THE EFFECTS OF LOCATION ELEMENTS ON HOME PURCHASE PRICES AND RENTS: Evidence from the San Francisco Bay Area

by

Kara Maria Kockelman

The following paper is a pre-print and the final publication can be found in
Transportation Research Record No. 1606: 40-50.
Presented at the 76th Annual Meeting of the Transportation Research Board, January 1997

Abstract

THE EFFECTS OF LOCATION ELEMENTS ON HOME PURCHASE PRICES AND RENTS:
Evidence from the San Francisco Bay Area

Author: Kara M. Kockelman

Land-market theory emphasizes travel savings as well as access to amenities as the underlying determinants of land prices (*e.g.*, von Thünen 1826 and Alonso 1964). This research investigates a series of hedonic models (which postulate a good's price to be a linear function of its multiple attributes) for implicit estimation of land values, along with explicit estimation of median housing price and monthly contract rent across the San Francisco Bay Area's census tracts. The benefits and valuation of location are assessed by including a variety of travel-based explanatory variables (including average trip characteristics and automobile ownership) as well as measures of local land-use patterns - after controlling for a variety of the dwelling units' structural characteristics. In many of the models, lot size is interacted with location attributes in order to elucidate the direct dependence, if any, of land values on location.

The results indicate that changes in accessibility and travel costs affect land and dwelling-unit values in highly statistically and economically significant ways. For example, the coefficient estimates associated with travel-time reductions suggest values of travel-time savings to be roughly five dollars per hour across all adult-traveller trip types (in 1989$). Furthermore, a high proximity to industrial land uses and very local commercial activity depress land (and dwelling unit) values, often substantially; and simple distance-to-CBD variables prove to be effective price predictors. The results also reveal that heteroskedasticity (a non-constant variance of the dependent variables' mean values) plays an important role in consistent and efficient estimation of model parameters and should thus be accommodated explicitly.

INTRODUCTION:

Regional scientists, land economists, and transportation planners have long asserted that a household's location decision depends, to a large degree, on access to opportunity sites. Subject to budget, time, and other constraints, it is a common assumption that households maximize their "utility" by locating in as desirable a home as possible - as near to necessary and desired activities (such as work, shopping, and recreation) as possible. The utility maximization is profoundly complex, dependent on far more environmental attributes than those that can be observed and quantified; for example, people often care about subtle neighborhood qualities (1), public services (2, 3), proximity to relatives, and other "goods" a location and its environment offer. Rather than attempt to elucidate all determinants of location choice and home valuation, this paper limits its focus to the variation of travel costs, accessibility, and land use patterns across an urban region and how these are reflected in housing prices.

If access is a critical determinant of land prices by allowing people to reduce travel expenses, then one may reasonably expect that reductions in travel expenses will be reflected in higher bids for homes. For example, John Holtzclaw (4) has argued that reduced automobile use (as evidenced through odometer tabulations) for more population-dense environments should translate to more income available for home purchases (and loan guarantees). Holtzclaw estimates that an "efficient location" like Nob Hill (in San Francisco) allows a household to save $6,000 a year in transportation expenses relative to a similar household in a city such as San Ramon. The American Automobile Association (5) estimates that the daily *fixed* cost of owning the average automobile is roughly ten dollars, representing over $3,500 a year - or a present value of over $40,000 on a 30-year eight-percent loan. However, in a study of so-called "wasteful" commuting, Giuliano and Small (6) conclude that factors other than location vis-a-vis employment sites affect household location choice to a much greater extent. And Wheaton (7) argues that the evidence for location choice with respect to transportation amenities is not a major factor in location decisions. The research pursued here aims to explore and test some of these conclusions.

It is well accepted that people value time and generally prefer to minimize delays in accessing opportunities, *ceteris paribus*. A great number of studies have been undertaken to estimate travellers' value of time (*e.g.*, 8, 9, 10); these rely almost exclusively on discrete-choice random-utility models where the ratio of coefficients estimated on time and fare variables is expected to provide a value of time. Applying this methodology, a varied assortment of value-of-time estimates have emerged, with values ranging from six percent (11) to 86% of wage (12) for inter-city car travelers and from 20% (13) to 180% (14) for auto commute trips. The variation in these estimates is one indication that the relationship is a highly complex one, not yet well modeled. However, the dependence of time valuation on income is widely accepted, given that one's opportunity cost in many cases may be based on income-producing pursuits.

This research's methodology takes a very different tack in approaching travel-time valuation, relying instead on home prices as a function travel costs, with time representing an important dimension of travel expenses in many of the early models. The general hypothesis tested here - that home prices (and rental rates) fall with expected travel dis-benefits - is premised on the underlying theories of location-choice models developed by economists such as Alonso (15) and Anas (16): *i.e.*, that there is a trade-off between land value and travel expenses. The basic model developed, the data used, and the analytic methods pursued - along with results and conclusions - follow here.

THE MODEL:

Despite the land market's complexities and model limitations, many researchers have attempted to estimate residential land values, often with respect to transportation provision (*e.g.*,17, 18, 19, 20). One of the functionally simplest methods of evaluating transportation's impact on land valuation is the hedonic price model, in which sales price is modeled as a function of a good's multiple attributes. (See Griliches [21] for a discussion of this method.) This is the method used here, but one should note that the coefficient estimates on individual attributes cannot be directly interpreted as the marginal values of these attributes except under highly restrictive assumptions (*e.g.*, all households are identical in preferences and in income). Under substantially less restrictive assumptions, McMillan *et al.* (22) demonstrate that, for households which do not vary too greatly in their tastes, the biases implied using these coefficients as estimates of marginal value may be very minor. In any case, our interest here lies more in the general magnitude, rather than the exact value, of transport costs and locational benefits as they are incorporated into housing prices.

DATA ANALYZED AND METHODS USED:

The Data:

In order to study the trade-off between housing costs and locational benefits, the data used here are the following: the 1990 Census of Population and Housing data; 1990 Bay Area Travel Survey (BATS) data (23); and 1990 Association of Bay Area Governments (ABAG) land-use data (24). The two *dependent* variables examined are median respondent-estimated price of an owner-occupied dwelling unit (OODU) and the median monthly contract rent of renter-occupied dwelling units (RODU), both provided in 1989 dollars in the 1990 Census.

In terms of explanatory variables, the Census data provide some relatively basic and aggregate information on dwelling-unit attributes across the region’s roughly 1,300 census tracts. Data such as number of bedrooms and rooms and median structure are available, along with information on vehicle ownership, commuting times, and number of renter- and owner-occupied dwelling units per tract. Coupled with the Association of Bay Area Governments' data on residentially developed land area by tract, dwelling-unit numbers allow one to estimate average lot size. Multiple computations are used to estimate the square footage of owner-occupied versus renter-occupied dwelling units.[1]

Other explanatory variables tested in the models include BATS-based travel "costs", Census Transportation Planning Package (CTPP)-based accessibility measures, and ABAG-based land-use measures. The BATS over-60,000 individual trip records were aggregated by tripmakers’ home tracts to provide estimates of average travel time per trip, average vehicle miles traveled (VMT) per trip[2], and other travel-related variables. By counting the number of adults (19 years and over) and vehicles across surveyed households in each tract, one is able to estimate the average number of vehicles per adult in each tract. Note that the normalization of travel characteristics with respect to adults and their trips is hoped to provide maximum comparability across zones. Per household normalizations, for example, would create difficulties since household sizes and their composition change across tracts.

