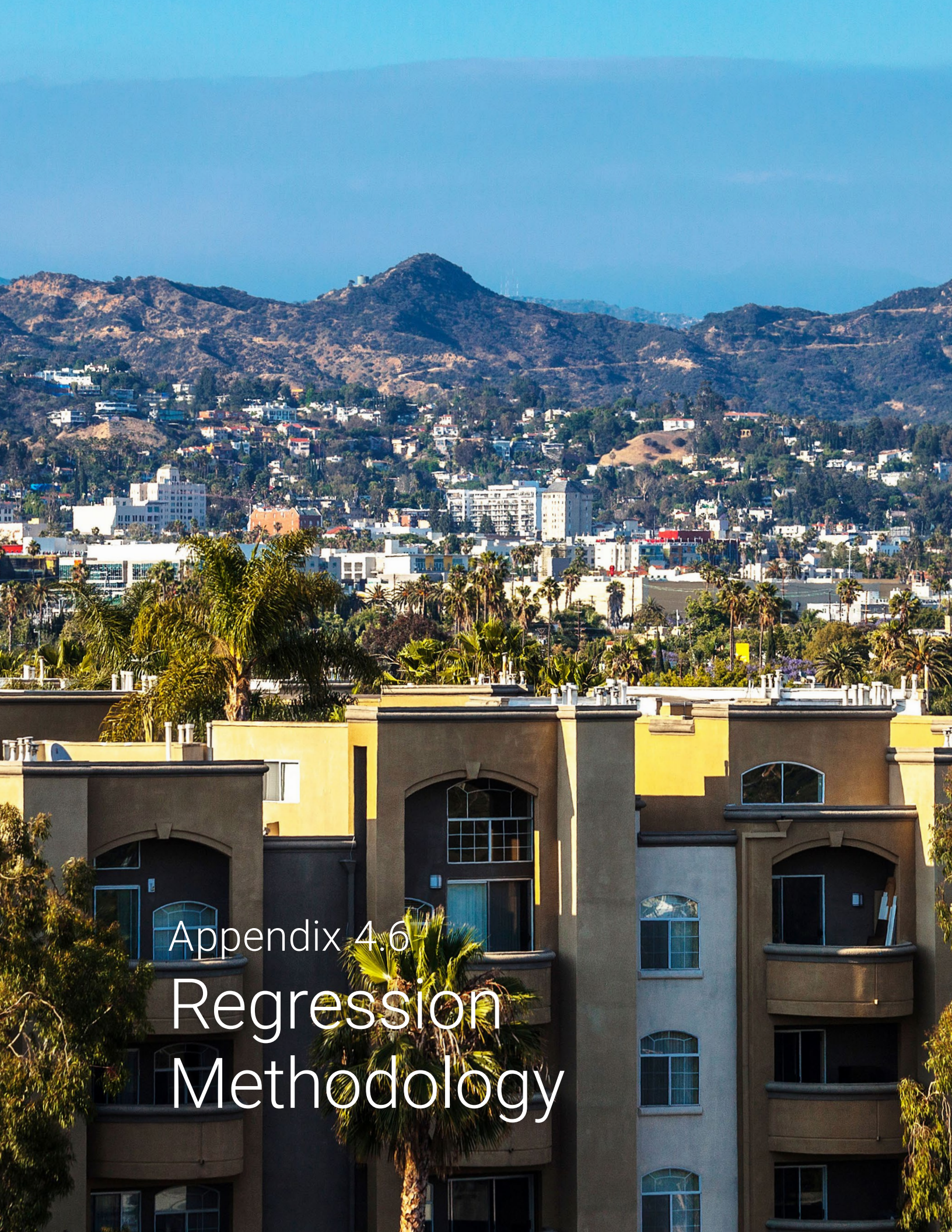


Appendix 4.6

Regression Methodology

Appendix 4.6

Technical Summary of Econometric Approach to Estimating Housing Element Site Capacity in the City of Los Angeles

Prepared by: Issi Romem, Ph.D.; MetroSight and UC Berkeley Terner Center for Housing Innovation

---

1. Overview:

1.1. Background:

The City of Los Angeles is required by law to submit to the State of California its Housing Element for the upcoming RHNA cycle later this year. AB 1397 (2017) now requires that, for each site in the Housing Element, the City specify the site's realistic capacity for new housing units during the upcoming cycle. It also requires the City to explain the methodology used to determine that capacity, and for the methodology to consider a list of factors related to existing uses, past experience, the market environment and regulatory incentives.

MetroSight has been hired by the Terner Center to assist the City by developing and implementing an econometric approach for estimating site capacity based on the City's past experience, as reflected in data on housing production. This memo describes the methodology developed and its results.

1.2. Zoned capacity for housing in the City of Los Angeles:

The City of Los Angeles currently has about 1.38m existing housing units in the sample of parcels considered for its housing element. If each parcel that allowed for housing was redeveloped as 100% residential to the maximum number of allowable (base) units, the number of housing units in the City would more than double. The City's zoned capacity has increased since 2010, especially after accounting for bonuses, likely due to the introduction of the City's Transit-Oriented Communities (TOC) program.

