

THE FEDERAL RESERVE BANK *of* KANSAS CITY
RESEARCH WORKING PAPERS

A Quantitative System of Monocentric Metros

Jordan Rappaport

July 2014

RWP 14-03

A Quantitative System of Monocentric Metros*

Jordan Rappaport
Federal Reserve Bank of Kansas City

July 2014

Abstract

The monocentric city framework is generalized to comprise a system of metros. A “representative” closed metro calibrates parameters and establishes a reservation utility and perimeter land price that must be matched by an indeterminate number of open metros. The open metros have exogenous productivity different from that in the representative metro. Large changes to the parameterization and to key model assumptions often affect land use patterns in the representative metro only minimally but have first-order effects on land usage in open metros. For a given level of productivity, transportation technology is the most important determinant of population, land area, population density, and house prices across and within metros. Changes in highway capacity significantly affect these quantities while leaving commute speeds unchanged. Telecommuting one workday per week increases the population of already crowded metros by as much as one third. Open metro land area asymptotes to a maximum at only moderately high relative productivity. A high threshold productivity is required for potential metro land to exceed is agricultural opportunity cost. The framework yields a number of quantitative insights into how preferences, production technologies, and transportation technologies shape outcomes within and across metros.

Keywords: Monocentric City, Urban Systems, Commuting Congestion, Value of Time

JEL classifications: R10, R41, C68

1 Introduction

The Alonso-Muth-Mills (AMM) model of circular metro area has been a workhorse model of urban land use for more than 50 years (Alonso, 1964; Mills, 1967; Muth, 1969). At its core, AMM gives insight into the fundamental tradeoffs by which decreasing house prices compensate workers for the monetary and time costs of commuting. From a theoretical viewpoint, it illuminates how the slopes of bid-rent functions determine income segregation within a metro area. From an empirical

*The views expressed herein are those of the author and do not necessarily reflect the views of the Federal Reserve Bank of Kansas City or the Federal Reserve System. The calibration herein is superseded by that in “Monocentric City Redux,” Federal Reserve Bank of Kansas City RWP 14-09. Thank you to David Albuoy for extensive advice and feedback. Thank you to Martina Chura, Kathleen Navin, and Daniel Molling for excellent research assistance.

viewpoint, AMM presents numerous testable hypotheses. For example, it implies that suburbanization will depend closely on the speed with which workers can commute to the CBD and so on highway infrastructure (Baum-Snow 2007a, 2007b). From a quantitative viewpoint, it gives insight into the degree to which population density and price gradients depend on deep structure such as the ability to substitute between land and structure in the production of housing (Muth, 1975; Arnott and MacKinnon, 1977a) and individuals' willingness to tradeoff housing and non-housing consumption (King, 1977; Richter, 1978).

In a world of seemingly self-evident polycentricity, it is natural to question how much the monocentric framework can teach us. From early on, it was criticized as being largely irrelevant to modern metro areas (Wheaton, 1979). But such criticism misses the necessity in all economic modeling to starkly simplify a complicated world. The benefit of such simplifications is that they often yield considerable insight. The assumption of a monocentric metro is likely no worse than the assumption of a world populated by homogeneous individuals and firms.

A stylized monocentric allocation of land may also not be as egregious as it would seem. Baum-Snow (2014) shows that the suburbanization of urban residents from 1960 to 2000 was nearly three times as large as the suburbanization of urban jobs. For industries in which there is likely to be large local agglomerative spillovers—such as finance, insurance, and real estate—employment decentralization was relatively minimal. Similarly, Brinkman (2013) documents the sharp fall in employment densities near the city center in each of the three metros he studies (Columbus, Houston, and Philadelphia, each in 2000).

A more salient concern is the extent to which assumptions such as monocentric land use obscure or rule out outcomes of first-order interest. In the quantitative implementation below, assumed monocentricity rules out generating a megalopolis such as the New York CBSA without resorting to implausible parameterizations of minimal traffic congestion, a high leisure content of commute time, and very high substitutability between numeraire consumption and each of housing and leisure. In one sense, this failure is reassuring. A model that suggested that 10 million workers could enjoyably commute at constant speed to a single CBD should provoke considerable skepticism. But the failure also emphasizes the importance of polycentric production. To incorporate this, a number of models exogenously specify two or more production locations (Brueckner, 1978, 1979; Helsley and Sullivan, 1991; Sivitanidou and Wheaton, 1992). Additionally, some models endogenize heterogeneous production land use within one or more exogenously-specified locations (Sullivan, 1983a,b,c; Sullivan, 1986).

More recently, economic models have successfully endogenized the location of multiple work places. Anas and Kim (1996) show that with sufficient scale economies in the retail sector, metro form is characterized by multiple stable equilibria with a discrete number of workplaces. Lucas and Rossi-Hansberg (2002) develop a more general framework in which the interaction of local agglomerative spillovers among firms together with commuting time costs can support a wide array of stable residential and business land use patterns. Chatterjee and Eyigungor (2013) and

Brinkman (2013) build on this framework by introducing land use restrictions and commuting congestion. Solving these models requires several important tradeoffs. For example, individual utility is assumed to be Cobb Douglas and housing production is assumed to be a linear function of land.

Reverting to an assumed exogenous pattern of business location allows for a more richly specified quantitative model. A closed monocentric metro with exogenously set population, CBD land area, and residential land area calibrates several key structural weighting parameters. This “representative” metro corresponds exactly to the metro typically modeled in the existing literature on urban form. It additionally establishes the reservation utility and perimeter land price that must be matched in an indeterminate number of open metros. Each of these open metros may exogenously differ from the representative metro with respect to one or more characteristics, such as productivity, consumption amenities, transportation technology, and public policy. The resulting system of metros represents a general equilibrium in the sense that all outcomes across and within metros are mutually consistent. But the system retains a partial equilibrium flavor because it does not endogenize the number of metros as in the system of cities developed in Henderson (1974).

From a quantitative perspective, land use in the representative metro proves fairly insensitive to changes in key assumptions and parameterized values, such as on transportation technology and on elasticities of substitution in utility and production. The reason is that endogenously calibrated weighting parameters in utility and production adjust shift in order to continue to hit key calibration targets.

The remainder of the paper is structured as follows: Section 2 presents some motivating empirics on variations in population density and commute times across and within metro areas. Sections 3 and 4 respectively describe the model and its calibration. Section 5 shows how gradients for population density, house and land prices, and number of other important outcomes vary within and across the representative metro and two open metros, one with total factor productivity moderately below that of the representative metro and one with total factor productivity moderately above that of the representative metro. Section 6 shows how summary measure outcomes, such as total population, total land area, and mean house and land prices, vary with open metro productivity. Section 7 illustrates how outcomes across and within metros are affected by perturbations to the baseline calibration and assumptions.

2 Empirical Motivation

A key goal of the quantitative system of monocentric metros is to approximately match the large variations in population, land area, and population density across and within metro areas. To do so, it is critical that observed and quantitative outcomes be measured comparably.

Beginning in the 1950s, the Office of Management and Budget has used an evolving set of technical criteria to designate the existence and borders of metro areas. In general, these have

sought to define contiguous land areas that constitute integrated labor markets within which people both live and work. The current generation of criteria was first promulgated in 2003 based on Census 2000 data. It defines a set of “Core-Based Statistical Areas” (CBSAs), the delineations of which are updated each year. Compared to the previous generation delineation of MSAs, PMSAs, and CMSAs, the current generation considerably improves the correspondence to metro labor markets. A continuing weakness is that by delineating metro areas to be composed of one or more complete counties, CBSAs are primarily made up of land that is agricultural or otherwise unoccupied.¹

This incongruity between formal and functional metro boundaries makes comparison of “raw” population density—total CBSA population divided by total CBSA land area—a highly misleading measure. For example the Phoenix CBSA spans two largely unoccupied desert counties. Its raw density in 2000 was under 600 persons per square mile, which is on par with the raw density of census tracts near the border of suburban settlement and agricultural land use. A more meaningful measure of metropolitan population density is to use the population-weighted mean of the raw density of small geographic units such as census tracts (Glaeser and Kahn, 2004; Rappaport, 2008a). This alternative measures average population density as experienced by metro residents rather than as experienced by metro land parcels. No judgement is required on which which parcels of CBSA land are truly “metropolitan” as the non-metropolitan parcels, whichever they might be, receive essentially no weight. Similarly, the aggregate population across all CBSA land parcels should be a reasonably accurate proxy for metropolitan population because non-metropolitan parcels will by definition be sparsely settled.

For measuring metropolitan land area, however, a more subjective approach is required. The most obvious way to do so is to measure the raw population density of small geographic units and then arbitrarily specify a lower bound threshold to be considered part of a functional metro area. Some experimentation suggests that 500 persons per square mile is a reasonable lower threshold. Aggregating over all census tracts in all CBSAs that had a population of at least 100 thousand in 2000, a 500 person per square mile threshold includes census tracts that account for 88 percent of aggregate CBSA population but only 16 percent of aggregate CBSA land area. For some high-population CBSAs, the 500 threshold captures an even higher share of CBSA population and a much higher share of CBSA land. For the New York CBSA, tracts with density of 500 or more account for 99 percent of CBSA population and 75 percent of CBSA land. For the Boston CBSA, the respective values are 98 and 80 percent. For the Phoenix CBSA, the 500 threshold captures 90 percent of population but only 9 percent of land.²

Figures 1 and 2 show the density variation by Census tract within two intermediate population

¹But a considerable benefit of delineating CBSAs to be the union of entire counties is the ability to match metros to numerous sources of aggregate and micro data.

²For comparison, the Census Bureau uses a threshold of 1000 persons per square mile to delineate the borders of urbanized areas. Using 500 instead captures a significant additional share of the aggregate population of the counties that together make up CBSAs while adding relatively little land area. As such, a 500 threshold implies geographic borders relatively similar to those of Census-delineated urbanized areas.

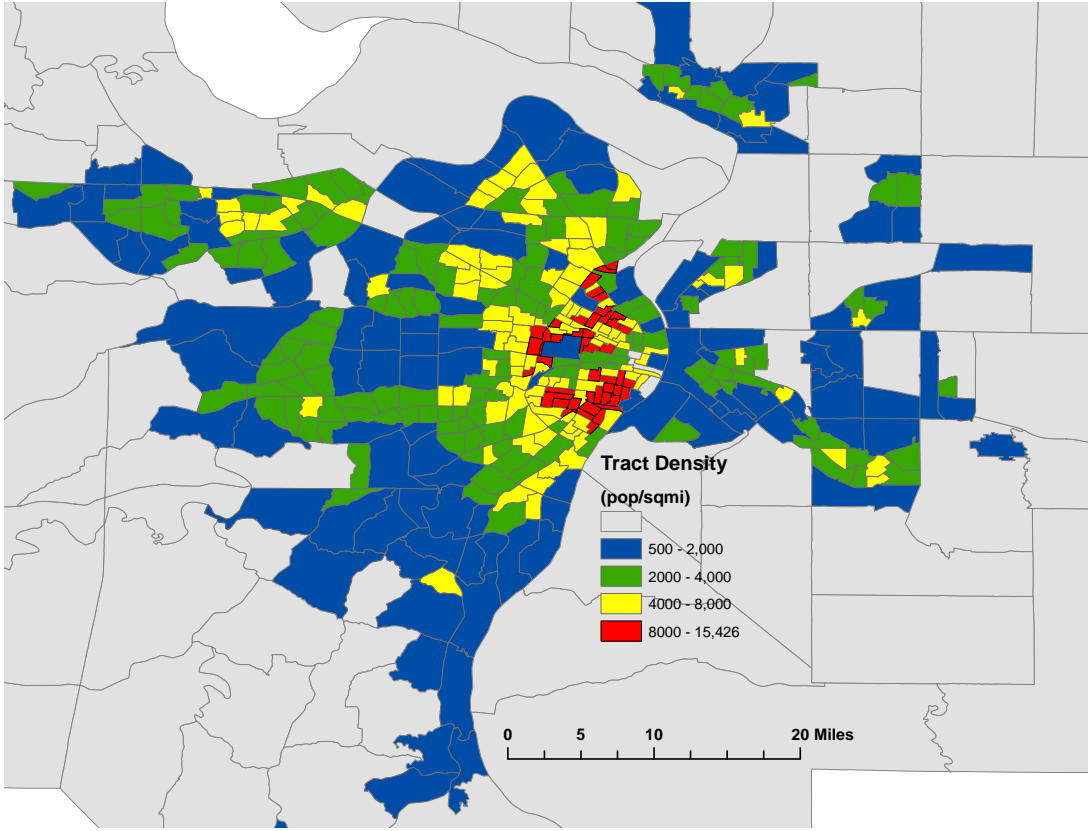


Figure 1: St. Louis Tract Population Density. Map shows tract population density in 2000 within the counties that formally constitute the St. Louis Core Based Statistical Area (CBSA). Tracts with population density of at least 500 persons per square mile account for 85 percent of St. Louis CBSA population but only 18 percent of CBSA land area.

CBSAs, St. Louis and Pittsburgh. The color-shaded cells have raw population density of 500 or higher and so illustrate the functional land area within formal CBSA boundaries. Consistent with the monocentric stylization, residential density rapidly decreases moving away from a moderately tight central location.

Population, land area, and summary density measures across and within selected functional metro areas are enumerated in Table 1. The metros are ranked by population. The second column shows the lower bound percentile of metro population as “experienced” by all residents in U.S. functional metro areas. For example, the 18.1 million residents of the functional New York metro area account for 9 percent of U.S. aggregate metro population. And so New York metro residents make up the 91st through 100th percentiles of aggregate U.S. metro population. Similarly, the 12.3 million residents of the functional Los Angeles metro area make up the the 85th through 91st percentiles in terms of experienced metro population.

Functional metro land area and population-weighted mean density are generally decreasing

rank	pop pctile (lwr bnd)	metro area	population	land area (sq.mi)	mean density (pop- wghtd)	density percentile across census tracts (pop-weighted)						
						99th	95th	90th	75th	50th	25th	10th
		Aggregate Metro	186 million	74,400	8,300	93,300	29,000	16,100	7,500	4,000	2,000	1,000
1	90.7%	New York	17,623,000	3,360	33,930	172,700	122,410	94,960	50,940	15,410	4,380	1,940
2	84.3%	Los Angeles	12,172,000	1,830	12,400	49,530	29,980	23,580	15,560	9,800	6,230	3,390
3	79.8%	Chicago	8,548,000	2,490	10,270	50,480	33,650	26,290	14,060	5,590	2,710	1,390
		Large (mean)	4,230,000	1,550	6,700	32,900	19,000	14,600	8,700	4,600	2,500	1,300
4	77.1%	Philadelphia	5,151,000	1,860	8,850	39,940	29,140	23,510	12,500	4,590	2,160	1,100
5	74.5%	Miami	4,856,000	1,180	6,970	29,330	16,800	12,520	8,580	5,760	3,730	2,060
6	69.9%	Dallas	4,476,000	1,750	4,660	20,900	10,870	8,510	5,750	3,940	2,180	1,120
7	69.9%	Washington DC	4,181,000	1,380	6,780	37,460	22,670	15,350	7,370	4,450	2,680	1,330
8	67.8%	Detroit	4,051,000	1,380	4,960	13,030	10,990	9,530	6,710	4,440	2,520	1,320
9	63.6%	Houston	4,024,000	1,550	4,650	19,580	10,080	8,700	6,150	3,910	2,190	1,230
10	63.6%	Boston	3,956,000	1,810	8,360	46,610	28,380	21,710	11,200	4,000	1,470	790
:	:	:	:	:	:	:	:	:	:	:	:	:
		Intermediate (mean)	1,800,000	680	5,000	17,000	11,500	9,500	6,600	4,300	2,500	1,300
		Mean	1,730,000	670	4,300	15,400	9,500	7,700	5,700	3,800	2,300	1,300
13	56.3%	Phoenix	2,931,000	880	5,370	15,890	10,810	9,250	6,850	5,180	3,180	1,680
14	54.9%	Seattle	2,754,000	1,040	4,910	27,030	10,510	8,560	6,010	4,010	2,520	1,100
15	53.4%	Riverside	2,713,000	1,070	4,470	12,480	10,020	8,540	6,260	3,830	2,120	1,110
16	52.1%	San Diego	2,584,000	640	7,590	28,190	17,790	15,010	9,620	6,140	3,790	2,020
17	50.8%	Minneapolis	2,431,000	1,010	4,330	18,810	11,860	8,690	5,250	3,250	2,020	1,160
18	49.6%	Baltimore	2,222,000	830	6,360	29,090	19,430	15,530	7,910	4,180	2,220	1,130
19	47.4%	Tampa	2,161,000	930	3,590	8,360	6,760	6,010	4,820	3,620	1,930	1,040
20	47.4%	St. Louis	2,113,000	990	3,760	13,510	9,410	7,340	4,840	3,150	1,650	960
21	46.3%	Denver	1,957,000	580	5,410	17,010	10,670	9,100	6,960	5,050	3,130	1,660
22	44.4%	Cleveland	1,911,000	740	5,210	16,760	12,880	11,240	7,470	4,200	2,200	1,060
23	44.4%	Pittsburgh	1,778,000	860	4,330	17,210	11,410	8,900	5,950	3,500	1,540	820
:	:	:	:	:	:	:	:	:	:	:	:	:
		Small (mean)	715,000	340	4,200	15,100	10,800	8,500	5,500	3,400	1,800	990
38	32.6%	Buffalo	997,000	430	5,340	17,610	13,800	11,500	7,760	4,310	1,980	900
39	33.1%	Charlotte	970,000	710	2,130	7,310	4,330	3,970	2,910	1,830	910	650
40	32.6%	Memphis	936,000	400	3,400	8,980	6,790	6,110	4,530	3,120	1,900	1,160
41	31.6%	Austin	903,000	390	4,260	17,360	10,850	8,320	5,250	3,520	1,790	950
42	31.6%	Hartford	888,000	500	4,140	20,910	14,860	10,430	5,090	2,280	1,240	750
43	31.2%	Jacksonville FL	886,000	450	2,740	7,250	5,800	4,700	3,530	2,520	1,700	1,110
44	30.3%	Salt Lake City	871,000	240	5,230	14,070	9,090	8,020	6,640	5,150	3,550	2,040
45	30.3%	Louisville	835,000	390	3,530	10,570	7,850	6,670	4,580	3,030	2,060	910
46	29.8%	Oklahoma City	819,000	380	3,290	7,860	6,300	5,540	4,370	3,340	1,810	950
47	29.0%	Bridgeport	811,000	420	5,360	23,980	18,910	12,500	7,680	3,280	1,220	810
48	29.0%	Nashville	810,000	510	2,490	8,950	5,380	4,930	3,400	1,980	1,180	770
49	28.6%	Honolulu	804,000	240	11,590	69,760	42,760	32,530	12,030	7,150	2,730	1,220
50	28.1%	Richmond	792,000	450	2,890	14,410	7,900	5,010	3,390	2,300	1,300	830
51	27.7%	New Haven	766,000	430	4,110	16,840	12,240	9,590	5,670	2,830	1,200	710
52	27.4%	Rochester NY	745,000	350	4,260	14,590	11,860	10,570	6,140	2,630	1,580	950
:	:	:	:	:	:	:	:	:	:	:	:	:

Table 1: Metro Population Density in 2000. Metros are constructed as the combination of all census tracts located within each 2003 Core Based Statistical Area that have a population density of at least 500 persons per square mile. Density percentiles are weighted by tract population.