The CTPP data include jobs (by type) across the region's 1099 traffic analysis zones (TAZs), and inter-zonal travel times (by car) were obtained by running the program MIN-UTP. Together, these data were used to construct a variety of accessibility-to-jobs indices, based on the popular gravity model, as shown in Equation 1. The functional form assumes that accessibility exhibits a direct proportionality to the number of opportunities and an inversive relation relative to the cost of accessing those opportunities (with "cost" proxied by travel time). The specific indices were constructed in a number of ways, based on exponential time functions incorporating coefficients estimated by Levinson and Kumar (25). The form depended on trip type (work versus non-work, for example) and mode used (SOV versus walk trips, for example). Total jobs as well as sales and service jobs per traffic analysis zone (TAZ) were used as measures of

$$Accessibility_i = \sum_j \frac{A_j}{f(t_{ij})},$$

$$where\ A_j = Attractiveness\ of\ Zone\ j\ and\ t_{ij} = Travel\ Time\ from\ Zone\ i\ to\ j.$$

attractiveness.

In addition to jobs-access measures, distance-to-CBD variables, in actual network miles to the region's three central business districts (in San Francisco, Oakland, and San Jose) were computed using data provided by the region's MPO, the Metropolitan Transportation Commission (MTC).

Finally, the land-use explanatory variables come from ABAG's hectare-level data set, which classifies each of the region's hectares (about the size of two football fields) as having a "dominant" use. The fractions of a tract's developed land designated as "industrial", "park/outdoor recreational", and "public/community" were computed from the hectare-level data, and relatively sophisticated measures of land-use *balance* and *mix* (or dispersion) were computed using Arc/Info on a geographical information system. All variables used or considered in the models are described briefly in Table 1.

Observe that land-use balance is quantified using the formula for entropy (as shown in Equation 2), which was originally defined for the energy state of a system in the Second Law of

Thermodynamics and proven by Ludwig Boltzmann in the 1870s. It is normalized with respect to the natural log of the number of distinct uses considered and thus varies between zero and one (with one signifying "perfect" balance of the uses considered). The six (J=6) land-use types considered distinct and used in the computation of this index are the following: residential, commercial, public, offices and research sites, industrial, and parks and recreation. Furthermore, to avoid bias against smaller tracts, in which there is relatively little area to allow for a variety of land-use types and to more adequately represent the concept of "neighborhood," a "mean

$$\textit{Mean Entropy} = \sum_k \frac{\sum_j \frac{(P_{jk} x \ln(P_{jk}))}{\ln(J)}}{K},$$

$$\textit{where } K = \textit{Number of Actively Developed Hectares in Tract}$$
$$\textit{and } P_{jk} = \textit{Proportion of Use Type } j \textit{ within a Half - Mile Radius}$$
$$\textit{of Developed Area Surrounding the kth Hectare.}$$

entropy" was constructed (in contrast to a tract-bounded entropy measure).

As shown in Equation 2, the mean entropy is the average of neighborhood entropies computed for all developed hectares within each census tract, where "neighborhood" is defined to include all developed area within one-half mile of each, relevant, active hectare.

While entropy helps quantify the degree of balance across distinct land uses, the degree to which these land uses come into contact with one another is also expected to be of importance since distances can be further minimized between distinct use types if they are more dispersed within a given area.

The land-use mix index may also be called a "dissimilarity index" since it is based on "points" awarded to each actively developed hectare based on the dissimilarity of its land use from those of the eight adjacent hectares. (Equation 3) The average of these point accumulations across all active hectares in a tract is the dissimilarity or mix index for the tract. (Please see Kockelman [26] for further description and use of the land-use balance and mix variables in

$$\begin{gathered}\textit{Dissimilarity Index} = \textit{Mix Index} = \sum_k \frac{1}{K} \sum_i^8 \frac{X_{ik}}{8},\\ \textit{where } K = \textit{Number of Actively Developed Hectares in Tract and}\\ \textit{and } X_{ik} = 1 \textit{ if Central Active H'ectares Use Type differs from}\\ \textit{that of a Neighboring Hectare } (X_{ik} = 0 \textit{ otherwise}).\end{gathered}$$

models of traveller behavior.)

ANALYSIS AND RESULTS:

Several stages of analysis were necessary for full model development. Early models relied on ordinary least squares (OLS) for estimation and tested a variety of explanatory variables. The use of the median price and rent statistics by tract introduces heteroskedasticity, wherein the variance of observed prices (or rents) is not constant across different values of the dependent variable. Given tract-median price data, observations' variances are approximately inversely related to the number of observations per tract. (27)

Moreover, many housing attributes can be reasonably expected to give rise to additional heteroskedasticity; for example, depending on how well a home is maintained and/or how flat its parcel is, housing price can be expected to fluctuate more across older homes and those on larger lots. For these reasons, heteroskedasticity was investigated. The null hypotheses of the ensuing tests for homoskedasticity (*i.e.*, constant variance) were soundly rejected, so the method of feasible generalized least squares (FGLS) replaced simple ordinary least squares and was used to obtain what are, under common assumptions, asymptotically efficient estimates with unbiased standard errors. The detailed results of these stages of statistical analysis are more fully described below.

Fundamental Models:

The first and most fundamental models tested rely on the hypothesis that the value of land is an explicit linear function of "accessibility", as measured using variables characterizing