However, the presence of additional zoned capacity does not guarantee that development of that capacity will occur within a given timeframe, or at all. In the 11 years from 2010 to 2020, only about **128k new units** were permitted in the city.

What prevents zoned capacity from materializing into new housing production? First, even when zoned capacity exists, it does not guarantee that development is financially feasible. Regulatory costs and barriers that are not reflected by zoned capacity come into play, including additional requirements such as parking minimums, an ever-more-complex building code, fees, and legal obstacles posed via the California Environmental Quality Act (CEQA), as well as uncertainty around the entitlement timeframe and approval (including the final unit count). In the case of Accessory Dwelling Units (ADUs), financing barriers as well as constraints on splitting lots can limit production. In addition, construction costs and other market factors also influence the financial feasibility of development in conjunction with regulatory factors.

Second, the take up of worthwhile development opportunities does not generally occur right away. The pursuit of such opportunities typically requires the acquisition of the property by a developer, and such sales are generally driven and timed by property owners' extraneous circumstances.[1]

In practice, the take-up of development opportunities suggested by the existence of available zoned capacity has been very gradual. From 2010 to 2020 only about **0.2 percent** of parcels in the City had new units permitted each year.

On the flip side, zoned capacity is not always the limiting factor for take-up of development opportunities. Even after accounting for bonuses, only about **55 percent** of new units permitted from 2010 to 2020 fell within their parcels' zoned capacity.[2] The rest either involved changes in zoned capacity or exceptions to it.

1.3. Modeling the permitting of new units empirically:

This study uses data on permitting from 2015 to 2019 to model the likelihood of new units being permitted on each parcel and their number. The model accounts for parcel's zoned capacity before and after bonuses,

[1] There are many additional complexities involved here. For example, the knife-edge financial feasibility of development that arises due to regulatory and other costs may result in developers' only being able to pursue development opportunities on land banked during opportune times in the real estate cycle. Another example is that the complex and often political nature of the entitlement process (in general, not just in Los Angeles) raises the bar for new developers to pursue opportunities outside their familiar territory, which likely limits the pool of developers and their bandwidth for pursuing opportunities in the city. Yet another example involves the availability of capital for development, which in recent decades has (anecdotally) skewed away from small- and mid-scale development to focus on larger opportunities--a development sometimes linked with the reduced number and role of local and regional lending institutions.

[2] See table enumerating total new units permitted in Section 3.

as well as the market conditions and various other factors which the city is required to address by law in its specification of site capacity for the upcoming RHNA cycle.

The study then applies the model to current data, including changes in zoned capacity, market conditions and other factors, in order to predict that likelihood and unit number going forward. Crucially, the model is constrained such that the predicted number of units on a site may never exceed its bonus-zoned capacity. This aligns the predicted site capacity with legal obligations under RHNA.

The model predicts that, within the bonus-zoned capacity, the parcels considered for the city's housing element would result in 29,505 new units permitted within 5 years. Adjusting that number for an 8-year period such as the RHNA cycle raises the number to **47,208** new units. (Applying simple adjustments to approximate the corresponding numbers of new units that would be permitted beyond the bonus-zoned cap and including parcels outside the set considered for the city's housing element yields estimates that exceed the city's past permitting performance.)

The model's prediction of approximately 47,000 new units being permitted in the city within the bonus-zoned cap in the span of 8 years falls an order of magnitude short of the city's upcoming cycle RHNA allocation of **456,643** units.

However, two important caveats to the estimates above are in order. The first is with respect to the city's TOC program. While that program's implications for bonus-zoned capacity have been accounted for in the model, its implications for other aspects of development have not been. For example, the relaxation of minimum parking requirements in projects subject to TOC bonuses may raise the likelihood of new units permitted on a parcel beyond the impact of higher bonus-zoned capacity alone. This study includes analysis gauging the additional impact of other concessions in the TOC program on the likelihood of new units being permitted. We estimate that other TOC related benefits would raise the 8-year prediction for new units permitted within the bonus-zoned cap from 47,208 to **61,158**.

The second caveat is that the nature of the model is inherently predictive, not causal. That means that the model does *not* estimate the causal relationship whereby a change in zoned capacity induces a change in the permitting of new units. Rather, it predicts the permitting of new units that, under the set of observed circumstances, is likely to correlate with different levels of zoned capacity. The distinction is important because it means that policy changes that alter zoned capacity may not result in

corresponding changes to permitting per the model's prediction.[3] In reality, the model is likely capturing a mixed causal and correlational relationship between zoned capacity and the permitting of new units, so its predictions can be taken--cautiously--as indications of permitting levels that would be obtained under different zoned capacity scenarios. Confidence in the model's predictions should vary inversely with how far future circumstances are assumed to stray from those observed in data used for estimation.

