment in the places of work.

To work around the limitation in Merced County with the few, large census tracts typical of a rural region, we purchased point-location microdata that permit the fine-scale detection of changes in human settlement patterns and the distribution of economic activities that can be attributed to the highway intervention. The employment data, which come from Dun & Bradstreet, identify the street address of each establishment in the county, the industry code for the principal product it produces, the number of workers it employs, and when relevant, its total sales for the year; we purchased these proprietary data for each of the three decade years 1980, 1990, and 2000. To proxy for population not reported at fine scale in a rural county, we obtained a complete residential inventory for the county from the Merced County Assessor. The Assessor's data identify the parcel number for every residential structure in the county, the year the structure was built, and the square footage of living space. These two sources of microdata enabled the creation of variables measured at fine spatial scales by imposing a grid over the county and using standard quadrat methods to record the intensity of employment and residential building square footage in each grid cell (Bailey and Gatrell, 1995).

The areal unit for the Orange County and Santa Clara County models is the census tract, while the areal unit for the Merced County models is a one square kilometer cell from a regular grid encompassing the county and snapped to the parallels of the Universal Transverse Mercator (UTM) zone 10N. The regular grid for Merced County overcomes a common small sample problem for Census data in rural areas (there were only 24 census tracts in Merced County in 1980) and takes full advantage of our microdata, which identify specific addresses of county business establishments and places of residence in 1980, 1990, and 2000. Merced County business establishments are simply geocoded to longitude and latitude coordinates and then job totals for each year are aggregated to the grid cell in which the point location falls. In similar fashion, we geocode the Assessor's data and calculate 1980, 1990, and 2000 values for the total amount of residential building space, the percent of building square footage built before 1960, and the share built before 1940 in each cell.

For the Orange County models, we include data for several land use classifications in 1990 that proxy for local land use policies. The data identifying the number of acres in each use within a census tract are known to improve the performance of estimated lag parameters under an assumption of dynamic stability; the data comes from Aerial Information Systems (See Boarnet, Chalermpong, and Geho, 2005). The land use classifications include Single Family Residential (LU1110), Multi-Family Residential (LU1120), Mixed Residential (LU1140), General Office Use (LU1210), Retail Stores and Commercial Services (LU1220), Other Commercial (LU1230), Public Facilities (LU1240), Light Industrial (LU1310), Heavy Industrial (LU1320), Wholesaling and Warehousing (LU1340) Agriculture (LU2000), and Vacant (LU3000).

DIFFERENCES-IN-DIFFERENCES ESTIMATOR

When the assumptions of classical research design hold as in a pure experiment where cases are assigned randomly to experimental and control groups prior to intervention, the simple difference-in-differences estimator is an unbiased estimator of δ (Card and Krueger, 1994). The difference-in-differences estimator δ_{DD} is the difference in average outcome in the intervention group before and after treatment minus the difference in average outcome in the control group before and after intervention

$$\hat{\delta}_{DD} = \overline{Y}^{I=1}_{after} - \overline{Y}^{I=1}_{before} - \left(\overline{Y}^{I=0}_{after} - \overline{Y}^{I=0}_{before}\right)$$

or equivalently

$$\hat{\delta}_{DD} = \overline{Y}^{I'=1}_{after} - \overline{Y}^{I'=0}_{after} - \left(\overline{Y}^{I'=1}_{before} - \overline{Y}^{I'=0}_{before}\right)$$

where $\overline{Y}^{I'=1}_{after}$ and $\overline{Y}^{I'=0}_{after}$ are the sample means of the outcome variable after the period of intervention for the intervention and control group, respectively, and $\overline{Y}^{I'=1}_{before}$ and $\overline{Y}^{I'=0}_{before}$ are the sample averages before the period of intervention for the corresponding groups. For a valid selection of comparison groups $I'=1$ with corresponding $I'=0$, $\overline{Y}^{I'=0}_{after} - \overline{Y}^{I'=0}_{before}$ provides an unbiased estimator of the counterfactual, the change in mean growth for spatial units in the experimental group had those spatial units never received the new highway. The appropriate null and alternative hypotheses for testing the impacts of a highway investment on either the employment growth or the population growth outcome variable is $\begin{cases} H_0: \ \delta = 0 \\ H_a: \ \delta \neq 0 \end{cases}$.

In a natural experiment such as the location of a new highway, experimental observations receiving the intervention have been predetermined by project siting and policy, not by random assignment. Lacking the ability to randomly assign cases and restrict which group receives the intervention, the researcher can use quasi-experimental techniques to mimic the research design of controlled experiments, allowing the possibility of easy-to-understand inferences about the impact of the investment (Boarnet, 2001). If the matching technique sufficiently controls for other factors besides the new highway that might shift, expand, or contract population or employment growth within the spatial units of observation, then a simple comparison of means across experimental and control groups such as the difference-in-differences estimator $\hat{\delta}_{DD}$ would be meaningful and unbiased. However, any differences between the experimental and control groups in important variables that affect the value of a home could bias $\hat{\delta}_{DD}$ as an estimator of δ. In instances when the matching technique is not so perfect, as often is the case, a common solution to this problem is to combine quasi-experimental methods with regression analysis to further control for possibly confounding factors.

Another possible source of bias particular to population and employment growth forecasting pertains to the adjustment process underlying both the employment and the population data we observe. Like most behaviors we observe, human settlement patterns and the location of economic activities are observations of an endogenous equilibrium that reflects the underlying changes in a locality's demand for workers and its labor supply over space and time. In urban economics, the idea of a spatial labor market equilibrium has roots in the monocentric model (Alonzo, 1964; Mills, 1967; Muth, 1969). At its roots, the model and its extensions demonstrate how the willingness to pay for various locations throughout a region adjust in response to changes in population and employment, in turn, adjusting population and employment until a new equilibrium is attained. Also important is the fundamental role in these models for improved access as a shock that lowers transportation costs, flattens the rent gradient, and works to disperse human settlement and economic activities from a central core.

LAGGED ADJUSTMENT MODEL

Our evaluation model adapts the two-equation endogenous growth regression system developed in Boarnet (1992 and 1994) and in Boarnet, Chalermpong, and Geho (2005) that follows a long tradition of intraurban population and employment location models (Bradford and Kelejian, 1974; Steinnes and Fisher, 1974; Carlino and Mills, 1987). We integrate a quasi-experimental selection of controls into the evaluation equations such that

$$POP\Delta_{i,t} = \mathbf{X}_{i,t-1}\hat{\mathbf{a}}_1 + b_1(\mathbf{I}+\mathbf{W})EMP_{i,t-1} + \frac{b_1}{\lambda_e}(\mathbf{I}+\mathbf{W})EMP\Delta_{i,t} - \lambda_p POP_{i,t-1} + c_1 I' + u_{i,t}$$

$$EMP\Delta_{i,t} = \mathbf{Y}_{i,t-1}\hat{\mathbf{a}}_2 + b_2(\mathbf{I}+\mathbf{W})POP_{i,t-1} + \frac{b_2}{\lambda_p}(\mathbf{I}+\mathbf{W})POP\Delta_{i,t} - \lambda_e EMP_{i,t-1} + c_2 I' + v_{i,t}$$

$$I_{i,t} = \mathbf{X}_{i,t-2}\mathbf{d}_1 + \mathbf{Y}_{i,t-2}\mathbf{d}_2 + f_1 POP_{i,t-2} + f_2 EMP_{i,t-2} + \xi_{i,t}$$

where the subscript t refers to time, POP_i and EMP_i are population and employment in spatial unit $i = 1, 2, 3, ..., n$, $\mathbf{X}$ is a matrix of characteristics that affect equilibrium population, $\mathbf{Y}$ is a matrix of characteristics that affect equilibrium employment, $\mathbf{I}$ is an n x n identity matrix, $\mathbf{W}$ is an n x n spatial weights matrix, and I' is an indicator variable identifying membership in the experimental group or matched control group, based on propensity scores estimated from the selection regression of I on predetermined values of all observed characteristics known or expected to influence growth. I is a dummy variable equal to one if the observation is located within a threshold distance of an access point to a highway improvement developed during the mid-1990s and equal to zero otherwise. λ_e and λ_p are adjustment parameters that take on values from the interval $(0,1)$ if the adjustment process is stable, b_1, b_2, c_1, c_2, f_1, and f_2 are scalar parameters to be estimated, $\hat{\mathbf{a}}_1$, $\hat{\mathbf{a}}_2$, $\mathbf{d}_1$, and $\mathbf{d}_2$ are column vectors of parameters to be estimated, and u, v, and ξ are random i.i.d. error terms.

The variables included in the matrices $\mathbf{X}$ and $\mathbf{Y}$ are identified in an extensive body of empirical evidence about the determinants of employment and population growth (Levernier and Cushing, 1994; Clark and Murphy, 1996; Bollinger and Ihlanfeldt, 1997; Mark, McGuire,and Papke, 2000; Edmiston, 2004):

(1) Variables included in both $\mathbf{X}$ and $\mathbf{Y}$ are the land area of a census tract (AREA), the spatial unit's per capita income (*IncomePerCapita*), percentage of adult population with at least a bachelor's degree *(%bachelors*), percentage of adult population with less than a high school diploma (%*noHighschool*), percentage of tract population that is black (%*black*), percentage that is Hispanic (%*hispanic*), population density (*popdensity*), two land use classifications compatible with residential, industrial, and commercial development (LU2000 and LU3000), and a dummy variable indicating if the spatial unit had prior access to a highway in pre-test 1980 (HIGHWAY);

(2) Variables included in $\mathbf{X}$, but excluded from $\mathbf{Y}$ are share of the housing stock built before 1960 (%*pre*60*homes*), share of the housing stock built before 1940 (%*pre*40*homes*), percentage of tract population in poverty (*povertyrate*), percentage of costburdened owners (%*costburdened*); and three land use classifications compatible with residential

development, but incompatible with industrial and commercial development (LU1110, LU1120, and LU1140); and

(3) Variables included in $\mathbf{Y}$, but excluded from $\mathbf{X}$ are share of employment in retail (%*retail*), share of employment in agriculture (%*agriculture*), share of tract employment in manufacturing (%*manufacturing*), share of employment in wholesale trade (%*wholesale*), dollar value of sales per worker (*OutputPerWorker*), and seven land use classifications compatible with industrial or commercial development, but incompatible with residential development (LU1210, LU1220, LU1230, LU1240, LU1310, LU1320, and LU1340).

Although one is tempted to temporally stack, or pool, the regressions, empirical studies have shown that the lagged adjustment parameters λ_e and λ_p vary over time and there is no theoretical basis to suggest that the adjustment process from shocks toward new equilibriums would be time-invariant. The central parameters of interest are c_1 and c_2; although constant for all observed spatial units i for the particular time period estimated ($t-1$ to t), these intervention impact measures will also vary in the "after" test period from the "before" test period if the new infrastructure has significantly affected growth. Insofar that the regression equations are correctly specified, a consistent and unbiased estimator of d is $\hat{c}_{after} - \hat{c}_{before}$, where the subscripts $after = 2000$ and $before = 1990$ indicate whether measurement of the dependent growth change variables is taken for the pre-test (before) or post-test (after) cycle of the natural experiment. The appropriate null and alternative hypotheses for testing the impacts of a highway investment on growth is

$$\begin{cases} H_0 : c_{1,after} - c_{1,before} = 0 \\ H_a : c_{1,after} - c_{1,before} \neq 0 \end{cases} \text{for population and } \begin{cases} H_0 : c_{2,after} - c_{2,before} = 0 \\ H_a : c_{2,after} - c_{2,before} \neq 0 \end{cases} \text{for employment.}$$

PROPENSITY SCORE MATCHING

For our selection of controls, we use a logistic regression of a dichotomous indicator of group membership (experimental or potential control) on the 1980 values of all predetermined variables. The estimated propensity scores for members of the experimental group defined as near a new highway investment are then matched to propensity scores of potential controls located further away from the intervention, based on minimizing distance or cost criteria. The propensity scores are estimated from the logit link function for receiving a highway investment within a threshold distance of the census tract or grid cell centroid $\Pr(I=1)$. Our use of propensity score matching in this instance differs from more common applications that examine the effect of specific programs such as welfare or job training programs for individuals or enterprise zone programs for geographic areas. Rather than predicting a spatial unit's selection to receive a new highway investment, we use the selection regression simply to choose as controls the census tracts or grid cells located farther than a one- or two- mile threshold distance from the improvement that are most similar to those geographic units that comprise the experimental group. The indicator variable I' is constructed based on matched pairs from the selection equation.

For each of the three study regions, Merced, Orange, and Santa Clara Counties, we have identified four alternative sets of matched pairs, two alternative control groups for each of the two distance thresholds (1 or 2 miles) defining receipt of the intervention. Each comparison group has been identified by matching propensity scores calculated by way of the logit link function, given the coefficients to be estimated and values of the independent variables in a logistic regression. Equations are estimated independently for each county, Merced, Orange, and Santa Clara, for two alternative sets of each county's data: (1) the entire set of areal units in the county and (2) a data set comprised of the treatment areal units and a superset of potential controls that are located within 2 miles of a future highway for the $I'(\text{Treatment} \leq 1\,\text{mi.})$ threshold and within 4 miles of a future highway for the $I'(\text{Treatment} \leq 2\,\text{mi.})$ definition.