TABLE 1. DESCRIPTION OF VARIABLES USED

Median Price OODU	OODU Median Price for Census Tract (1989$)
Median Rent RODU	RODU Median Rent (1990$)
#Bedrooms	#Bedrooms/DU (OO or RO)
#Rooms	#Rooms/DU (OO or RO)
Log(Age)	Natural Logarithm of Median Age (for OODU or, in the case of RODUs, all DUs)
Lot Size	Estimate of Square Footage per DU parcel (OO or RO)
SF xTime/Trip	Lot Size (sf) x Average Time per BATS trip (100ths of a minute)
SF xVehicles/Adults	Lot Size (sf) x Average Number of Vehicles per Adult (19+ years) across BATS Households
SF xAvg. Time to Work	Lot Size (sf) x Average Commute Trip Time (minutes)
SF xWork Accessibility (30 min.)	Lot Size (sf) x Accessibility to All Jobs within 30 minutes (uncongested travel time by automobile)
SF xSales&Service Walk Accessibility	Lot Size (sf) x Accessibility to Sales and Service Jobs by Walking
SF xIndustrial Fraction	Lot Size (sf) x Fraction of Tract's Developed Area in Industrial Use
SF xPublic Fraction	Lot Size (sf) x Fraction of Tract's Developed Area in Public & Community Uses
SF xParkspace Fraction	Lot Size (sf) x Fraction of Tract's Developed Area in Park & Outdoor Recreational Space
Land Use Balance	Normalized Entropy across six distinct use types, averaged across 1/2-mile radius neighborhoods
Mix of Land Uses	Dissimilarity Index quantifying the number of dissimilar neighboring hectare-defined land uses
Distance SF CBD	Network (*i.e.*, non-Euclidean) Miles to San Francisco's City Hall
Distance SJ CBD	Network Miles to San Jose's City Hall
Distance Oakland CBD	Network Miles to Oakland's City Hall
Sales & Service Walk Accessibility	Accessibility to Sales and Service Jobs using exponential decay function for non-work trips with walk times
Work Accessibility (30 min.)	Accessibility to All Jobs within 30 minutes using exponential decay function for work trips
Time/Trip (100ths of hour)	Mean Time per Trip Recorded by Tract Dwellers (from BATS data set, in 100ths of an hour)
Avg. Time to Work	Mean number of Minutes spent commuting to work
Vehicles/Adults	Mean number (across BATS households) of vehicles in household per household member age 19 years or more
VMT/Trip	Mean number of Euclidean miles per Trip made by BATS surveyed household members living in that Census tract
Industrial Fraction	Fraction of Developed Area in Tract where Industrial Uses Dominate (from ABAG data set)
Public Fraction	Fraction of Developed Area in Tract where Public & Community Uses Dominate (from ABAG data set)

Parkspace Fraction	Fraction of Developed Area in Tract where Park & Outdoor Recreational Uses Dominate (from ABAG data set)
Vacant Fraction	Fraction of Tract's Dwelling Units that are Vacant
1/(#OODU)	Inverse of Number of OODUs in census tract
1/(#RODU)	Inverse of Number of RODUs in census tract
Other Variables Studied but found to be Insignificant and/or of "Incorrect" Sign:	
SF xPay-Parking Fraction	Lot Size (sf) xFraction of Trips taken by Tract Residents where one had to Pay for Parking (from BATS)
SF xTransit Trips	Lot Size (sf) x#Transit Trips per Person

Note: OODU and RODU stand for Owner-Occupied and Renter-Occupied Dwelling Unit.

average trip costs and other locational qualities. Much previous research incorporates only purely additive models, with no interaction between parcel size and locational characteristics (*e.g.*, 17, 18, 19, 20, 28). An improvement in these research efforts may arise through the

$$Price = C_1 + \mathrm{B} * Structural\ Characteristics + Parcel\ Size * Value\ of\ Land + \varepsilon_1,$$
$$where\ Value\ of\ Land = C_2 + \Gamma * Locational\ Characteristics + \varepsilon_2$$

following set-up:

The constant C_2 is expected to represent the value of minimally accessible land in the Bay Area, something akin to the land's agricultural value. Note the interaction between parcel size and all locational characteristics. As long as *Parcel Size* and ε_2 are not correlated, the introduction of a *Parcel Size* $* \varepsilon_2$ term in the first equation should not create a situation where OLS estimates are biased or inconsistent.

Investigation of Heteroskedasticity and Estimation Using Feasible Generalized Least Squares (FGLS):

The presence of heteroskedasticity can be ascertained from OLS output, assuming that the model's residuals are consistent estimates of the underlying error terms. By regressing the square of these residuals on a set of variables expected to affect error-term variance (such as home age and lot size), one can test for the significance of such a regression (or the presence of *homo*skedasticity). The regression of the owner-occupied (OO) and renter-occupied (RO) models' squared residuals on a variety of explanatory variables, including the inverse of the initial weighting strategy (*i.e.*, the inverse of the number of OO and RO dwelling units [DUs] per tract, respectively), produced (adjusted) R^2's of 0.266 and 0.213, respectively - along with p-values for model insignificance of 0.0000, allowing one to reject, in both cases, null hypotheses of homoskedasticity.[3]

Fortunately, feasible generalized least squares is as asymptotically efficient in estimation as is the method of maximum likelihood. (29) Moreover, FGLS does not require any assumptions as to the error terms' distribution (*e.g.*, normal versus something else). This method uses the inverse of the fitted values (or estimates) of the variance model to weight the least-squares regression of the primary model. The results of the sample size-weighted OLS models and their FGLS counterparts, both relying on interactions with parcel size, are shown in Tables 2 and 3.

In reviewing the model results, note the counteracting effects of bedrooms and rooms, which appear in all hedonic home-price models that the author has tested here and previously; rather clearly, people tend to prefer non-bedrooms to bedrooms, but both are valued overall (since, in adding a bedroom, a "room" is also added, and the coefficient on the variable #Rooms more than counteracts the negative coefficient of the bedroom variable).

One of the more striking results of the models is the change in estimates that results from taking into account the significant heteroskedasticity that is neglected in the sample-size-weighted OLS results. Differences of 50 percent in estimates' levels are not uncommon (*e.g.*, #Bedrooms and Mix in the OODU models, and interactions with parcel size in the RODU models).

TABLE 2. OWNER-OCCUPIED DWELLING UNIT (OODU) PRICE REGRESSIONS

Primary Model [$y=X\beta$]:

Dependent Variable: Median Respondent-Estimated Home Price (1989$)

Explanatory Variables:	*OLS Coefficient*	*(SE)*	*p-value*	*FGLS Coefficient*	*(SE)*	*p-value*
Constant	-1.204e+5	(3.331e+4)	.0003	-3.903e+4	(3.061e+4)	.2027
#Bedrooms	-7.710e+4	(2.355e+4)	.0011	-4.752e+4	(1.176e+4)	.0001
#Rooms	+9.858e+4	(1.273e+4)	.0000	+7.642e+4	(6.375e+3)	.0000
Log(OO Age)	+6.784e+3	(5.771e+3)	.2401	-1.852e+4	(5.055e+3)	.0003
Lot Size (sf)	+5.541	(1.681)	.0010	+7.595	(0.9930)	.0000
SF xAvg.Time to Work (sf*min)	-0.2792	(0.0721)	.0000	-0.3428	(0.0415)	.0000
SF xWork Accessibility (30 min.)	+5.153e-5	(5.194e-6)	.0000	+6.268e-5	(3.972e-6)	.0000
SF xSales&Service Walk Access.	-3.285e-3	(1.292e-5)	.0112	-3.231e-3	(9.529e-4)	.0007
SF xIndustrial Fraction	-6.3482	(1.665)	.0001	-6.040	(1.266)	.0000
SF xPublic Fraction	+8.597	(2.480)	.0006	+8.943	(3.904)	.0222
SF xParkspace Fraction	-2.763	(2.493)	.2681	-1.132	(1.671)	.4982
Mix of Land Uses	-2.076e+5	(4.426e+4)	.0000	-1.062e+5	(3.635e+4)	.0036
N=864[4]	**R^2= 0.506**			**R^2=0.563**		
	Weights=#OO DUs			Weights=Inverse of Predicted Values of Variance Model		