1.4. This memo:

The remainder of this memo is structured as follows:

- Section 2 describes the data sources used to inform the model.
- Section 3 provides information on the city's past performance with respect to zoning capacity and the permitting of new units as it emerges from the data.
- Section 4 introduces the model, as follows:
  - Section 4.1 presents the methodology.
  - Section 4.2 relates the model to the requirements set forth in the law.
  - Section 4.3 addresses the model's estimation and accuracy
  - Section 4.4 characterizes the model's predictions
  - Section 4.5 discusses the approach's limitations
- Section 5 introduces the Transit-Oriented Communities (TOC) exercise

[3] For example, suppose that zoned capacity is higher in areas where market conditions render development more worthwhile or where neighbors' effective opposition to development is weaker, and it is for those reasons--as opposed to greater zoned capacity per se--that larger numbers of new units are permitted in areas with higher zoned capacity. In that case, because raising zoned capacity in an area does not change market conditions or neighbors' opposition, changing zoned capacity would not influence the permitting of new units.

2. **Data**

The study incorporates the following data:

2.1. **Data on "base-zoned units:"** *The number of housing units allowed on each city parcel (PIN) circa 2010 and 2020*. These numbers reflect base zoning as well as additional land use policy that applies to each property (such as overlays, specific plans, General Plan Footnotes, D limitations, and Q conditions,), compiled by city staff and accompanied by data on the underlying land use policies used to determine them. The numbers exclude the influence of bonuses.

Those data are available for samples of approximately 719k and 727k parcels in 2010 and 2020, respectively. The 2010 Sample is essentially contained within the 2020 Sample.

The 2020 Sample comprises the set of parcels that may be considered for the city's housing element, as they reflect parcels which are zoned to permit residential uses.

2.2. **Data on "bonus-zoned units:"** *Estimate of the number of housing units allowed on each city parcel with bonuses*, in 2010 and 2020. Bonuses can be applied in different ways, e.g. depending on the share of subsidized units and their rent level, so rather than there being a single number that reflects allowed capacity with bonuses there is in reality a menu of such numbers. The numbers provided here reflect assumptions as to which bonus structure to apply to each parcel, based on eligibility for a range of affordable housing incentive programs (including Density Bonus, TOC, and Community Plan Implementation Overlay Zones).

Those data are available for the same samples as the data on base-zoned units.

2.3. **Building permit data:** *An extract of all building permits reflecting an addition to the number of housing units on a site,* issued by the city for the 11-year period from Jan 1, 2010 to Dec 31, 2020. These data include information on the change in unit count, as well as the timing of permits. Associated demolition permits were not included in the extract. The data are at the city parcel (PIN) level, and in cases in which a single parcel is associated with multiple permits, we estimated the relevant overall change in unit count and assigned it to a single point in time.

2.4. **Assessor data:** *Los Angeles County assessor rolls* from 2010, 2015 and 2019 (the latest currently available). These were used to obtain information on existing unit counts, structure ages and square footage (for obtaining Floor Area Ratios (FAR)), as well as information on existing

use categories. These data are at the county parcel level (APN/BPP) and were reconciled with city parcel data.

2.5. **Spatial data:** *Spatial data on the full universe of city parcels*, obtained from the city's public-facing GeoHub website, used for observing that full universe (as opposed to just the 2010 and 2020 Samples), as well as for spatial matching with supplementary data.

2.6. **Supplementary data:** *A variety of data from third-party sources,* including the U.S. Census' American Community Survey (rental vacancy rates, household income and demographic information), publicly-available data on home values and rents from Zillow, and proprietary data on commercial property leases and their terms acquired by the Terner Center from Compstak.

3. Past performance:

3.1. The universe of parcels and the 2010 and 2020 Samples:

The full universe gleaned from the spatial data includes 851,095 city parcels. The 2010 and 2020 Samples, i.e. the sets of parcels with base- and bonus-zoned units data as of 2010 and 2020, respectively, include 718,691 and 727,301 city parcels.

The overwhelming majority of parcels have a zoned capacity of 1-4 units, with the remainder consisting of those with a zoned capacity of 5-50 units. Parcels with zoned capacity greater than 50 units have in the past comprised less than one percent of parcels. However, changes since 2010, including the TOC program, have raised the latter's share as measured by bonus-zoned capacity above the one percent mark.

Parcel counts by number of base- and bonus-zoned units						
Sample	**Base vs bonus**	**0 units**	**1-4 units**	**5 to 50 units**	**>50 units**	**Total**
2010 Sample	Base	19,966 (2.8%)	579,269 (80.6%)	115,011 (16.0%)	4,445 (0.6%)	718,691
	Bonus	9,968 (1.4%)	555,023 (77.2%)	148,011 (20.6%)	5,689 (0.8%)	
2020 Sample	Base	20,580 (2.8%)	581,851 (80.0%)	119,589 (16.4%)	5,281 (0.7%)	727,301
	Bonus	1,231 (0.2%)	553,987 (76.2%)	162,859 (22.4%)	9,224 (1.3%)	

3.2. Upzoning and downzoning:

Over the course of a decade, approximately one percent of parcels were downzoned with respect to base- and bonus-zoned capacity, and just 1.3 percent was upzoned with respect to base-zoned capacity. However, as much as 16 percent of parcels were upzoned with respect to bonus-zoned capacity, presumably reflecting the TOC program.