After experimenting with several matching techniques, we settled on two algorithms; one is a version of caliper matching and the other is a version of kernel matching (Smith and Todd, 2000). Under caliper matching, potential matches are included in the comparison group only if the difference in propensity scores is less than a designated tolerance level, thus permitting control over the quality of the match. In kernel matching, all members of the superset of potential controls (all county census tracts or grid cells in our research) are used, but each observation in the comparison group is weighted proportionately to the quality of its match, or difference in propensity scores, with its closest treatment counterpart. Thus, perfect matches are maximally weighted and lower quality matches are penalized accordingly through smaller expansion weights in this sampling design.

Mainly our two approaches to matching differ in terms of their optimization criteria. Based on our experience, we believe that the greedy algorithm, which minimizes a local cost (difference between propensity scores) as its objective, is more appropriate when the set of potential controls is geographically restricted to be within a threshold distance of the intervention. In the case when controls are not geographically restricted other than coming from the same county, the Hungarian algorithm, which minimizes a global cost as its objective, is preferred because it maximizes degrees of freedom when there are few observations in the intervention group at the start.

LOCALLY OPTMAL MATCHING

We use a 5-to-1 greedy matching algorithm to successively find the best local match among potential controls for each case in the experimental group. The 5-to-1 digit greedy matching algorithm attempts to match the experimental group observations to controls based on five digits of the propensity score. For those that did not match, cases are then matched to controls based on four digits of the propensity score. This continues down until a one digit match on propensity score for those that remained unmatched. Therefore, the process ensures that best matches occur first, second-best matches next, and so on in a hierarchical sequence until no more matches can be made. Best matches are those with the highest digit match on propensity score, in other words, those that have the least absolute difference in propensity scores.

Incomplete matching may result due to two reasons: cases have missing data or there are disjoint ranges of treatment and control propensity scores. Data must be complete for all covariates in the multiple logistic regression analysis used to calculate the propensity score.

If any covariate data are missing, the case is eliminated from the analysis and a propensity score is not calculated. Incomplete matching will result and the cases with missing data will be excluded. Alternatively, the treatment cases and the controls may contain a disjoint range of propensity scores. Incomplete matching will result and the treatment cases with the highest propensity scores and the controls with the lowest propensity scores will be excluded.

The inability of the greedy algorithm to match some observations that are near highway investments to potential controls located further away often results in fewer degrees of freedom in models already confronting small samples. An alternative to locally optimal matching is the Hungarian matching algorithm, which minimizes the sum of all distances between propensity scores and thus forces matches for all observations in the experimental group.

GLOBALLY OPTIMAL MATCHING

The Hungarian matching algorithm assigns a control observation to each case in the experimental group such that the sum of all distances between propensity scores, the global cost, is minimized. Therefore, it has the advantage of including all experimental group observations in the analysis, whereas locally optimal matching necessarily results in the omission of potentially important data. For this reason, we generally prefer the globally optimal Hungarian method to the locally optimal nearest neighbor matching of the greedy algorithm when no maximum distance restriction is imposed on the superset of potential controls. In our assessments of match quality that follow, Matches (> 1 mile) and Matches (> 2 miles) are identified by way of the Hungarian algorithm, which maximizes the number of matched pairs for evaluation in the before and after tests. Therefore, any difference between the number of observations in the intervention group and the number of observations in the matched control that is not geographically restricted (other than coming from the same county) is due to missing values on one or more of the key matching variables in the logistic regression.

Table 1 Mean (Standard Deviation) of Matching Variables, Santa Clara County Intervention Tracts and Matches, 1980 (Intervention ≤ 1 mile)

Characteristic	Intervention	Matches (> 1 mile)	Matches (1-2 miles)	Remainder of County
Census Tract Area (meters^2)	3,098,163.75 (3,347,995.20)	10,284,101.95 (25,765,891.61)	2,887,177.36 (2,393,709.48)	15,228,904.74 (107,716,459.70)
Total Population	5,593.52 (2,420.45)	5,049.50 (1,634.05)	5,413.23 (1,808.10)	4,865.32 (2,214.36)
Total Employment	2,276.04 (3,255.09)	1,657.59 (2,142.97)	1,200.00 (1,163.84)	2,649.38 (6,188.42)
Total Income	51.638,248.43 (26,926,569.90)	48,205.945.14 (18,365,948.97)	54.022,695.77 (22,506,847.15)	46,477,365.34 22,092,478.17
% Bachelors Degree	0.23913461 (0.11193327)	0.2500134 (0.1175262)	0.28201731 (0.14618319)	0.26481410 (0.20066108)
% No High School Diploma	0.21125551 (0.13392410)	0.1991543 (0.0995758)	0.19607488 (0.14428800)	0.22419093 (0.14808729)
% Black	0.01818087 (0.01826051)	0.0138285 (0.0101221)	0.01591781 (0.01692616)	0.03384976 (0.03492634)
% Hispanic	0.14778526 (0.16307937)	0.1381257 (0.1227666)	0.14324343 (0.17623477)	0.17924794 (0.16313575)
% Poverty	0.07570664 (0.06893090)	0.0700183 (0.0407652)	0.06496894 (0.05687257)	0.06960526 (0.04973386)
% Costburdened Homeowners	0.13023751 (0.05375908)	0.1338385 (0.0505224)	0.12879594 (0.03582001)	0.12383179 (0.06649925)
% Agricultural Employment	0.01833431 (0.02356486)	0.0231019 (0.0330868)	0.01946779 (0.03133660)	0.01794113 (0.04311662)
% Construction Employment	0.06642825 (0.06557251)	0.0734020 (0.0708684)	0.06642732 (0.07613016)	0.05201874 (0.05823762)
% Manufacturing Employment	0.15213535 (0.18540279)	0.1194709 (0.1339682)	0.11710457 (0.12168483)	0.16475934 (0.19787209)
% Retail Employment	0.21746046 (0.12897181)	0.2074810 (0.1062872)	0.20957969 (0.09615656)	0.21450928 (0.14210237)
Number of observations	48	44	26	211

Although the globally optimal matching works well when there are many spatial units in the superset of potential controls, the greedy algorithm is appropriate when the superset of potential controls is geographically restricted to be 1-2 miles or 2-4 miles of the new highway investment. In essence, the inability of the greedy algorithm to find a suitable match for an experimental case reflects intolerance for observed differences in characteristics deemed important for our analysis (Smith and Todd, 2005). An outcome of globally optimal matching is that every observation in the experimental group will be matched to a control regardless of the quality of the match (distance between propensity scores). When there are fewer observations in the superset of potential controls from which to match to cases experiencing the intervention, the sum of propensity score distances in the objective function of the Hungarian algorithm, although minimized, will be large, reflecting the fact that matches are forced in globally optimal matching where no good local match would be found under the greedy approach. In the tables assessing match quality that follow, Matches (1-2 miles) and Matches (2-4 miles) are identified by way of the greedy algorithm, which explains why the number of matched pairs for these geographically restricted definitions of potential controls are substantially fewer than the number of cases in the accompanying intervention groups.

MATCH QUALITY

Tables 1 through 6 assess the quality of our matches by comparing descriptive statistics for each definition of an intervention (within 1 or 2 miles of a new highway) to the mean and standard deviation for their matches. Tables 1 and 2 are the results for Santa Clara County, for the two intervention definitions, Tables 3 and 4 are the results for Merced County, and Tables 5 and 6 are the results for Orange County.

The “Intervention” column in each table provides the mean and standard deviation for each variable for those areal units falling within the intervention definition. For example, the centroids of 44 Santa Clara County census tracts in 1980 are located within one mile (Table 1) of a future location of a new highway constructed between 1990 and 2000 and the centroids of 87 census tracts are located within two miles of a future highway (Table 2). The equivalent number of observations for the intervention groups within one mile of the Livingston Bypass in Merced County is 15 grid cells (Table 3) and 16 Orange County census tracts are within a mile of the new toll roads (Table 5). Similarly, 49 cells are within two miles of the Livingston Bypass in Merced County (Table 4) and 38 census tracts are within two miles of a new toll road in Orange County (Table 6). In the case of Merced County, our microdata allow us to precisely locate population and employment, which results in some cells having no activity. Thus, we have one unusable intervention cell in Table 3 and nine unusable intervention cells in Table 4 due to zero values for all key matching variables. Including these in the selection process would not be valid because they would have multiple perfect twins with propensity scores arising completely from the estimated intercept.

Table 2 Mean (Standard Deviation) of Matching Variables, Santa Clara County Intervention Tracts and Matches, 1980 (Intervention ≤ 2 miles)

Characteristic	Intervention	Matches (> 2 miles)	Matches (2-4 miles)	Remainder of County
Census Tract Area (meters²)	2,808,453.23 (2,778,185.26)	5,244,102.85 (12,100,339.02)	3,319,068.16 (6,473,568.86)	18,490,723.31 (120,569,142.40)
Total Population	5,387.95 (2,120.14)	5,028.31 (1,951.52)	4,842.61 (1,516.01)	4,790.29 (2,321.69)
Total Employment	1,950.33 (2,710.77)	2,688.63 (5,898.04)	1,911.16 (1,971.39)	2,922.99 (6,850.53)
Total Income	50,972,666.46 (24,116,879.77)	48,443,744.92 (18,861,160.35)	47,369,358.79 (19,918,518.70)	45,439,312.54 (22,291,667.09)
% Bachelors Degree	0.25134337 (0.12093510)	0.27792875 (0.25595560)	0.24514262 (0.11270649)	0.26508028 (0.21632666)
% No High School Diploma	0.21318648 (0.13491529)	0.21192476 (0.12786174)	0.22419911 (0.13592396)	0.22664807 (0.15116726)
% Black	0.01874814 (0.01873770)	0.02123180 (0.01719502)	0.01929256 (0.01744143)	0.03755298 (0.03702796)
% Hispanic	0.15475248 (0.16598109)	0.16052482 (0.14180159)	0.16977476 (0.16048904)	0.18352698 (0.16137708)
% Poverty	0.07383015 (0.06362870)	0.06683397 0.04244142	0.06609554 (0.04032995)	0.06897801 (0.04727254)
% Costburdened Homeowners	0.12816760 (0.04926214)	0.12701045 (0.04769063)	0.12195468 (0.04797536)	0.12331344 (0.07118707)
% Agricultural Employment	0.01758896 (0.02465244)	0.01399221 (0.02360569)	0.01234169 (0.02590784)	0.01824604 (0.04656404)
% Construction Employment	0.06759669 (0.06559838)	0.05212845 (0.05491524)	0.05410483 (0.06286748)	0.04767181 (0.05534021)
% Manufacturing Employment	0.13921553 (0.16664863)	0.16243937 (0.18776105)	0.13273611 (0.13526448)	0.17504997 (0.20869314)
% Retail Employment	0.22121120 (0.12451793)	0.21178877 (0.12984059)	0.22544612 (0.11010837)	0.21170558 (0.14729966)
Number of observations	91	87	38	168

Table 3 Mean (Standard Deviation) of Matching Variables, Merced County Intervention Grid Cells and Matches, 1980 (Intervention ≤ 1 mile)

Characteristic	Intervention	Matches (> 1 mile)	Matches (1-2 miles)	Remainder of County
Residential Building Area (feet^2)	85,670.00 (157,264.20)	50,100.92 (90,357.25)	6,722.89 (7,940.75)	38,075.75 (117,596.59)
Total Population	472.5155103 (867.3955119)	211.5696873 (377.8187459)	57.3211858 (72.3498889)	153.9648410 (419.8442571)
Total Employment	355.5714286 (1,239.90)	116.4615385 (367.5120441)	83.3333333 (199.8061561)	27.8760529 (120.5754587)
Total Income	2,514,255.03 (4,615,411.52)	1,441,449.18 (2,815,766.69)	318,278.26 (403,750.59)	1,041,359.83 (2,871,453.55)
Median Age of Housing Stock	24.6071429 (14.4477554)	21.8076923 (11.8225393)	29.5000000 (11.2361025)	26.0080764 (14.9819355)
% Bachelors Degree	0.0826653 (0)	0.1253911 (0.0763391)	0.0454016 (0.0324008)	0.0703056 (0.0526061)
% No High School Diploma	0.5786484 (0)	0.4786873 (0.1212435)	0.4084861 (0.2356031)	0.3694660 (0.1976316)
% Black	0.0100762 (0)	0.0192417 (0.0388920)	0.0078132 (0.0044298)	0.0270076 (0.0391416)
% Hispanic	0.4856230 (0)	0.4373636 (0.1317304)	0.2615977 (0.1913668)	0.2152925 (0.1707763)
% Poverty	0.1418039 (0)	0.1535508 (0.0422521)	0.1084815 (0.0615321)	0.1236202 (0.0773681)
% Costburdened Homeowners	0.0880170 (0)	0.0953971 (0.0172734)	0.0840936 (0.0503813)	0.0908525 (0.0606989)
% Agricultural Employment	0.2184148 (0.3665105)	0.2747253 (0.4420943)	0.3333333 (0.5000000)	0.1363141 (0.3334954)
% Construction Employment	0.0063291 (0.0236814)	0.000518583 (0.0018698)	0 (0)	0.0317615 (0.1613183)
% Manufacturing Employment	0.2659816 (0.3820120)	0.3050994 (0.4423994)	0.1094691 (0.3284072)	0.1043808 (0.2735732)
% Retail Employment	0.0935155 (0.2529533)	0.1347882 (0.3324559)	0 (0)	0.0402393 (0.1596560)
Output per Worker	36,289.51 (32,712.20)	37,903.68 (32,471.29)	27,360.49 (27,867.32)	30,085.15 (60,543.39)
Number of observations	14[a]	13	9	831

Notes: a The actual number of grid cells within 1 mile of the new bypass is 15, but one observation had neither housing nor employment in 1980 so it is excluded due to zero values for all key matching variables.