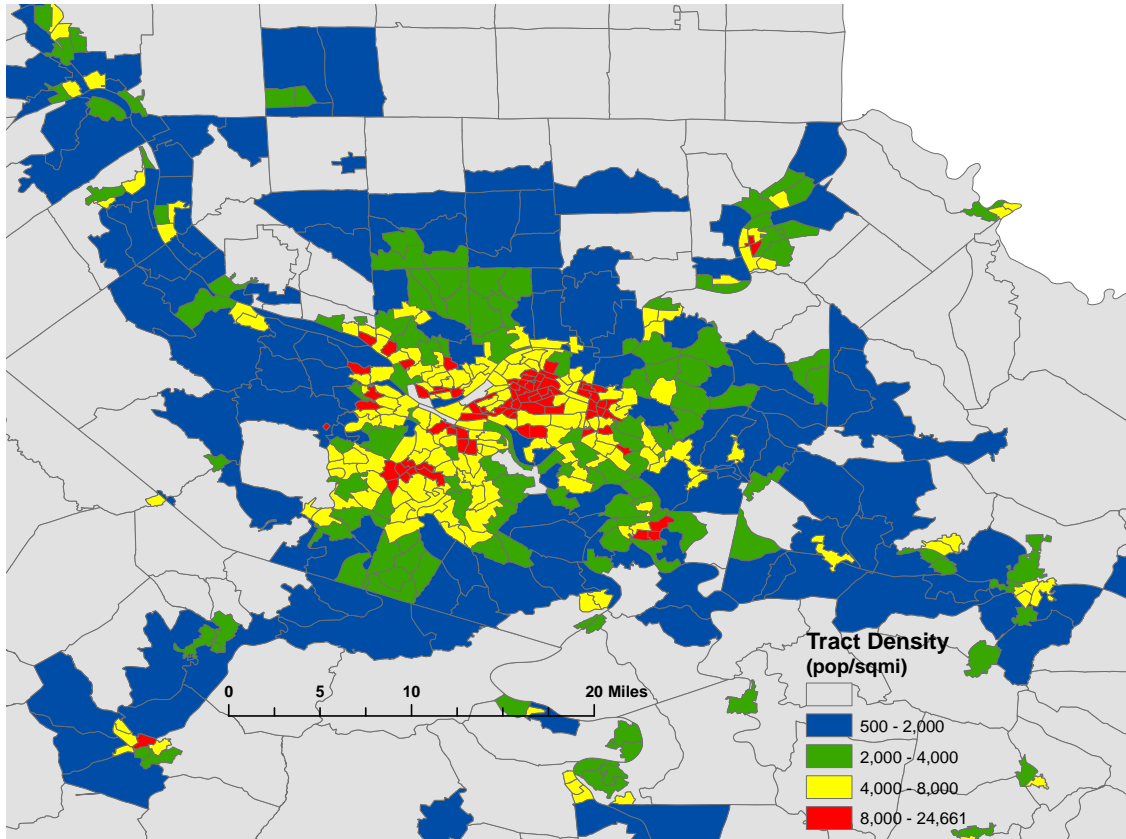


Figure 2: Pittsburgh Tract Population Density. Map shows tract population density in 2000 within the counties that formally constitute the Pittsburgh CBSA. Tracts with population density of at least 500 persons per square mile account for 87 percent of Pittsburgh CBSA population but only 38 percent of CBSA land area.

as functional metro population decreases. On average, so too are the 99th and 90th percentile population density *within* each metro area. These percentiles are constructed in a similar way to the metro population percentiles. Census tracts within each specific functional metro are ordered according to raw density. For example, the 90th percentile within density of a metro is the density experienced by a person for whom at least 10 percent of metro residents live in tract with at least as high population density and at least 90 percent live in a metro with no higher population density.

Metros are grouped by population into “large,” “intermediate,” and “small” categories. The summary rows for each of these report mean values over the metro-specific values immediately below. In contrast, the aggregate population-weighted mean density and density percentiles enumerated in the topmost row are constructed by stacking all functional metro tracts and then applying the same calculations that are used for each specific metro.

The functional New York metro stands out with several extreme values. Its total population is 45 percent greater than the next highest; its population mean density is more than 2.5 times that

of the next highest.³ Its 99th percentile within density, 173 thousand persons per square mile, is more than three times that of the next highest 99th percentile within density. Its 95th, 90th and 75th percentile within density are similarly more than three times higher than the next highest of each.

Commute times across and within selected functional metros are enumerated in Table 2. Mean one-way commute time in 2000 for all residents in functional U.S. metros, calculated by stacking census tracts in all metros, was 29 minutes. The corresponding 10th through 90th percentile commute times across all metro residents were 11 and 51 minutes. These percentiles are constructed in two steps. Within tract commute-time percentiles are constructed by interpolation applied to the number of workers in each tract (who did not work at home) based on reported commute times within each of 11 bins, 0 to 5 minutes through 90 or more minutes. The reported 10th and 25th percentiles for each metro are *across* the 25th *within* tract percentiles. The rationale for using the 25th percentile within percentile rather than a lower one is to exclude the short commutes of suburban residents who live near their workplace. The reported 50th percentile for each metro is the 50th percentile commute time across tracts of the 50th percentile commute time within each tract. The reported 75th, 90th, and 95th commute times are measured across 75th percentile within tract commute times. The rationale for using the 75th within tract percentile rather than a higher one is the goal of measuring the location choice of a worker with approximately “representative” preferences.

Constructed percentile commute times across metros are sensitive to the within percentile choice. For example, the 90th percentile across-metro commute time measured by the 90th percentile within commute time is 69 minutes or higher for each of the nine highest population metros. For most of these metros, this represents a 20 minute increase from the 90th percentile across-metro time measured by the 75th percentile within commute time. For the New York metro area, the 90th percentile across tracts of the 90th percentile within tract lies in uppermost bin (90 minutes or more).

3 The Model

The system of monocentric metros is composed of a closed “representative” metro and an indeterminate number of open metros. The former corresponds to the single metro that has been the subject of most AMM theoretical research to date. It is closed in the sense that its population and land area are specified exogenously. Utility is required to be equal across a finite number of discrete residential rings, but the actual level at which this equalization takes place is endogenous.⁴

³Note that neither of these ratios is consistent with Zipfs Law.

⁴The representative metro can be interpreted as open when it is calibrated by exogenously setting the level of utility and perimeter land price. Doing so has sometimes proved necessary to solve models (Sullivan, 1983a,b,c). But the targeted levels of utility and perimeter land price are necessarily arbitrary because utility is ordinal and the numeraire with which the price of land is measured has no inherent units. With the exponential increase in computing power over the past 30 years, this solution approach is typically no longer necessary.

Pop Rank				Mean Cmt Time	one-way cmt time pctiles across tracts (min)					
	Metro Area	population	Land Area (sq.mi)		10th	25th	50th	75th	90th	95th
	Aggregate			29	11	13	23	42	51	56
1	New York	18,100,000	1,900	38	13	15	31	62	72	76
2	Los Angeles	12,300,000	830	32	14	15	25	43	49	52
3	Chicago	8,800,000	1,300	34	13	15	28	51	56	59
	Large (mean)	4,010,000	800	31.7	13.0	14.9	26.2	45.5	51.3	54.6
4	Philadelphia	5,500,000	1,200	31	12	14	24	44	49	53
5	Miami	5,000,000	580	31	13	15	25	43	49	53
6	Dallas	4,760,000	1,030	30	13	14	24	42	48	51
7	Washington DC	4,470,000	870	36	15	17	31	53	59	62
8	Houston	4,340,000	1,000	31	12	15	27	46	52	55
9	Boston	4,300,000	1,080	31	12	14	26	45	51	54
:										
	Intermd (mean*)	2,090,000	540	26.9	11.7	13.2	21.9	35.1	40.5	45.1
17	Minneapolis	2,640,000	640	26	12	13	22	35	42	47
18	Baltimore	2,440,000	600	32	14	15	27	46	51	54
19	Tampa	2,310,000	540	28	12	13	22	37	44	47
20	St. Louis	2,300,000	650	24	10	11	18	30	33	36
21	Pittsburgh	2,100,000	760	28	11	13	23	39	44	45
22	Cleveland	2,050,000	470	26	12	13	22	34	38	41
23	Denver	2,050,000	350	29	13	15	24	37	42	50
24	Cincinnati	1,760,000	540	27	12	14	22	34	41	49
25	Portland, OR	1,730,000	370	27	12	13	22	34	40	46
:										
	Small (mean)	590,000	160	24.0	10.2	11.5	19.3	31.4	37.1	40.7
63	Omaha	650,000	150	22	10	12	19	28	34	37
64	Albuquerque	640,000	160	25	11	13	21	33	42	50
65	El Paso	631,000	120	22	9	9	17	28	31	32
66	Springfield, MA	586,000	220	24	10	11	19	32	36	41
67	Toledo	571,000	170	22	10	11	19	29	33	34
68	Sarasota	562,000	140	24	11	12	19	32	34	38
69	Grand Rapids	545,000	170	24	11	12	19	33	41	46
70	Bakersfield	545,000	100	25	9	11	18	32	35	35
71	Baton Rouge	536,000	220	28	11	13	22	36	49	54
:										

Table 2: Metro One-Way Commute Times in 2000. The 10th and 25th percentile commute times across tracts are with respect to the 25th percentile commute time within each tract. The 50th percentile commute times are with respect to the interpolated 50th percentile commute time within each tract. The 75th, 90th, and 95th commute times are with respect to the 75th percentile commute times within each tract. Aggregate and individual metro mean times are the population-weighted mean of tract means.
 *Mean values over the intermediate-sized metros exclude Baltimore due to its adjacency to Washington D.C.

The representative metro has two purposes. First, it serves as the sole basis for calibrating structural weighting and cost parameters that are determined by targeting endogenous outcomes. For example, the weighting of the housing component of assumed CES utility is calibrated by targeting a mean housing expenditure share in the representative metro. Similarly, the per mile numeraire commute cost is calibrated by targeting a per mile-share of wage income in the representative metro. Second, the representative metro establishes the reservation utility and perimeter land price that must be matched by each open metro.

As in the existing AMM literature, the representative metro generates patterns of land usage such as gradients of population density, land price, and housing price that are downward sloping as one moves outward from the CBD. The house price gradient is simply the bid-rent curve that allows utility to remain constant as commute time increases. The population density gradient reflects the interaction of housing demand and supply required to support the house price gradient. Because utility and housing production are modeled as constant elasticity of substitution rather than Cobb-Douglas or linear, closed form solutions are not possible. For this reason, each metro is constructed using rings with discrete width, as is typical in the quantitative AMM literature.

There are several large payoffs to the additional modeling of the open metros. Most important is that the effect of varying key assumptions and deep structural parameters can often be seen only in the open metros. For example, the assumption of whether commuting speed is subject to congestion helps determine the calibrated weight on leisure in utility. As will be illustrated in Section 7, the effect on land use from assuming no traffic congestion rather than significant traffic congestion is essentially “undone” in the representative metro by an offsetting decrease in the weight on leisure. This offset is needed to match the calibration target that residents in the representative metro on average desire to work 40 hours per week. Without the offset, residents in the representative metro unsurprisingly want to use some of their saved commute time to increase their leisure time. In sharp contrast, the same assumption of no traffic congestion can boost several fold the population of an open metro with moderately high TFP.

Two additional payoffs concern the potentially “black box” nature of quantitative research. One way of illuminating the black box is to quantitatively model the effect of characteristics that *vary across metro areas*. The present paper focuses on variations in the total factor productivity with which numeraire is produced. Unsurprisingly, open metro population and land area increase as open metro productivity increases relative to productivity in the representative metro. More surprising is that open-metro land area asymptotes to an upper bound as open-metro productivity increases from a moderately above-average level. Further plausible increases in relative productivity increase open-metro population via population infill. Also surprising is that the effect of metro productivity on metropolitan population proves highly asymmetric. Increases in productivity to even implausibly high relative levels increase population log linearly. But decreases in productivity to levels only moderately below that in the representative metro cause open-metro population to plunge to near zero. Importantly, such quantitative results can be matched to underlying theoretical

considerations such as the high convexity of marginal product and marginal utility that are implied by CES specifications with elasticities of substitution less than 1.

A second way of illuminating the potential black box is to quantitatively model how various assumptions and calibration choices that hold in *both the representative and open metros* mediate the effects of exogenous variations across metros, such as to productivity. For example, the baseline setup assumes that commute speed endogenously slows from 65 mph to 10 mph as traffic volume increases. An alternative assumption is that commute speed is constant at its baseline average in the representative metro (37mph). For an open metro with TFP moderately higher than TFP in the representative metro, the absence of congestion increases population threefold. As a second example, a baseline assumption is that individuals are required to work 40 hours per week. Instead allowing individuals to choose their own hours increases the population of an open metro with moderately high TFP by more than 10 percent. In addition, the fixed hour requirement causes individuals with long commutes to have a marginal value of leisure time that far exceeds their wage.

More generally, quantitative analysis can be used to sign first and second *total* derivatives across exogenous model parameters.

3.1 Setup

The setup is static and so should be interpreted as a long-run steady state. Each metro, representative and open, consists of a central business district where production of a numeraire good takes place and a finite number of concentric residential rings surrounding it. The CBD has exogenous radius, \hat{d}_0 , both in the representative and in each open metro. The width of each ring except the outermost one is also exogenous and identical across metros, $\hat{\tilde{d}}_j$. The number of rings in each metro, J , and the width of the outermost ring in each, \tilde{d}_J is endogenously determined. For the representative metro, J and \tilde{d}_J are immediately implied by the assumed land area. But for the open metros, land area is itself endogenously determined.⁵

Rather than being a complete circle, each metro has an identical exogenous span of occupation, $\hat{\theta}_m \leq 360^\circ$. Calibrating metro settlement as occupying considerably less than a full circle is essential to simultaneously matching observed land areas and commute distances.

Utility is assumed to be a function of numeraire consumption, housing consumption, and leisure. The explicit inclusion of leisure in utility is essential to the calibration. More typically,

⁵A quick guide to notation: Decorative hats, such as in \hat{d}_0 , denote variables with exogenous values that are identical across the representative and open metros. Decorative tildes denote “local” variables that apply within a ring but that are also experienced by residents in other rings. For example, \tilde{s}_j denotes the speed of commute traffic across ring j . The one exception to this is \tilde{U} , which is the housing-numeraire sub-component of utility. The combination of a hat on top of a tilde denotes a local variable that is exogenous and identical across metros. For example, $\hat{\tilde{d}}_j$ is the identical width of the j th ring in each metro. For the case of ring width, the exogenous and identical characteristics no longer hold for the outermost ring of each metro, whatever ring number that may be. In other words, the width of the outermost residential ring for each metro is \tilde{d}_J . A decorative overbar denotes a population-weighted mean value. For example, \bar{n} denotes the population-weighted mean number of work hours ($(\sum_{j=1}^J POP_j \cdot n_j)/POP$). In order to not further complicate notation, subscripts to denote the metro to which a variable applies are typically dropped. But occasionally they are included. For example, $r_{m,j}^L$ denotes the price of land in ring j of metro m .

leisure is implicitly modeled by assuming there is a numeraire opportunity cost per unit of commute time. Labor mobility is perfect and so utility must be equal across and within metros. Endogenously determined prices and commute times determine individuals' desired labor supply. However, a baseline assumption is that individuals are required to supply a fixed number of hours.

Production of the numeraire good combines land, capital, and aggregate work hours using a Cobb Douglas production technology with total factor productivity that exogenously varies across metro areas. Housing production combines land and capital in each residential ring using a CES technology. Capital is residually determined by an exogenously specified rental rate. As will be explained immediately below, a key baseline assumption is that capital and land rents are paid to absentee owners. Alternatively, capital and land rents can be rebated lump sum to individuals without respect to where they live.

Figure 3 illustrates the setup. It includes additional variables that will be described in the subsections below.

3.2 Production

Numeraire production is Cobb Douglas in land, capital, and aggregate hours of labor input,

$$X = A \cdot L_0^{\alpha_L} K_0^{\alpha_K} N^{1 - \alpha_L - \alpha_K} \quad (1)$$

Aggregate labor hours is the sum of labor hours supplied by residents in each residential ring, j ,

$$N = \sum_{j=1}^J POP_j \cdot n_j \quad (2)$$

Each factor of production is paid its marginal product, r_0^L , r_0^K , and w . CBD capital is determined residually by achieving an exogenously-specified required rent, $r_0^K = \hat{r}^K$.

Housing in each residential ring is produced with a constant elasticity of substitution between land and capital. For housing, capital can alternatively be interpreted as structure. Productivity is assumed to be equal across and within metros and so it is not explicitly parameterized,

$$H_j = \left(\eta_L L_j^{\frac{\sigma_L - 1}{\sigma_L}} + (1 - \eta_L) K_j^{\frac{\sigma_L - 1}{\sigma_L}} \right)^{\frac{\sigma_L}{\sigma_L - 1}} \quad (3)$$

Labor is excluded from housing production so that all work remains located in the CBD. Land and capital are each paid their marginal revenue product,

$$r_j^L = p_j \cdot \frac{\partial H_j}{\partial L_j} \quad r_j^K = p_j \cdot \frac{\partial H_j}{\partial K_j} \quad (4)$$

As in the CBD, the quantity of capital in each ring is residually determined such that $r_j^K = \hat{r}^K$.

All payments to land and structure are paid to absentee owners. For housing, this is equivalent to paying rent to an absentee landlord. In the quantitative literature, the rent a household pays is

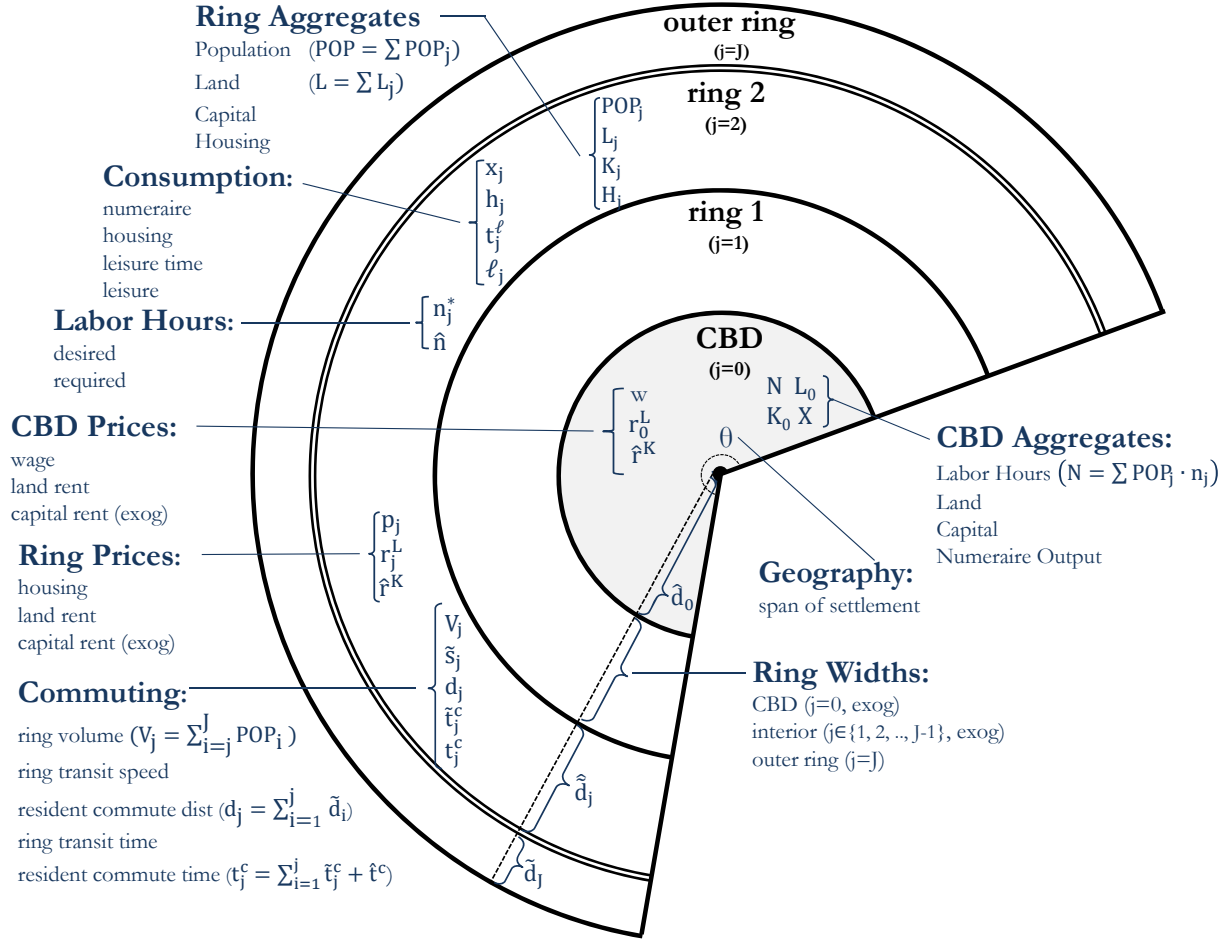


Figure 3: Metro Structure. Figure shows the layout of typical metro, either representative or open. (Footnote 5 explains notation conventions.)

often instead assumed to be rebated to that household on a nominally lump sum basis. But while the household takes the rebate as predetermined in choosing *how much* housing to consume, the rebate does continue to affect the household's decision of *where* to live. In particular, the household considers that by moving to a location where the price of housing is high, it will receive a larger rebate. In consequence, significantly more households choose to live in high-priced locations than would be the case if there were no positive wealth transfer from doing so.⁶

⁶Alternatively, all households can be assumed to have equal shares in a single firm or mutual fund that owns all land and capital in all metros. Households, wherever they live, would receive a lump sum transfer equal to per household aggregate land and capital payments across all metros in the economy. A disadvantage of this alternative assumption is that it lessens the importance of wage income in determining the distribution of population across metros. In particular, a system-wide rebate of this sort would obviate the result, described in Section 6, that there is a relatively high threshold productivity level, A_m , for a metro to have positive land area.