Parcels upzoned or downzoned by 1 or more units between 2010 and 2020[4]				
Base vs bonus	**Downzoned**	**Unchanged**	**Upzoned**	**Total**
Base	7,331 (1.0%)	702,195 (97.7%)	9,165 (1.3%)	718,691
Bonus	9,263 (1.3%)	594,532 (82.7%)	114,896 (16.0%)	718,691

3.3. Existing units and zoned capacity:

As noted in the overview, the City of Los Angeles currently has about 1.38m existing housing units in the sample of parcels that can be considered for its housing element. If each parcel in the sample was redeveloped as 100% residential to the maximum number of allowable (base) units, the number of housing units would more than double. However, the likelihood of seeing new units permitted on a given parcel each year is very low (see Section 3.5 below).

[4] Upzoning and downzoning at the parcel level can only be observed for parcels observed both in 2010 and 2020.

3.4. Total new units permitted:

The number of new units permitted each year emerged from the previous decade's housing bust and increased throughout much of the last decade, peaking in 2017.[5] Note that neither the 2010 nor the 2020 Sample include the full scale of new unit permitting in the city. A substantial amount of new units was permitted outside of those samples. Note also that only about 54 and 56 percent of the new units permitted in the 2010 and 2020 Samples, respectively, fall within the bonus-zoned cap.

Time period	All parcels	2010 Sample		2020 Sample	
		All permits	Within bonus-zoned cap[6]	All permits	Within bonus-zoned cap[7]
2010	4,271	2,892	1,850	2,960	1,920
2011	6,917	3,088	1,825	3,553	2,351
2012	7,103	5,133	2,550	5,163	2,635
2013	10,112	7,060	3,717	7,285	3,968
2014	12,460	6,913	3,647	7,498	4,137
2015	16,122	10,615	5,471	10,790	5,673
2016	15,337	10,919	5,332	11,611	5,943
2017	15,404	10,965	6,209	11,888	6,744
2018	15,078	10,519	6,042	12,152	7,115
2019	14,110	9,987	5,600	10,696	6,316
2020	11,518	9,085	5,197	10,157	6,114
2015-2019	**76,051**	**53,005**	**28,654**	**57,137**	**31,791**

[5] New units in Los Angeles likely skew heavily towards large multifamily rental developments--a sector which experienced a lot of new supply coming online in the latter part of the decade.

[6] This column sums up the number of units permitted after assigning every city parcel a number of units that is the lower of (i) the number of permits actually issued, and (ii) the number of bonus-zoned units for the parcel (circa 2010). In other words, this column reports the number of units permitted after omitting units permitted *in excess* of the bonus-zoned units for the parcel.

[7] The previous footnote applies here as well, with the number of bonus-zoned units for the parcel being circa 2020, instead of 2010.

2010-2020	**128,432**	**87,176**	**47,440**	**93,753**	**52,916**

3.5. The likelihood of permitting new units:

The following table provides backward-looking figures with respect to parcels' likelihood of seeing new units permitted, the number of new units conditional on such permitting and the composition of the two, which can serve as comparisons for forward-looking predictions of these figures from the later model.

Although it varies over time, the average parcel's average annual likelihood of seeing new units permitted was approximately 0.2 percent per year.

Time period	**Share of parcels yielding permits for new units[8]**		**Average number of new units per parcel, conditional on yielding permits**		**Average number of new units per parcel, unconditional**	
	2010 Sample	**2020 Sample**	**2010 sample**	**2020 sample**	**2010 Sample**	**2020 Sample**
2010	0.101%	0.101%	3.99	4.04	0.0040	0.0041
2011	0.102%	0.101%	4.22	4.83	0.0043	0.0049
2012	0.125%	0.125%	5.70	5.69	0.0071	0.0071
2013	0.154%	0.153%	6.38	6.55	0.0098	0.0100
2014	0.186%	0.185%	5.17	5.58	0.0096	0.0103
2015	0.234%	0.232%	6.32	6.39	0.0148	0.0148
2016	0.233%	0.238%	6.52	6.71	0.0152	0.0160
2017	0.316%	0.321%	4.83	5.10	0.0153	0.0163
2018	0.233%	0.232%	6.27	7.21	0.0146	0.0167
2019	0.197%	0.204%	7.06	7.22	0.0139	0.0147

[8] Note that, in principle, the share of parcels yielding permits for new units over the course of a multi-year period may be smaller than the sum of that share calculated separately for each year, because some parcels may yield permits repeatedly in different years. In this case, the annual shares do add up to the multi-year share, because of the way in which all permits issued for a parcel are reduced to having a single date. See earlier footnote on this matter in the data section.