Variance Model [$\sigma_i^2=z_i\beta$]:

Dependent Variable: Squared Residuals of Primary OLS Model

Explanatory Variables:	OLS Coefficient	(SE)	p-value	FGLS Coefficient	(SE)	p-value
Constant	+5.363e+9			+4.213e+9	(6.288e+9)	.5030
#Bedrooms	n/a			+8.865e+9	(1.305e+9)	.0000
#Rooms	n/a			-3.109e+9	(7.452e+8)	.0000
Log (OO Age)	n/a			-3.386e+9	(1.601e+9)	.0347
Age (Median, of OODUs)	n/a			+1.773e+8	(7.736e+7)	.0221
Lot Size (sf)	n/a			+0.2373	(0.4850)	.6247
Vacant Fraction	n/a			+5.295e+10	(9.454e+9)	.0000
1/(#OODU)	-			+5.412e+10	(1.472e+9)	.0003
Sales & Service Walk Access.	n/a			+8.153e+6	(1.001e+6)	.0000
Work Accessibility (30 min.)	n/a			-1.098e+4	(8261)	.1843
Distance SF CBD (mi.)	n/a			-2.363e+8	(3.975e+7)	.0000
Dist. Oakland CBD (mi.)	n/a			+1.334e+8	(4.196e+7)	.0015
Time/Trip (100ths of hour)	n/a			+3.106e+7	(3.711e+7)	.4030
Avg. Time to Work (min.)	n/a			-9.397e+8	(8.524e+8)	.3305
Vehicles/Adults	n/a			+7.669e+8	(8.524e+8)	.3685
VMT/Trip (euclidean miles)	n/a			-1.671e+8	(1.668e+8)	.3168
Industrial Fraction	n/a			-2.899e+9	(2.350e+9)	.2176
Fraction of DUs Boarded Up	n/a			-1.300e+11	(6.403e+10)	.0427
Mix of Land Uses	n/a			-7.571e+9	(4.520e+9)	.0944
				R^2=0.266 (variance model)		

Notes: OODUs differed from RODUs for the following explanatory variables: #Bedrooms, #Rooms, Square Footage estimates (SF), and Structure Age (where OO median age and all-structures median age were used, respectively).

The OLS model's standard errors have been estimated using White's robust estimator, under heteroskedasticity. [5]

All R^2's are adjusted for degrees of freedom.

Vehicle ownership (vehicles/adult) and average trip time (for all trip types) were rejected as explanatory variables due to a complete lack of statistical significance in essentially all OODU models where commute times were already included.

TABLE 3. RENTER-OCCUPIED DWELLING UNIT RENT REGRESSIONS

Primary Model [y=Xβ]: *OLS Coefficient (SE) & p-value* *FGLS Coefficient (SE) & p-value*

Dependent Variable: Median Contract Rent/Mo. (1989$)

Explanatory Variables:

Variable	OLS Coefficient	(SE)	p-value	FGLS Coefficient	(SE)	p-value
Constant	+390.9	(57.96)	.0000	430.5	(49.29)	.0000
#Bedrooms	-217.2	(37.65)	.0000	-85.11	(28.33)	.0027
#Rooms	+232.8	(23.68)	.0000	160.8	(18.42)	.0000
Log(Age)	-62.89	(12.06)	.0000	-68.93	(9.461)	.0000
Lot Size (sf)	-0.0044	(6.017e-3)	.4647	-7.466e-4	(3.941e-3)	.8498
SF xTime/Trip (sf*100ths of hr.)	-9.346e-5	(7.722e-5)	.2265	+4.858e-5	(6.409e-5)	.4487
SF xAvg.Time to Work (sf*min)	-2.775e-4	(2.024e-4)	.1709	-5.106e-4	(1.396e-4)	.0003
SF xWork Accessibility (30 min.)	+1.370e-7	(2.693e-8)	.0000	+1.782e-7	(1.323e-8)	.0000
SF xSales&Service Walk Access.	-6.333e-6	(4.198e-6)	.1318	-9.827e-6	(2.956e-6)	.0009
SF xVehicles/Adult	+5.118e-3	(2.675e-3)	.0561	+1.498e-3	(1.206e-3)	.2145
SF xIndustrial Fraction	-0.02441	(5.869e-3)	.0000	-1.968e-2	(5.658e-3)	.0005
SF xPublic Fraction	+0.01731	(6.738e-3)	.0104	+1.609e-2	(1.318e-2)	.2223
Land Use Balance	-76.64	(72.79)	.2928	-80.92	(41.23)	.0500
Mix of Land Uses	-1.995e+2	(45.19)	.0000	-105.5	(66.68)	.1140
N=787	**R^2= 0.414**			**R^2=0.530**		
	Weights=#RO DUs			Weights=Inverse of Predicted Values of Variance Model		

Variance Model [$\sigma_i^2=z_i\beta$]:

Dependent Variable: Squared Residuals of Primary OLS Model

Explanatory Variables:

Variable	OLS			FGLS Coefficient	(SE)	p-value
Constant	+14,650			+35,036	(10,174)	.0006
#Bedrooms	n/a			+17,581	(4,984)	.0004
#Rooms	n/a			-10,399	(3,386)	.0022
Age (Median years of all DUs)	n/a			-225.9	(75.28)	.0028
Lot Size (sf)	n/a			+1.399e-5	(1.509e-6)	.0000
1/(#RODU)	-			+3.241e+5	(9.987e+4)	.0012
Sales&Service Walk Access.	n/a			-2.080	(1.658)	.2100
Distance SF CBD (miles)	n/a			-269.9	(51.65)	.0000
Time/Trip (100ths of hour)	n/a			+163.6	(82.36)	.0473
Avg. Time to Work (min.)	n/a			-284.1	(212.7)	.1820
Industrial Fraction	n/a			-4,859	(5,266)	.3565
Fraction of DUs Boarded Up	n/a			+3.478e+5	(1.347e+5)	.0100
Vacant Fraction	n/a			+33,164	(21,200)	.1181
VMT/Trip (euclidean miles)	n/a			-790.8	(368.9)	.0324
Land Use Balance	n/a			+10,455	(5,980)	.0808
				R^2=0.213 (variance model)		

Notes: OO DUs differed from RO DUs for the following explanatory variables: #Bedrooms, #Rooms, Square Footage estimates (SF), and Structure Age (where OO median age and all-structures median age were used, respectively).

The OLS model's standard errors have been estimated using White's robust estimator, under heteroskedasticity. (Please see endnote numbe 5 for further explanation.)

All R^2's are adjusted for degrees of freedom.

Here, the neglect of (non-sample-size) heteroskedasticity may lead one to erroneous conclusions regarding magnitude, and even sign, of explanatory-variable effects (*e.g.*, age in the OODU model). Observe that differing DU sizes (via number of rooms), ages, accessibilities, land-use patterns, and population travel characteristics exert important effects on the perceived variance in these models; the efficiency of estimates appears to be substantially compromised when OLS is the method of analysis used.