3.3 Individuals

Utility is nested constant elasticity of substitution. Housing services and numeraire consumption are combined in an inner nesting. Leisure, ℓ , and the housing-numeraire hybrid are then combined in an outer nesting to give actual utility,

$$\tilde{U}_j = \left(\eta_h h_j^{\frac{\sigma_h-1}{\sigma_h}} + (1 - \eta_h) x_j^{\frac{\sigma_h-1}{\sigma_h}} \right)^{\frac{\sigma_h}{\sigma_h-1}} \quad (5a)$$

$$U_j = \left(\eta_\ell (\ell_j - \ell^{min})^{\frac{\sigma_\ell-1}{\sigma_\ell}} + (1 - \eta_\ell) \tilde{U}_j^{\frac{\sigma_\ell-1}{\sigma_\ell}} \right)^{\frac{\sigma_\ell}{\sigma_\ell-1}} \quad (5b)$$

The parameters weighting housing and leisure, η_h and η_ℓ , are constrained to lie strictly between 0 and 1. Modeling leisure as entering utility in a nested CES specification follows Aguiar and Hurst (2007). Including a required minimum (i.e., a Stone-Geary specification) serves as a normalization to allow for necessities such as sleeping and eating. Equivalently, individuals' time budget could be reduced by ℓ_{min} .

Leisure is derived both from leisure time, t_j^ℓ , and from commute time, t_j^c ,

$$\ell_j = t_j^\ell + \lambda \cdot t_j^c \quad \lambda \leq 1 \quad (6)$$

The leisure component to commuting time makes it possible to match empirical estimates that, on average, individuals are willing to pay only half their wage rate to marginally shorten their commutes (Small and Verhoef, 2007). A possible explanation is that drivers enjoy listening to their radio and talking on their cell phones (hand free). *If individuals can choose their work hours*, (6) implies that they will value their commute time at $(1 - \lambda)$ times their wage. An alternative interpretation of individuals' unwillingness to pay their full wage to marginally shorten their commute is that they dislike working time more than they dislike commuting time. To the extent that this latter interpretation is correct, including a positive leisure component to commuting will overly flatten bid-rent curves.

Individuals face the budget constraint that their consumption expenditure not exceed their disposable income,

$$x_j + p_j \cdot h_j \leq w \cdot n_j - \delta \cdot d_j \cdot \text{trips} \quad (7)$$

The first right-hand-side term is simply the wage times weekly work hours. The second right-hand-side term represents weekly commute costs, which are the product of the cost per mile times the distance of each one-way trip times the number of weekly trips.

Similarly, individuals face the weekly time constraint that work time plus commute time plus leisure time not exceed total weekly hours.

$$n_j + t_j^c \cdot \text{trips} + t_j^\ell \leq 24 \cdot 7 \quad (8)$$

Under the baseline set of assumptions, weekly work hours are required to be supplied inelastically. Specifically, individuals in all rings of all metros work $n_j = \hat{n}$ hours per week. Abstracting from household labor force participation, a fixed number of work hours is likely to be more realistic than allowing individuals to offset longer commutes with shorter work hours. Under the latter assumption, the marginal valuation of leisure will be identical within a metro, regardless of commute time. If there is no leisure component to commuting, reduced work time offsets commute time one for one.

3.4 Commuting

Individuals drive directly from the outer perimeter of their residential ring to the outer perimeter of the CBD, whatever their circumference location. The model thus abstracts from the number and placement of highway rays and the arterial segments of commutes. Having commutes end at the perimeter of the CBD lessens the commute distance driven at the most congested speeds and so minimizes the sensitivity of outcomes to the calibration of CBD radius.⁷

As described above, the numeraire cost of a one-way commute is the per-mile cost, δ , times the distance from the outer perimeter of one's residential ring to the edge of the CBD, d_j . The per mile cost is calibrated to achieve a targeted mean ratio of commute cost relative to wage income in the representative metro. Distance is simply the sum of the ring widths required to reach the perimeter of the CBD. More realistically, numeraire commute costs also depend on congestion: sitting in traffic uses more gasoline per mile than driving at a free-flow speed. Including a numeraire cost to congestion would steepen the bid rent curve across rings near the CBD, which is where speeds are slowest.

A baseline assumption is that commute speed endogenously decreases with the volume of traffic. Let V_j denote the volume of commuters passing through residential ring j during each daily commute,

$$V_j \equiv \sum_{i=j}^J POP_i$$

The speed through any residential ring depends on highway capacity through that ring, \tilde{V}_j , and free-flow traffic speed, \hat{s}^f , according to the following standard formula subject to minimum and maximum limits (Small and Verhoef, 2007) ,

$$\frac{1}{\tilde{s}_j} = \frac{1}{\hat{s}^f} \cdot \left(1 + a \cdot \left(\frac{V_j}{\tilde{V}_j} \right)^b \right) \quad a, b > 0 \quad (9a)$$

$$\tilde{s}_j = \min \left(\max \left(\tilde{s}_j, s^{min} \right), s^{max} \right) \quad (9b)$$

⁷The requirement that commutes begin at the outer perimeter of residential ring is for technical reasons. Alternatively assuming that households begin their commutes from the interior of a ring makes it more difficult to quantitatively solve for the number of rings and the width of the outermost one.

Highway capacity is assumed to endogenously depend on commute volume according to the specification,

$$\tilde{V}_j = \hat{V} \cdot \left(\frac{V_j}{\hat{V}} \right)^{\sigma_V} \quad 0 \leq \sigma_V \leq 1 \quad (10)$$

The term \hat{V} is an exogenously specified value at which road capacity volume exactly equals actual commute volume. The term σ_V is the elasticity at which commute capacity is adjusted to meet demand. Parameterizing σ_V to equal 1 is equivalent to assuming that speed is constant. Parameterizing σ_V to equal 0 is equivalent to assuming that road capacity is constant. While a fixed highway capacity through a given residential ring may be realistic in an intermediate-time context, it is difficult to rationalize in a static model meant to capture long-run outcomes (a temporal statement). Similarly, it is difficult to rationalize an assumption that highway capacity does not increase with the volume of traffic as commuters approach the CBD (a location statement). Intermediate values of σ_V imply that speed decreases with commute volume.

The time to commute through any ring, \tilde{t}_j^c , is the inverse of the speed through that ring, $\frac{1}{s_j}$. Total commute time is the sum of the times required to pass through each residential ring, from the innermost one through ring one in which one lives, plus a ring-invariant fixed component,

$$t_j^c = \sum_1^j \tilde{t}_j^c + \hat{t}^c \quad (11)$$

4 Baseline Calibration

Key baseline parameters, which apply to all metros, and calibration targets, which apply only for the representative metro, are shown in Table 3.⁸ An extensive discussion on the robustness of outcomes to these is the subject of Section 7.

4.1 Population and Geography

The population and land area of the representative metro are respectively assumed to be 2.2 million and 700 square miles. These are the rounded average values for the functional St. Louis and Pittsburgh metro areas, which together span the 48.9 through 51.1 percentiles of aggregate U.S. metro population in 2000 as experienced by residents and so include the “median person” (second column of Table 1).⁹ The fixed CBD radius and the span of occupancy are respectively assumed to be 5 miles and 90°. Together these parameters calibrate the share of representative metro land area occupied by the CBD and the distance of the outermost commute. For the present case, these are respectively 2.8 percent and 24.9 miles. The latter outbound commute consists

⁸The baseline calibration herein is superseded by that in “Moncentric City Redux,” Federal Reserve Bank of Kansas City RWP 14-09.

⁹A revised calibration will likely use the mean values of the intermediate sized metros listed in Table 1. The corresponding representative metro population would remain 2.2 million but the representative land area would be 550 square miles rather than 700 square miles.

of driving through five inner rings with width 2 miles each, four intermediate rings with width 3 miles each, and an outermost ring with width 2.9 miles. The assumed span of occupancy is small compared to quantitative monocentric models, which typically assume a full 360° span.¹⁰ But for the nine intermediate-sized metros enumerated in Table 1, Saiz (2010) estimates that an average of 25 percent of land area within a 30 mile radius of the CBD is not developable due to geographic constraints such as steep slopes and bodies of water. Within land area that can be feasibly developed, suburban population density tends to be clustered along major highways, which have been a main conduit for suburbanization (Baum-Snow, 2007a, 2007b). For example, in St. Louis tracts with density above 500 persons per square mile approximately occupy only 120° around the CBD (Figure 1). In Pittsburgh, the corresponding span of occupancy is approximately 180° (Figure 2).

4.2 Utility

The utility specification, (5a) and (5b), requires the calibration of five parameters. The elasticity of substitution between housing and the numeraire good, σ_h , is set to 0.85. This is the approximate maximum likelihood point estimate based on median annual housing expenditure shares and median gross rents in 2000 across 50 large metropolitan areas using data constructed by Davis and Ortalo-Magné (2011).¹¹ Several studies using micro data suggest that σ_h may be considerably lower. Based on results from subsidizing rents for treatment groups of low income households, Hanushek and Quigley (1980) report point estimates for σ_h of 0.45 from an experiment in Phoenix and 0.64 from an experiment in Pittsburgh. Using simulated method of moments applied to a structural model of life-cycle housing consumption, Li et al. (2012) estimate σ_h to be 0.32.

The elasticity of substitution between leisure and the housing-numeraire hybrid, σ_ℓ , is set to 0.35. The criteria for this choice was to achieve a mean compensated elasticity of *desired* weekly work hours, n_j^* , with respect to wages in the representative metro of approximately 0.20, which is the value suggested by a comprehensive survey of empirical literature reported in Reichling and Whalen (2012). This targeted Frisch elasticity applies to the intensive margin only and so does not include the effect of wages on labor force participation. A baseline assumption is that actual labor hours, n_j , are fixed and so the elasticity of *observed* labor hours with respect to the wage is zero.

The weighting on housing, η_h , is calibrated such that the mean housing share of consumption expenditures in the representative metro, $\mu \equiv (p \cdot h)/(x + p \cdot h)$, equals 18 percent. This is consistent with housing expenditure shares reported in the U.S. consumer expenditure survey and in U.S. NIPA accounts but is on the low side compared to the range of values across metros reported in Davis and Ortalo-Magné (2011).

The weighting on leisure, η_ℓ is calibrated such that the mean number of desired hours in the

¹⁰An exception is Arnott and MacKinnon (1977b), who assume a semicircular metro.

¹¹Excluding San Jose, which is a considerable outlier, considerably lowers the maximum likelihood estimate of σ_h . So too does using contract rents, which exclude imputed utility expenses.

Description	Notation	Value/Target	Rationale
Population & Geography			
Population	POP	2.2 million	median U.S. metro resident
Land Area	L	700 sq.mi.	median U.S. metro resident
CBD Radius	d_0	5 mi	arbitrary
Span of Settlement	θ	90°	s.t. outer commute ≈ 25 mi
Utility			
CES, h and x	σ_h	0.85	Davis and Ortalo-Magne (2011); Li, Haiyong, Yang, Yao (2012)
CES, h - x hybrid and ℓ	σ_ℓ	0.35	s.t. mean Frisch elasticity ≈ 0.20 (Reichling and Whalen, 2012)
weight on housing	η_h	$\overline{\left(\frac{p \cdot h}{x + p \cdot h}\right)} = 0.18$	Consumer Expenditure Survey
weight on leisure	η_ℓ	$\bar{n}^* = \hat{n}$	s.t. mean desired = required
min wkly leisure	ℓ^{min}	70 hrs	normalization
Production			
housing			
CES, L and K	σ_L	0.85	literature survey
weight on land	η_L	$\overline{\left(\frac{r^L \cdot L}{r^K \cdot K + r^L \cdot L}\right)} = 0.35$	Davis and Heathcote (2007)
numeraire			
land factor share	α_L	0.016	Jorgenson et al.
capital factor share	α_K	0.328	Jorgenson et al.
req wkly wkr hrs	\hat{n}	40	Robinson and Godfrey (1999); Aguiar and Hurst (2007)
Commuting			
wkly one-way commutes	trips	10	
per mile cost	δ	$\frac{\delta}{w \cdot \bar{n}} = 0.0004$	s.t. $\overline{\left(\frac{\delta \cdot d \cdot \text{trips}}{w \cdot \bar{n}}\right)} \approx 0.05$ (Albouy and Lue, 2014)
leisure content	λ	0.50	Small and Verhoef (2007)
speed			
free/max/min	$\hat{s}^f / \hat{s}^{max} / \hat{s}^{min}$	75/65/10 mph	arbitrary
fixed time component	\hat{t}^c	10 min	s.t. mean commute ≈ 30 min
road capacity			
benchmark capacity	\hat{V}	750 ths	arbitrary
elasticity w/ volume	σ_V	0.80	arbitrary

Table 3: Calibrated Values and Targets for the Representative Metro. A bar decoration denotes a population-weighted mean. The implied values for the structural parameters, η_h , η_ℓ , η_L , and δ , are used to calculate outcomes in open metros.

representative metro, \bar{n}_{rep}^* , is 40. This is consistent with average hours for working males in 1985 and 2003 respectively reported in Robinson and Godbey (1999) and Aguiar and Hurst (2007). In any case, representative and open-metro outcomes are relatively insensitive to the calibration of η_ℓ . Much more important is the relationship between desired work hours and required work hours, \hat{n} . The latter is set to equal \bar{n}_{rep}^* .

4.3 Production

Assumed Cobb Douglas production of the numeraire good requires setting the factor income shares accruing to land and to capital, α_L and α_K . The factor income share for labor is determined residually. The land share is assumed to be 1.6%. This value is a weighted average across a large number of industries using intermediate input shares estimated by Jorgenson, Ho, and Stiroh (2005).¹² It is nearly identical to the 1.5% land share that Ciccone (2002) suggests is reasonable for the manufacturing sector. One third of remaining factor income is assumed to accrue to capital; two thirds are assumed to accrue to labor (Gollin, 2002).

Production of housing services requires calibrating the elasticity of substitution between land and structure, σ_L , and the relative weight on land, η_L . The former is set to 0.85, which is meant to balance the wide range of empirical estimates. A survey by McDonald (1981) reports preferred estimates from twelve different studies ranging from 0.36 to 1.13. Updating this research, Jackson, Johnson, and Kaserman (1984) estimate the elasticity to lie somewhere between 0.5 and 1. More recently, Thorsnes (1997) argues that a unitary elasticity of substitution cannot be rejected. The weight on land is calibrated such that the household-weighted mean share of housing factor income in the representative metro, $\nu \equiv (r^L \cdot L)/(r^L \cdot L + r^K \cdot K)$, equals 0.35 as suggested in Davis and Heathcote (2007).¹³

4.4 Commuting

The number of weekly one-way trips under the baseline calibration is assumed to be 10. The quantitative implications of making fewer weekly trips are described in Section 7. The per mile commute cost (which applies in all metros), δ , is calibrated to equal 0.04 percent of weekly wage income in the representative metro. This target is chosen so that the implied mean weekly commute cost in the representative metro approximately equals 5 percent of wage income there, which is the estimated value reported in Albouy and Lue (2014).¹⁴

The leisure content of commuting time, λ , is set to 0.5. If individuals choose their work hours,

¹²The industry-specific intermediate input estimates, which are not included in the publication, were kindly provided by the authors.

¹³Davis and Heathcote find that between 1975 and 2004, land accounted for an average of 47 percent of the sales value of the aggregate U.S. housing stock. Adjusting for the fact that structures depreciate but land does not brings the land share down to approximately 35 percent.

¹⁴The per mile numeraire cost is not explicitly targeted to match a mean share of wage income because doing so makes it significantly more difficult to get solutions.

this calibration implies that all individuals across and within metros are willing to pay at exactly 50 percent of their wage rate to marginally shorten their commute. This marginal valuation of commute time matches the benchmark estimate reported in Small and Verhoef (2007). Under the baseline assumptions, however, workers are required to supply a fixed number of weekly hours that is calibrated to match mean desired hours in the representative metro ($\hat{n} = \bar{n}_{rep}^*$) and so most workers' willingness to pay will not exactly equal 50 percent. But $n_{m,j}^*$ remains close to \bar{n}_{rep}^* across and within metros that have TFP similar to or below that of the representative metro. Residents of these will have a marginal willingness to pay close to 50 percent of their wage (whatever w_m may be). Similarly, individuals living in the inner rings of metros with TFP significantly higher than that of the representative metro have $n_{m,j}^*$ close to \bar{n}_{rep}^* and so value marginal commute time at close to 50 percent of their wage.

For the equation determining commute speeds, (9b), the technical parameters, a and b are respectively set 0.2 and 10, which are standard values for highway travel (Small and Verhoef, 2007). "Free flow" speed, the speed at which traffic is *capable* of flowing absent any congestion is set to 75 mph. This implies that commute speed when actual commute volume through a ring equals assumed highway capacity will be 62.5 mph. Commuters are law abiding and so their maximum actual speed is set to 65 mph. The minimum commute speed is set to 10 mph. To urban commuters accustomed to sitting in gridlock traffic, this minimum might seem overly optimistic. But in metros where 10 mph commuting traffic is typical, other transportation modes such as walking and subways are likely to be (endogenously) available.¹⁵ Gradients and total population of open metros with TFP at least moderately above that of the representative metro are especially sensitive to the assumed minimum speed.

The benchmark level at which assumed highway capacity matches actual highway volume, \hat{V} , and the elasticity with which (long-term) highway capacity adjusts to commute volume, σ_V , are arbitrarily set to 750 thousand and 0.80. Figure 4 shows the implied relationship between commute speed and commute volume. Also shown is the implied speed for a lower assumed elasticity, $\sigma_V = 0.65$. Outcomes under this lower elasticity are described in Section 7.

5 Baseline Quantitative Results

The first subsection below describes how population density, land prices, house prices, and house expenditure vary within a system composed of one representative metro, one low-TFP metro, and one high-TFP metro. The second subsection then measures the fit of the calibrated model to observed metro outcomes.

The differences in open-metro productivity are modeled as exogenous, which superficially contrasts with the large urban economics literature on agglomeration. In a deeper sense, however, the exogenous modeling of productivity complements models of agglomeration. From a quanti-

¹⁵Modeling multi-modal commuting is a priority for future research.

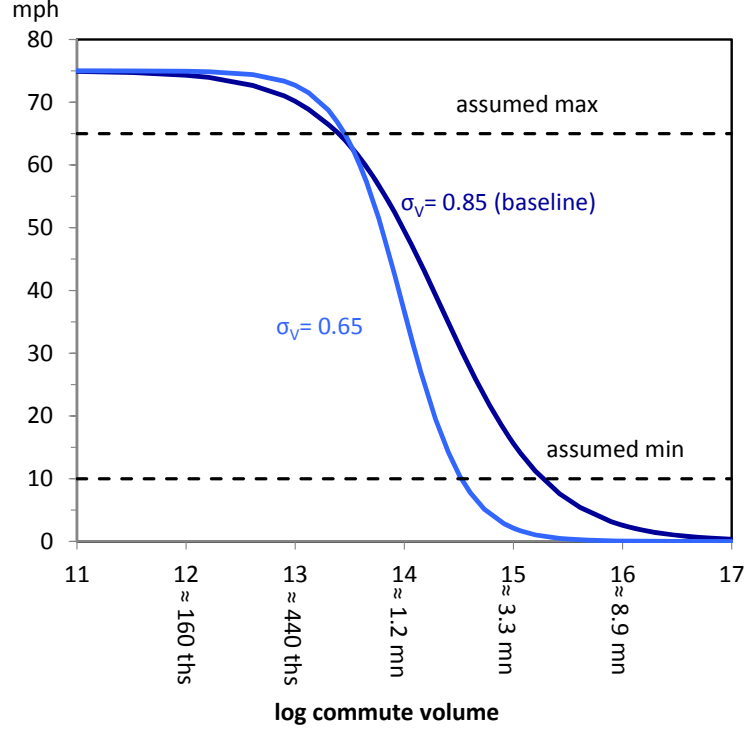


Figure 4: Commute Speed, Commute Volume, and Highway Capacity. Figure shows commute speed as a function of commute volume when highway capacity is itself assumed to be a function of commute volume. The curves are drawn for two alternative elasticities. The baseline calibration sets σ_V to 0.80. An alternative calibration, discussed in Section 7, sets σ_V to 0.65. In both cases, capacity is assumed to equal volume at a daily level of 750 thousand commuters.

tative perspective, the exogenous productivity levels should be interpreted as what is required to support observed differences in population, land area, and population density. In the real world, such required differences almost surely arise from non-agglomerative metro attributes multiplied by agglomerative ones. The production functions for numeraire and housing can easily be generalized to allow for endogenous components to productivity.¹⁶

5.1 A System of Three Metros: Representative, Low-TFP, and High-TFP

Comparing outcomes within and across the three metros yields a number of quantitative and qualitative insights. As stressed in the introduction, caution is required in interpreting the three metros below as comprising a system. The three satisfy all general equilibrium criteria for mutually consistent outcomes in the presence of perfect mobility and linked perimeter land prices. But the

¹⁶In a quantitative framework without internal metro structure, Rappaport (2008b) illustrates how agglomerative TFP multiplies differences in consumption amenities.