2020	0.151%	0.155%	8.35	9.00	0.0126	0.0140
2015-2019	**1.213%**	**1.226%**	**6.08**	**6.41**	**0.0738**	**0.0786**
2010-2020	**2.032%**	**2.046%**	**5.97**	**6.30**	**0.1213**	**0.1289**

4. Model:

4.1. Methodology:

The model predicts two outcomes of interest:

- *The likelihood of new units being permitted on each parcel*
- *The number of new units permitted on each parcel, conditional on new units being permitted*

The model is estimated ("trained") using permits issued on parcels in the 2010 Sample from the beginning of 2015 to the end of 2019 (5 years), and conditional on a variety of parcel attributes observed prior to the onset of that period, including base- and bonus-zoned capacity circa 2010. The parcel attributes included as covariates in the model correspond to the various factors which AB 1397 (2017) requires cities to consider in their methodology for determining site capacity (more on that below).

The estimated model is then used to predict the future outcomes of interest for parcels in the 2020 Sample from the beginning of 2021 to the end of 2025 (5 years), conditional on the same variety of parcel attributes used in estimation--but updated to their values as of 2020--including base- and bonus-zoned capacity.

The model consists of two steps:

- **Step 1:** The likelihood of new units being permitted on a parcel is estimated for the full 2010 Sample using a **logit regression model**. The logit model ensures that predicted probabilities of new units being permitted fall within the [0,1] range.
- **Step 2:** The conditional number of new units permitted on a parcel is estimated for the subset of parcels in the 2010 Sample which had new units permitted, using a **fractional logit regression model**.[9]

  The observed outcomes informing such a model must be shares, i.e. numbers on the [0,1] interval, and its predicted values are numbers on this

[9] Whereas the logit regression model is ubiquitous in economics and machine learning, the fractional logit regression model is esoteric. Both the logit regression model and the fractional logit regression model are estimated by maximizing the same log-likelihood function: $\log L = \Sigma_j (y_j \log(\Lambda(\mathbf{x}_j'\boldsymbol{\beta})) + (1-y_j) \log(1-\Lambda(\mathbf{x}_j'\boldsymbol{\beta})))$, where y_j is the outcome variable for parcel j, $\mathbf{x}_j$ is a k x 1 vector of covariates, $\boldsymbol{\beta}$ is a k x 1 vector of coefficients, and $\Lambda(z)$ is the logistic cumulative distribution function, $\Lambda(z) = e^z / (1+e^z)$. The difference is that in the logit regression model y_j takes on only the discrete values 0 or 1, whereas in the fractional logit regression model, y_j takes on continuous values in the [0,1] interval. A brief introduction to the fractional logit regression model can be found in Wooldridge, J. M. 2002. *Econometric Analysis of Cross Section and Panel Data.* 1st ed. Cambridge, MA: MIT Press, p. 661-663.

interval as well. In this case, the outcome is defined as the ratio of the number of new units permitted and the parcel's bonus-zoned capacity.[10] Thus, the fractional logit regression model ensures that the predicted number of new units permitted is never (below zero or) above the bonus-zoned capacity.

In many cases, the number of new units permitted on a parcel between 2015 and 2019 exceeds the bonus-zoned capacity circa 2010, suggesting that zoned capacity was either altered by the time the permits were issued, or that some type of exception was made. In such cases the numerator of the outcome variable is truncated at the bonus-zoned capacity, e.g. if 11 new units were permitted on a parcel with a bonus-zoned capacity of 10, the outcome for that parcel is taken to be 1 = 10/10, instead of 1.1 = 11/10, which be outside the [0,1] range.

The separation into steps addressing the two outcomes--the likelihood of new units permitted and their conditional number--is useful for several reasons. The permitting of new units is a fairly sparse outcome, and estimating the emergence of a sparse outcome *and* its gradations (the number of units) in a single model would be challenging. Additionally, the separation into steps allows the set of covariates on which each step is conditioned to differ. For example, the age of the existing structure on a parcel at the onset is likely to matter for whether new units are permitted, but less likely to matter for how many new units are permitted.

[10] In rare instances, the base-zoned capacity observed in the data for a parcel exceeds the bonus-zoned capacity (for clarity, the bonus-zoned capacity includes the base-zoned capacity and any further capacity owing to bonuses). To address such cases, the denominator is defined as the greater of the base- and bonus-zoned capacities.