Interestingly enough, *VMT*-per-trip and auto ownership do not come into statistically significant play in the primary models. Yet, travel *times* to work consistently are economically and statistically significant. For example, if a person in an OODU makes an average of two one-way work trips per weekday, 250 days per year, and there are *two* such time-valuing persons per typical household living on the median OODU quarter-acre lot, the -0.3428 coefficient translates to a net present value of $3.73 per yearly commute minute saved (or $224 for each commute hour saved), over the entire course of the home ownership. After accounting for future time benefits (via discounting), this result translates to a value of time of roughly $20 per hour - a rather sizable figure when compared with many of the results of studies cited at the beginning of this report, particularly when one considers that the model already controls for two other measures of accessibility. However, the commute-time variable may be picking up time savings found in other, non-work trips, causing the work-trip time to have a coefficient that is biased-high. For example, if all adult trips are valued the same (per minute spent) and the same time savings found in the work trip is exhibited in all other trips made, the $20-per-travel-hour figure may be closer to $5 per hour (assuming roughly three non-work trips for every commute trip).

In addition to travel times to work, the work-accessibility variable proves highly useful.

Plot 1 provides a view of owner-occupied-dwelling-unit price estimates across different work-accessibility levels and at distinct work travel times. Note the rather dramatic dependencies.

There is a positive, rather than negative, coefficient on vehicles-per-adult in the RODU models; this somewhat unexpected result is likely an indication of the value to a renter of parking provision by the landlord. In many areas renters are required to pay $100 or more per month for a single parking space, so this variable's sign shouldn't be taken too seriously as a measure of travel costs.

To give one a better idea of the influence of explanatory variables, several of the elasticity estimates are provided here:

TABLE 4. ELASTICITY ESTIMATES FOR OO & RO DU MODELS

	OODU Price	**RODU Rent/Month**
With respect to:		
#Rooms	+1.71	+0.980
log(Age)	-0.235	-0.331
Lot Size (sf)	+0.321	-0.009
SF x Avg.Time to Work	-0.343	-0.108
Mix of Land Uses	-0.049	-0.019

Note: Elasticities are computed at means.

From the elasticity estimates one may infer that the number of rooms (a proxy for home size) is highly influential in its economic impact on home price and rent; and, while lot size figures very prominently in the owner-occupied market, dwelling unit age is a more important determinant of rental-market prices. Note also the significant elasticity levels associated with the travel time to work variable (interacted with parcel size); evidently, travel considerations can impact a sizable share of home value (as well as rents).

Use of Purely Additive Models:

The models based on the interactive model, discussed above, incorporate a variety of interaction terms between trip costs (and accessibility indices) and square-footage estimates; however, their adjusted coefficients of determination (R^2) are substantially lower than those achieved using purely additive models, whose results are shown in Table 5 (*e.g.*, R^2s of 0.56 versus 0.86). The substantial goodness-of-fit differences may be suggestive of a marketplace for dwelling units where the buyer (and/or seller) considers attributes as relatively marginal in their impact on price (or rent). This makes some sense considering the rigidities of the land market; for example, parcel sizes are rather fixed over the near term (~10-25 years) thanks to zoning laws

TABLE 5. GENERAL ADDITIVE MODELS' FGLS RESULTS

	***Owner*-Occupied Dwellings OODU**			***Renter*-Occupied Dwellings RODU**		
Primary Model [y=Xβ]:	***FGLS Coefficient (SE) & p-value***			***FGLS Coefficient (SE) & p-value***		
Dependent Variable:	**Median Respondent-Estimated Home Price (1989$)**			**Median Contract Rent/Mo. (1989$)**		
Explanatory Variables:						
Constant	-1.189e+5	(2.450e+4)	.0000	+506.2	(48.92)	.0000
#Bedrooms	-7.027e+4	(1.029e+4)	.0007	-82.38	(22.67)	.0003
#Rooms	+1.128e+5	(5.317e+3)	.0000	+189.5	(14.38)	.0000
Log(Age)	-1.605e+4	(3.012e+3)	.0000	+189.5	(14.38)	.0000
Square Feet of Parcel	+2.441	(0.236)	.0000	*		
Dist.SF CBD (mi.)	-7.502e+3	(213.2)	.0000	-9.940	(0.5339)	.0000
Dist. SJ CBD (mi.)	-1.632e+3	(113.6)	.0000	-3.280	(0.2083)	.0000
Dist. Oakland CBD (mi.)	+7.888e+3	(184.5)	.0000	+9.958	(0.4798)	.0000
Work Accessibility (30 min.)	+0.3750	(0.0596)	.0000	+8.603e-5	(1.167e-4)	.4612
Sales & Service Walk Access.	+7.770	(6.827)	.2554	*		
Vehicles/Adult	-8,543	(3,867)	.0275	*		
Time/Trip (100ths of hour)	-307.3	(156.4)	.0499	*		
Industrial Fraction	-4.938e+4	(1.442e+4)	.0007	*		
Land Use Balance (Mean Entropy)	+1.062e+4	(1.332e+4)	.4254	*		
Mix of Land Uses	-9.957e+4	(2.426e+4)	.0000	*		
	R^2=0.865			**R^2=0.725**		
	N=737			N=770		
	(Weights=Inverse of Predicted Values of Variance Models)					
Variance Model [$\sigma_i^2=z_i\beta$]:						
Dependent Variable: Squared Residuals of Primary OLS Model						
Constant	-3.258e+9	(2.149e+9)	.1299	+18,383	(8,379)	.0503
#Bedrooms	+6.269e+10	(6.148e+8)	.0000	+1.814e+4	(4,114)	.0000
#Rooms	-3.056e+7	(4.702e+5)	.0000	-1.091e+4	(2,783)	.0000

Log (Age)	+2.663e+8	(3.432e+8)	.4379	-3,179	(1,339)	.0178
Lot Size (sf)	*			*		
1/(#OODUs)	+4.962e+10	(6.143e+9)	.0000	+8.459e+4	(7.434e+4)	.2555
Work Accessibility (30 min.)	+1.269e+4	(4,630)	.0063	+.03872	(0.02482)	.1192
Sales & Service Walk Access.	+3.454e+6	(4.702e+5)	.0000	-0.6045	(1.498)	.6867
Dist.SF CBD (mi.)	*			-170.1	(97.11)	.0803
Dist. SJ CBD (mi.)	+3.921e+7	(3.017e+6)	.0000	+69.21	(41.55)	.0959
Dist. Oakland CBD (mi.)	*			+65.88	(86.17)	.4448
Vehicles/Adult	+5.609e+8	(4.083e+8)	.1699	+2354	(1594)	.1401
Time/Trip (100ths of hour)	*			+140.5	(68.05)	.0393
Industrial Fraction	+3.487e+9	(1.152e+9)	.0025	+2,554	(4,364)	.5585
Land Use Balance (Mean Entropy)	-3.169e+9	(1.476e+9)	.0321	+7320	(5901)	.2151
Mix of Land Uses	-3.140e+9	(2.564e+9)	.2209	-1.543e+4	(9571)	.1073
	R^2=0.280 (variance model)			R^2=0.0584 (variance model)		

Notes: OODUs differed from RODUs for the following explanatory variables: #Bedrooms, #Rooms, Square Footage estimates (SF), and Structure Age (where OO median age and all-structures median age were used, respectively).
* These variables were dropped early due to lack of significance and "improper" sign.