Table 4 Mean (Standard Deviation) of Matching Variables, Merced County Intervention Grid Cells and Matches, 1980 (Intervention ≤ 2 miles)

Characteristic	Intervention	Matches (> 2 miles)	Matches (2-4 miles)	Remainder of County
Residential Building Area (feet^2)	35,015.00 (98,669.84)	46,174.28 (123,132.72)	11,025.30 (24,777.51)	39,055.56 (119,339.09)
Total Population	199.2226452 (543.2196873)	209.4576204 (455.1541788)	88.3301859 (223.1743839)	157.2560176 (425.9988063)
Total Employment	147.0000000 (738.7718152)	80.5128205 (302.5161041)	9.8260870 (27.6859862)	27.6559006 (120.6458660)
Total Income	1,064,061.15 (2,890,010.46)	1,342,687.35 (3,332,724.01)	489,082.79 (1,244,140.99)	1,068,179.53 (2,917,438.22)
Median Age of Housing Stock	27.1973684 (12.4642947)	29.5135135 (16.3908369)	25.0434783 (10.0926538)	25.9094368 (15.0994852)
% Bachelors Degree	0.0721555 (0.0226974)	0.0833494 (0.0491370)	0.0718240 (0.0192427)	0.0704404 (0.0532988)
% No High School Diploma	0.5356848 (0.1289338)	0.5005394 (0.1547037)	0.5036669 (0.0938700)	0.3644045 (0.1971340)
% Black	0.0095643 (0.0022222)	0.0105382 (0.0098088)	0.0199300 (0.0191512)	0.0275798 (0.0396345)
% Hispanic	0.4221550 (0.1358758)	0.3976631 (0.1747079)	0.3496634 (0.1462889)	0.2097150 (0.1682989)
% Poverty	0.1341027 (0.0311917)	0.1436999 (0.0475451)	0.1391081 (0.0031232)	0.1234026 (0.0783693)
% Costburdened Homeowners	0.0888933 (0.0242262)	0.0852541 (0.0311166)	0.1135807 (0.0307506)	0.0909005 (0.0614409)
% Agricultural Employment	0.2014452 (0.3853388)	0.1129700 (0.3095027)	0.1538789 (0.3469414)	0.1345056 (0.3311921)
% Construction Employment	0.0022152 (0.0140101)	0.000102564 (0.000640513)	0.0562660 (0.2146552)	0.0327874 (0.1638032)
% Manufacturing Employment	0.1229873 (0.2907782)	0.1386367 (0.3318768)	0.0889694 (0.2373948)	0.1062667 (0.2755989)
% Retail Employment	0.032730 (0.1528692)	0.0521460 (0.1789397)	0.0048309 (0.0231683)	0.0415390 (0.1620503)
Output per Worker	24,778.10 (31,179.04)	20,080.98 (27,169.12)	21,777.37 (37,254.25)	30,456.75 (61,264.31)
Number of observations	40[b]	39	23	805

Notes: b The actual number of grid cells within 2 miles of the new bypass is 49, but nine observations had neither housing nor employment in 1980 so they are excluded due to zero values for all key matching variables.

In columns three and four, we provide descriptive statistics for the best matches from two alternative supersets of potential controls. The “Matches (> 1 mile)” column summarizes the closest twins based on propensity scores, using all remaining areal units in the same county that are greater than one mile from a future highway as the set from which potential controls are matched to the intervention observations. The “Matches (1-2 Miles)” column restricts potential matches to be within 2 miles of a future highway. Similarly for the intervention group defined as an areal unit within 2 miles of a future highway, the “Matches (> 2 miles)” column summarizes the best matches with no restriction on the distance from a future highway other than that it exceeds the two-mile intervention threshold. The

"Matches (2-4 Miles)" instead restricts potential matches to be within four miles of a future highway.

Table 5 Mean (Standard Deviation) of Matching Variables, Orange County Intervention Tracts and Matches, 1980 (Intervention ≤ 1 mile)

Characteristic	Intervention	Matches (> 1 mile)	Matches (1-2 miles)	Remainder of County
Census Tract Area (meters^2)	51,604,463.12 (99,538,618.41)	22,400,700.68 (28,264,785.09)	9,351,993.65 (3,011,841.75)	3,086,194.07 (6,987,724.32)
Total Population	5,040.38 (3,140.30)	6,007.06 (2,952.85)	7,412.00 (2,674.55)	4,607.12 (1,702.01)
Total Employment	1,328.19 (1,322.90)	1,849.00 (3,008.46)	1,315.00 (1,516.69)	2,224.28 (4,471.98)
Total Income	51,424,700.88 (41,247,398.89)	82,384,455.13 (46,146,758.63)	78,137,304.00 (28,195,108.66)	42,912,422.72 (23,895,742.81)
% Housing Stock Built Before 1960	0.06068383 (0.10625504)	0.07486338 (0.10527719)	0.03033687 (0.05496131)	0.20929890 (0.19843102)
% Housing Stock Built Before 1940	0.02405393 (0.04364181)	0.02276417 (0.06818161)	0.00292665 (0.00611663)	0.03891426 (0.08282402)
% Bachelors Degree	0.32854097 (0.13058306)	0.34911827 (0.12833715)	0.38621748 (0.13041720)	0.20599966 (0.12388499)
% No High School Diploma	0.11969090 (0.09619244)	0.09431806 (0.04721060)	0.08945045 (0.05218357)	0.20487796 (0.13932011)
% Black	0.01494058 (0.03810519)	0.02126795 (0.04280576)	0.00437282 (0.00339262)	0.01265296 (0.01953268)
% Hispanic	0.07553059 (0.04162065)	0.06219114 (0.03404987)	0.04047393 (0.02441810)	0.15222667 (0.16195991)
% Poverty	0.03737764 (0.03208206)	0.03472761 (0.01569287)	0.02795752 (0.00827762)	0.07072444 (0.05168657)
% Costburdened Homeowners	0.27274125 (0.08152558)	0.29290177 (0.20821920)	0.20843082 (0.04811554)	0.15966703 (0.08385520)
% Retail Employment	0.23318853 (0.21789803)	0.27199231 (0.18934105)	0.26852482 (0.18670799)	0.25689017 (0.17358076)
Number of observations	16	16	5	402

The last column in Tables 1 through 6 describes the remainder of the areal units in the county, after excluding those that meet the intervention group definition. The number of observations in this balance of the county category is therefore the total number of usable areal units in the county minus the number of observations in the intervention group. In

1980, there were 260 census tracts in Santa Clara County; however, two tracts were combined to account for reporting discrepancies, making the total number of observations 259. In Merced County, 845 one km² cells had nonzero-valued employment or residential building space in 1980 although the county contains 5,332 one km² cells snapped to the UTM zone 10N grid. In 1980, Orange County had 418 census tracts with complete and usable data.

Table 6 Mean (Standard Deviation) of Matching Variables, Orange County Intervention Tracts and Matches, 1980 (Intervention ≤ 2 miles)

Characteristic	Intervention	Matches (> 2 miles)	Matches (2-4 miles)	Remainder of County
Census Tract Area (meters^2)	13,818,294.69	19,461,683.48	23,759,794.74	4,055,858.49
	(24,504,831.17)	(65,626,705.23)	(81,373,689.31)	(21,726,986.99)
Total Population	4,761.87	4,750.16	5,794.83	4,609.89
	(2,893.82)	(2,447.94)	(2,207.72)	(1,625.81)
Total Employment	1,699.37	3,737.08	3,028.92	2,239.04
	(2,509.88)	(11,339.64)	(10,051.09)	(4,540.90)
Total Income	57,042,216.34	60,599,171.41	68,244,171.04	41,857,855.07
	(42,986,196.03)	(37,683,494.75)	(32,651,144.50)	(21,749,033.33)
% Housing Stock Built Before 1960	0.05643031	0.06891246	0.06862379	0.21800889
	(0.10465293)	(0.12755876)	(0.11829408)	(0.19901367)
% Housing Stock Built Before 1940	0.00997815	0.00637415	0.01036971	0.04105557
	(0.02738328)	(0.01242524)	(0.01914348)	(0.08472155)
% Bachelors Degree	0.32188134	0.32850287	0.31758875	0.19784197
	(0.16116042)	(0.12274762)	(0.11124475)	(0.11684684)
% No High School Diploma	0.09055821	0.09364724	0.10428287	0.21209316
	(0.07242732)	(0.04450442)	(0.05302742)	(0.13953169)
% Black	0.01218828	0.01444585	0.01291406	0.01279575
	(0.02834622)	(0.03334780)	(0.02752409)	(0.01956721)
% Hispanic	0.06893603	0.06804681	0.07200608	0.15732642
	(0.08067399)	(0.03551173)	(0.06092458)	(0.16344275)
% Poverty	0.03362997	0.03709027	0.03893549	0.07302981
	(0.02558479)	(0.01983683)	(0.01874983)	(0.05202942)
% Costburdened Homeowners	0.21022687	0.21148641	0.23082491	0.15793659
	(0.17306888)	(0.07096827)	(0.06696219)	(0.07135272)
% Retail Employment	0.27038047	0.28496370	0.28971066	0.25454318
	(0.20055119)	(0.17782218)	(0.16581922)	(0.17271880)
Number of observations	38	37	24	380

A common yardstick for match quality is that, on net, the matches described in columns three and four should have values for the variables that are closer to the values for the intervention group in column two than the remainder of county group in column five. We also examine the tradeoff between geographic proximity and closeness among the matching variables. Naturally, the values of the variables for matches in column three will be closer to the intervention group than matches in column four because the latter are drawn from a restricted subset of the observations included in matching the former. Although we give up some similarity in values of key variables in column four, we gain geographic similarity by restricting the superset of potential controls to be within two or four miles of a future highway. One drawback to further geographically restricting the superset of potential controls beyond being in the same county is a smaller sample size and fewer degrees of freedom in the evaluation equations. In the next section, we present our impact analyses and findings.

Figure 7 Santa Clara County Census Tracts in Experimental Group Defined as within 1 Mile of New Highway Extension and Matched Controls Located Further Away (Globally Optimal Matching)

LOCATION OF MATCHED CONTROLS

As noted above, globally optimal matching is preferred when there are no additional geographic restrictions imposed on the superset of potential controls other than the requirement that the spatial unit comes from the same county. This enables the approach to find the best match based upon the set of attributes hypothesized to affect growth regardless of the distance to the new highway intervention. Figures 7, 8, and 9 map the locations of the intervention group and the matched control group for Santa Clara, Orange, and Merced Counties, respectively. As indicated by the maps, several of the matched controls are located just beyond the intervention cases and this is particularly true for the two metropolitan counties, Santa Clara and Orange. For the most part, then, the set of attributes deemed important to growth apart from the intervention tend to also identify matches that are spatially near as well as close in terms of values of the matching variables; thus the approach also captures to some extent regional similarity, or spatial regimes.