	Low-TFP Open Metro $A_{low} = 0.92$		Representative Metro $A_{rep} = 1$		High-TFP Open Metro $A_{high} = 1.16$	
Wage (relative)	0.91		1.00		1.21	
Population	510,000		2,200,000		6,890,000	
Radius						
cbd metro	5.0 mi	18.4 mi	5.0 mi	29.9 mi	5.0 mi	36.9 mi
Land Area						
cbd metro	20 sq.mi	265 sq.mi	20 sq.mi	700 sq.mi	20 sq.mi	1,070 sq.mi
Metro Pop Density (pop-wghtd mean)	2,300 pop/sq.mi		4,700 pop/sq.mi		20,600 pop/sq.mi	
Res Ring Pop Density max min	3,700 pop/sq.mi	1,300 pop/sq.mi	9,900 pop/sq.mi	1,200 pop/sq.mi	50,400 pop/sq.mi	1,000 pop/sq.mi
Gradients (fitted)						
density p	-0.093	-0.037	-0.090	-0.038	-0.128	-0.058
r^2 p-h	-0.114	-0.009	-0.110	-0.009	-0.153	-0.011

Table 4: Summary Metro Outcomes. Italics denote directly calibrated values. (CBD land areas follow immediately from the assumed 5 mile cbd radius and 90° span of occupancy.) Gradients are measured as semi-elasticities based on a regression of the log of the outcome on the distance from the perimeter of the CBD.

combined population of the three metros is determined residually.

The low-productivity open metro is assumed to have numeraire TFP that is 0.92 times that of the representative metro. The high-productivity open metro is assumed to have numeraire TFP that is 1.16 times that of the representative metro. Under the baseline calibration, these imply respective wages 0.91 and 1.21 times the wage in the representative metro. This range of wage variation is well within that estimated in Albuoy (2009) after controlling for observable human capital. As described in the previous section, the representative metro is assumed to have 2.2 million residents and occupy 700 square miles. The endogenous population and land area of the low-TFP metro are 510,000 and 275 square miles. The same for the high-TFP metro are 6.9 million and 1,100 square miles. Table 4 summarizes these and other aggregate metro outcomes.

Population density, house price, land price, and housing expenditure gradients under the baseline calibration and assumptions are shown in Figure 5. The horizontal axis in each panel measures the distance from the perimeter of each residential ring to the border of the CBD. For the representative metro, the assumed 700 square mile total land area, 90° span of occupancy, and 5-mile CBD radius together imply an outer commute distance of 24.9 miles. The low-TFP and high-TFP metros are similarly assumed to have a 90° span of occupancy, and 5-mile CBD radius. The respective endogenous outer commutes are 13.4 miles and 31.9 miles.

Within each of the four panels, the low-productivity gradient is primarily a downward shift of the representative gradient together with a shorter outermost commute. Each of the high-

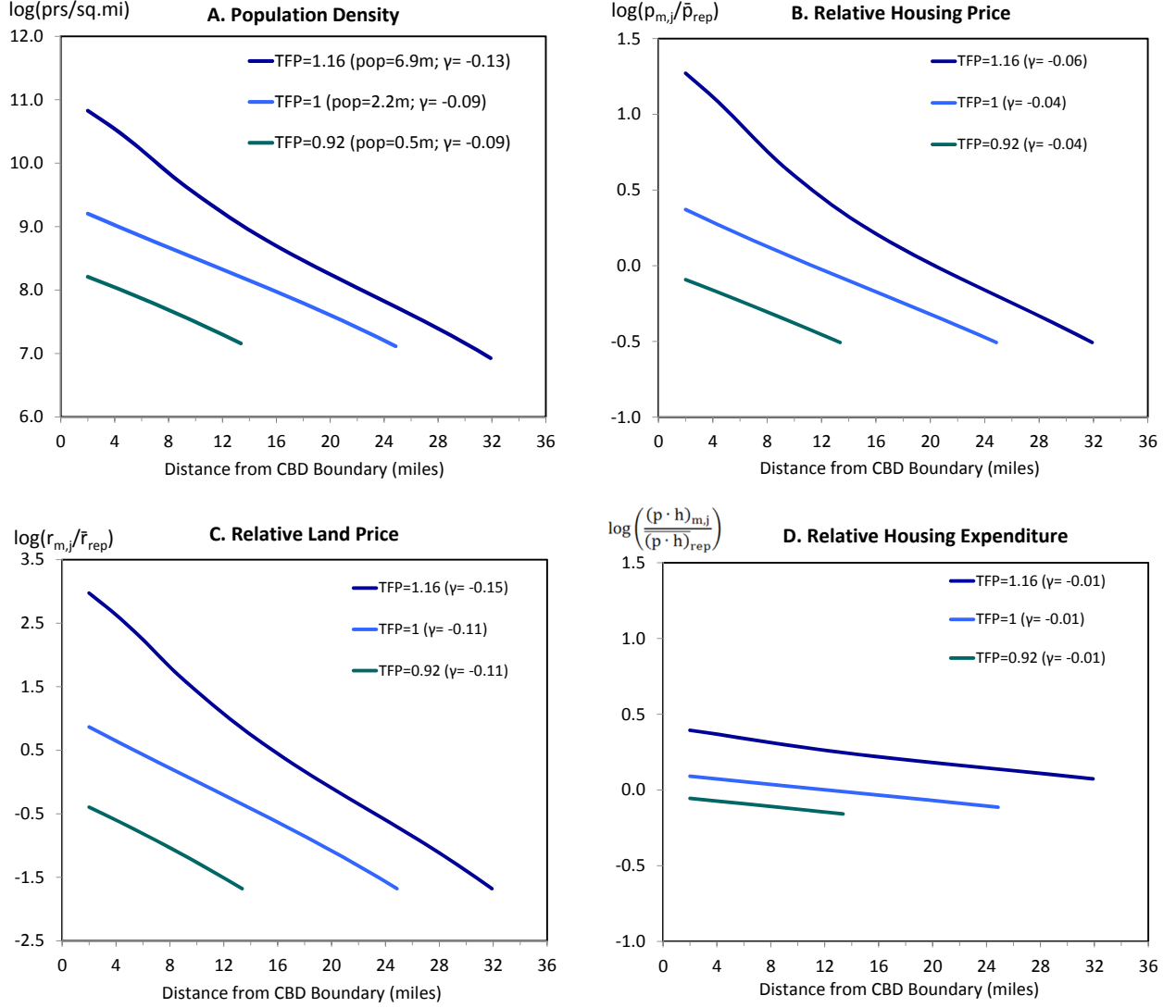


Figure 5: Population Density and Price Gradients. Figure shows log population density, log relative prices, and log total housing expenditure for each residential ring under the baseline calibration and set of assumptions. The middle line in each panel corresponds to the representative metro. The lower and upper lines respectively correspond to open metros with low and high relative TFP in the production of the numeraire good. Metro population and fitted gradients are reported in parentheses.

productivity gradients is primarily an upward shift of the respective representative gradient together with a longer outermost commute. At closer commute distances, each higher productivity gradient is also at least moderately steeper than the corresponding representative metro gradient. This is especially true for the house price gradient, reflecting slow commute speeds at near distances and so large incremental increases in commute time from living one additional residential ring away from the CBD. By assumption the low- and high-TFP metros have outermost-ring *land* prices equal to that of the representative metro (Panel C: identical lowermost point of the three gradients). By implication, outermost-ring *house* prices are equal across metros because both input prices are the same (Panel B).

For each metro type in each of the four panels, gradients are always steepest for land prices and next steepest for population density. For example the representative metro has land price gradient, γ_r , and population density gradient, $\gamma_{density}$, respectively equal to -0.11 and -0.09. The representative-metro house price gradient is considerably flatter ($\gamma_p = -0.04$). Its more gradual falloff reflects households' willingness to substitute from housing to numeraire consumption and developers' ability to substitute from land to structure inputs. The housing expenditure gradients, which correspond to *observed* rents and sales prices, are nearly flat. This reflects the assumed near unitary elasticity of substitution between housing services and numeraire consumption in utility ($\sigma_h = 0.85$). The small downward slope that remains primarily arises from the decrease in disposable income associated with the per mile commute cost.

Ring segment commute speeds, \tilde{s}_j , and one-way commute times, t_j , are shown in Figure 6. For the commute speeds, it is helpful to think of the inbound commute (moving from right to left in Panel A). In the representative metro, speed through the outermost 9 miles of the commute, which correspond to the outermost three residential rings, is at its maximum value of 65 mph. The "entry" of additional commuters at the perimeter of the fourth-from-outer residential ring increases volume relative to highway capacity sufficiently to bring representative-metro commute speed down to 59 mph. Moving inward toward the CBD, commute volume continues to increase faster than does highway capacity thereby slowing commute speed. Across the innermost residential ring of the representative metro, speed has slowed to 28 mph. In the low-TFP metro area, volume remains sufficiently below capacity for speed to be at its maximum value for the entire commute. In the high-TFP metro area, commute speed is at its maximum for the outer 7 miles of the commute. Continuing toward the CBD, speed rapidly slows. Speed during the final 6 miles is at its assumed floor of 10 mph.

Of particular importance for one-way commute times is the variation in commute speeds across the three metros *at identical distances from the CBD*. For the low-TFP and representative metros, innermost-ring commutes are respectively 12 and 14 minutes, most of which is accounted for by the assumed 10 minute fixed component. In the high-TFP metro, the innermost commute time is 22 minutes. More generally, longer relative commute times in the high-TFP metro primarily arise from slower commute speeds rather than commutes from further away. For example, the one-way

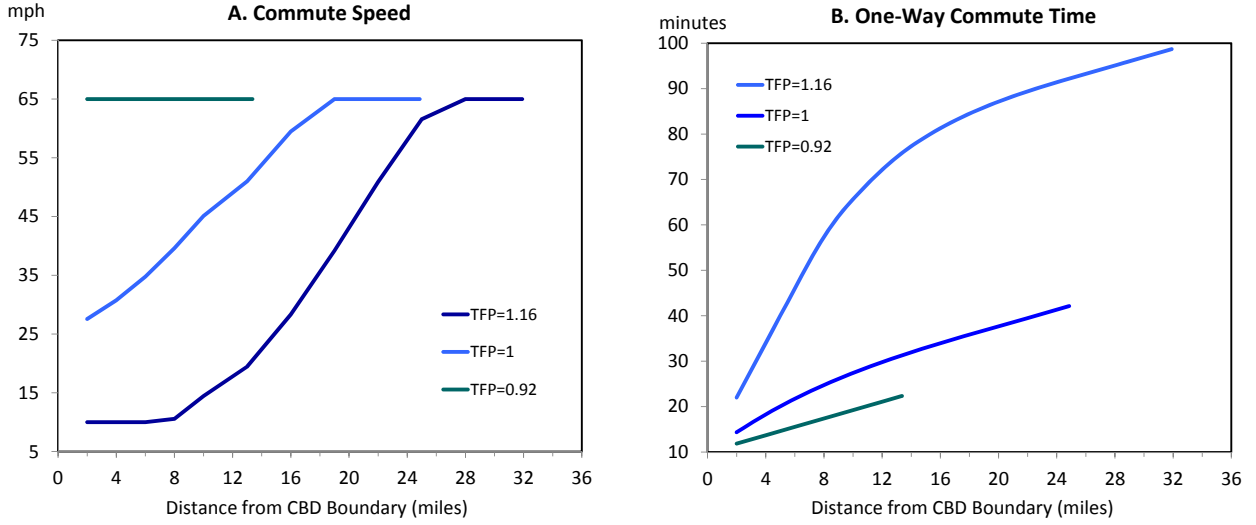


Figure 6: Commute Speeds and Time. Figure shows commute speed through each residential ring and one-way commute times.

commute from a distance 10 miles from the CBD takes 27 minutes in the representative metro but 66 minutes in the high-TFP metro.

Gradients for numeraire consumption, housing consumption, and leisure time are shown in Figure 7. Numeraire and housing consumption values are normalized by the mean level of consumption in the representative metro, \bar{x}_{rep} and \bar{h}_{rep} .

Numeraire consumption at any given commuting distance is always highest in the high-TFP metro, reflecting higher numeraire wages there (Panel A). In each metro, the numeraire consumption gradient slopes downwards. In part this reflects the hit to disposable income from numeraire commuting costs. For example, disposable income (total income minus numeraire commute costs) of someone living in the outer ring of the representative metro is 9 percent below that of someone living in the innermost ring of the representative metro. In part the downward slope reflects substitution from numeraire to housing consumption as the price of housing decreases.

Conversely, housing consumption at any given commuting distance is highest in the low-TFP metro (Panel B). In each metro, the housing consumption gradient slopes upward. As is intuitive, this increase in housing consumption with commute distance is driven by the steep falloff in the price of housing and the associated income and substitution effects. These dominate the negative effect from the falloff in disposable income. Because σ_h is calibrated to be less than 1, the housing share of consumption expenditure, $\mu_{m,j}$, falls with the price of housing and so with commute distance (not shown). For the representative metro, this decrease is moderate: from 18.9 percent in the innermost ring down to 17.0 percent in the outer ring. (The baseline calibration targets $\bar{\mu}_{rep}$

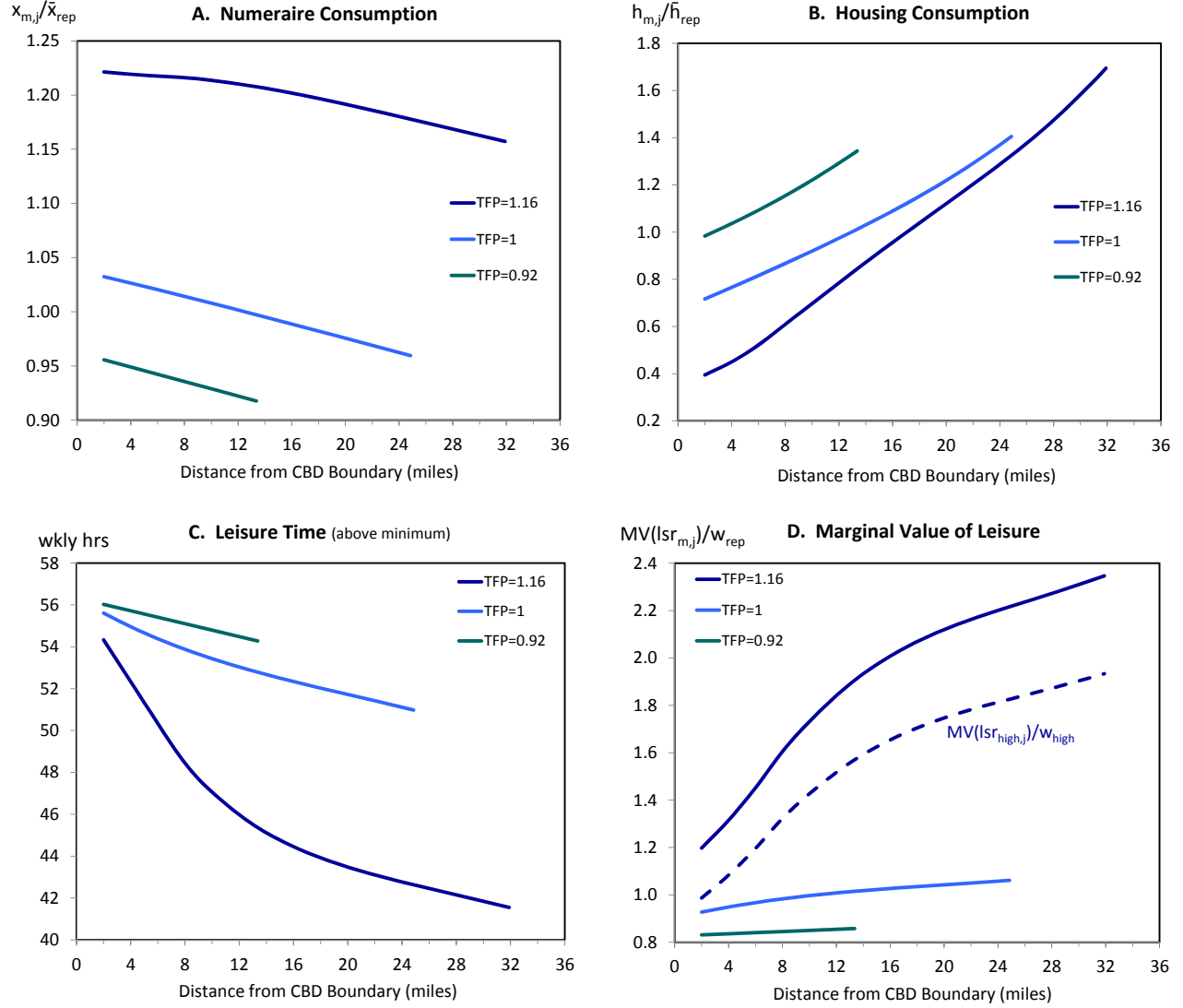


Figure 7: Consumption Gradients. Panels A and B show numeraire and housing consumption relative to the mean value of each in the representative metro. Panel C shows weekly leisure time above its assumed minimum value of 70 hours. Panel D shows the marginal value of leisure and leisure time relative to the wage in the representative metro (solid lines) and relative to the wage in the high-tfp metro (dashed line).

to be exactly 18 percent). In the high-TFP metro, the decrease in μ is larger: from 21.1 percent in the innermost ring down to 17.0 percent in the outer ring.

Leisure time at any commuting distance, t_j^ℓ , is always highest in the low-TFP metro and lowest in the high-TFP metro (Panel C). Leisure time falls off especially sharply in the high-TFP metro due to the traffic congestion there. Because commuting time is calibrated to include a 50 percent leisure component, total leisure, ℓ_j , falls off only half as steeply as does leisure time (not shown).

The solid-line gradients in Panel D show individuals' marginal valuation of leisure and leisure time normalized by the wage in the representative metro.¹⁷ If individuals could work their desired hours, the marginal valuation of leisure would be pinned down by a metro's wage. In consequence, total leisure and its marginal valuation would be equal within metros. Each of the displayed gradients would then be exactly horizontal. The gradient for the representative metro would be horizontal at a value of 1. But with hours required to be supplied inelastically, leisure decreases and its marginal valuation increases with commute time and so the gradients in Panel D are upward sloping. The lower marginal valuations of leisure in the low-TFP metro and higher valuations in the high-TFP metro in part reflect differences in numeraire wages and hence differences in $n_{m,j}^*$ for a given commute time. In addition, commute times at any given commute distance are shortest in the low-TFP metro and longest in the high-TFP metro, which further pushes apart marginal valuations.