The covariates included in the model are as follows:

Covariate	Included in Step 1	Included in Step 2
The number of base-zoned units[11]	✓	✓
The number of bonus-zoned units (the sum of base-zoned units and any additional units allowed per bonus)[12]	✓	✓
The ratio of existing units to base-zoned units	✓	
A set of indicators for each of Los Angeles' four market area types[13]	✓	✓
A set of indicators for broad existing-use categories: Commercial, Industrial, Institutional, Recreational and Residential (as well as Miscellaneous and Missing), drawn from county assessor records	✓	
A set of indicators for existing structure age buckets: 0-25 years, 25-50 years, and >50 years, as well as a "missing" group (which largely captures vacant lots)[14]	✓	
A set of indicators for Floor-to-Area Ratio (FAR) buckets: 0%, 0-50%, 50-100%, 100%+, as well as a "missing" FAR group[15]	✓	
An indicator for existing structures subject to the Los Angeles' Rent Stabilization Ordinance (RSO)	✓	
The Community Plan Area's (CPA) ratio of total permitted units to total base-zoned units over the prior 5-	✓	✓

[11] In Step 2, the coefficient on base-zoned units is allowed to vary for each of Los Angeles' four market area types.
[12] Similarly, in Step 2, the coefficient on bonus-zoned units is also allowed to vary for each of Los Angeles' four market area types.
[13] For the grouping of Community Plan Areas into four market area types, see the map on Page 5 here.
[14] To be accurate, this "missing" age group captures a combination of parcels which have no structure on them, even if they are not strictly vacant (e.g. a parking lot), and ones that have a structure whose age information is missing in the records.
[15] The 0% FAR group captures vacant lots, similar to the "missing" age group, whereas the "missing" FAR group likely captures a mixture of vacant lots (whose square footage is recorded as missing rather than zero), and of parcels with structures whose square footage information is missing in the records.

year period		
The log of the typical estimated home value in the zip code area, drawn from the Zillow Home Value Index (ZHVI)[16]	✓	✓
The log of the typical estimated asking rent in the zip code area, drawn from the Zillow Observed Rent Index (ZORI)[17]	✓	✓
The average rental vacancy rate in the Census Public Use Microdata Area (PUMA) during the prior 5-year period	✓	✓
The average remaining lease duration for commercial properties in the Community Plan Area (CPA), drawn from Compstak data	✓	
Separate intercepts for parcels with 1-4, 5-50 and >50 base-zoned units	✓	✓

In addition, both steps of the model are designed such that the influence of the covariates can differ freely depending on a parcel's base-zoned capacity category. Specifically, both steps of the model interact with each of the above covariates with indicators for base-units zoned falling in the 1-4 unit range, the 5-50 unit range, and the >50 unit range.[18] That means that each step of the model actually consists of three independent sub-models: One for each of the three base-zoned capacity categories.

4.2. The model's relation to the requirements set forth in the law:

As noted in the overview, AB 1397 (2017) now requires that cities specify the realistic capacity for new units during the upcoming RHNA cycle and explain the methodology used to determine it, which must consider a list of factors.

[16] The model is estimated using the average monthly ZHVI (or ZORI) value for the zip code area for 2015, and predictions use the corresponding information for 2020. Zillow often suppresses information at the zip code level when it is informed by too few observations or exhibits suspicious abnormalities (this occurs more often with respect to ZORI than with respect to ZHVI). In zip code areas whose ZHVI (or ZORI) is unavailable, the ZHVI (or ZORI) for the nearest zip code area with available ZHVI (or ZORI) data was used (using Census zip code tabulation area centroids). Incorporating home values and rents with a log transformation is common practice, and improves the model's fit.

[17] See previous footnote, which applies to ZORI as well.

[18] According to the planning department's staff, site plan review is required once a threshold of 50 proposed units is reached, whereas a by-right standard applies for projects proposing fewer than 50 units.

Specifically, AB 1397 (2017) requires the following (item enumeration, bracketed text and emphasis added).

> *"Section 65583.2 (g) (1) For [relevant sites], the city or county shall specify the additional development potential for each site within the planning period and shall provide an explanation of the methodology used to determine the development potential. The methodology shall consider factors including[:]*
>
> *[i] the extent to which **existing uses** may constitute an impediment to additional residential development,*
> *[ii] the city's or county's **past experience** with converting existing uses to higher density residential development,*
> *[iii] the current **market demand** for the existing use,*
> *[iv] an analysis of any **existing leases** or other contracts that would perpetuate the existing use or prevent redevelopment of the site for additional residential development,*
> *[v] **development trends**,*
> *[vi] **market conditions**, and*
> *[vii] regulatory or other **incentives** or standards to encourage additional residential development on these sites."*

The correspondence between the model and factors *i* through *vii* required in the law is as follows.

I. "The extent to which **existing uses** may constitute an impediment to additional residential development" is reflected by the conditioning of the model on existing use categories and on existing structures' age and FAR, as well as the Rent Stabilization Ordinance and the remaining local commercial existing lease duration variable.

II. "The city's or county's **past experience** with converting existing uses to higher density residential development" is captured by the basic premise of the exercise: Predicting future permitting based on an empirical estimate that draws on the city's recent (5-year) experience. That experience consists primarily of converting existing uses--including less-dense residential use--to higher density residential development.