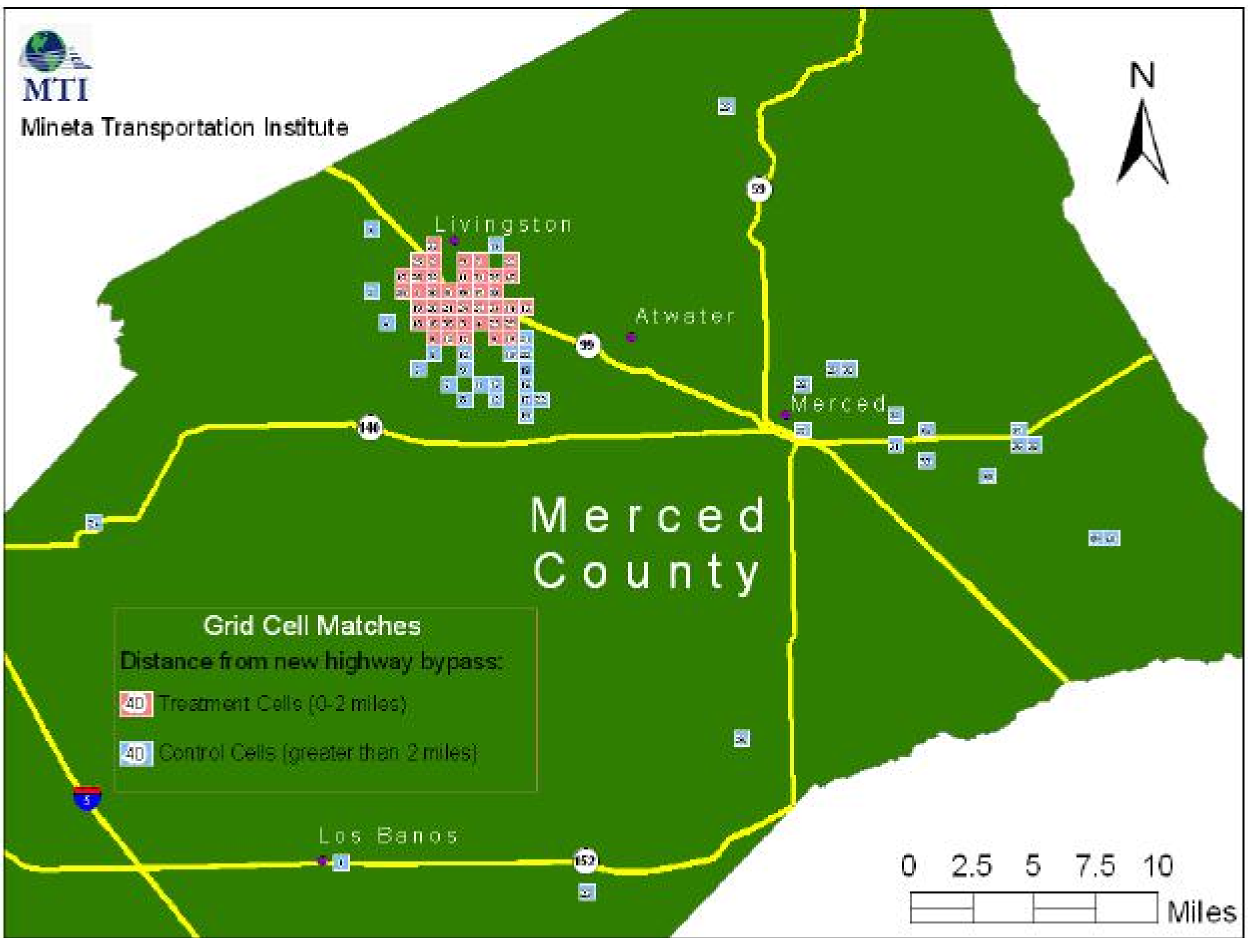


Figure 8 Merced County Grid Cells in Experimental Group Defined as within 2 Miles of New Highway Bypass and Matched Controls Located Further Away (Globally Optimal Matching)

On the other hand, fewer matched controls will be spatially close to the intervention when potential controls nearby are very different from members of the intervention group in terms of values for the key matching variables. Indeed, this is the case displayed in Figure

8 for the one rural county, Merced. It is much more common in this instance to find controls located along key highways in the county or in the outskirts of Merced, the largest city in the county. Naturally this occurs because there are few good substitutes, given the values of the variables, located in the outskirts of the small town of Livingston.

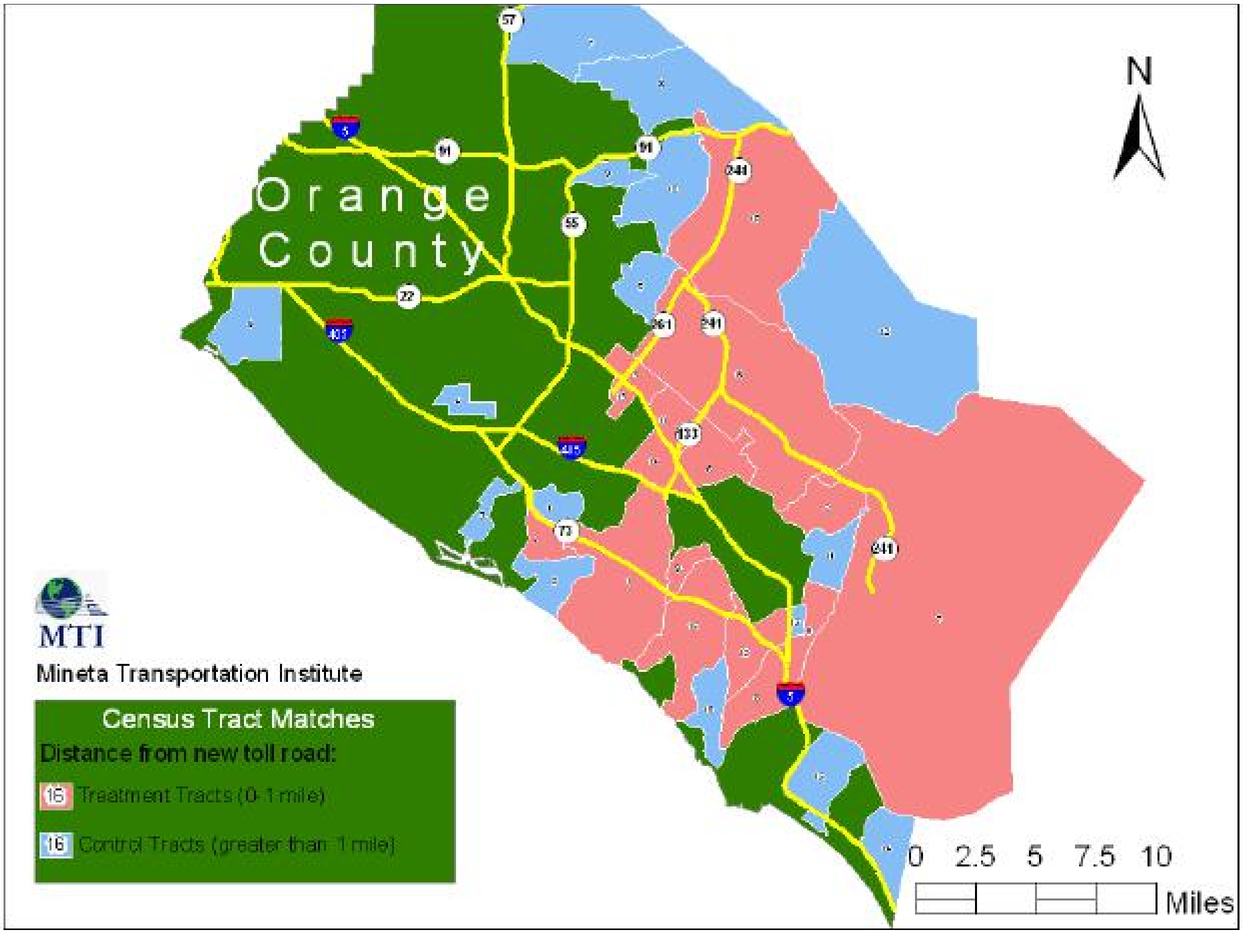


Figure 9 Orange County Census Tracts in Experimental Group Defined as within 1 Mile of New Toll Road and Matched Controls Located Further Away (Globally Optimal Matching)

In addition to displaying the general location of members of the intervention group in light pink and members of the control group in light blue, Figures 7, 8, and 9 also identify the specific location of each matched pair. For Santa Clara County, Figure 7 maps each of the 44 census tracts that have centroids within one mile of one of the two new highway extensions (State Routes 85 and 87) that opened during the mid-1990s. Each of these tracts is assigned a number 1 through 44 and its particular matched control shares the same number. In a similar manner, Figure 8 shows the location of each of the 40 grid cells located within two miles of the Livingston Bypass and its corresponding matched control. Although we conducted the matching and mapping of each of four scenarios (for the two definitions of the intervention group and the two definitions of the superset of potential controls), we provide the figures for only the specification in each of the three counties that had the most interesting results in the analyses that follow. In other words, we find that the Livingston Bypass had a statistically significant employment effect on an intervention group defined as the set of 1 km^2 grid cells that are located within two miles of the new extension; whereas within a distance of one mile, we find no significant employment effect. In the case of Orange County, we find no significant employment effect when the intervention is defined to reach census tracts within two miles; however, we find an employment effect when the intervention is assumed to be more localized at a distance less than or equal to one mile. Thus, Figure 9 locates each of the 16 Orange County census tracts that are within one mile of a new toll road and its matched control. None of the scenarios had interesting results for Santa Clara County in the analyses that follow; Figure 7 identifies the matched pairs given the same definitions in Figure 9 for Orange County to enable comparison between the two metropolitan counties. In the analysis that follows, we offer some potential explanations for the apparent discrepancy in the reach of an intervention in the rural county when compared to the two urban counties.

ANALYSIS

We begin our analysis by investigating summary differences in population and employment growth between experimental and matched control groups before and after the opening of the new highway extensions. Table 7 summarizes results from paired t tests of the difference-in-differences estimator for the highway intervention(s) in each county and the four scenarios involving two alternative definitions for the experimental group and two alternative definitions for the superset of potential controls.

Table 7 Difference-in-Differences Estimator Results

	Scenario 1[a] DIDE (t-statistic)	Scenario 2[b] DIDE (t-statistic)	Scenario 3[c] DIDE (t-statistic)	Scenario 4[d] DIDE (t-statistic)
Santa Clara County				
Total employment	185.84 (0.2766)	-97.105 (-0.1073)	224.80 (0.3900)	104.588 (0.2376)
Total population	30.70 (0.1510)	-90.526 (-0.3192)	228.46 (0.9900)	221.147 (1.2246)
Merced County[e]				
Total employment	90.615 (0.6947)	45.500 (1.0377)	-45.778 (-2.7595)***	-6.478 (-1.3641)
Residential living space (sq. ft.)	40,691.85* (1.8161)	50,224.75 (1.6777)	10,165.14 (1.0924)	15,151.61 (1.4747)
Orange County				
Total employment	6,092.250 (3.6439)***	2,752.000 (2.5085)*	-4,669.963 (-1.1043)	-3,593.292 (-0.7043)
Total population	3,382.563 (1.1019)	824.200 (0.4090)	33.189 (0.0210)	233.833 (0.0994)

Notes. Paired t-test. * and *** indicate the value is significant at a $p < 0.10$ and $p<0.01$ level of significance, respectively.

[a]Scenario 1: Treatment ≤1 mile; Control >1 mile. Degrees of freedom = 37 for Santa Clara County, 12 for Merced County, 15 for Orange County (Hungarian algorithm matches).

[b]Scenario 2: Treatment ≤1 mile; Control 1-2 miles. Degrees of freedom = 18 for Santa Clara County, 8 for Merced County, 4 for Orange County (Greedy algorithm matches).

[c]Scenario 3: Treatment ≤2 miles; Control >2 miles. Degrees of freedom = 75 for Santa Clara County, 38 for Merced County, 36 for Orange County (Hungarian algorithm matches).

[d]Scenario 4: Treatment ≤2 miles; Control 2-4 miles. Degrees of freedom = 33 for Santa Clara County, 22 for Merced County, 23 for Orange County (Greedy algorithm matches).

[e]All Merced County scenarios use Greedy algorithm matches.

DIFFERENCE-IN-DIFFERENCES ESTIMATOR

Table 7 provides the difference (after intervention minus before intervention) in mean differences (treatment minus matched control) for employment growth $\Delta EMP_t = EMP_t - EMP_{t-1}$ and population growth $\Delta POP_t = POP_t - POP_{t-1}$ in each of the three counties and the corresponding t statistic in parentheses, given the estimated difference in differences and its standard error. Scenarios 1 and 2 provide the results when the intervention group is defined as within one mile of the new highway and Scenarios 3 and 4 are the results when the reach of the intervention is defined to be within two miles. The superset of potential controls in Scenarios 1 and 3 are similar in that no distance restriction is imposed other than the fact that controls are drawn from the same county; thus, potential controls are further than 1 mile from the new highway extension in Scenario 1 and further than 2 miles in Scenario 3. In Scenarios 2 and 4, the superset of potential controls is restricted to being 1-2 miles and 2-4 miles, respectively, from the new extension.