At commute distances of 19 miles or higher, the marginal valuation of leisure in the high-TFP metro is at least twice that at the same distances in the representative metro. These large differences in marginal valuation occur despite the baseline assumption that commute time includes a 50 percent leisure component. The dashed-line gradient in Panel D shows the marginal valuation of leisure in the high-TFP metro relative to the high-TFP wage. The downward shift from the solid high-TFP line reflects the higher denominator wage. At the outermost commute distance (which corresponds to a commute time of 99 minutes), the marginal value of leisure is almost twice the high-TFP wage. Individuals making such a commute would be willing to pay 1.5 times their wage rate to marginally shorten it.¹⁸

A more comprehensive set of quantitative outcomes within and across each of the metros is enumerated in Table 5. Italicized numbers represent moments that are directly targeted in the calibration or else immediately implied by it.¹⁹ Flagging these helps distinguish between assumptions

¹⁷The marginal valuation of leisure is equivalent to $(\partial U/\partial \ell)/(\partial U/\partial x)$. The numerator also equals the marginal value of leisure time because $\partial \ell/\partial t^c = 1$. Individuals' desired number of hours, n_j^* , is determined by equating the marginal utility of leisure time divided by its price (i.e., the wage) to the marginal utility of numeraire consumption.

¹⁸For comparison, Small, Winston, and Yan (2005) estimate that freeway commuters in Orange County CA have a median marginal value of commute time that is approximately equal to their wage. If commute time indeed included a 50 percent leisure component, they would have a marginal value of leisure 1.5 times their wage. For the calibrated high-TFP metro, a one-way commute from the sixth residential ring (13 miles, 55 minutes) has a marginal value of leisure equal to 1.5 times the high-TFP wage and so a marginal valuation of commute time approximately equal to the high-TFP wage. Small et al. note that commuters using the specific freeway segment on which they base their estimate typically have long commutes.

¹⁹An example of a moment that is "immediately implied" is the mean unitary marginal value of leisure relative to

and results. Four outcomes particularly stand out.

First, the land share of housing factor income, $\nu_{m,j}$, varies considerably across and within metros (first horizontal block, top row). Within each metro, ν_j decreases moving outward from the CBD as the price of land decreases. Across metros, the mean land income share, $\bar{\nu}_m$ is highest in the high-TFP metro, where the average price of land is highest. Both of these patterns follow directly from the assumed complementarity of land and structure in housing production. More surprising is the broad quantitative range of the land factor share: from 0.30 in the outer ring of all three metros up to 0.47 in the innermost residential ring of the high-TFP metro. This wide range occurs notwithstanding the relatively high assumed elasticity of substitution between land and structure ($\sigma_L = 0.85$). The “explanation” is the extremely wide variation in land prices within and across metros. In the high-TFP metro, for example, the price of land in the innermost residential ring is 100 times the price of land in the outermost residential ring (fifth horizontal block, middle row).

Second, the price elasticity of housing supply, ϵ_h^s , also varies considerably across and within metros (first horizontal block, bottom row). Within the high-TFP metro, ϵ_h^s increases from 0.97 in the innermost residential ring up to 1.96 in the outermost ring. Across metros, the mean housing supply elasticity, $\bar{\epsilon}_h^s$, ranges from 1.27 in the high-TFP metro up to 1.78 in the low-TFP metro. Across the three metros, ϵ_h^s is identical in the outermost ring, reflecting identical land and house prices. More generally, the endogenous variation in ϵ_h^s reflects the endogenous variation in land prices which in turn follows from the assumed sharply decreasing marginal product of structure in housing production.

For comparison, several empirical studies estimate that *across metro areas*, ϵ_h^s ranges from close to zero up to at least 3 (Green, Malpezzi, and Mayo, 2005; Saiz, 2010; Davidoff, 2013; Wheaton et al., 2014). The quantitative range under the baseline calibration thus falls in the middle of this estimated range. Achieving a near-zero supply elasticity, which estimates suggest characterizes some especially crowded U.S. metros, would likely require the imposition of land-use constraints such as those described in Glaeser, Gyourko, and Saks (2005). In addition, the static nature of the current model upwardly “biases” supply elasticities. In a dynamic context, acquiring and tearing down existing houses to make room for new construction can be expensive, even without land-use restrictions.

Third, the price elasticity of housing demand, ϵ_h^d , varies relatively little across and within metro areas (second horizontal block, bottom row). This reflects the relatively high calibrated elasticity of substitution between housing and numeraire ($\sigma_h = 0.85$) together with the endogenous variation of house prices across and within metros that is an order of magnitude smaller than the variation in land prices (fifth horizontal block, top row).

Fourth, the “Frisch” compensated elasticity of desired work hours with respect to wages, $\epsilon_w^{n^*}$,

the wage in the representative metro, $\overline{((\partial U / \partial \ell) / (\partial U / \partial x)) / w_{rep}}$. The direct calibration target is that the required number of work hours \hat{n} equals mean desired hours in the representative metro, \bar{n}_{rep}^* . A first order condition for determining n^* is that the marginal value of leisure equal the wage.

	Low-TFP Open Metro $A_{low} = 0.92, w_{low}/w_{rep} = 0.91$			Representative Metro $A_{rep} = 1$			High-TFP Open Metro $A_{high} = 1.16, w_{high}/w_{rep} = 1.21$		
	mean	ring 1	outer ring	mean	ring 1	outer ring	mean	ring 1	outer ring
1. HOUSE PRODUCTION									
land factor share (v)	0.32	0.34	0.30	<i>0.35</i>	0.39	0.30	0.41	0.47	0.30
capital per land (rel)	0.46	0.74	0.25	<i>1.00</i>	2.15	0.25	5.25	12.9	0.25
h srvc per land (rel)	0.62	0.86	0.41	<i>1.00</i>	1.70	0.41	2.57	4.73	0.41
supply elasticity (ε_p^s)	1.78	1.61	1.96	<i>1.59</i>	1.34	1.96	1.27	0.97	1.96
2. CONSUMPTION									
numeraire (rel)	0.94	0.96	0.92	<i>1.00</i>	1.03	0.96	1.21	1.22	1.16
housing (rel)	1.15	0.98	1.34	<i>1.00</i>	0.72	1.41	0.76	0.39	1.69
housing expnd (rel)	0.90	0.95	0.85	<i>1.00</i>	1.10	0.89	1.33	1.48	1.08
h expnd share (μ)	0.174	0.179	0.170	<i>0.180</i>	0.189	0.170	0.195	0.211	0.170
h dmnd elasticity (ε_p^d)	-0.70	-0.70	-0.71	-0.70	-0.69	-0.71	-0.68	-0.67	-0.71
3. LEISURE									
lsr hrs (wkly, above min)	55.1	56.0	54.3	53.2	55.6	51.0	48.0	54.3	41.6
lsr (wkly, above min)	56.6	57.0	56.1	55.6	56.8	54.5	53.0	56.2	49.8
MV lsr (rel to rep wage)	0.84	0.83	0.86	<i>1.00</i>	0.93	1.06	1.70	1.20	2.35
MV lsr (rel to own wage)	0.92	0.91	0.94	<i>1.00</i>	0.93	1.06	1.40	0.99	1.93
desired work hrs (wkly)	40.6 hr	40.8 hr	40.5 hr	<i>40.0 hr</i>	40.6 hr	39.5 hr	37.6 hr	40.1 hr	35.3 hr
Frisch elstcty (desired hrs)	0.19	0.19	0.19	<i>0.20</i>	0.19	0.20	0.24	0.20	0.28
4. ONE-WAY									
distance	7.7 mi	2.0 mi	13.4 mi	12.4 mi	2.0 mi	24.9 mi	11.4 mi	2.0 mi	31.9 mi
hwy segment speed*	65 mph*	65 mph	65 mph	51 mph*	28 mph	65 mph	11 mph*	10 mph	65 mph
avg commute hwy speed	65 mph	65 mph	65 mph	37 mph	28 mph	46 mph	12 mph	10 mph	22 mph
avg speed (incl fixed time)	25 mph	10 mph	36 mph	23 mph	8 mph	35 mph	10 mph	5 mph	19 mph
time	17 min	12 min	22 min	29 min	14 min	42 min	60 min	22 min	99 min
weekly cost (rel to w incm)	0.034	0.009	0.058	<i>0.050</i>	0.008	0.099	0.038	0.007	0.105
5. PRICES									
housing (rel)	0.75	0.91	0.60	<i>1.00</i>	1.45	0.60	2.05	3.57	0.60
residential land (rel)	0.39	0.67	0.19	<i>1.00</i>	2.38	0.19	7.28	19.6	0.19
CBD land (rel)	0.21			<i>1.00</i>			3.80		

Table 5: Detailed Metro Outcomes. Relative values are with respect to the population-weighted mean across the residential rings of the representative metro. Italics denote moments that are directly targeted by the calibration or else immediately implied by it. Also italicized are unitary values that represent a representative-metro outcome normalized by itself. *Highway segment speed: the starred value is the speed through the segment in which the individual with the median-commute distance lives.

is relatively constant across and within the representative and low-TFP metros. But it varies meaningfully within the high-TFP metro: from 0.20 in the innermost residential ring to 0.28 in the outermost one (third horizontal block, last row). This might seem counterintuitive given the high marginal utility of leisure of outer ring residents in the high-TFP metro. But the elasticity is measured for desired work hours rather than actual work hours. In the high-TFP metro, n_j^* ranges from 40.1 in the innermost residential ring to 35.3 in the outermost ring (third horizontal block, next-to-last row). The higher Frisch elasticity in the outer rings primarily reflects that an identical absolute increase in n^* represents a larger percentage increase when n^* is lower. The higher Frisch elasticity in the outer rings is also partly driven by the lower disposable income there, due to the numeraire per mile cost (fourth horizontal block, bottom row). Numeraire consumption thus decreases and its marginal utility increases with commute distance. Partly offsetting this is that housing consumption increases and its marginal utility decreases with commute distance. Under the baseline calibration, the higher numeraire marginal utility from numeraire consumption quantitatively dominates the lower marginal utility from housing consumption thereby causing a larger marginal increase in n^* in response to a compensated marginal increase in the metro wage.

5.2 Model Fit

As the description of results above and below is meant to illustrate, the present quantitative framework can yield considerable qualitative insight into the determination of outcomes across and within metropolitan areas. But for one to have much confidence in such insights, it is important that the calibrated model sufficiently match corresponding observed outcomes.

By construction, representative-metro mean outcomes equal their targeted values for the land share of housing factor income, the housing share of consumption, and desired work hours. By trial and error, the representative-metro mean commute cost almost exactly hits its implicit target. And as argued immediately above, the calibrated model implies a range for the price elasticity of housing supply that is within the empirically-estimated range.

The calibrated model also does a reasonably good job at matching the variation in population density and commute times across and within metros.

Figure 8 shows density percentiles along the representative and high-TFP metro gradients. For example, the right-most marker along the representative-metro gradient indicates the commute distance and population density of the household experiencing the 10th percentile population density within the representative metro. The left-most markers on the left-hand-side of the figure show the average of the St. Louis and Pittsburgh tract densities at the labeled population percentile. In other words the displayed 10th percentile marker is at average of the St. Louis 10th percentile density and the Pittsburgh 10th percentile density. These serve as benchmarks for the representative metro density gradient. Density percentiles for each of Chicago, Los Angeles, and New York City serve as benchmarks for the high-TFP metro density gradient.

The interdecile log population density range in the representative metro lies entirely within

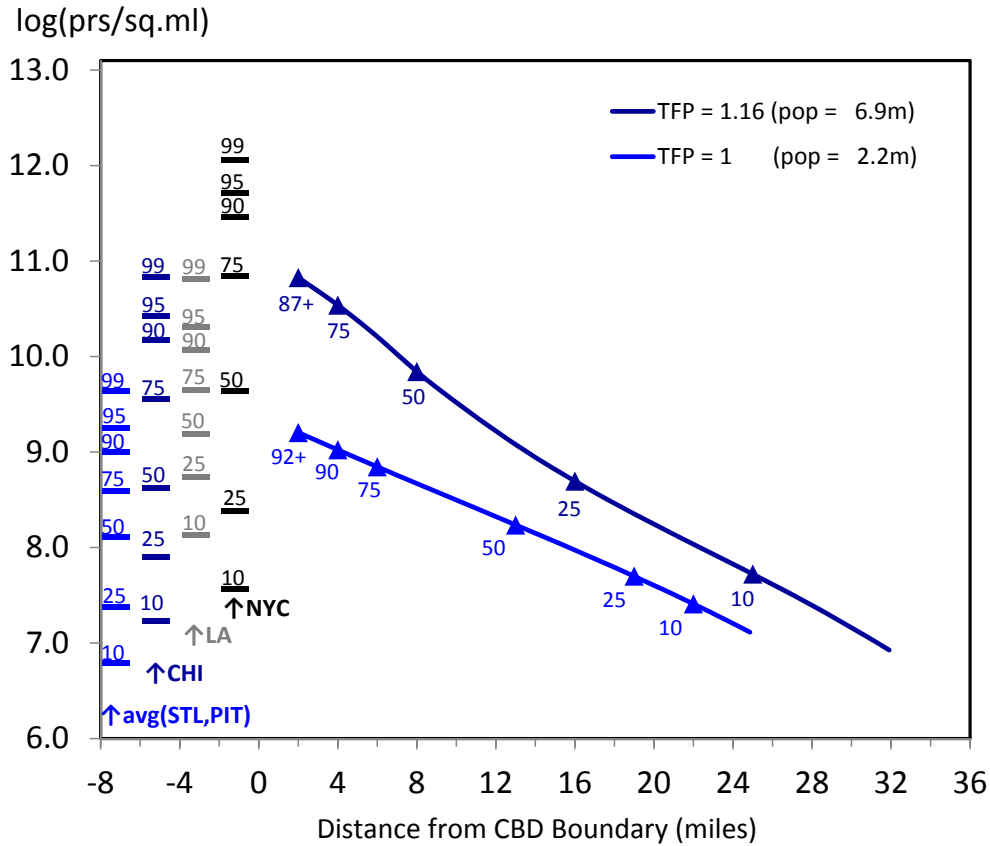


Figure 8: Quantitative versus Observed Population Density. Population density gradients on the right are for the representative and high-TFP metros. Marked density percentiles on the left are for actual metros. The average of the indicated percentile densities in Saint Louis and Pittsburgh benchmark the representative metro. The percentile densities within Chicago, Los Angeles and New York City benchmark the high-TFP metro.

and is more compact than the interdecile range of the Pittsburgh-St. Louis hybrid. This greater compactness suggests that variations in other outcomes within the representative metro may be similarly compact in comparison to the range of observed results within an intermediate-sized real metro. A possible reason is the considerable homogeneity within the quantitative model, such as ex ante identical households and the immediate adjacency of all households to radial commuter highways. The interdecile log population density range in the high-TFP metro is slightly shifted down from the corresponding inter-decile range in Chicago. Assuming a high relative TFP level of 1.19 rather than 1.16 shifts the high-TFP density gradient upwards to approximately match that of Chicago and implies a total metro population of 8.5 million, which also approximately matches that of Chicago.

Figure 9 shows analogous percentile commute times. For the representative metro, commute

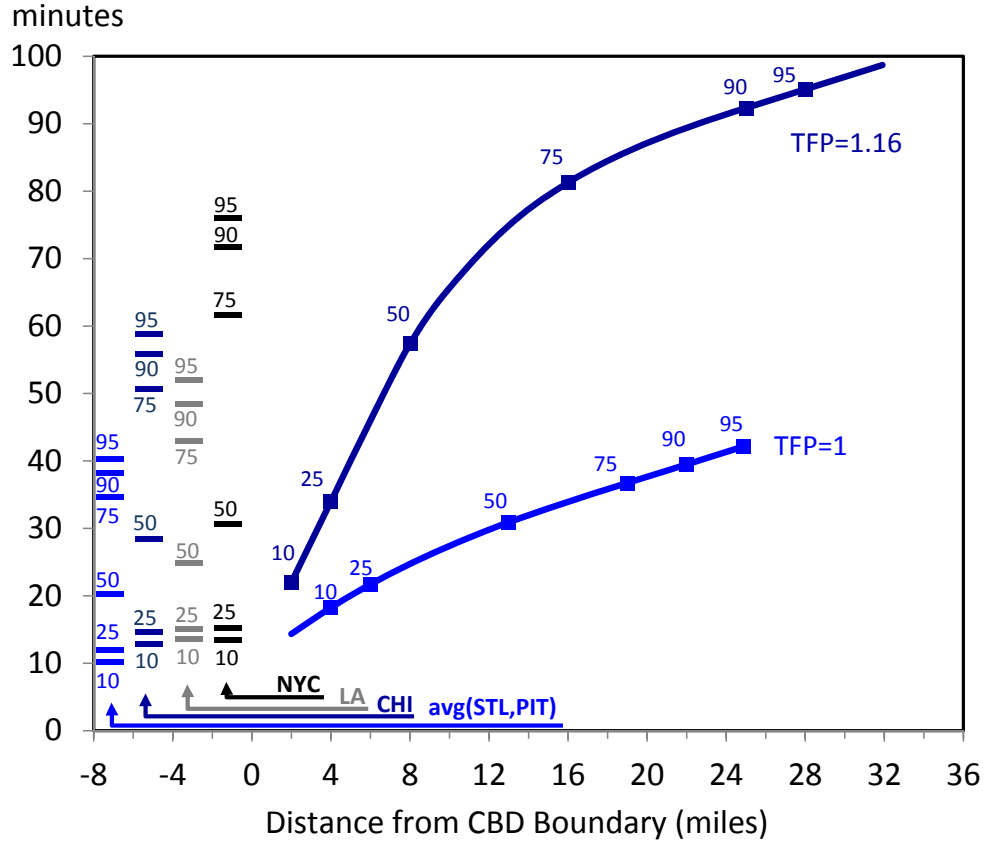


Figure 9: Quantitative versus Observed Commute Times. Commute time gradients on the right are for the representative and high-TFP metros. Marked commute time percentiles on the left are for actual metros. The average of the indicated percentile densities in Saint Louis and Pittsburgh benchmark the representative metro. The percentile commute times within Chicago, Los Angeles and New York City benchmark the high-TFP metro.

times at lower percentiles moderately exceed commute times at the same percentile averaged across St. Louis and Pittsburgh. For example, the respective modeled and observed 10th-percentile commute times are 18 minutes and 10 minutes. The shorter observed times may reflect commutes from residences to work locations close to each other but far from the CBD. They may also reflect individuals who walk to work. Recalibrating the fixed time component of commutes to be 5 minutes rather than 10 minutes should achieve a better fit. At higher percentiles, representative-metro commute times approximately match the average values across St. Louis and Pittsburgh.

Commute times in the high-TFP metro significantly exceed commute times even in the New York City metro area. For shorter-time commutes this is unsurprising. The New York metro area is especially polycentric with large portions matching the interspersed residential and commercial land use modeled in Lucas and Rossi-Hansberg (2002). Hence a large share of commutes will be

very short, both in distance and time. More surprising is that the modeled longer-time commutes significantly exceed long observed commutes in the New York City metro and elsewhere. In this case, the measurement of observed commutes partly accounts for the difference. As described in the empirical motivation section above, observed commute times at the 75th percentile and above are measured by the 75th percentile commute time *within* each census tract. In other words, the reported 90th percentile time is the 90th percentile across the tracts in a metro of the 75th percentile commute time within in each tract. Alternatively measuring the 90th percentile time within each metro by the 90th percentile time within each tract implies an observed value of 81 minutes in Chicago and at least 90 minutes in New York (the uppermost reported bin is for commute times that exceed 90 minutes). The longer high-TFP commute times approximately match this alternative observed measure.

6 The Effect of Productivity on Aggregate and Mean Metro Outcomes

The previous section focused on how productivity affects outcomes both across and within metro areas. The present section focuses only on outcomes across metros. Using the same baseline calibration and set of assumptions as above, it shows how continuously varying productivity below and above that of the representative metro affects *aggregate* metro population and land area and how it affects *mean* population density, prices, and consumption.

Doing so yields two important results. First, open-metro land area asymptotes to a value about twice that of the representative metro as its productivity increases without bound. Further increases in productivity increase population through “infill.” Second, there is a relatively high productivity threshold for a metro area to exist. At productivity levels below this threshold, the price of potentially residential land falls below its agricultural value, as proxied by the perimeter land price in the representative metro.

6.1 Quantities

The top two panels of Figure 10 show open-metro population, land area, and mean population density as open-metro relative TFP varies from 0.76 to 1.42. As above, open-metro population is pinned down by the equal utility requirement and open-metro land area is pinned down by equal perimeter price requirement. The dashed vertical lines are drawn at relative TFP levels that correspond to the benchmark low- and high-TFP metros described in the previous section.