(*e.g.*, density caps) and the durability of dwellings. In other words, parcel sizes cannot adapt quickly to changes in underlying land valuation, so the dwelling becomes a package of relatively separable qualities. Lot size may therefore, in the short run, be relatively separable from access considerations.

From the additive models' results, one observes substantial value associated with reduced trip lengths and vehicle ownership needs in the OODU model, but these become negligible in the RODU model. Perhaps tracts with a significant sample of rental units (which are then heavily weighted in the regressions) offer *more* than "threshold" levels of viable travel-mode choice (such as rapid transit and high-frequency bussing) and within-walking-distance commercial uses. Such a situation is credible, given that most transit lines serve commercial centers and high-density residential uses generally are zoned alongside commercial uses. So, if travel-choice and local-access threshold levels are met for most rental units, then at the margin the differences in travel times, trip lengths, auto ownership, and walk accessibility across different rental units will not appear as economically significant. Additionally, collinearity across these explanatory variables (*e.g.,* +0.36 between vehicle ownership and VMT-per-trip, and +0.47 between VMT-per-trip and time-per-trip) can mask an individual variable's contribution.

In both model types, it is rather clear that major premiums accompany proximity (in miles) to downtown San Francisco; distances to San Jose and Oakland are influential as well. After controlling for five measures of regional accessibility in these models (*i.e.*, two accessibility indices and three distance-to-CBD variables), the estimated value of travel time for a household making ten one-way trips per day, 365 days a year, yields a discounted value of time of roughly just $1 per hour in the OODU additive model, across all trip types. Clearly then, access - not just observed behavior (*i.e.*, time per trip) - is very important in location valuation. This relatively low value-of-time estimate is probably largely due to the fact that the already-controlled-for variable of "access" allows greater choice within the same travel-time radius, enabling more valued trip in high-access environments than those of similar travel time in lower-access areas.

Note the positive coefficient for the log of Age in the additive RODU model - in contrast to the *negative* (and influential) coefficient on this variable in the interactive RODU model. A change in model set-up, even after controlling for a variety of access measures here (which are correlated with structure age, since the most central areas of the region are the oldest), has led to an unanticipated result.

ADDITIONAL OBSERVATIONS:

Other observations and conclusions that may be drawn from the results of this research as well as from the series of models that led to the final interactive and additive models' results are the following:

- Local built-environment diversity does not seem to be appreciated in the housing and rental markets. Coefficients on land use mix and local accessibility to sales and service jobs were consistently negative[6], after controlling for more regional accessibility; and land use balance's coefficient was almost always negative. Additionally, the fraction of a tract's developed area that was industrial was a highly statistically significant and negative-coefficient variable for these regressions, and the fractions of park and public space were often not of statistical importance.[7] Such results suggest that single-use zoning may protect property values.

- CBD distances, while highly simplistic, turn out to very effectively *predict* housing prices and rents - even in the presence of far more sophisticated measures of access and travel costs. Models used elsewhere which rely on such simplistic variables are perhaps not as weak as they may first appear, at least for predictive (if not theoretical) purposes.

The consistently highly positive coefficients associated with distance to Oakland's downtown, competing in absolute value with the (negative) coefficients on distance to San Francisco's CBD, are probably indicating the relative abundance of land in the East Bay, versus that on the peninsula, as well as the comparative dearth of strong cultural, entertainment, shopping, and work attractions housed in Oakland's downtown - coupled with relatively high crime rates.

- Work proximity appears to be very relevant in location valuation, diverging somewhat from Guiliano and Small's wasteful-commuting conclusions (6). The coefficients on travel time to work and work accessibility measures were consistently of expected sign and highly statistically significant, in contrast to the other measures of travel behavior and access. This result suggests a great value placed by households on locations close to their work; but it may also be an indication of other land use types in competition with households for these locations.[8] For example, the economies to businesses and industry from proximity to job hubs - which house buyer and supplier markets as well as skills and information - can be sizable, so these locators are likely to bid high for such locations.

- Also of note is the fact that the variable of lot size was not useful in the interactive RODU model and was dropped in the additive RODU models, since its coefficient appeared as statistically insignificant (and often with a negative sign). The levels of this variable may be suspect, since they had to be indirectly inferred from the aggregate data; and/or the need for such a variable in RODU models may be questioned, since renters do not control the land their units lie on and since the square footage of a rented unit may be substantially independent of the structure's footprint. However, several interactions of parcel size with travel variables held out as statistically significant and useful to the interactive RODU model.
- Lack of detail in dwelling unit attributes remains a problem, although the adjusted R^2's produced here in the additive models with distance-to-CBD variables are actually higher than those in previous models run by the author for individual Alameda County homes, controlling for a great variety of structural characteristics (*e.g.*, R^2's of 0.86 versus 0.73). Much of the rise in R^2's may be attributed to the

more varied data set, thanks to a far wider set of locations.[9]

Model Limitations:

Given that some of the FGLS model's results were unexpected, as discussed in the above sections, one may want to scrutinize the model and acknowledge its limitations as well as weaknesses. Many of these are discussed here. They are:

- The reliance on purely additive models of home valuation, while offering a better measure of model fit (R^2), is not as theoretically "pure" as is, for example, an explicit acknowledgement of the underlying market for land via interaction terms with lot size. Furthermore, interactions with household size for variables such as the number of bedrooms, structure age, *etc.*, may be relevant since the benefits accrue to several individuals, perhaps additively over individuals (and/or with declining returns), rather than in fixed, marginal amounts.
- A "true" equilibrium in the housing market is essentially impossible to achieve for the following reasons: housing is a highly durable good, the costs of acquiring information on homes for sale and the costs of moving are substantial, and local entities impose zoning and other constraints on home construction. For this reason there may be relatively short-run excess demand and/or supply for different housing attributes (including location) which will then impact the hedonic "prices" of the model (*i.e.*, the model coefficients).
- The BATS trip data are limited to *weekday* travel, when work travel is relatively dominant; thus, only proxies of estimated travel savings have been constructed here. Moreover, the explanatory variables used come from sample zonal

averages, which are not as precise for estimating impacts of marginal changes in attributes.

- The reliance on per-trip (or per-adult) measures of travel costs avoids the difficulty in interpreting *total* household travel expenditures, due to travel minimizing behaviors (such as trip avoidance and trip chaining) present in low-opportunity environments. However, the use of per-trip and per-adult variables makes the implicit assumption that the value of a trip (whose value arises from purposes served at the trip destination) in one area of the region is the same as a trip made in a very different environment. For example, the value of a trip to a large supermarket in the suburbs is implicitly taken to be the same as a trip made in San Francisco's Nob Hill to a mini-market.