None of the four scenarios resulted in estimating a significant employment growth or population growth effect for Santa Clara County. In addition to these definitions, we experimented with a small variation in the identification of experimental group tracts that excluded or included large tracts with centroids further than the given distance threshold from the new highway, yet traversed by it. Including all tracts traversed by the highway in the experimental group (or excluding the larger ones) made no difference in results for Santa Clara County. In the end, we preferred defining tracts intersected by the highways as within zero distance due to the uneven distribution of population and employment over land. In other words, the fact that land within a large census tract, on average, is further than one or two miles of new highway access is likely to be less important to that tract's future growth than the fact that any land within the tract is less than one or two miles of new access. Figure 7 illustrates the distinction well. Note that the experimental tract and the control tract comprising Matched Pair 11 are adjacent to one another and near State Route 85. Although Treatment Tract 11 (in pink) and Control Tract 11 (in blue) are similar in size, the new extension runs through the treatment tract, but not the control tract. Since both tracts are fairly large, their centroids are further than one mile from the new highway. Thus, the tract intersected by the highway was assigned to experimental group while the tract not intersected was assigned to the superset of potential controls and it just so happens that the two were the best match. Given the peculiarity in this particular instance, we also tested models that included Control Tract 11 as part of the intervention group and other models that excluded Control Tract 11 from both the intervention group and the superset of potential controls so that a different match would be found for Treatment Tract 11.

In the case of Merced County, there is some evidence that the Livingston Bypass has resulted in new construction of housing within one mile of an on- or off-ramp for the new section of freeway. In Scenario 1, the estimate for our population proxy suggests that 40,692 more square feet of housing per square kilometer was constructed near the bypass that might be attributed to the intervention and the estimate significantly differs from zero with 90 percent confidence. In addition, there appears to be a large gain in employment very close to access points for the new bypass relative to counterfactuals, although the effect cannot be measured precisely and so the effect is not significant. If we include as the intervention group all 1 km² cells that are within two miles of the new bypass, the effect on our population proxy is much smaller and insignificant. The difference in mean differences, for employment change, however, becomes negative and significantly different from zero with 99 percent confidence. An explanation for the apparent discrepancy is that the bypass induces a positive yet extremely localized effect on growth in cells near the intervention when compared to pertinent counterfactuals such that housing and enterprises serving the freeway spring up very near the new access points; however, the bypass also introduces a countervailing negative employment effect in the town's traditional business district, much of which is located within a range of 1-2 miles of the bypass. Some of the traditional businesses are also located within one mile of the new bypass and these establishments also shed jobs, which leads to the imprecision that causes the estimate of positive employment effects for Scenario 1, although large, to be insignificant. On net then, the Difference in Differences Estimator (DIDE) suggests that the typical 1 km² cell within two miles of the bypass added 46 fewer jobs than would be expected had the bypass not been constructed.

In the case of Orange County, a typical census tract within one mile of a new toll road added 6,092 more jobs after the highway investment than comparable tracts located further away, and the estimate is significantly different from zero with 99 percent confidence. When intervention group tracts are defined to have centroids within two miles of a new toll road, the estimated employment growth effect is negative although not significantly different from zero. This result suggests that employment growth is shifting toward the regions that gained access as opposed to the new access inducing new employment in the county that would not have occurred elsewhere but for the highway investment. When controls are restricted to be within two miles of the intervention as in Scenario 2, the employment growth effect is smaller at 2,752 jobs, although still statistically significant with 90 percent confidence. Together the total employment estimates for Scenarios 1 and 2 suggest a highly localized positive effect on growth that can be attributed to the new highway.

On one level, there is no contradiction between the results from the DIDE analyses in Merced County and Orange County. Both tend to support a highly localized positive effect on employment growth from new highway access. Where they differ is on what happens just beyond the distance within which the positive effect occurs. In the case of the small town of Livingston, the bypass has on net depressed its overall employment growth, while the estimate for Scenario 3 lacks precision for Orange County to suggest any similar net negative overall effect for tracts within a two-mile corridor along both sides of the new toll road system.

Notwithstanding findings in Scenario 1 for Merced, population growth appears to be less responsive to new highway access than employment growth. Furthermore, Scenario 2 for total employment growth in Orange County was the only instance where restricting the superset of controls to within a two or four-mile distance of the intervention resulted in a statistically significant result. Again, this result should not be surprising; particularly Scenario 2 takes some of the members of the intervention group defined in Scenario 3 and treats them as controls in the Scenario 2 estimations and tests. In addition, limiting the scenarios with geographically restricted controls to locally optimal matches results in small sample sizes and tests that lack statistical power. Only 19, 9, and 5 matched pairs were identified, respectively, for Santa Clara, Merced, and Orange Counties when controls were restricted to within 1-2 miles of the new highway in Scenario 2 and only 34, 23, and 24 matched pairs were identified for the same respective counties when controls were restricted in Scenario 4 to 2-4 miles of the intervention.

We now turn our attention to our evaluation models that subject these initial findings to the context of a spatial labor market equilibrium where human settlement patterns and the location of economic activities are observed as the result of an endogenous equilibrium that reflects the underlying changes in a locality's demand for workers and its labor supply over space and time.

LAGGED ADJUSTMENT MODEL

In this section of the paper, we focus our efforts on subjecting the initial findings from the paired t tests of the Difference in Differences Estimator to the adjustment process and control variables, assuming the matching was less than perfect to mimic a controlled experiment. In that spirit, we restrict our discussion to the analyses of interventions conceived in Scenario 1 for the two metropolitan counties, Santa Clara and Orange and the intervention tested in Scenario 3 for Merced County. Specifically, we subject to the additional tests the assumptions of the two-mile reach of the bypass for the one small town case and a more localized one-mile reach of the highway extensions in Orange and Santa Clara counties.

Full Specifications

Our starting point for modeling the two-equation simultaneous growth system presented in Section 3 was to replicate the results from the 1980 to 1990 model for Orange County in Boarnet, Chalermpong, and Geho (2005). With some minor additions of control variables as suggested by Edmiston (2004) and inclusion of the key impact variable $I'(\text{Treatment} \leq 1\,\text{mi.})$ that identifies census tracts within one mile of the new toll road network, we estimated the 1980 to 1990 model, covering the period before the toll road opened, and a model covering 1990 to 2000, the period after the toll road opened. Results from our conventional model of the lagged adjustment process in Orange County are presented in Table 8. By conventional, we mean that the lagged adjustment models do not incorporate the quasi-experimental selection of controls, the third equation in the system described in the Methods Section. Instead, the conventional models still implicitly assume that all observations (tracts or cells) in the county are valid comparisons for the subset that receives the intervention.

The results for the 1980-1990 model in Table 8 are qualitatively similar to those reported in Boarnet et al. (2005) when **W** is defined as a non-normalized contiguity matrix. Although we performed each of our estimations in the current paper using both a non-normalized contiguity matrix and an inverse distance matrix in alternative specifications of **W**, we found that this choice of weights made no material difference in our estimates of the key impact variable. Therefore, in all tables that follow, we report only the results from models with **W** defined as the non-normalized first-order contiguity matrix for the spatial units in a county.

Table 8 Conventional Lagged Adjustment Model Estimates, Orange County, California

	ΔPopulation					
	1980-1990			1990-2000		
Variable	Estimate		StdErr	Estimate		StdErr
Intercept	181.4118		451.6341	-963.2653		872.8227
$\mathrm{LU1110}_i$	6.8845	***	0.3975	1.0289		0.8287
$\mathrm{LU1120}_i$	15.0975	***	1.0425	8.8872	***	1.9120
$\mathrm{LU1140}_i$	16.8323	*	7.9112	13.0466		11.2777
$\mathrm{LU2000}_i$	1.6692	***	0.2585	1.2343	**	0.3774
$\mathrm{LU3000}_i$	-0.2279		0.1258	0.9764	***	0.1791
AREA_i	0.0000		0.0000	0.0001	***	0.0000
$IncomePerCapita_{i,t-1}$	-0.0546		0.0289	-0.0270		0.0205
$\%bachelors_{i,t-1}$	-1615.8723		1122.9842	2518.6281		2883.0281
$\%noHighschool_{i,t-1}$	-47.5710		1338.7249	-735.9203		2619.7358
$\%black_{i,t-1}$	11179.9020	**	4011.9076	10365.5987		7105.8808
$\%hispanic_{i,t-1}$	1233.9847		1045.1873	2058.7498		1938.6261
$popdensity_{i,t-1}$	203645.1153	**	75073.5964	29909.0120		85047.1786
$\%pre60homes_{i,t-1}$	-230.4785		454.4415	315.5446		659.1750
$\%pre40homes_{i,t-1}$	614.9856		950.8411	-3521.2213		1845.5910
$povertyrate_{i,t-1}$	4706.1813	*	2134.9473	2044.8564		2166.5298
$\%rentburdened_{i,t-1}$	503.9568		887.2090	995.1246		1387.0367
$(\mathbf{I}+\mathbf{W})EMP_{i,t-1}$	0.0825	**	0.0267	-0.0475		0.0320
$(\mathbf{I}+\mathbf{W})EMP\Delta_{i,t}$	0.1064	*	0.0448	-0.0048		0.0222
$POP_{i,t-1}$	-0.4718	***	0.0535	-0.1191		0.0742
$\mathrm{HIGHWAY}_i$	189.7703		152.5888	-445.1089	*	219.9253
$I'(\text{Treatment} \le 1\text{ mi.})$	520.4676		446.5823	2165.2175	***	632.3054
n	415			415		
k	57			57		
F	27.8626			23.3083		
prob>F	0.0000			0.0000		
R-squared	0.8165			0.7882		
Adj R-squared	0.7872			0.7544		
Root MSE	1217.1856			1725.5588		
Over-Id ~ F	4.8015			1.3378		
Over-Id (prob>F)	0.0001			0.2159		

	ΔEmployment					
	1980-1990			1990-2000		
Variable	Estimate		StdErr	Estimate		StdErr
Intercept	-729.4412		913.8423	421.2627		899.1504
$\mathrm{LU1210}_i$	-11.5735	**	3.7888	66.4402	***	2.0249
$\mathrm{LU1220}_i$	25.1189	***	5.1553	3.5442		4.2457
$\mathrm{LU1230}_i$	13.0646	**	4.2752	18.5947	***	3.2537
$\mathrm{LU1240}_i$	9.7573		12.9179	12.1863		11.6056
$\mathrm{LU1310}_i$	-1.0806		2.4782	22.2435	***	1.5863
$\mathrm{LU1320}_i$	90.3418	***	23.5536	-124.8818	***	17.4785
$\mathrm{LU1340}_i$	4.4823		7.2104	-3.9805		5.2841
$\mathrm{LU2000}_i$	0.2296		0.5151	1.5349	***	0.3922
$\mathrm{LU3000}_i$	0.0838		0.2704	0.3381		0.1997
AREA_i	0.0000		0.0000	0.0000	*	0.0000
$IncomePerCapita_{i,t-1}$	-0.0022		0.0574	-0.0136		0.0209
$\%bachelors_{i,t-1}$	4647.8819	*	2093.6670	706.8994		2976.1371
$\%noHighschool_{i,t-1}$	-1001.5476		2385.0908	1450.3955		2764.1832
$\%black_{i,t-1}$	-22137.7819	**	7786.3330	26515.3680	***	7436.7392
$\%hispanic_{i,t-1}$	3937.3019		2044.5030	-252.9398		1963.9796
$popdensity_{i,t-1}$	-499781.5796	**	152396.3813	-149272.5428		93031.6173
$\%retail_{i,t-1}$	-715.8557		865.9462	-258.2973		928.6904
$(\mathbf{I}+\mathbf{W})POP_{i,t-1}$	0.0686		0.1023	0.0578		0.0703
$(\mathbf{I}+\mathbf{W})POP\Delta_{i,t}$	-0.0997		0.1205	-0.0401		0.1424
$EMP_{i,t-1}$	-0.5548	***	0.0708	-0.6637	***	0.0380
$\mathrm{HIGHWAY}_i$	53.9587		308.6017	-245.7578		230.2720
$I'(\text{Treatment} \le 1\text{ mi.})$	-1900.9911	*	898.5431	1668.0832	*	675.6152
n	415			415		
k	58			58		
F	9.6287			60.4351		
prob>F	0.0000			0.0000		
R-squared	0.6107			0.9078		
Adj R-squared	0.5473			0.8928		
Root MSE	2453.7323			1822.9566		
Over-Id ~ F	2.2973			3.0566		
Over-Id (prob>F)	0.0208			0.0024		

Note: Although not included in the table, census designated place dummy variables are included in the models.

Another interesting similarity with the results in Boarnet et al. (2005) is how inclusion of the land use classification variables is necessary to stabilize the coefficients on $POP_{i,t-1}$ and $EMP_{i,t-1}$ in the population growth and employment growth equations, respectively. As mentioned above, λ_e and λ_p are adjustment parameters that take on values from the interval $(0,1)$ if the adjustment process is stable. Since the coefficients on $POP_{i,t-1}$ and $EMP_{i,t-1}$ are the negative of λ_p and λ_e, respectively, the coefficients are expected to take on values from the interval $(-1,0)$ when adjustment is stable. As Boarnet and others (2005) point out, λ_p can also be inferred from the coefficient on $(\mathbf{I} + \mathbf{W})EMP_{i,t-1}$ and λ_e from the coefficient on $(\mathbf{I} + \mathbf{W})POP_{i,t-1}$. Notwithstanding that our results for both the 1980 to 1990 and 1990 to 2000 models conform to the hypothesized values of lambda suggesting a stable adjustment process, it has been speculated that a missing variable correlated with the temporally lagged endogenous variable, while biasing the estimator of the lagged parameter, does not necessarily bias parameter estimators for other variables in the model (Boarnet, 1992, 1994). This becomes an important assumption in our paper because we lack similar data for the other two counties that quantify the amount of land available for development in the various types of commercial, industrial, and residential uses. Therefore, the models for the other two counties and the sparse variable specifications described below tend to estimate coefficients on $POP_{i,t-1}$ and $EMP_{i,t-1}$ that are outside the interval suggested by a stable adjustment process.