Open-metro population, residential land area and mean population density each increase as open-metro TFP increases. These positive relationships were implicit in the corresponding benchmark gradients shown in the previous section. More interesting is that open-metro land area and outermost commute distance approximately asymptote to upper bounds as relative TFP rises from

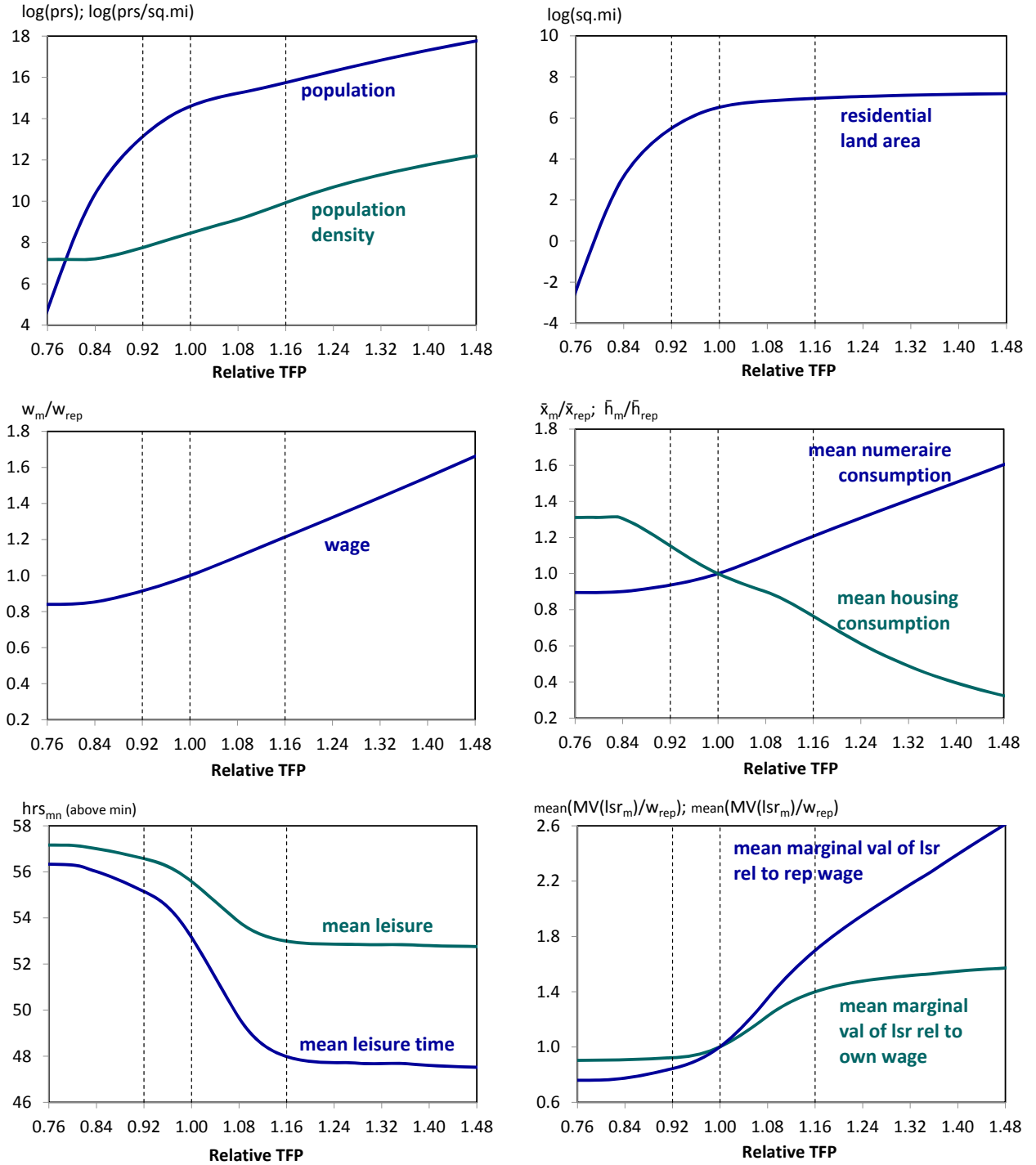


Figure 10: The Effect of Productivity on Aggregate and Mean Quantities. Figure shows the aggregate and mean open-metro outcomes as TFP in producing the numeraire good varies from its level in the representative metro. Vertical dashed lines are drawn at the representative-metro and at the benchmark low- and high-TFP relative values discussed in Section 5.

a level moderately below 1.²⁰ The outermost commute time proves to be the key characteristic limiting land size. As speed becomes less sensitive to commute volumes, the asymptotic land area increases. This underscores the importance of traffic congestion in driving outcomes.

Conversely, as relative productivity decreases moderately below 1, population and residential land area quickly fall off to near zero. To support a population of at least 100 thousand people—a reasonable threshold to merit “metropolitan” status—open-metro productivity must be no more than 14 percent below that of the representative metro ($A_m/A_{rep} \geq 0.86$). At this threshold, metro residential land area is 52 square miles. At a productivity just 20 percent below that of the representative metro, population is just 2,700 and residential land area is just 5 miles.

This sharp falloff in metro size follows from the required matching of perimeter land prices. Because the price of capital is equal everywhere, perimeter house prices must be the same across metros as well. And so there is a binding limit on how far house prices can fall to compensate for low wages (Figure 11, top right panel, lower left). Conversely, there is a binding limit on how far relative wages can drop because residents must still achieve the representative-metro level of utility (Figure 10, middle left panel, lower left). To keep wages from falling below this limit, numeraire production becomes extremely intensive in land input thereby driving down the marginal product of CBD land and its price to near zero (Figure 11, bottom left panel, lower left).²¹

An additional salient characteristic of the open-metro outcomes is that mean leisure, mean leisure time, and the marginal value of open-metro leisure relative to the open-metro wage each transition from an approximately fixed level at relatively low productivities to a different approximately fixed level at high productivities (Figure 10, bottom panels). The leveling at low relative productivities reflects that at the maximum commute speed, the differences in commute time from the outer residential ring (whatever it may be) to the CBD are a matter of just a few minutes. The plateau at high relative productivity reflects that the majority of the population infill is to rings from which commute speed is already at its minimum.

6.2 Prices

Unsurprisingly, mean residential land prices, house prices, and house expenditure increase as TFP increases (Figure 11). Similar to the case of price variations within metros, the variation in means across metros is an order of magnitude larger for land prices is an order of magnitude larger than the variation in means for house prices and expenditures. As described immediately above, mean prices are insensitive to productivity at levels below the threshold for metropolitan land use.

²⁰Less approximately, land area and outermost commute distance eventually attain maximum values and then decrease with further increases in TFP. Under the baseline calibration, these occur at a relative TFP of 1.80 and respectively equal 1,390 square miles and 37 miles. As the open-metro population at such a high relative TFP is 207 million, it seems most appropriate to describe land area and outermost commute distance as asymptoting to an upper bound over plausible relative TFP levels.

²¹With externally-funded transfers—such as Social Security, Medicare, and Medicaid—the threshold productivity for metro land use would be lower. Threshold productivity levels are also inversely affected by the level of metro consumption amenities (Rappaport, 2008b).

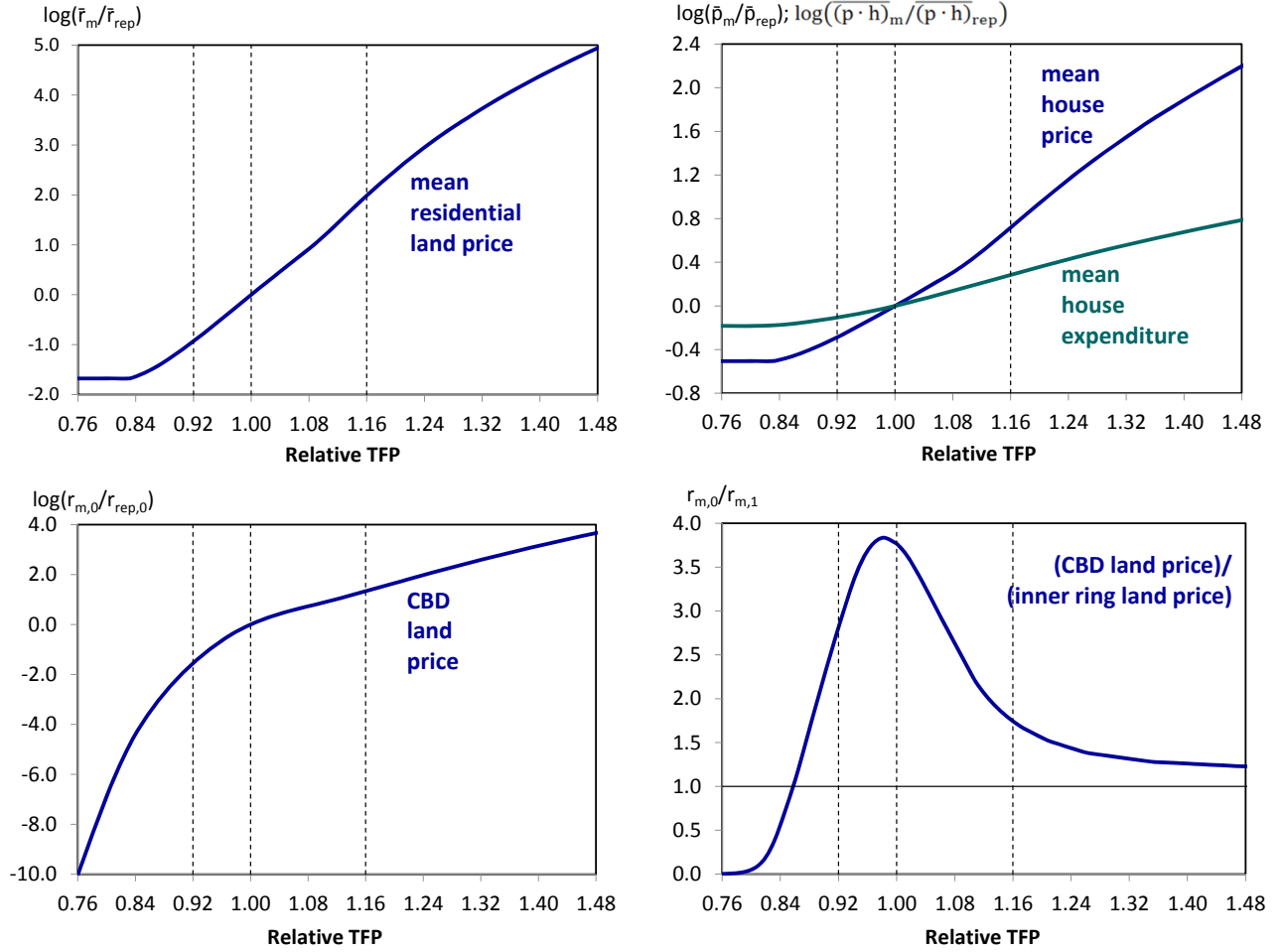


Figure 11: The Effect of Productivity on Mean House and Land Prices. Figure shows aggregate, mean, and CBD relative prices as TFP in producing the numeraire good varies from its level in the representative metro. Vertical dashed lines are drawn at the representative-metro and at the benchmark low- and high-TFP relative values discussed in Section 5.

A baseline assumption is that the CBD in each metro has an identical fixed radius and land area. As a result, the price of land in each CBD will differ from the price of land in the surrounding innermost residential ring (Figure 11, bottom right panel). In the representative metro, the price of CBD land is nearly four times the price of innermost residential land ($r_{rep,0}^L \gg r_{rep,1}^L$).

As open-metro TFP increases above the representative level, the two prices quickly converge. The primary reason is the sharp rise in $r_{m,1}^L$ as people crowd into the inner residential ring, due both to the increase in metro population and to the increase in traffic congestion. This increase is partly offset by an increase in $r_{m,0}^L$ as the marginal product of CBD land increases (both as a direct consequence of the increase in TFP and as an indirect consequence from the implied increases in labor and capital input). But because land is calibrated to have a very small share of numeraire

factor income, the residential crowding dominates.²²

7 Alternative Calibrations and Assumptions

A common criticism of quantitative models is that they are a “black box” in the sense that it is difficult to divine the underlying mechanisms driving outcomes. To the extent that this is true, quantitative models will fail to deliver intuition. In order to further remove opaqueness and build intuition, this section describes population density and house price gradients for a number of perturbations from the baseline calibration and assumptions. These perturbations include changing the various elasticities of substitution, the calibration targets determining weights in the CES housing production and utility functions, the requirement that hours be inelastically supplied, and several key characteristics of commuting. For each perturbation, gradients are shown both for the representative metro and for an open metro and the benchmark high-TFP metro described in Section 5. In many cases the perturbations leave outcomes in the representative metro virtually unchanged from baseline outcomes. In contrast, the perturbations typically have first-order effects on outcomes in the high-TFP metro.

7.1 Perturbations to Utility and Production Elasticities

Population density and house price gradients from perturbing the elasticities of substitution in utility and housing production are shown in Figure 12. Counterintuitively, the gradients for the perturbed representative metros are indistinguishable from the gradients for the baseline representative metro (various colored lines versus thick navy line). For example, intuition suggests that increasing the elasticity of substitution in utility between housing and the numeraire good, σ_h , should steepen the population density gradient because individuals in the inner rings are more willing to substitute away from housing. This does not happen in the perturbed representative metro because the calibrated weight on housing, η_h , adjusts upward in order to continue to hit the unchanged calibration target for the housing consumption share, $\bar{\mu}_{rep}$.²³

In contrast, the intuition on steepening does indeed hold in the high-TFP metro. For an increase in σ_h to 0.95 (from 0.85), innermost density increases by 20 percent from its baseline value (left panel, light blue gradient).²⁴ In addition, the high-TFP population increases to 7.5 million (versus 6.9 million under the baseline). Notice that there is no decrease in population density at further distances. The reason is that residents there are already consuming five times the quantity

²²The CBD radius and land area can be endogenized by requiring the price of CBD land to equal the price of inner-ring residential land. Doing so generalizes the present setup to include the Rosen (1979)-Roback (1982) compensating differential model with quantities described in Rappaport (2008a,b).

²³Four parameters are calibrated by matching representative-metro targets: the weight on land in producing housing, the weight on housing in utility, the weight on leisure in utility, and the per mile commute cost (η_L , η_h , η_ℓ , and δ). For a change to either σ_h or σ_ℓ but not both, only the corresponding weight, η_h or η_ℓ , adjusts. For a change to either σ_L or the per-mile commute cost target share of wage income, all four of these calibrated variables adjust.

²⁴Because density is displayed as a logarithm, the vertical increase in the density gradient is relatively modest.

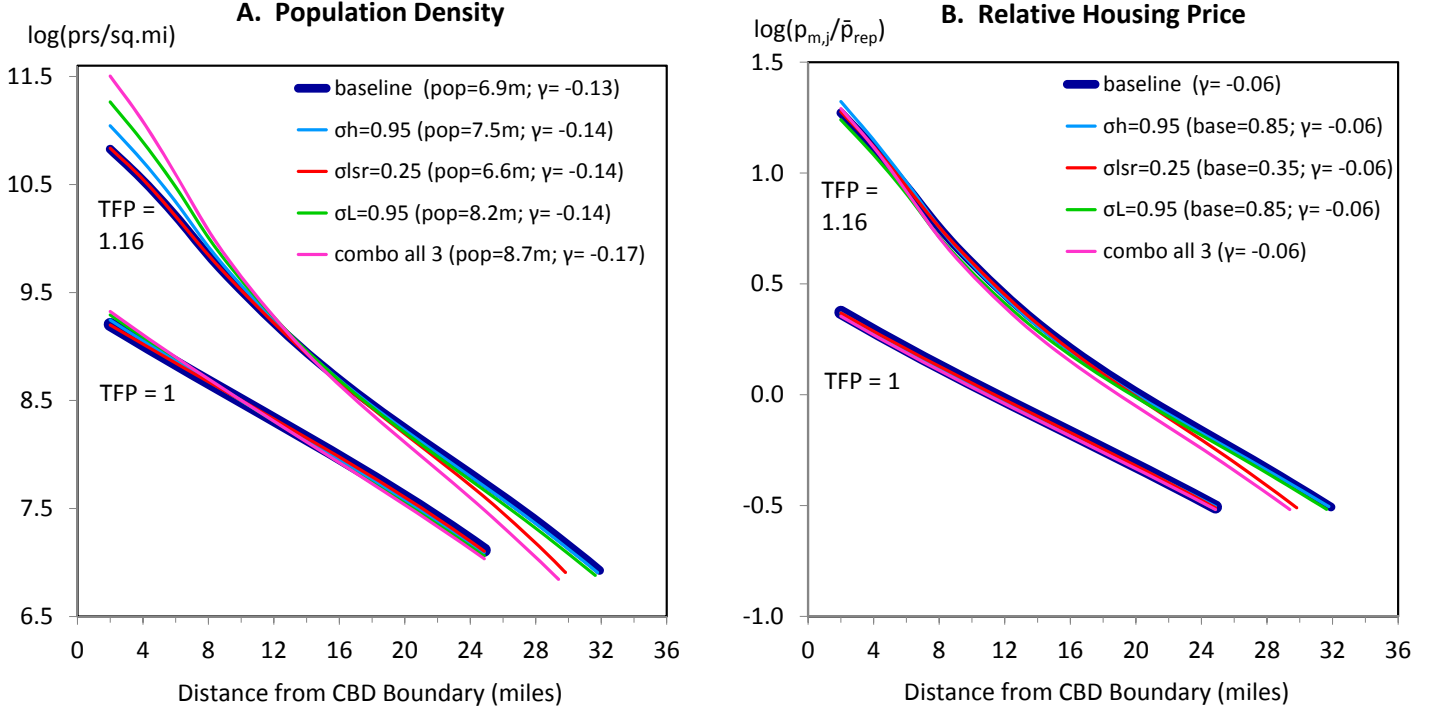


Figure 12: Population Density and House Prices with Alternative Elasticities. Population and gradients enumerated in parentheses are for the high-productivity open metro. The combination calibration combines $\sigma_h = 0.95$, $\sigma_\ell = 0.25$, and $\sigma_L = 0.95$. As described in the text, representative metro gradients are essentially unaffected because of the adjustment of weighting parameters to match calibration targets.

of housing consumed by individuals in the innermost ring. Even with less curvature of utility than under the baseline calibration, further increases in housing consumption yield considerably diminished marginal utility. Notice also that the price gradient is unchanged. This simply reflects that as σ_h increases, individuals need less compensation for decreases in housing consumption from already-low levels.

Decreasing the leisure elasticity from 0.35 to 0.25 steepens the the population-density gradient for the high-TFP metro at far commute distances (left panel, high-TFP red line). For commute times that are already fairly long, a lower value of σ_ℓ makes individuals less willing to tradeoff even longer commutes for higher levels of housing consumption. Correspondingly, the price gradient also steepens at far commute distances (right panel, high-TFP red line). In addition, the outermost commute distance and total metro population modestly decrease (respectively, to 6.6 million and 29.8 miles from 6.9 million and 31.9 miles).

Increasing the housing-production elasticity, σ_L , makes it easier to shift away from scarce land in producing housing and so accommodate a higher population at short commute distances

without upward pressure on house prices. For the high-TFP metro, innermost-ring population density increases by 50 percent and total population increases to 8.2 million. But the price gradient is unchanged.

Combining all three perturbations to the baseline elasticities causes some additional steepening in the population density gradient (left panel, high-TFP pink line). High-TFP innermost density doubles from its baseline value and total metro population increases to 8.7 million.

7.2 Perturbations to Utility and Production Weights

Figure 13 illustrates the effects of perturbations to the targets that calibrate the structural weighting variables in utility and housing production. A first perturbation targets the mean housing share of consumption expenditure in the representative metro, $\bar{\mu}_{rep}$, to be 0.15 rather than its baseline value of 0.18 (red lines). A second perturbation targets the mean factor income share of labor from housing production in the representative metro, $\bar{\nu}_{rep}$, to be 0.30 rather than its baseline value of 0.35 (light blue lines). A third perturbation combines both of these ($\bar{\nu}_{rep} = 0.30$ and $\bar{\mu}_{rep} = 0.15$; orange lines). A fourth perturbation targets mean desired work hours in the representative metro, \bar{n}_{rep}^* , and required hours in all metros, \hat{n} , to each equal 45 rather than 40 (magenta lines). This is equivalent to an decrease in desired leisure hours and so a decrease in the weighting on leisure. A final perturbation sets the land factor income share for numeraire production in the representative metro to near zero ($\alpha_L = 0.001$) rather than its already-low baseline value of 0.016 (green lines).