III. "The current **market demand** for the existing use" is reflected in the existing use indicators, as well as the remaining local commercial existing lease duration variable and the local information on rental vacancy rates and on residential property values and rents.

IV. "An analysis of any **existing leases** or other contracts that would perpetuate the existing use or prevent redevelopment of the site for additional residential development" is not directly addressed by the

exercise, however it is indirectly broached by the inclusion of the local remaining existing commercial lease duration variable. City staff have incorporated subsequent steps outside of the Model to remove any sites from the Sites Inventory that have an existing regulatory agreement or other regulatory protection related to affordable housing units that would preclude the redevelopment of the site during the planning period.

V. "**Development trends**" are captured by the basic premise of the exercise as explained in item *II*, and also by the local ratio of total permitted units to total base-zoned units over the prior 5-year period. That variable essentially captures the recent level of permitting for housing in the area. (The division by total base-zoned units is necessary for that variable to not simply convey the size of the Community Plan Area, and to account for regulation-imposed differences in past permitting, as opposed to market-driven development trends.)

VI. "**Market Conditions**" are captured by the local ratio of total permitted units to total base-zoned units over the prior 5-year period that was just mentioned, as well as the local information on rental vacancy rates and on residential property values and rents, the remaining local commercial existing lease duration variable, and the set of existing land use indicators, as well as the City's identification of four residential market area types.

VII. "regulatory or other **incentives** or standards to encourage additional residential development on these sites" are reflected by the distinction between base-zoned units and bonus-zoned units, as well as the inclusion of their ratio. In particular, bonus-zoned units capture--to a limited extent--the influence of the Transit-Oriented Communities (TOC) program.[19]]

4.3. Model estimation and accuracy:

- **In total:**

  From the beginning of 2015 to the end of 2019, the parcels observed in the 2010 Sample yielded permits for **28,654** new units within the bonus-zoned cap.[20]

  That is the benchmark which the model ideally ought to predict for those same parcels over that period, and indeed the model gets

[19] The reflection of TOC in the bonus-zoned units variable corresponds *only* to the program's influence on the bonus-zoned unit count; it does *not* reflect other distinct aspects of the program, such as its influence on minimum parking requirements or the entitlement process. This is captured in a latter step (see Section 5 of this Appendix).

[20] That is the number of units permitted on those parcels after omitting units permitted *in excess* of the bonus-zoned units for each parcel.

very close, producing a backward-looking *prediction* of **28,542** new units permitted.[21]

Thus, the model seems to perform well with respect to *total* new units permitted, but how do the two steps of the model perform with respect to the accuracy of individual parcel predictions?

- **At the parcel level:**

The likelihood of new units being permitted on a parcel in a given period is affected by its location, by development trends and market circumstances, and by characteristics of the site and the existing use. However, it is also subject to a great deal of idiosyncrasy: The owner's financial and life circumstances (e.g. heirs selling a deceased parent's property), developers' past successful land banking, lot assembly, and so forth.

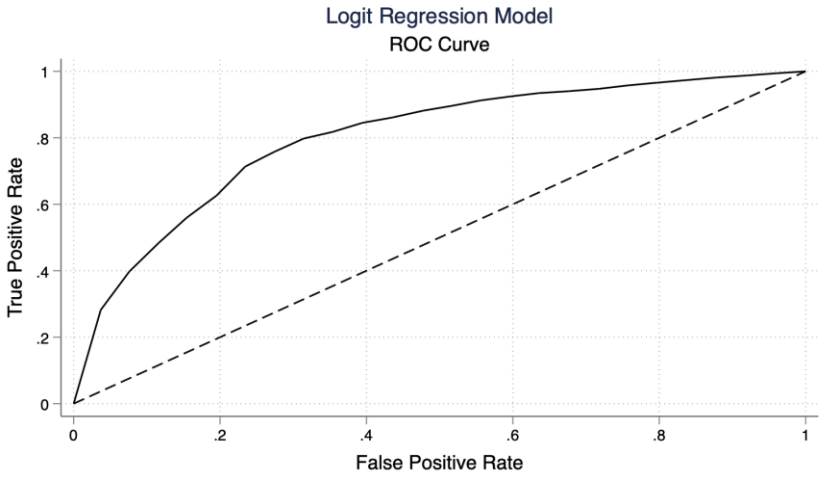
As a result, accurately predicting new units being permitted on a *specific* parcel during a given period is more challenging than predicting the total number of new units permitted in a large set of parcels. This challenge is reflected in Step 1 of the model, which explains only a fairly limited share of variation in outcomes at the parcel level.[22]

Yet despite that, Step 1 of the model has substantial predictive power--meaning that it performs substantially better than a random guess--even at the individual parcel level. To see this, consider the predictions made in Step 1. Those predictions consist of a set of numbers between 0 and 1, one for every parcel in the 2010 Sample, that reflect the (backward-looking) predicted likelihood that the parcel will have seen new units permitted. Suppose also that when a parcel's predicted likelihood is greater than some threshold, say 0.5, we interpret that as a prediction that the parcel will have seen new units permitted and, vice versa, we interpret a likelihood below 0.5 as a prediction of no new units being permitted. The performance of the resulting parcel-level predictions can be compared with the *actual* observations for those parcels, i.e. whether they did or didn't in fact see new units permitted.