The key hypothesis tests of our interest in the lagged adjustment models involve the estimates of the impact parameters $c_{1,before}$ and $c_{1,\mathrm{after}}$, which are the coefficients on $I'(\text{Treatment} \leq 1\,\text{mi.})$ in the population equations for 1980-1990 and 1990-2000, respectively and $c_{2,before}$ and $c_{2,\mathrm{after}}$, which are the coefficients on $I'(\text{Treatment} \leq 1\,\text{mi.})$ for the 1980-1990 and 1990-2000 employment equations. While controlling for the simultaneous adjustment process as shifts in supply and demand bring about new equilibriums and while controlling for other factors that affect growth, $c_{1,\mathrm{after}}$ and $c_{2,\mathrm{after}}$ quantify any residual gain or loss not elsewhere explained by the model that is directly attributable to being close (in this case within one mile) to a new highway. Since the new highway did not open until the after period, in the same spirit as the DIDE we also test null hypotheses involving $c_{1,before}$ and $c_{2,before}$ to ensure such gains or losses are not spurious and therefore related to the new highway and also ensure that signals of new land development did not precede the after period. An appropriate test in this instance simply examines whether the confidence intervals overlap for $c_{1,before}$ and $c_{1,\mathrm{after}}$ and whether they overlap for $c_{2,before}$ and $c_{2,\mathrm{after}}$. In the event that confidence intervals for the parameters in the two time periods do not overlap, then a change in growth trajectory can truly be ascribed to highway intervention.

In the results of the population equations in Table 8, the impact variable indicates that tracts near the new highway on average added 2,165 more people after the new toll roads than the average census tract in Orange county and the population gain is significantly different from zero. The 95 percent confidence intervals are thus $-358 < c_{1,before} < 1{,}399$

and $922 < c_{1,after} < 3{,}409$, which indeed overlap so that we may not ascribe the gain to the new highway with a lot of confidence. In the case of the employment equations, however, the confidence intervals do not overlap; we have as 95 confidence intervals $-3{,}667 < c_{2,before} < -133$ and $339 < c_{2,after} < 2{,}997$. Now, since the opening of the new toll roads, there has been a clear reversal in the trajectory for new employment in the census tracts near the new extensions and the impact for these tracts in the experimental group is clearly positive. But these estimates are in comparison to the grand mean for all census tracts in Orange County, and not necessarily an appropriate comparison group. The conventional model therefore prohibits one from making statements about a no-build scenario or the so-called counterfactual. Only by limiting the regression to the group of census tracts that are near the toll roads and their relevant comparisons can one interpret the coefficients in terms of the impact that can be ascribed to the state with the highway when compared to the state without the highway. The conventional model may therefore introduce two sources of error that belie the true impact about which we wish to infer from the models: (1) it artificially increases the degrees of freedom and statistical power of a test and (2) it could substantially alter the grand mean by which we will make a comparison to the conditional mean where $I' = 1$ and all other variables in the model are held constant. We thus turn our efforts toward the models that integrate a quasi-experimental selection of controls into the lagged adjustment models.

Sparse Specifications

As with the pure quasi-experimental approach that tests the Difference in Differences Estimator (DIDE), the integrated models suffer a substantial loss of statistical power due to fewer degrees of freedom. Consequently, we make two adjustments to our analyses in the integrated models due to the fewer number of observations, restricted to those that are relevant to the build versus no-build scenarios of interest. As the first adjustment, we find it necessary to specify far fewer control variables than possible when all spatial units are used as observations. We therefore maximize the degrees of freedom for each hypothesis test by restricting the variables included in the integrated models to the structural components of the lagged adjustment process and the variables instrumenting for ($\mathbf{I}$ + $\mathbf{W}$)EMP$\Delta_{i,t}$ and ($\mathbf{I}$ + $\mathbf{W}$)POP$\Delta_{i,t}$ in the first-stage employment and population equations, respectively. The other variables required by the structure of the model include ($\mathbf{I}$ + $\mathbf{W}$) EMP$_{i,t-1}$, ($\mathbf{I}$ + $\mathbf{W}$)POP$_{i,t-1}$, POP$_{i,t-1}$ and EMP$_{i,t-1}$. As the second adjustment, our criterion for ascribing an impact uniquely to the new highway investment is non-overlapping 90 percent confidence intervals for the before and after coefficients on the impact variable. Although not reasonable for the hypothesis tests in the conventional models where the degrees of freedom are ample for a powerful test, the far fewer observations in the integrated models makes us less suspicious of the potential for a type I error; moreover, we have estimates from the conventional model and difference in difference estimator to guide us through comparison.

Table 9 Quasi-Experimental Integrated Model Estimates, Orange County, California

	ΔPopulation					ΔEmployment			
	1980-1990		1990-2000			1980-1990		1990-2000	
Variable	Estimate	StdErr	Estimate	StdErr	Variable	Estimate	StdErr	Estimate	StdErr
$(\mathbf{I}+\mathbf{W})EMP_{i,t-1}$	-0.0245	0.0899	0.1230	0.2173	$(\mathbf{I}+\mathbf{W})POP_{i,t-1}$	0.0463	0.0313	0.0517	0.0349
$(\mathbf{I}+\mathbf{W})EMP\Delta_{i,t}$	0.2082	0.1846	-0.3162	0.3000	$(\mathbf{I}+\mathbf{W})POP\Delta_{i,t}$	0.0598	0.0382	-0.0562	0.0630
$POP_{i,t-1}$	0.3233	0.4659	0.5679	0.3500	$EMP_{i,t-1}$	0.2094	0.1897	-0.0663	0.2850
HIGHWAY_i	-1500.1333	2647.6331	-290.5739	5308.9336	HIGHWAY_i	115.8006	869.9481	845.5029	1963.7669
$I'(\text{Treatment} \le 1\text{ mi.})$	2949.7066	2706.6549	4286.1238	5415.2169	$I'(\text{Treatment} \le 1\text{ mi.})$	-1945.4932 *	922.8180	3775.6067	2223.1679
n	30		30		n	30		30	
k	5		5		k	5		5	
F	1.0967		1.4745		F	1.7259		2.2254	
prob>F	0.3879		0.2350		prob>F	0.1670		0.0848	
R-squared	0.1860		0.2350		R-squared	0.2645		0.3168	
Adj R-squared	0.0164		0.0756		Adj R-squared	0.1112		0.1744	
Root MSE	7154.1283		14135.4141		Root MSE	2283.7913		5058.2167	
Over-Id ~ F	0.7977		0.3324		Over-Id ~ F	0.0721		1.1599	
Over-Id (prob>F)	0.5841		0.9110		Over-Id (prob>F)	0.9981		0.3698	

Table 9 presents the results from the integrated model that incorporates the quasi-experimental selection of controls into the lagged adjustment model for Orange County. Despite estimates for the lag coefficients on $POP_{i,t-1}$ and $EMP_{i,t-1}$ that are outside the interval for a stable adjustment process, the estimates produced by this model infer impacts attributable to the new toll roads that are qualitatively similar to those produced by the conventional model. The 90 percent confidence intervals for parameter measuring employment growth impacts from the new highway are $-3{,}524 < c_{2,before} < -366$ and $-28 < c_{2,after} < 7{,}579$. Therefore, we can infer that a reasonable estimate of the employment growth impact in census tracts near the toll road that can be ascribed to the new investment lies in the interval -28 + 366 = 338 to 7,579 + 3,524 = 11,103. Had the toll roads not been built, then we might have anticipated 338 to 11,103 fewer jobs in these census tracts than the current levels. Even at the lower level, the difference is statistically and economically significant. Given that average employment in 1990 (before the new toll roads) was 1,868 for census tracts near where a toll road would later open, the estimates suggest upwards of an 18 percent shift in employment growth to the tracts that gained access when compared to their relevant counterfactuals. Note also that the difference in mean differences of 6,092 falls neatly between the range of probable impacts that can be ascribed uniquely to the opening of the toll roads. Again, in the case of population, however, we are unable to ascribe an affect because the 90 percent confidence intervals overlap.

Merced County

Once the conventional lagged adjustment model for Orange County was calibrated and producing estimates for the 1980-1990 before-intervention period that are equivalent to findings in Boarnet et al. (2005), we adapted the specification to the other two counties to the extent possible. Table 10 provides results from the conventional lagged adjustment model for Merced County. Including all 5,332 cells for the county in the conventional regression approach results in an estimated impact on population growth in the 1980-1990 period that suggests cells near the new bypass on average added 7,271 fewer square feet of housing than a typical 1 km^2 cell in the county, a statistically significant effect. However, the difference in new living space precedes the opening of the highway and the 95 percent confidence intervals for the population change effect overlap for the before- and after-intervention periods. Therefore, we may not ascribe a change in residential construction to the new bypass.

Table 10 Conventional Lagged Adjustment Model Estimates, Merced County, California

	ΔPopulation								ΔEmployment					
	1980-1990			1990-2000					1980-1990			1990-2000		
Variable	Estimate		StdErr	Estimate		StdErr		Variable	Estimate		StdErr	Estimate		StdErr
Intercept	-321.7924		312.0011	-903.0601		497.8396		Intercept	0.0196		0.5347	3.2964	**	1.0040
$\%pre60homes_{i,t-1}$	-0.5720	***	0.0402	-0.5826		0.0468	***	$\%agriculture_{i,t-1}$	-8.2544	*	3.8316	6.7627		4.8446
$\%pre40homes_{i,t-1}$	-0.2384	***	0.0580	-0.3734		0.0958	***	$\%manufacturing_{i,t-1}$	6.3480		5.8500	9.0561		11.7533
$(\mathbf{I}+\mathbf{W})EMP_{i,t-1}$	8.1146	***	1.0996	3.7371		1.0978	***	$\%wholesale_{i,t-1}$	29.9929	***	7.2806	40.9974	***	10.9926
$(\mathbf{I}+\mathbf{W})EMP\Delta_{i,t}$	88.2574	***	12.5689	83.1488		14.9464	***	$OutputPerWorker_{i,t-1}$	0.0000		0.0000	0.0000	**	0.0000
$POP_{i,t-1}$	0.2545	***	0.0100	0.2444		0.0094	***	$(\mathbf{I}+\mathbf{W})POP_{i,t-1}$	0.0001	***	0.0000	0.0000	***	0.0000
$I'(\text{Treatment} \le 2\text{ mi.})$	-7270.7651	*	3046.0268	4745.5297		3820.4184		$(\mathbf{I}+\mathbf{W})POP\Delta_{i,t}$	0.0000		0.0000	-0.0002	***	0.0000
								$EMP_{i,t-1}$	-0.1542	***	0.0066	0.4070	***	0.0110
								$I'(\text{Treatment} \le 2\text{ mi.})$	-6.0378		5.2173	27.9983	***	8.5294
n	5332			5332				n	5332			5332		
k	6			6				k	8			8		
F	187.6355			165.2497				F	180.1550			249.9343		
prob>F	0.0000			0.0000				prob>F	0.0000			0.0000		
R-squared	0.1745			0.1570				R-squared	0.2131			0.2731		
Adj R-squared	0.1736			0.1560				Adj R-squared	0.2119			0.2720		
Root MSE	19981.1603			25917.0099				Root MSE	35.8118			58.7645		
Over-Id ~ F	11.0504			9.5729				Over-Id ~ F	90.0063			169.2159		
Over-Id (prob>F)	0.0001			0.0001				Over-Id (prob>F)	0.0001			0.0001		

More interesting, the coefficient on the impact variable in the 1990-2000 employment equation suggests that the new bypass induced a gain of 28 jobs per square kilometer near the investment and the difference is statistically significant. Moreover, the 95 percent confidence intervals are $-16 < c_{2,before} < 4$ and $11 < c_{2,before} < 45$ and therefore do not overlap.

The problem, of course, is that using all 5,332 one km^2 cells to implicitly develop the counterfactual artificially increases the degrees of freedom and statistical power of the test yielding the narrow confidence bounds above and all the zero-valued cells in a rural county substantially reduce the mean from which the contrast is derived for estimating the coefficient on the impact variable. Now, Table 11 provides results for our quasi-experimental integrated approach.