Similar to the perturbations to elasticities, the perturbations to \bar{n}_{rep}^* and to α_L have essentially no effect on the representative-metro gradients. In contrast, the lower targets for $\bar{\mu}_{rep}$ and $\bar{\nu}_{rep}$ each moderately steepens the representative population density gradient. This is intuitive as the scarce factor and the good in which it is used intensively have been down weighted and so there is less aversion to crowding. Importantly, there is no offset in the weighting parameters as they are now the object of the perturbation. The third perturbation, which lowers the targets for both $\bar{\mu}_{rep}$ and $\bar{\nu}_{rep}$, causes an even larger steepening of the representative-metro density gradient. Each of these three perturbations also steepens the density gradient in the high-TFP metro in proportion to the corresponding steepening of each in the representative metro. For the high-TFP metro, the perturbations to $\bar{\nu}_{rep}$ and $\bar{\mu}_{rep}$ respectively increase innermost-ring density by 40 and 50 percent from its baseline high-TFP value. The combined perturbations more than double innermost density. Correspondingly, high-TFP population rises to 8.0 million and 8.3 million for the respective single perturbations (versus 6.9 million under the baseline). The combined perturbations increase high-TFP population to 10.1 million.

The increased weighting on housing also steepens the price gradient in each metro. As individuals care less about housing, a sharper price drop is necessary to compensate for longer commutes. This steepening primarily occurs at further distances. The reason is that at nearer distances, households consume considerably less housing than under the baseline calibration, which is just

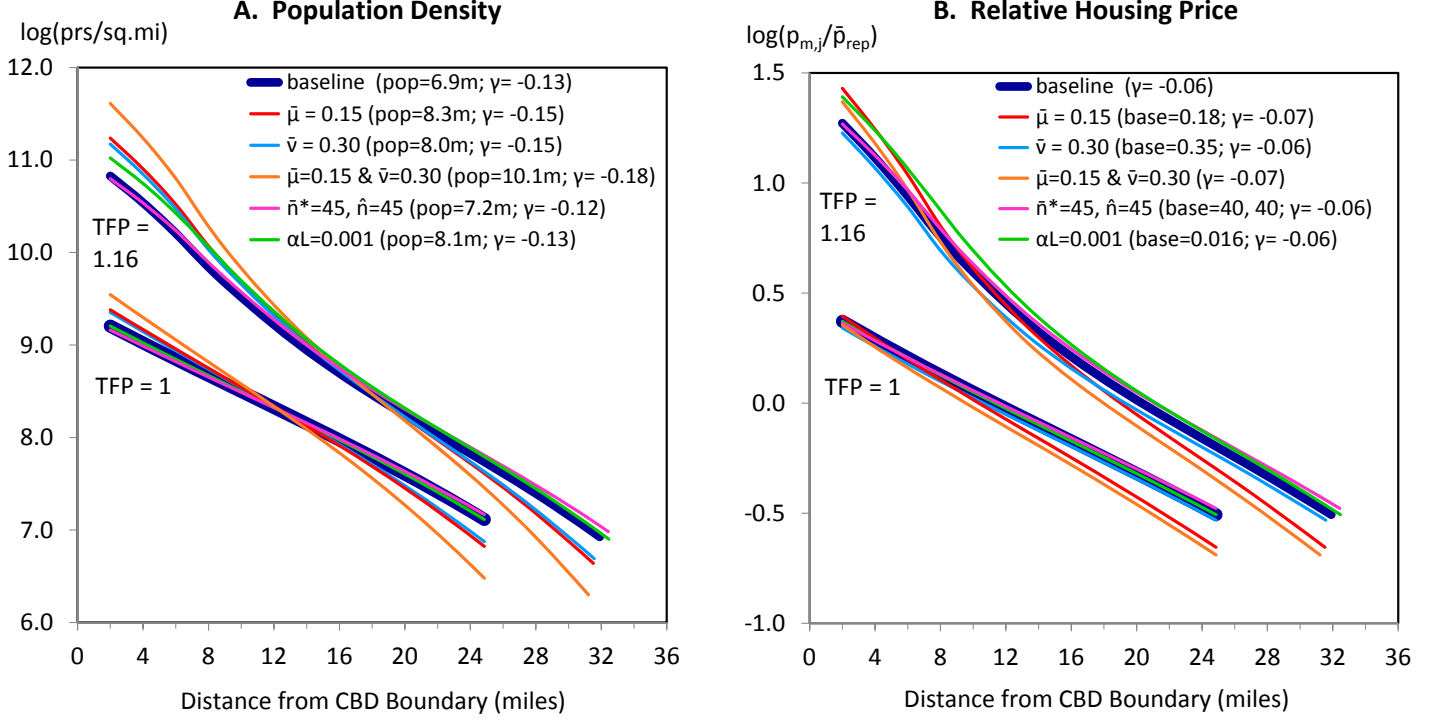


Figure 13: Population Density and House Prices with Alternative Weights. Population and gradients enumerated in parentheses are for the high-productivity open metro.

the flip-side of the increase in densities at near distances. This pushes up the marginal utility from housing consumption at near distances to a level similar to under the baseline and so the price tradeoff for longer commutes is approximately the same.

The lower weighting of leisure very modestly flattens the density and price gradients in the high-TFP metro and nudges its population up to 7.2 million.

Setting the factor income share of land in numeraire production to near zero significantly increases the population of the high-TFP metro to 8.1 million. As is intuitive, this increase reflects that the fixed land input no longer exerts as much downward pressure on the marginal product of labor. Correspondingly, the high-TFP relative wage rises modestly from its baseline value.

7.3 Alternative Key Assumptions

Gradients and high-productivity metro population under four alternative key assumptions are shown in Figure 14. The first alternative assumes that the loss from commuting arises exclusively from the per mile numeraire cost (magenta line). Conversely, the second alternative assumes that the loss from commuting arises exclusively from reduced leisure (red line). The third alternative sets required hours (in both metros) to be above targeted mean desired work hours in the representative metro (blue line). The fourth alternative allows individuals to choose their own work

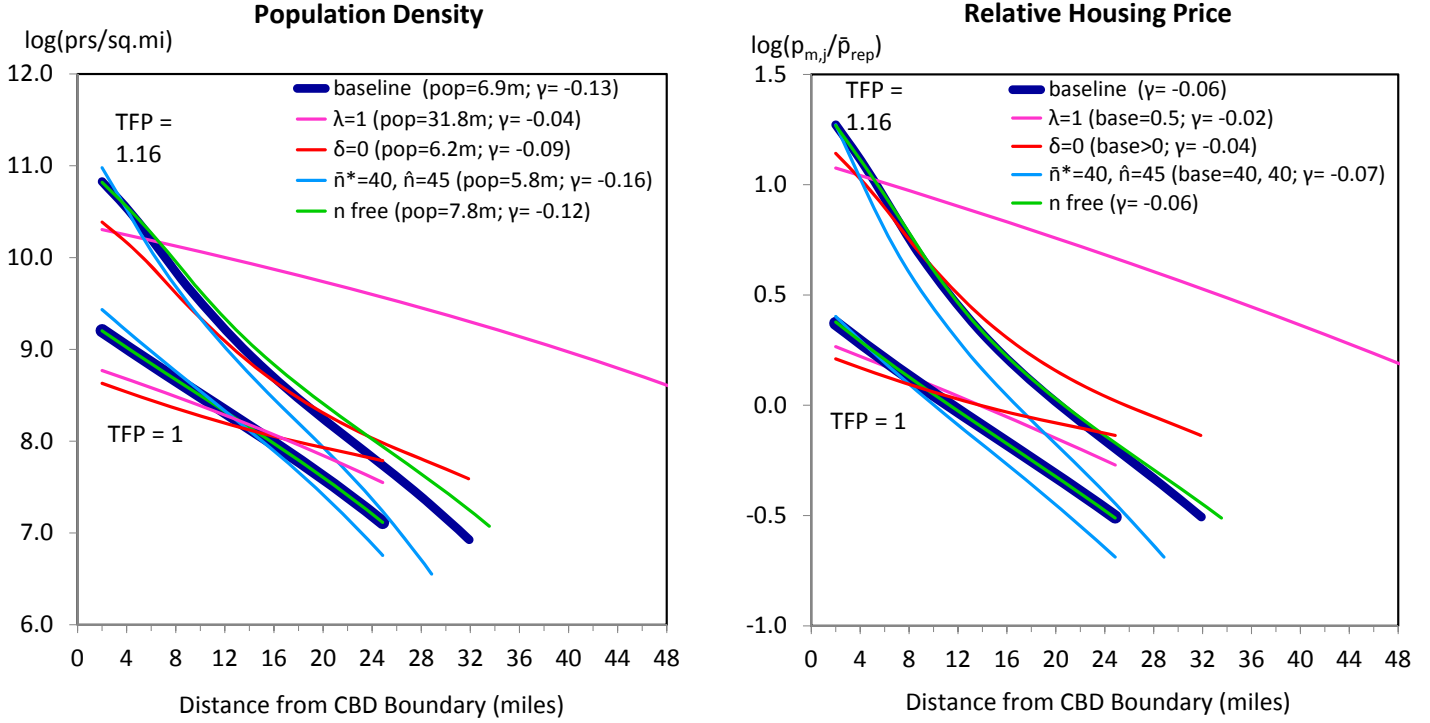


Figure 14: Population Density and House Prices with Alternative Key Assumptions.

Population and gradients enumerated in parentheses are for the high-productivity open metro. For the alternative assumption that there is no leisure cost to commuting ($\lambda = 1$), the outermost commute is 67 miles.

hours (green line).

Eliminating the leisure cost to commuting can be accomplished by parameterizing commute time to have 100 percent leisure content ($\lambda=1$). Doing so dramatically affects metro outcomes (pink lines). In the representative metro, the density and price gradients significantly flatten. In the the high-TFP metro, they flatten by an order of magnitude. The high-TFP metro expands to have an outer radius of 67 miles and a total population of 32 million! Even doubling the per mile commute cost to equal 0.08 percent of representative-metro wage income still implies a high-TFP metro with an outer radius of 48 miles and a total population of 21 million.

Eliminating the per mile numeraire cost to commuting also significantly affects metro outcomes (red lines). Under the baseline assumptions and calibration, the numeraire cost to commuting depresses disposable income in the representative metro by as much as 10 percent. And so for the representative metro, eliminating the numeraire cost considerably lessens the compensation needed for longer commutes. A partial offset comes from an implicit increase in the leisure cost to commuting. This occurs because the increase in disposable income decreases the marginal utility of

numeraire and housing relative to the marginal utility of leisure.²⁵ On net, the representative-metro gradients considerably flatten. Population density in the innermost residential ring decreases by more than one third. population density in the outermost ring nearly doubles.

In the high-TFP metro, the flattening of gradients is more moderate and occurs primarily at further commute distances. At near distances, where commute speeds are slow, the implicit increase in the weight on leisure approximately offsets the elimination of the numeraire distance cost leaving the density and price gradients similar to what they were under the baseline. At further commute distances, where speed is near its maximum, there is less offset and so gradients flatten significantly. On net, the implicit increase in the weight on leisure slightly dominates the elimination of the numeraire commute cost in the sense that the high-TFP metro population moderately decreases (to 6.2 million from 6.9 million).

The third alternative assumption increases required weekly work hours, \hat{n} , to 45 while continuing to target mean desired work hours, \bar{n}^* , to equal 40. Doing so pushes individuals up their marginal utility of leisure curve thereby making the leisure cost of commuting higher. In consequence, the population density and house price gradients significantly steepen in both the representative and high-TFP metros and the high-TFP metro population decreases to 5.8 million.

Allowing individuals' to choose their own work hours ($n_{m,j} = n_{m,j}^*$) leaves the gradients in the representative metro unchanged and modestly flattens the gradient in the high-TFP metro at further distances. Especially for individuals living at these further distances, partly offsetting long commutes by decreasing work hours is more efficient than bearing the cost through decreased leisure alone. With freely chosen work hours, high-TFP metro population increases to 7.8 million.

7.4 Alternative Commuting Technology

Gradients and high-TFP metro populations under variations in the commuting technology are shown in Figure 15. The first alternative assumes no traffic congestion. Instead, highway commute speed is constant at 37 mph, its mean value in the representative metro under the baseline calibration and assumptions (blue lines).²⁶ The second alternative assumes highway capacity is less elastic with respect to commute volume ($\sigma_V = 0.65$ rather than 0.80; red lines). As in the baseline setup, capacity is assumed to equal volume when the latter is 750 thousand. (Figure 4 above illustrates the implied relationship between commute speed and volume under both the baseline and this alternative.) The third alternative assumes that highway capacity is less elastic with respect to volume only in the high-productivity metro (magenta lines).

Commuting without traffic congestion has only a small effect on representative-metro outcomes

²⁵The explicit weight on leisure in utility, η_ℓ , actually decreases. But this is dominated by the realignment of marginal utilities so that on net, variations in leisure exert more “leverage” on utility than they do under the baseline.

²⁶The representative mean speed is constructed as follows: For each ring, total travel distance is divided by total variable travel time to determine average highway speed (reported in Table 5). The mean average highway speed is the populated-weighted mean over the ring average highway speeds.

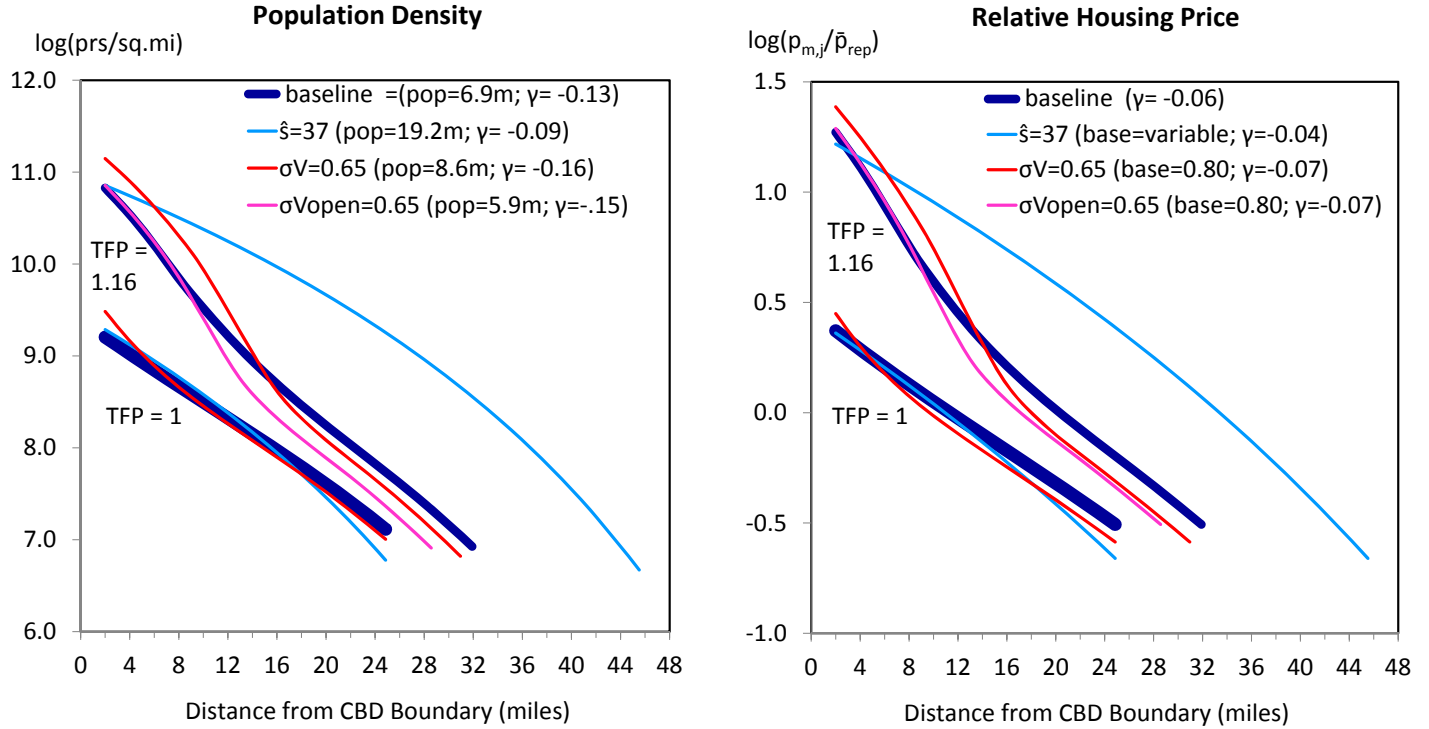


Figure 15: Population Density and House Prices with Alternative Commuting Technology. Population and gradients enumerated in parentheses are for the high-productivity open metro. For the alternative in which the elasticity of highway capacity is lowered only in the high-TFP metro, the representative metro gradients are the same as under the baseline and so are not shown.

both because it causes the calibrated weight on leisure, η_ℓ , to be adjusted downward and because traffic congestion in the representative metro was in any case relatively moderate. But removing congestion has a huge effect on outcomes in the high-productivity metro. As is intuitive, gradients at short commute distances are considerably flattened because commute speeds are now much faster than under the baseline. Conversely, gradients at far commute distances are modestly steepened because commute speeds are now slower than under the baseline. The faster near-in commute speeds easily dominate. High-TFP metro population increases to 19.2 million. The furthest commute increases to 46 miles and high-TFP metro area increases to 2000 square miles (versus 32 miles and 1100 square miles).

Decreasing the elasticity of highway capacity in both metros counterintuitively increases the population of the high-TFP metro (to 8.6 million versus 6.9 million under the baseline). One reason is an implied decrease in the weight on leisure in order to continue to hit the targeted mean desired work hours. Equally important, is that the lower elasticity of capacity disproportionately decreases speed in the representative metro. In particular, highway speed through the innermost residential ring of the representative metro falls to its assumed minimum of 10 mph (versus 28

mph under the baseline). As a result, there is no longer a leisure cost to living in the innermost ring of the high-TFP metro compared to living in the innermost ring of the representative metro. The compensation for the higher numeraire wage in the former must be accomplished solely by an increase in house prices which in turn requires an increase in population density to be supported. For this reason, the density and price gradients shift upward at near distances. Population density in the innermost residential ring increases by more than a third from its baseline value.

At intermediate distances, the high-TFP gradient slopes steepen relative to the baseline because of slower commute speeds. At further distances, the high-TFP slopes flatten again because commute speeds at traffic volumes below the assumed benchmark capacity ($\hat{V} = 750$ thousand) are unchanged from the baseline.²⁷

By assumption, decreasing the assumed elasticity of highway capacity volume *in the open-metro only* does not affect representative-metro outcomes and so no representative-metro gradients are shown for this alternative. In the high-TFP metro, population decreases to 5.9 million and there is no upward shift in the density and price gradients at near distances. Consistent with Duranton and Turner (2011), a change in highway capacity in one or a few metros primarily affects population while leaving average commute speeds largely unchanged.

7.5 Alternative Commuting Leisure Content

Based on estimates of commuters' willingness to pay to marginally shorten their commutes, the baseline calibration assumes that commuting time includes a 50 percent leisure component. But scores of anecdotes and surveys suggest that commuting is often a source of negative rather than positive leisure. A Google search on the duplet, {'road rage', 'commuting'}, brings up more than one million links.

Figure 16 shows outcomes under two alternative assumptions on the leisure content of commuting. Under the first, commuting lacks any consumption component (green line). Under the second, the leisure component is assumed to be an increasing function of speed, $\lambda(s)$ with $\lambda'(\cdot)$ positive (blue lines).²⁸ More specifically, λ is assumed to be 0.50 at speeds of 55 mph or higher. As speed falls from 55 to 10, λ is assumed to fall linearly to zero.

Unsurprisingly, removing the leisure content to commuting significantly steepens the density and price gradients in both the representative and high-TFP metros. To keep desired work hours in the representative metro at its target, the calibrated weight on leisure decreases. In the high-productivity metro, the gradients are characterized by pronounced negative second derivatives. These reflects the sharply rising marginal of utility at longer commutes. Correspondingly, the high-TFP outermost commute distance decreases to 27 miles and and the high-TFP population

²⁷Just as capacity adjusts upward proportionally less than increases in volume above the benchmark, it adjusts down proportionally less than decreases in volume below the benchmark. At the benchmark capacity, speed is already near its maximum value and so essentially speed at further distances is the same as under the baseline.