An alternative to these types of normalizations is *household* trip-consumption variables; for example, *total* travel time per household. A major weakness of such models is that foregone trip-making, wherein a household suffers some disutility by not making costly trips or by chaining trips, is not observed. Thus, if relatively constant travel time budgets exist (such as those hypothesized by Zahavi [30] and others), they can mask the variation in a location's value to travelers. Per-trip and per-person normalizations help avoid such difficulties.

- Household tastes vary in unobservable ways which can dampen expected changes in home valuation. For example, those whose value of travel time is less (because, for example, they enjoy driving) will tend to locate in more distant areas and may bid up home prices more so than would be expected, even after controlling for income. In contrast, people who feel they need to be near plenty of activity or who dislike driving will tend to compete for more central locations.

Moreover, travel time and VMT estimates may be poor in tracts where the sampled BATS households are in some way unusual, relative to the general population. This may occur through sorting (*e.g.*, where people who love to drive place themselves in very isolated tracts) or by interest and occupation (such as communities for the elderly or full of college students). It would be best to control for such factors in the analysis, perhaps through interactions of travel behavior and population attributes.

CONCLUSION:

The implications of this research are many, although model limitations (as discussed above) do exist. First, non-constant error-term variance (*i.e.*, heteroskedasticity) appears to play an important role in the models of home and rental-units' valuation and should not be neglected in estimation, particularly since generalized least squares results can diverge substantially from those of the OLS (as evidenced here).

Use of travel-cost variables, such as average trip time, car ownership, and per-trip VMT do not negate the role and impact of other travel-related measures such as "accessibility" and distance to major downtowns. In fact, even after controlling for a variety of detailed travel measures and accessibility indices, the rather simplistic distance-to-CBD measures are very strong predictors of housing price and rental rates, as seen in the additive models presented here.

The importance of *work* access seems clear from the model results, while the coefficients on other forms of access (such as park space and local sales and service access) and travel "costs" (such as auto ownership) are often insignificant or of unanticipated sign. Average commute-trip time was found to be highly economically and statistically significant in the theory-based lot size-interactive OODU model, at a rate of approximately $20 per commute-trip hour, which *may* then translate to $5 per hour across *all* trip types (1989$). In the purely additive OODU home-price model, the average trip time computed across *all* trip types was found to be statistically significant - although not nearly as economically significant, due in large part to this model's inclusion of a variety of accessibility measures. In any case, these models suggest new methods for arriving at value-of-time estimates, which can be critical in planning transportation improvements.

Interestingly, local land-use diversity generally appears to reduce dwelling unit values, lending credibility to a perception which has long fueled the implementation of single-use zoning. However, accessibility, land-use mix, and land-use balance do often appear to dampen the variability inherent in expected values. These land-use explanatory variables, like the observed-travel-behavior variables, have been absent in the great majority of property valuation models; yet their significance here suggests that they should be included in future planning models and policy consideration.

Overall, given the variety of explanatory variables and heteroskedastic specifications probed, the possibilities for model interpretation are many and their implications complex. Yet the models examined here do appear to offer many insights and hope for future regional land-modelling efforts of this kind.

ACKNOWLEDGMENTS:

The author is grateful to many persons who provided support in the pursuit of this research: Robert Cervero for the funding of BATS data manipulation and land-use variables computation (via a University of California Transportation Center grant), Alexander Skabardonis for interzonal travel times, Charles Purvis and Alfred Round for BATS data provision and definition, and Gordon Ye for his GIS expertise. Thanks also go to the reviewers of this paper for their highly constructive suggestions.

REFERENCES:

(Abs.) von Thünen, J.H. 1826. *Der Isoline Staat in Beziehung auf Nationale-Konomie und Landwirkschaft*. Stuttgart, Germany: Gustav Fischer. (1966 reprint)

(1) Krumm, Ronald J. 1980. "Neighborhood Amenities: An Economic Analysis." *Journal of Urban Economics*. Vol. 7: 208-224.

(2) Tiebout, Charles. 1964. "A Pure Theory of Local Expenditures." *Journal of Political Economics*. Vol. 64: 416-424.

(3) Sonstelie, Jon C., and Paul R. Portney. 1980. "Gross Rents and Market Values: Testing the Implications of Tiebout's Hypothesis." *Journal of Urban Economics*. Vol. 7: 102-118.

(4) Holtzclaw, John. 1994. "Using Residential Patterns and Transit to Decrease Automobile Dependence and Cost." Paper prepared for California Home Energy Efficiency Ration Systems. San Francisco: natural Resources Defense Council.

(5) American Automobile Association. 1993. *Your Driving Costs, 1993 Edition*. Heathrow, FL: AAA.

(6) Giuliano, Genevieve, and Kenneth A. Small. 1992. "Is the Journey to Work Explained by Urban Structure?" University of California Transportation Center Working Paper #107.

(7) Wheaton, William C. 1977. "A Bid Rent Approach to Housing Demand." *Journal of Urban Economics*. Vol. 4 (April): 200-217.

(8) Waters, W.G. 1992. "Values of travel time savings used in road project evaluation: a cross-country/jurisdiction comparison." *Australian Transport Research Forum* (Vol. 17, Part 1). Canberra: Bureau of Transport and Communications Economics.

(9) Small, Kenneth A. 1992. *Urban Transportation Economics*. Chur, Switzerland: Harwood.

(10) Kraft, J. and A. Kraft. 1974. "Empirical Estimation of the Value of Travel Time Using Multi Mode Choice Models." *Journal of Econometrics*, Vol. 2: 317-326.

(11) Morrison, Steven A., and Clifford Winston. 1985. "An Econometric Analysis of the Demand for Intercity Passenger Transportation." In *Research in Transportation Economics: A Research Annual*, Volume 2, Theodore E. Keeler (Ed.). Greenwich, Conn.: JAI Press.

(12) Thomas, Thomas C., and Gordon I.Thompson. 1970. "The value of time for commuting motorists as a function of their income level and the amount of time saved." *Highway Research Record, No. 314*: 1-19.

(13) Bruzelius, Nils. 1979. *The Value of Travel Time*. London: Croom Helm.

(14) Train, Kenneth. 1980. "A Structured Logit Model of Auto Ownership and Mode Choice." *Review of Economic Studies*, Vol. 47: 357-370.

(15) Alonso, William. 1964. *Location and Land Use*. Cambridge, Mass.: Harvard University Press.

(16) Anas, Alex. 1982. *Residential Location Markets and Urban Transportation*. New York: Academic Press.

(17) Anas, Alex. 1979. "The Impact of Transit Investment on Housing Values: A Simulation Experiment." *Environment and Planning A*. Vol. 12: 747-764.