Comparing results from the conventional regression to the model that integrates a quasi-experimental selection of controls reveals the extent that failing to carefully develop an appropriate comparison group biased the estimation of the new bypass's impact on local growth. As now suggested by the impact coefficient in Table 11 for the 1990-2000 employment equation, opening of the new bypass is associated with 22 fewer jobs per square kilometer on average, which not only agrees with results from the Difference-in-Differences test in Table 7, but also appears to be consistent with at least one qualitative account in Livingston five years after the bypass opened to traffic (Mello, 2001).

The new 90 percent confidence intervals for parameter measuring employment growth impacts from the new highway are $-1 < c_{2,before} < 53$ and $-30 < c_{2,after} < -13$, which also do not overlap. Although, the wider 95 percent confidence bounds are non-overlapping, we report the 90 percent intervals for consistency with the analysis in Orange County. Here, the hypothesis test involves 73 degrees of freedom, while in Orange County the test involves 24 degrees of freedom, which explains the much narrower confidence bounds in the current test. These estimates suggest a fair level of confidence that the employment growth ascribed to the bypass lies in the interval -30 - 53 = -83 to -13 + 1 = -12. Had the bypass not been built, then we might have anticipated 12 to 83 more jobs per square kilometer in the cells near the bypass. These differences are statistically and economically significant, representing a growth rate that is 18 percent lower, relative to the 1990 average of 110 jobs in cells within a two-mile distance of ramps accessing the new freeway. Again, the Difference-in-Differences Estimator reported in Table 7 produces an estimated employment impact of -46 that falls neatly within this range.

Table 11 Quasi-Experimental Integrated Model Estimates, Merced County, California

	ΔPopulation					
	1980-1990			1990-2000		
Variable	Estimate		StdErr	Estimate		StdErr
Intercept	3321.9337	**	3411.7365	6446.1479		8979.2749
$(\mathbf{I}+\mathbf{W})EMP_{i,t-1}$	6.2890		1.9237	6.9644		5.0714
$(\mathbf{I}+\mathbf{W})EMP\Delta_{i,t}$	-10.1977	*	6.7953	-22.3255		15.7282
$POP_{i,t-1}$	0.0536		0.0240	0.1410	*	0.0674
$I'(\text{Treatment} < 2\text{ mi.})$	-3762.3422		5018.9976	10573.0971		11158.3268
n	78			78		
k	4			4		
F	7.9896			3.9775		
prob>F	0.0000			0.0056		
R-squared	0.3045			0.1789		
Adj R-squared	0.2664			0.1340		
Root MSE	19951.6501			46966.1285		
Over-Id ~ F	0.9754			0.8010		
Over-Id (prob>F)	0.4566			0.5893		

	ΔEmployment					
	1980-1990			1990-2000		
Variable	Estimate		StdErr	Estimate		StdErr
Intercept	-30.3693		16.6199	8.7042		4.7101
$(\mathbf{I}+\mathbf{W})POP_{i,t-1}$	0.0000		0.0000	0.0001	***	0.0000
$(\mathbf{I}+\mathbf{W})POP\Delta_{i,t}$	0.0004		0.0002	-0.0001	**	0.0000
$EMP_{i,t-1}$	0.1672	**	0.0625	0.1482	***	0.0151
$I'(\text{Treatment} < 2\text{ mi.})$	25.6052		16.1693	-21.5584	***	5.1741
n	78			78		
k	4			4		
F	10.4200			258.0681		
prob>F	0.0000			0.0000		
R-squared	0.3635			0.9340		
Adj R-squared	0.3286			0.9303		
Root MSE	69.4712			22.5870		
Over-Id ~ F	3.0827			25.1173		
Over-Id (prob>F)	0.0071			0.0001		

Santa Clara County

In the case of Santa Clara County, none of our models produced satisfying results. We experimented with several alternative specifications and in each instance, we found no indication that the major highway extensions of State Routes 85 and 87 had any significant influence on growth. Table 12 reports our best fitting specification for Santa Clara County from the conventional model.

Table 12 Conventional Lagged Adjustment Model Estimates, Santa Clara County, California

	ΔPopulation							ΔEmployment					
	1980-1990			1990-2000				1980-1990			1990-2000		
Variable	Estimate		StdErr	Estimate		StdErr	Variable	Estimate		StdErr	Estimate		StdErr
Intercept	280.1684		971.9881	153.7642		812.5574	Intercept	1659.9797		3193.6997	-1801.3532		1625.9996
AREA_i	0.0000		0.0000	0.0000		0.0000	AREA_i	0.0000		0.0000	0.0000		0.0000
$IncomePerCapita_{i,t-1}$	-0.1304	*	0.0611	0.0626	**	0.0218	$IncomePerCapita_{i,t-1}$	0.0468		0.2090	0.0870	*	0.0435
$\%bachelors_{i,t-1}$	1535.4107	**	580.3992	-5258.0513	***	1243.1523	$\%bachelors_{i,t-1}$	-6309.5322	*	2567.8710	-5564.3798	*	2540.6359
[illegible]	194.4880		1500.5663	-2526.1171		1421.6915	[illegible]	1211.8458		5490.9118	-3415.4853		2894.9591
$\%black_{i,t-1}$	8598.5163	*	3824.8907	-6975.0978		4332.9776	$\%black_{i,t-1}$	-11293.8327		12591.5404	6642.0307		8561.3500
$\%hispanic_{i,t-1}$	190.2093		1271.8829	1597.4632		1222.7415	$\%hispanic_{i,t-1}$	-8060.8263		4313.2602	-2311.0846		2552.9106
$popdensity_{i,t-1}$	-200261.7330	*	96461.8391	-51926.5730		79487.2040	$popdensity_{i,t-1}$	-289243.1701		319967.2601	6092.7078		161660.9007
$povertyrate_{i,t-1}$	2928.6447		2493.0396	2453.2665		2053.0567	$\%manufacturing_{i,t-1}$	-2360.9206		1728.4109	-37.8307		2078.0086
[illegible]	1833.5516		1484.0422	1543.4400		1300.9579	$\%agriculture_{i,t-1}$	4423.9812		11580.8225	-3184.4574		9713.1365
$(\mathbf{I}+\mathbf{W})EMP_{i,t-1}$	0.0044		0.0044	0.0057		0.0034	$\%wholesale_{i,t-1}$	17770.6607		9585.3053	5318.0730		6734.7158
$(\mathbf{I}+\mathbf{W})EMP\Delta_{i,t}$	0.0110		0.0166	0.0523		0.0366	$\%utilities_{i,t-1}$	-1194.7365		3403.2159	3069.6988		4247.5786
$POP_{i,t-1}$	0.2230	***	0.0396	0.1183	***	0.0293	$\%businessServices_{i,t-1}$	657.4069		6062.3447	3857.5544		4453.3503
$\mathrm{HIGHWAY}_i$	-35.5411		190.8027	48.7034		189.8641	$(\mathbf{I}+\mathbf{W})POP_{i,t-1}$	-0.0001		0.0230	0.0034		0.0126
$I'(\text{Treatment} \le 1\text{ mi.})$	-532.1635	*	259.6219	-166.6971		267.6630	$(\mathbf{I}+\mathbf{W})POP\Delta_{i,t}$	0.2356		0.1206	0.1346		0.1627
							$EMP_{i,t-1}$	0.0530		0.0893	0.0416		0.0285
							$\mathrm{HIGHWAY}_i$	307.1157		594.8670	958.5379	*	372.2873
							$I'(\text{Treatment} \le 1\text{ mi.})$	129.0685		830.6984	410.0892		541.7615
n	244			259			n	244			259		
k	43			43			k	46			46		
F	7.4313			4.7566			F	3.4422			2.5315		
prob>F	0.0000			0.0000			prob>F	0.0000			0.0000		
R-squared	0.6151			0.4456			R-squared	0.4875			0.3545		
Adj R-squared	0.5323			0.3162			Adj R-squared	0.3850			0.2145		
Root MSE	1050.1277			3343.5089			Root MSE	1091.5875			2184.2349		
Over-Id ~ F	0.3761			1.2828			Over-Id ~ F	3.1936			3.0242		
Over-Id (prob>F)	0.8936			0.2664			Over-Id (prob>F)	0.0247			0.0306		

Note: Although not included in the table, the models include dummy variables for census designated places, planning areas, and an indicator of whether the tract is located in an urban growth boundary.

Only the coefficient on the impact variable in the pre-intervention population equations tests significantly different from zero, and its 95 percent confidence interval intersects the confidence bounds for the post-intervention population impact coefficient. Furthermore, neither the Difference-in-Differences Estimator nor the quasi-experimental integrated approach showed signs of any significant effect on growth in Santa Clara County. In the end, we came to suspect that the result may be attributed to an incorrect bounding of the study region to a single urban county. While modeling effects across urban counties was beyond the scope of our study, Santa Clara County's setting within the larger Silicon Valley commuter shed is expected to be at least part of the culprit behind the seemingly small growth effects attributable to the large investments in highway transportation undertaken in the region.

SUMMARY AND CONCLUSIONS

Understanding linkages of new highway construction or capacity expansion to regional growth patterns is crucial for transportation planners and policy makers. Our evaluation models adapt the two-equation endogenous growth regression system developed in Boarnet (1992 and 1994) to incorporate a quasi-experimental selection of controls. In this study, we directly model the effects of major highway investments during the mid-1990s on nearby land uses and their appropriate no-build counterfactuals in three California counties. Controlling for the simultaneous spatial interaction between population and employment location further incorporates into the forecasts the dependence of growth on local commuter sheds, the distances people choose to travel from home to work and vice versa.

The central finding of the paper is that, while improvements in surface transportation tend to have large impacts on growth patterns, the nature of the effects is materially dependent on the context of the highway investment. Our models estimate that, on average, a statistically and economically significant 338 to 11,103 new Orange County jobs occurred within a typical census tract in the County's formerly exurban region after gaining highway access when compared to no-build counterfactuals. On the other hand, our models predict a starkly different outcome as a result of a highway bypass built outside the small town of Livingston in Merced County where we find an economically and statistically significant 12 to 83 job losses per square kilometer that might be anticipated had the bypass not been built. We find no significant effects on population or employment growth that can be attributed to the new highway investments near the urban center of Santa Clara County.

The policy implications from this analysis are potentially significant, particularly as it relates to the environmental review process. Our results suggest that context is important and that the impacts on population and employment growth from infrastructure improvements are not necessarily consistent from one geographic region to another, nor from one type of project to another. As seen in the Illinois case (Sierra Club v. United States DOT, 1997), documenting the potential impact is an essential component of the review process and better models are needed to forecast changes.

This study scratches the surface of understanding the critical linkages between highway improvements and patterns of regional growth and land use. Our research represents a promising, yet initial, step in applying quasi-experimental analysis to the question of highway infrastructure and urban growth patterns. As demonstrated in our analysis of four major highway investments in three diverse California counties, there is strong reason to suspect that the linkages between highway infrastructure and growth patterns are dependent on identifiable characteristics and contexts such as the type of highway improvement (new extensions, new connections, or expanded capacity, for example) and characteristics of the location (for example, a rapidly growing urban area, exurban region, or a more rural context). Although we examine highway investments that vary across urban, exurban, and rural contexts, these results should be combined with other analyses to build a base of knowledge that can relate land use impacts to highway projects. In that light, this study should be viewed as a beginning step in applying matched pair methodology to the question of the growth impacts of highways and we hope that the present analysis encourages the study of other cases examining the question of how growth impacts relate to a broader range of projects and land use settings than studied here. In addition, we feel that this approach could be used to examine a wide range of infrastructure developments or other significant land use changes and may be particularly interesting for analyses of rail projects in light of the Obama Administration's plan for a nationwide high-speed rail network.

Replication of these methods for the study of other regions experiencing shocks to surface transportation infrastructure could be facilitated by the construction of a national inventory of highway improvements and measures of land development impact. In addition to providing a resource for estimating development effects that arise from a variety of recent highway projects throughout the country, a national database would be helpful in addressing two features of intraregional forecast models encountered in the current study that can limit their broader applicability: (1) insufficient power in models of rural highway improvements due to the prevalence of a few, large census tracts in rural counties and (2) large standard errors in models of urban impacts due to invisible boundary constraints. A potential solution for labor markets that are not contained within invisible boundaries may be to expand the study region to include larger commuter sheds across urban counties. In the case of rural and small town interventions, a potential solution for the limited degrees of freedom may be to pool intervention data across regions. A national database would facilitate testing of the feasibility of these potential solutions in addition to enabling the examination of a much broader set of cases essential to expanding our knowledge base about the nature of impacts from highway infrastructure.

Notwithstanding the two limitations described above, our findings in Santa Clara, Merced, and Orange counties support the notion that the nature of regional growth and land use effects is materially dependent on the context of the highway investment. We find here that a new highway investment led to growth in regions gaining new access in an exurban setting within Orange County, while a new investment shifted growth away from a small town bypassed by a freeway in a rural setting in Merced County, California. In the latter case, we further conclude that forecast approaches lacking explicit control selection can lead to erroneous estimates of an impact. Given that context appears to play an important role in determining the nature of an impact, and choosing the right comparison group, or counterfactual, appears to matter for estimating an impact correctly, analysis of additional cases is needed to move the research forward.