²⁸Empirical evidence that the valuation of commute time increases as congestion worsens is reported in Abrantes and Wardman (2011), Hensher and Greene (2011), and Rizzi, Limonado, and Steimetz (2012).

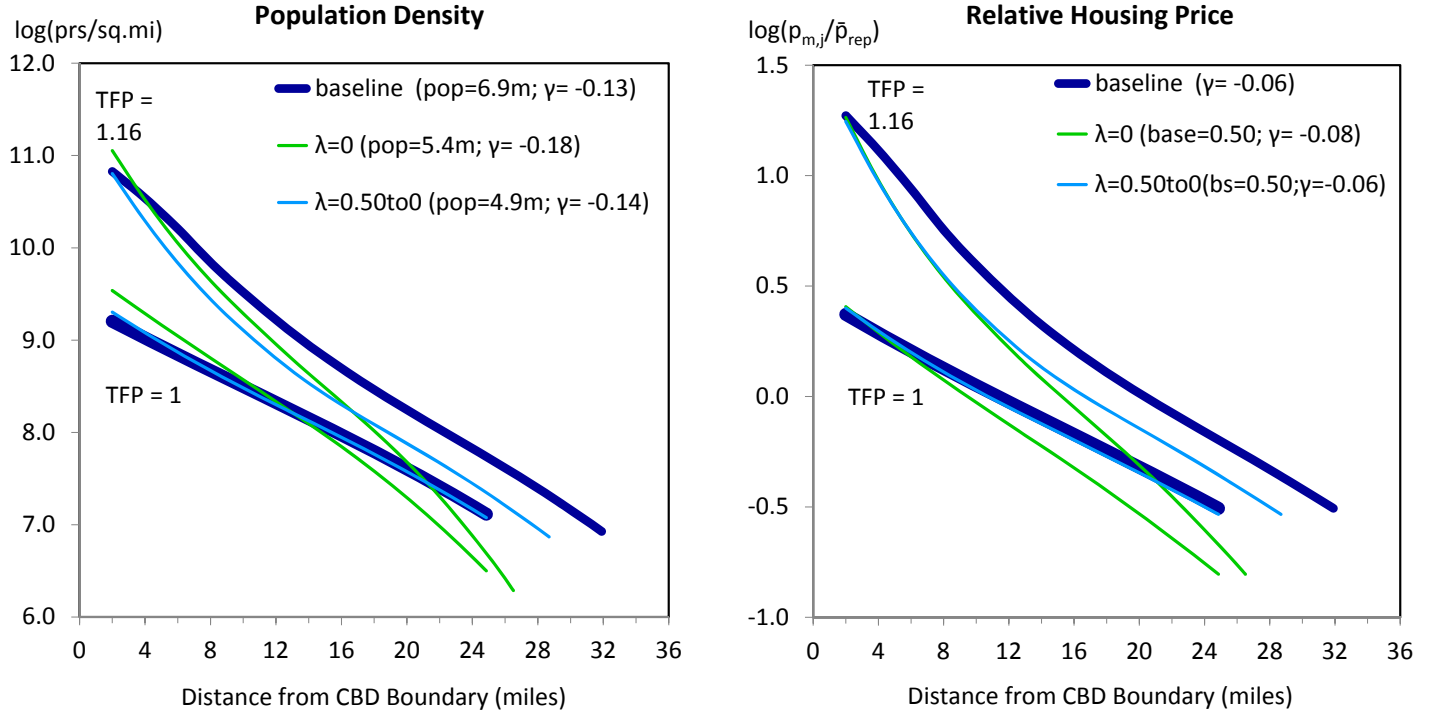


Figure 16: Population Density and House Prices with Alternative Commuting Leisure Content. Population and gradients enumerated in parentheses are for the high-productivity open metro. λ is the leisure content of commute time. In the second alternative scenario, λ is assumed to decrease linearly from 0.50 to 0 as commute speed decreases from 55 mph to 10 mph.

decreases to 5.4 million (versus, respectively, 32 miles and 6.9 million under the baseline).

Assuming that leisure content is a decreasing function of speed leaves representative-metro gradients essentially unchanged. In the high-TFP metro, gradients significantly steepen at near distances and then flatten at further distances. This pattern reflects that at further distances, commute speeds are higher and so the leisure content of commute time is closer or equal to its constant baseline value. High-TFP outermost commute distance decreases moderately to 29 miles. High-TFP population falls to 4.9 million, which is below its level when the commute time leisure content was zero at all speeds. The explanation is that commute speeds are considerably faster in the representative metro and so residents there continue to enjoy significant leisure content to their commutes. This boosts the reservation utility that must be matched in the high-TFP metro, where commute speeds are slower. The decrease in high-TFP metro population relative to the baseline keeps speeds from being even slower.

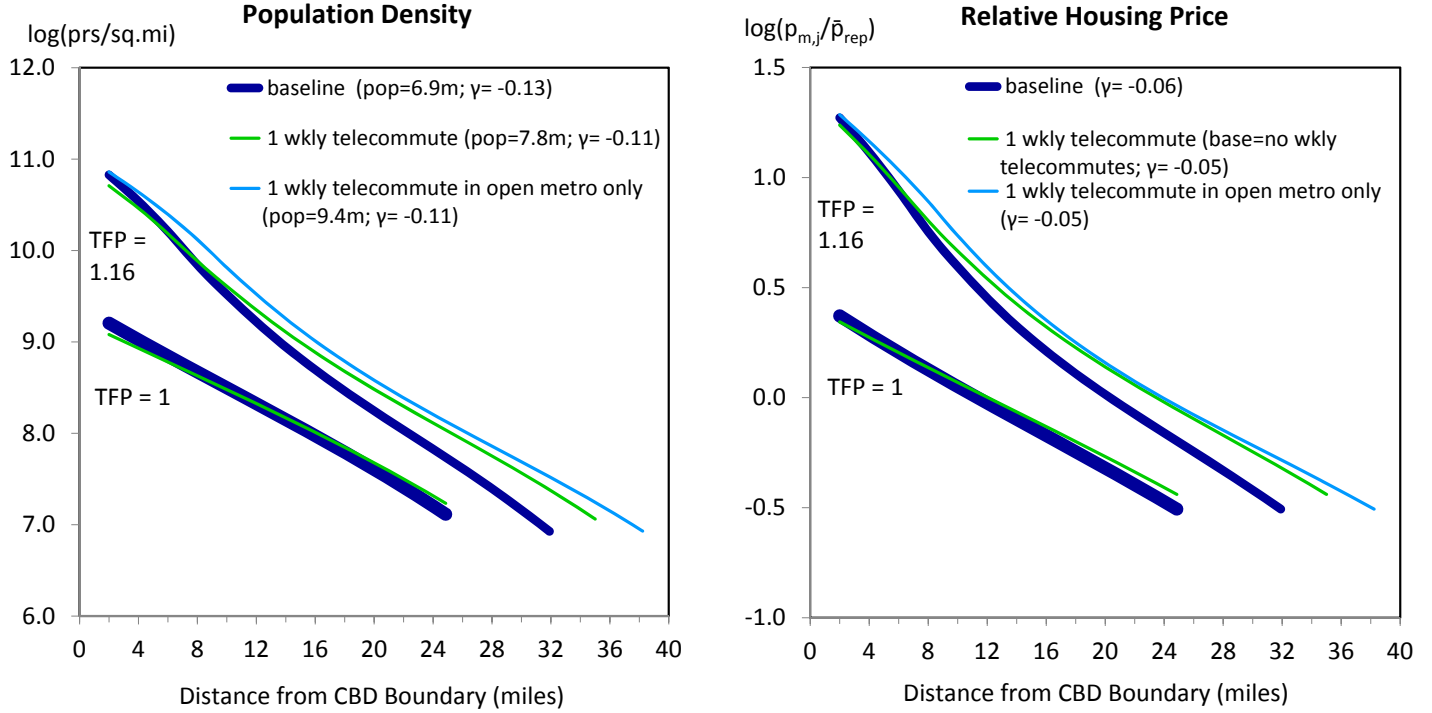


Figure 17: Population Density and House Prices with Telecommuting. Population and gradients enumerated in parentheses are for the high-productivity open metro. For the alternative scenario in which only high-TFP metro residents telecommute, the representative metro gradients are the same as under the baseline and so are not shown.

7.6 Telecommuting

A final set of alternative outcomes varies the number of weekly commutes (Figure 17). The first alternative assumes that individuals in both the representative and high-TFP metros telecommute one day per week. On any given workday, one fifth of individuals do so, and hence only 80 percent of residents in each metro will physically commute to work (green lines). The second alternative assumes that only individuals in the high-TFP metro telecommute, again with one fifth of them doing so on any given workday (blue lines). In this case, the calibrated parameters and representative-metro outcomes are identical to those under the baseline and so not shown. Under both scenarios, individuals continue to be required to work 40 hours per week.

When the weekly telecommute applies to both the representative and the open metros, the representative-metro gradients are essentially unchanged from those under the baseline. The high-productivity gradients are moderately flatter than under the baseline and the outermost commute increases to 35 miles (versus 32 miles with no telecommuting). In consequence, high-TFP metro population increases to 7.8 million (versus 6.9 million under the baseline).

When the weekly telecommute takes place only in the high-TFP metro, the high-TFP gradients

shift moderately up in addition to flattening. High-TFP metro population increases to 9.4 million. The ability to telecommute in metros that are especially crowded but not in other metros potentially increases the population of the crowded ones by more than one third.

8 Conclusions

A quantitative system of monocentric metros is comprised of a closed representative metro and an indefinite number of open metros. The representative metro serves to calibrate the model parameters and establish a reservation level of utility that must be matched across and within the open metros as well as the outermost-ring land price that must be matched across metros. By assuming moderate exogenous differences in total factor productivity across open metros, the model is able to approximately replicate the observed distribution of population, population density, and commute times across and within a number of intermediate and large U.S. metros. The model can similarly match a number of other moments such as the elasticity of housing supply with respect to price and the compensated elasticity of demand of labor supply with respect to the wage. The latter, “Frisch” elasticity is shown to vary moderately across homogeneous workers living at different locations within large metro areas.

The calibration of the the representative metro to hit exogenously-specified targets such as the average land share of factor income and the average housing share of consumption expenditure causes gradients and other outcomes in the representative metro to be somewhat insensitive to deep structural parameters. But the same deep structural parameters typically have first-order effects on open-metro outcomes.

Metro land area is shown to asymptote to a maximum as total factor productivity increases moderately above that in the representative metro. For a metro to have a perimeter land price above its alternative agricultural value, its total factor productivity can be only moderately below that of the representative metro. The requirement that individuals supply a fixed number of work hours tempers the effect of productivity on metro size. Fixed work hours also cause individuals with long commutes to have a marginal value of leisure time that is far above their wage.

More generally the framework developed herein gives significant quantitative and qualitative insight into the determination of quantity and price outcomes across and within metro areas. The framework can be easily extended to capture a number of other characteristics such as variations across and within metros in consumption amenities, taxes, land-use regulation, and increasing returns to scale in production. Less easily, it can also be extended to allow for fixed highway locations, monopolistic competition, heterogeneous households, and exogenously-specified polycentric work locations. Enriching (but complicating) the framework in these ways should deliver further insight into the determination of outcomes across and within metro areas.

Bibliography

- Abrantes, Pedro A.L. and Mark R. Wardman (2011). "Meta-Analysis of UK Values of Travel Time: an Update." *Transportation Research Part A*, 45(1), 1-17.
- Aguiar, Mark and Erik Hurst (2007). "Measuring Trends in Leisure: The Allocation of Time over Five Decades." *Quarterly Journal of Economics*, 122(3), 969-1006.
- Albuoy, David (2008). "Are Big Cities Bad Places to Live? Estimating Quality of Life across Metropolitan Areas." NBER Working Paper 14472.
- Albuoy, David (2009). "What Are Cities Worth? Land Rents, Local Productivity, and the Capitalization of Amenity Values." NBER Working Paper 14981.
- Albuoy, David and Bert Lue (2014). "Driving to Opportunity: Local Rents, Wages, Commuting Costs and Sub-Metropolitan Quality Of Life." NBER Working Paper 19922.
- Alonso, W. (1964) *Location and Land Use*. Cambridge: Harvard University Press.
- Anas, Alex and Ikki Kim (1996). "General Equilibrium Models of Polycentric Urban Land Use with Endogenous Congestion and Job Agglomeration." *Journal of Urban Economics* 40, 232-256.
- Arnott, R. J., and MacKinnon, J. G. (1977a). "The Effects of the Property Tax: A General Equilibrium Simulation. *Journal of Urban Economics*, 4(4), 389-407.
- Arnott, R. J., and MacKinnon, J. G. (1977b). "The Effects of Urban Transportation Changes: A General Equilibrium Simulation. *Journal of Public Economics*, 8(1), 19-36.
- Baum-Snow, Nathaniel (2007a). "Suburbanization and Transportation in the Monocentric Model." *Journal of Urban Economics* 62, 3, pp. 405-423.
- Baum-Snow, Nathaniel (2007b). "Did Highways Cause Suburbanization?" *Quarterly Journal of Economics* 122, 2, pp. 774-805.
- Baum-Snow, Nathaniel (2014). "Urban Transport Expansions, Employment Decentralization, and the Spatial Scope of Agglomeration Economies." Brown University Working Paper.
- Brueckner, Jan K. (1978). "Urban General Equilibrium Models with Non-central Production. *Journal of Regional Science* 18(2), 203-215.
- Brueckner, Jan K. (1979). "A Model of Non-central Production in a Monocentric City. *Journal of Urban Economics* 6(4), 444-463.
- Brinkman, Jeffrey C. (2013). "Congestion, Agglomeration, and the Structure of Cities." Federal Reserve Bank of Philadelphia Working Paper 13-25.
- Chatterjee, Satyajit and Burcu Eyigungor (2013). "Do Supply Restrictions Raise the Value of Urban Land? The (Neglected) Role of Production Externalities." Federal Reserve Bank of Philadelphia Working Paper 13-37.
- Ciccone, Antonio (2002). "Agglomeration Effects in Europe." *European Economic Review* 46, 213-227.

- Davidoff (2013). "Supply Elasticity and the Housing Cycle of the 2000s. *Real Estate Economics* 41(4), 793-813.
- Davis, Morris and François Ortalo-Magné (2011). "Household Expenditures, Wages, Rents." *Review of Economic Dynamics* 14, pp. 248-261.
- Duranton, Gilles and Matthew A. Turner (2011). "The Fundamental Law of Road Congestion: Evidence from U.S. Cities." *American Economic Review* 101, 6, pp. 2616-2652.
- Davis, Morris and Jonathan Heathcote (2007). "The Price and Quantity of Residential Land in the United States." *Journal of Monetary Economics* 2007, 54(8), 2595-2620.
- Glaeser, Edward L. and Matthew E. Kahn (2004). "Sprawl and Urban Growth." In: Henderson, J. Vernon, Thisse, Jacques Francoise (Eds.), *Handbook of Regional and Urban Economics*, vol. 4. Amsterdam: Elsevier North-Holland, 2481-2527.
- Glaeser, Edward, and Joseph Gyourko (2005), "Urban Decline and Durable Housing." *Journal of Political Economy* 113(2), 345-375.
- Glaeser, Edward, Joseph Gyourko, and Raven Saks (2005). "Why is Manhattan So Expensive." *Journal of Law and Economics* 48, 331-369
- Gollin, Douglas (2002). "Getting Income Shares Right." *Journal of Political Economy* 110 (2), 458-474.
- Helsley, Robert W. and Arthur M. Sullivan (1991). "Urban Subcenter Formation." *Regional Science and Urban Economics* 21, 255-275.
- Henderson, Vernon (1974). "The Sizes and Types of Cities." *American Economic Review*, 64(4), 640-657.
- Hensher, David A. and William H. Greene (2011). "Valuation of Travel Time Savings in WTP and Preference Space in the Presences of Taste and Scale Heterogeneity." *Journal of Transport Economics and Policy* 45, 3, pp. 505-525.
- Jorgenson, Dale W., Mun S. Ho, and Kevin J. Stiroh (2005). "Growth of U.S. Industries and Investments in Information Technology and Higher Education." In *Measuring Capital in the New Economy*, eds. Carol Corrado, John Haltiwanger, and Daniel Sichel. Chicago IL: University of Chicago Press.
- King, A. T. (1977). "Computing General Equilibrium Prices for Spatial Economies." *The Review of Economics and Statistics*, 59(3), 340-350.
- Li, Wenli, Haiyong Liu, Fang Yang, and Rui Yao (2012). "Housing over Time and over the Life Cycle: A Structural Estimation." Working paper.
- Lucas, Robert E. Jr. and Esteban Rossi-Hansberg (2002). "On the Internal Structure of Cities." *Econometrica* 70(4), pp. 1445-1476.
- Mayer, Christopher J. and C. Turiel Mayer (2000). "Residential Construction: Using the Urban Growth Model to Estimate Housing Supply." *Journal of Urban Economics* 48, 85-109.

- Mills, E.S. (1967). "An Aggregative Model of Resource Allocation in a Metropolitan Area." *American Economic Review*, 57, 197-210.
- Muth, R. (1969). *Cities and Housing*. Chicago: University of Chicago Press.
- Muth, R. F. (1975). "Numerical Solution of Urban Residential Land-Use Models." *Journal of Urban Economics*, 2(4), 307-332.
- Rappaport, Jordan (2008a). "A Productivity Model of City Crowdedness." *Journal of Urban Economics* 63, 715-722.
- Rappaport, Jordan (2008b). "Consumption Amenities and City Population Density." *Regional Science and Urban Economics* 38(6), 533-552.
- Reichling, Felix and Charles Whalen (2012). "Review of Estimates of the Frisch Elasticity of Labor Supply." Congressional Budget Office Working Paper 2012-13.
- Richter, D. K. (1978). "The Computation of Urban Land Use Equilibria." *Journal of Economic Theory*, 19(1), 1-27.
- Rizzi, Luis L., Juan Pablo Limonado, and Seiji S.C. Steimetz (2012). "The Impact of Traffic Images on Travel Time Valuation in Stated Preference Choice Experiments." *Transportmetrica* 8, 6, pp. 427-442.
- Roback, Jennifer (1982). "Wages, Rents, and the Quality of Life." *Journal of Political Economy* 90 (6), 1257-1278.
- Robinson, John and Geoffrey Godbey (1999). *Time for Life: The Surprising Ways Americans Use Their Time*. University Park PA: Penn State University Press.
- Rosen, Sherwin (1979). "Wage-Based Indexes of Urban Quality of Life." In Miezkowski and Straszheim, Eds., *Current Issues in Urban Economics*. Baltimore: Johns Hopkins University Press.
- Saiz, Albert (2010). "The Geographic Determinants of Housing Supply." *Quarterly Journal of Economics* 125, 3, pp. 1253-1296.
- Small, Kenneth A. and Erik T. Verhoef (2007). *The Economics of Urban Transportation*. New York City: Routledge
- Small, Kenneth A, Clifford Winston and Jia Yan (2005). "Uncovering the Distribution of Motorists' Preferences for Travel Time and Reliability." *Econometrica* 73(4), 1367-1382.
- Sivitanidou, Rena and William C. Wheaton (1992). "Wage and rent capitalization in the commercial real estate market." *Journal of Urban Economics* 31, 2, 206-229
- Sullivan, Arthur M. (1983a). "A General Equilibrium Model with External Scale Economies in Production." *Journal of Urban Economics* 13(2), 235-255.
- Sullivan, Arthur M. (1983b). "The General Equilibrium Effects of Congestion Externalities." *Journal of Urban Economics* 14(1), 80-104.
- Sullivan, Arthur M. (1983c). "Second-Best Policies for Congestion Externalities." *Journal of Urban Economics* 14(1), 105-123.

Sullivan, Arthur M. (1986). “A General Equilibrium Model with Agglomerative Economies and Decentralized Employment.” *Journal of Urban Economics* 20(1), 55-74.

Wheaton, William C. (1979) “Monocentric Models of Urban Land Use: Contributions and Criticism, in *Current Issues in Urban Economics*, P. Mieszkowski and M. Straszheim, Eds., Baltimore: Johns Hopkins University Press.

Wheaton, William C., Serguei Chervachidze, and Gleb Nechayev (2014). “Error Correction Models of MSA Housing Supply Elasticities: Implications for Price Recovery.” MIT Working Paper 14-05.