(18) Boyce, D.E., & B. Allen. 1973. "Impact of Rapid Transit on Suburban Residential Property Values and Land Development." Philadelphia, Penn: Department of Regional

Science, University of Pennsylvania.
(19) Mohring, Herbert. 1961. "Land Values and the Measurement of Highway Benefits." *Journal of Political Economy*. Vol. 49 (June): 236-249.
(20) Voith, Richard. 1991. "Transportation, Sorting and House Values." *AREUEA Journal*. Vol. 19, No. 2: 117-137.
(21) Griliches, Z. (Ed.). 1971. *Price Indices and Quality Changes: Studies in New Methods of Measurement*. Cambridge, Mass.: Harvard University Press.
(22) McMillan, Melville L., Bradford G. Reid, and David W. Gillen. 1980. "An Extension of the Hedonic Approach of Estimating the Value of Quiet." *Land Economics*. Vol. 56, No. 3 (August): 315-328.
(23) Metropolitan Transportation Commission. 1994. "San Francisco Bay Area 1990 Regional Travel Characteristics." Oakland, CA: MTC.
(24) Association of Bay Area Governments. 1993. *Description of ABAG Land Use File* (July 1993 Version). Oakland, CA: ABAG.
(25) Levinson, David M., & Ajay Kumar. 1995. "Multimodal Trip Distribution: Structure and Application." *Transportation Research Record 1446*. Transportation Research Board, NRC, Washington, D.C.
(26) Kockelman, Kara. 1996. "Travel Behavior as a Function of Accessibility, Land Use Mixing, and Land Use Balance: Evidence from the San Francisco Bay Area." Master's Thesis, Department of City and Regional Planning, University of California at Berkeley.
(27) Rice, John. 1995. *Mathematical Statistics and Data Analysis*. Belmont, CA: Duxbury Press.
(28) Landis, John, Subhrajit Guhathakurta, and Ming Zhang. 1994. "Capitalization of Transit Investments into Single-Family Home Prices: A Comparative Analysis of Five California Rail Transit Systems." The University of California Transportation Center Working Paper No. 246, U.C. Berkeley (July 1994).
(29) Oberhofer, W., and J. Kmenta. 1974. "A General Procedure for Obtaining Maximum Likelihood Estimates in Generalized Regression Models." *Econometrica*. Vol. 42: 579-590.
(30) Zahavi, Yacov. 1974. "Travel Time Budget and Mobility in Urban Areas." FHWA, USDOT. NTIS PB 234 145, Washington, D.C. (May 1974).
(31) White, H. 1980. "A Heteroscedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity." *Econometrica*, Vol. 48: 817-838.

ENDNOTES:

1. For estimation of owner-occupied and renter-occupied dwelling-unit (OODU & RODU) mean lot sizes, the total area used by the OODUs is assumed to equal the area covered by the lowest-density DUs until all owner-occupied DUs are "used up." All remaining DUs (*i.e.*, those at higher densities) are assumed to be renter occupied.

The "equivalent" number of single-unit detached DUs is estimated by assuming the relative footprint sizes of moderate- and high-density dwelling unit types (by structure "size" - *i.e.*, #DUs in structure) as a percentage of the single-unit detached DU footprint and is shown here:

Type	Description	Lot-Size%
1	1-unit detached	100%
2	1-unit attached	75%
3	2-unit	65%
4	3-4 units	50%
5	5-9 units	40%
6	10-20 units	15%
7	20-50 units	5%
8	50+ units	1%
9	Mobile Homes	5%
10	"Other" DU types	1% (expect ~dormitories here...)

The distributional and footprint assumptions translate to square foot estimates via the following equations:

Install Equation Editor and double-click here to view equation.

Note, for example, if the number of OODUs is less than n_1 then all of these are of square footage for Type 1. After determining/estimating how many OODUs fall into each category, an estimate of the average lot size per OODU is just the weighted average of the different lot sizes estimated to occur across the OODU sub-population.

2. A Euclidean version of VMT (*i.e.*, "as the crow flies") was estimated using each trip's origin and destination tracts' centroid coordinates and was adjusted for vehicle occupancy levels. If a trip's origin and destination census tract were the same, the trip was assigned a distance estimate of 0.2 Euclidean miles (which typically translates to a quarter mile).

3. Other parameterizations of the form of heteroskedasticity were investigated as well: an exponential (where $\sigma_i^2=\exp(z_i\beta)$), and a squared form (where $\sigma_i^2=(z_i\beta)^2$) - both of which ensure positivity of variance estimates. These models performed substantially poorer than the original heteroskedastic models (registering R^2s of less than 0.10). Thus, the initial parameterization of heteroskedasticity ($\sigma_i^2=z_i\beta$) was maintained. However, this form does not ensure positivity of variance estimates; thus, four percent of the owner-occupied and 0.4 percent of the renter-occupied variance estimates, being negative, were not of aid in the second stage of the FGLS regression process.

4. The number of usable records ended well under the maximum number possible (*i.e.*, the 1,382 census-tract population for the nine-county San Francisco Bay Area) because observations were removed for the following reasons: 1) some census tracts had no OO and/or no RO dwelling units; 2) median price and/or rent exceeded the census survey's upper bounds/caps of $500,000 and $1,000, respectively; 3) no BATS data were available for the tract; and 4) lot-size calculations gave unreasonable estimates.

5. Note that when heteroskedasticity is present, the standard error results of the early OLS regressions must be revised, because OLS error estimates are biased (downward) and may encourage improper model interpretation. With heteroskedasticity, $VC(\beta)$ no longer reduces to $\sigma^2(X'X)^{-1}$. Instead, one must estimate $(X'X)^{-1}(X'\Omega X)(X'X)^{-1}$, where Ω is the covariance matrix of the model's error terms. White's robust estimate of the coefficient estimates' covariance matrix (31) can be computed independent of the underlying form of heteroskedasticity and relies on a summation of variance estimates (*i.e.*, the squared OLS residuals) times the outer product of observation vectors. White's estimate of $X'\Omega X$, $SUM(e_i^2 \mathbf{x_i x_i'})$, was used here to robustly estimate the true standard deviations of the OLS coefficients.

6. Wheaton (7) observes a very similar result, where the coefficient for his measure of accessibility, based on highly local employment, garners a negative - rather than the expected positive - sign.

7. Note, however, that the most valued sites are too expensive for industrial uses, and perhaps for much park use or even public use. So there is collinearity and possible error-term correlation with explanatory variables occurring here.

8. The lesser variability and reduced measurement error associated with the Census- versus the BATS-based explanatory variables may be responsible for differences in statistical significance as well. For example, the average commute time variable is based on the long-form Census questionnaire, which is distributed to between ten and fifteen percent of the region's households. In contrast, the BATS forms went out to under one percent of the region's households.

9. The number of observations between the two data sets is roughly the same (~800+), but those for this sample are averaged across essentially all homes in the census tract. While the number and detail of structural explanatory variables fell substantially in moving from TRW-collected to Census data, the number of observed census tracts increased from under 300 in the Alameda County regressions to over 800 here; thus, the number of distinct levels (or values) of tract-averaged BATS-based explanatory variables (such as average trip time and VMT per trip) increased dramatically here.