The knowledge base regarding impacts in rural and small town environs is particularly small. There is a great need to extend a focus to small towns and rural regions and perhaps our finding of a negative employment effect attributed to new freeway access in the small town of Livingston will provide some impetus. Are negative employment effects prevalent in small towns? As the saying goes "the road runs both ways" and, indeed, analysis of highway transportation costs in developing regions suggests that enterprises most likely to avail from major highway investments are firms in urban conurbations enjoying scale economies and ready to tap new markets in the hinterland once new highways are constructed (Lall, Funderburg, and Yepes, 2004). If these costs are indeed common when freeway bypasses are constructed near small towns, then efforts should be taken to examine estimates from models similar to those described herein within the context of the benefits from bypasses such as increased safety in a rounded benefit-cost framework.

Post-Project Update

As part of ongoing research conducted by the authors after the conclusion of this project with the Mineta Transportation Institute, revisions were made to our original case study for Santa Clara County. New evidence was discovered to indicate that significant infrastructure improvements occurred along Highway 237 in the northern part of Santa Clara County, connecting I-880 in the eastern part of the county to Highway 101 in the west. These improvements took place during the same time frame as construction on Highways 85 and 87 and were, in fact, financed through the same mechanism, Measure A, passed in 1984. Our preliminary results are consistent with findings from our original study indicating no impact on population or employment growth as a result of the highway improvement in this well-developed county.

BIBLIOGRAPHY

Alonzo, William. Location and Land Use. Cambridge, MA: Harvard University Press, 1964.

Bailey, Trevor C. and Anthony C. Gatrell. Interactive Spatial Data Analysis. New York, NY: J. Wiley, 1995.

Baum-Snow, Nathaniel. "Did Highways Cause Suburbanization?" Quarterly Journal of Economics 122 (2007): 775-805.

Boarnet, Marlon G. "Intra-Metropolitan Growth Patterns: The Nature and Causes of Population and Employment Changes within an Urban Area." Unpublished Ph.D. diss., Princeton University, Princeton, NJ., 1996.

Boarnet, Marlon G. "An Empirical Model of Intrametropolitan Population and Employment Growth." Papers in Regional Science 73 (2001): 135-152.

Boarnet, Marlon G. "Highways and Intrametropolitan Employment Growth: A Spatial Econometric Approach." Unpublished Manuscript. Department of Planning Policy and Design, Irvine, CA: University of California, Irvine, 1996.

Boarnet, Marlon G. "Enterprise Zones and Job creation: Linking Evaluation and Practice." Economic Development Quarterly 15 (2001): 242-254.

Boarnet, Marlon and Saksith Chalermpong. "New Highways, House Prices, and Urban Development: A Case Study of Toll Roads in Orange County, California." Housing Policy Debate 12 (2001): 575-605.

Boarnet, Marlon and Saksith Chalermpong. "Urban Growth and Highway Infrastructure: A 'Before' and 'After' Test of the Influence of Highways on Metropolitan Growth Patterns." Paper presented at the North American Regional Science Association, San Juan, Puerto Rico, 2002.

Boarnet, Marlon G., Saksith Chalermpong, and Elizabeth Geho. "Specification Issues in Models of Population and Employment Growth." Papers in Regional Science 84 (2005): 21-46.

Boarnet, Marlon G., Joseph F. DiMento, and Gregg Macey. (2002). "Toll-Highway Finance Lessons from Orange County." Berkeley, CA: University of California Policy Research Center. http://www.ucop.edu/cprc/documents/tollroadrpt.pdf (accessed June 17, 2008).

Bollinger, Christopher R. and Keith R. Ihlanfeldt. "The Impact of Rapid Rail Transit on Economic Development: The Case of Atlanta's MARTA." Journal of Urban Economics 42 (1997): 179-204.

Card, David and Alan B. Krueger. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." American Economic Review 84 (1994): 772-793.

California Department of Transportation. California 99 Corridor Enhancement Master Plan. Sacramento, 2004.

Cervero, Robert. “Road Expansion, Urban Growth, and Induced Travel: A Path Analysis.” Journal of the American Planning Association 69 (2003): 145-163.

Chalermpong, Saksith. “Empirical Study of Economic Spillovers of Interstate 105 in Los Angeles.” Transportation Research Record, No. 1864 (2004): 94–102.

Clark, David E. and Christopher A. Murphy. “Countywide Employment and Population Growth: An Analysis of the 1980s.” Journal of Regional Science 36 (1996): 235-256.

Cook, T.D. and D.T. Campbell. Quasi-Experimentation: Design and Analysis Issues for Field Settings. Boston: Houghton-Mifflin, 1979.

Dehejia, Rajeev H. and Sadek Wahba. “Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs” Journal of the American Statistical Association 94 (1999): 1053-1062.

Dehejia, Rajeev H. and Sadek Wahba. “Propensity Score Matching Methods for Nonexperimental Causal Studies.” Review of Economics and Statistics 84 (2002): 51-161.

Edmiston, Kelly D. “The Net Effects of Large Plant Locations and Expansions on County Employment.” Journal of Regional Science 44 (2004): 289-319.

Fimrite, Peter. “Infamous Blood Alley to Disappear: New Bypass around Highway 99 Stoplight.” San Francisco Chronicle, November 4, 1996.

Greenwood, Michael J. “Human Migration: Theory, Models, and Empirical Studies.” Journal of Regional Science 25 (1985): 521-544.

Handy, Susan and Deb Niemeier. “Measuring Accessibility: An Exploration of Issues and Alternatives.” Environment and Planning A, Vol. 29 (1997): 1175-1194.

Hasan, N. “Measure A Replay.” Metroactive (April 12-18, 2006).

Haynes, Kingsley E. and Stewart Fotheringham. Gravity and Spatial Interaction Models. Beverly Hills, CA: Sage Publications.

Henry, Mark S., David L. Barkley, and Shuming M. Bao. “The Hinterland’s Stake in Metropolitan Growth: Evidence from Selected Southern Regions.” Journal of Regional Science 37 (1997): 479-501.

Henry, Mark S., Bertrand Schmitt, Knud Kristensen, David L. Barkley, and Shuming Bao. “Extending Carlino-Mills Models to Examine Urban Size and Growth Impacts on Proximate Rural Areas.” Growth and Change 30 (1999): 526-548.

Holl, Adelheid. “Transport Infrastructure, Agglomeration Economies, and Firm Birth: Empirical Evidence from Portugal.” Journal of Regional Science 44 (2004): 693-712.

Holzer, Harry J., John M. Quigley, and Steven Raphael. "Public Transit and the Spatial Distribution of Minority Employment: Evidence from a Natural Experiment." Journal of Policy Analysis and Management 22 (2003): 415-441.

Houston, Karyn. "Final 99 Stoplight Set to Go: Livingston Bypass Gets State Approval." Modesto Bee, September 21, 1990.

Lall, Somik V., Richard Funderburg, and Tito Yepes, "Transport Costs, Fiscal Incentives and Industrial Performance in Lagging Regions: Evidence from Brazil" (working paper, World Bank Policy Research, 2004).

Levernier, William and Brian Cushing. "A New Look at the Determinants of the Intrametropolitan Distribution of Population and Employment." Urban Studies 31 (1994): 1391-1405.

Mark, Stephen T., Therese J. McGuire, and Leslie E. Papke. "The Influence of Taxes on Employment and Population Growth: Evidence from the Washington, D.C. Metropolitan Area." National Tax Journal 53 (2000): 105-124.

Mello, Michael. "Stoplight Finds Home at Museum." Modesto Bee, October 25, 2001.

Mills, Edwin S. "An Aggregative Model of Resource Allocation in a Metropolitan Area." American Economic Review 57 (1967): 197-210.

Minnesota Population Center. National Historic Geographic Information System: Pre-release Version 0.1. Minneapolis, MN: University of Minnesota. (2004).

Muth, Richard F. Cities and Housing. Chicago, IL: University of Chicago Press, 1969.

Muth, Richard F. "Migration: Chicken or Egg?" Southern Economic Journal 37 (1971): 295-306.

O'Keefe, Suzanne. "Job Creation in California's Enterprise Zones: A Comparison Using a Propensity Score Matching Model." Journal of Urban Economics 55 (2004): 131-150.

Reese, J. "Highway through the Heartland." Via (March, 2004).

Richards, Gary. "Highway 85: Ready, Set, Go." San Jose Mercury News, October 19, 1994.

Rosenbaum, Paul R. and Donald B. Rubin. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." Biometrika 70 (1983): 41-55.

Schmitt, Bertrand and Mark S. Henry. "Size and Growth of Urban Centers in French Labor Market Areas: Consequences for Rural Population and Employment." Regional Science and Urban Economics 30 (2000): 1-21.

Sierra Club, Illinois Chapter, South Corridor Against the Tollway, Inc., Environmental Law and Policy Center of the Midwest, and Business and Professional People for the Public Interest v. U.S. Department of Transportation, Federal Highway Administration, Illinois Department of Transportation; and Illinois State Toll Highway Authority. U.S. District

Court for the Northern District of Illinois, Eastern Division, No. 96 C 768, Decided January 16, 1997.

Smith, Jeffrey A. and Petra E. Todd. "Does Matching Overcome LaLonde's Critique of Nonexperimental Estimators?" Journal of Econometrics 125 (2005): 305-353.

Steinnes, Donald N. "Causality and Migration: A Statistical Resolution of the Chicken or Egg Fowl-Up." Southern Economic Journal 45 (1978): 218-226.

Steinnes, Donald N. and Walter D. Fisher. "An Econometric Model of Intraurban Location." Journal of Regional Science 14 (1974): 65-80.

Thome, Joe. "Livingston Supports Plan to Do Away with Highway 99 Stoplight." Modesto Bee, October 23, 1990.

Wyman, B. "Highway Construction Puts Brakes on 800 Almaden Valley Homes." San Jose Business Journal, July 9, 1990.

ABOUT THE AUTHORS

RICHARD G. FUNDERBURG, Ph.D.

Dr. Funderburg is assistant professor of urban and regional planning at the University of Iowa. His research investigates an array of factors that attract economic activities to both developing and advantaged regions; it includes the identification and analysis of industry clusters and the estimation of impacts from public investments in transportation infrastructure, housing, and financial incentives on economic growth and patterns of development. He earned a Ph.D. in Planning, Policy, and Design from University of California, Irvine and he holds M.P.P.A. and B.A. degrees from California State University.

HILARY NIXON, Ph.D.

Dr. Nixon is an Assistant Professor of Urban and Regional Planning at San José State University. Her research and teaching interests in the field of environmental planning and policy focus on the relationship between environmental attitudes and behavior, particularly related to waste management, and linkages between transportation and the environment. She has a B.A. from the University of Rochester in Environmental Management and a Ph.D. in Planning, Policy, and Design from the University of California at Irvine.

MARLON G. BOARNET, Ph.D.

Marlon G. Boarnet is Professor of Planning, Policy, and Design and Economics at the University of California, Irvine, where he has been on faculty since 1991. From 2003 through 2006, Boarnet was Chair of the Department of Planning, Policy, and Design at UC Irvine. Boarnet earned a Ph.D. from Princeton University in 1992. Boarnet's research focuses on the inter- and intra-metropolitan distribution of economic activity, the links between transportation policy and economic development, local economic development policy, and land use . travel behavior interactions.

Boarnet's publications on these and other topics include over 40 refereed articles, book chapters, and reviewed reports. Boarnet is a co-author of Travel by Design: The Influence of Urban Form on Travel, a comprehensive study of the potential for linking land use and urban design to transportation policy published by Oxford University Press in 2001. Since 2002, Boarnet has served as managing Co-Editor of the Journal of Regional Science, a leading scholarly journal at the intersection of urban economics and quantitative geography. Boarnet is also an associate editor of the Journal of the American Planning Association, and he serves on the editorial boards of the Journal of Planning Literature and Papers in Regional Science. Boarnet has received grants from or conducted contract research for the U.S. Environmental Protection Agency, the U.S. and California Departments of Transportation, the Robert Wood Johnson Foundation, the World Bank, the Bay Area Economic Forum, and the Orange County Business Council, among others.

PEER REVIEW

San José State University, of the California State University system, and the MTI Board of Trustees have agreed upon a professionally acceptable research protocol.

Research projects begin with the approval of a scope of work by the sponsoring entities, with in-process reviews by the MTI research director and the project sponsor. Periodic progress reports are provided to the MTI research director and the Research Associates Policy Oversight Committee (RAPOC). Review of the draft research product is conducted by the Research Committee of the board of trustees and may include invited critiques from other professionals in the subject field. The review is based on the professional propriety of the research methodology.

SAN JOSÉ STATE
UNIVERSITY

Funded by U.S. Department of
Transportation and California
Department of Transportation