[21] That is an in-sample (training set) prediction, which could potentially reflect overfitting. A corresponding backward-looking out-of-sample prediction obtained using five-fold cross-validation came in at **28,726** new units permitted.

[22] The logit regression model has a (McFadden) pseudo-R^2 of 0.126. An OLS linear probability model presented later in this memo (and which also includes some explanatory variables reflecting household income and race/ethnicity) has an R^2 value of 0.038.

The following chart, known as an ROC curve, plots the true-positive rate (TPR) against the false positive rate (FPR), i.e. the share of those parcels that actually had new units permitted which were also correctly predicted to do so by the model, and the share of those parcels that had no new units permitted but were falsely predicted to have them by the model. The solid ROC curve reflects the combinations of TPR and FPR that would be obtained by setting different values between 0 and 1 for the discrimination threshold above, instead of just the 0.5 example.[23] In contrast, the dashed diagonal represents the TPR and FPR that would be obtained by assigning each parcel a *random* likelihood of seeing new units permitted, i.e. the diagonal is the ROC curve of an uninformed guess.[24]



Source: UC Berkeley Terner Center; Analysis by MetroSight.

Notes: The ROC curve is based on five-fold cross validation within the 2010 sample; the AUC is just over 0.8.

[23] To see how it works, suppose for example if we set a threshold of 0, implying a prediction of new units permitted on *all* parcels, we would obtain a TPR of 1 because every parcel that actually had new units permitted would be predicted to have them, but it would also have an FPR of 1, because every parcel that didn't see new units permitted would also be predicted to have them. The opposite would be true if we set the threshold at 1. The solid ROC curve results from setting a sufficient range of different thresholds between 0 and 1, and reporting the actual TPR and FPR obtained from the predictions and the data on actual permitting of new units in the 2010 Sample.

[24] To see why, suppose we set the threshold for interpreting a likelihood as a prediction of new units being permitted at 0.17, i.e. any parcel whose randomly assigned likelihood exceeds 0.17 would be predicted to have new units permitted. The TPR would be 0.83 because, among those parcels that actually had new units permitted, 17% would have randomly assigned likelihoods below 0.17 and 83% above it (remember that the random likelihood is assigned to each parcel independently of their actual outcome). Similarly the FPR would be 0.83 because, among those parcels that actually had no new units permitted, 17% would have randomly assigned likelihoods below 0.17 and 83% above it (just like the other group), and the latter's predictions would be false positives. The entire diagonal can be derived by assigning each parcel a random likelihood and applying all possible thresholds between 0 and 1.

The fact that the solid ROC curve produced by Step 1 of the model lies above the diagonal indicates that the predictions made by Step 1 make better predictions than a random guess.[25]

How much better? One measure used to qualify ROC curves is the area under the curve (AUC), and the ROC curve for step 1 has an AUC of 0.801. How good is that? While there is not strictly an answer to that question, one text has suggested the following rule of thumb (bullets and bracketed note added).[26]

> *"0.5 = This suggests no discrimination, so we might as well flip a coin. [Note the random guess diagonal has an AUC of 0.5.]*
> *0.5-0.7 = We consider this poor discrimination, not much better than a coin toss.*
> *0.7-0.8 = Acceptable discrimination*
> *0.8-0.9= Excellent discrimination*
> *>0.9 = Outstanding discrimination"*

The ROC curve used to establish the predictive power of Step 1 applies to binary classification models, of which a logit regression model is one, however the fractional logit regression model used in Step 2 does not fall into this category. Step 2 predicts the (conditional) number of new units permitted on each parcel *as a fraction of the bonus-zoned units for the parcel*. Thus, Step 2 predicts a number between 0 and 1 for each parcel, where a number of 0.9, for example, indicates a prediction that if new units were permitted on the parcel, they would amount to 90% of the bonus-zoned capacity for the site.

How well does Step 2 perform? To gauge that, its performance was compared to that of random guesses by assigning every parcel used in the Step a random prediction between 0 and 1, and observing how large an error was produced. An error, in this context, means the absolute difference between the predicted fraction of a parcel's bonus-zoned units that would be permitted, and the actual fraction that was observed to be permitted in the 2010 Sample. Repeating the process of randomly assigning predictions 1,000 times yielded a mean absolute error of **0.425**, i.e.

[25] For instance, using 0.12 as the threshold for interpreting the model's likelihoods as predictions of new units being permitted would yield a true positive rate of about 3 in 4 and false positive rate of about 1 in 4.
[26] Hosmer Jr, David W., Stanley Lemeshow, and Rodney X. Sturdivant. *Applied logistic regression*. Vol. 398. John Wiley & Sons, 2013, p.177.

an error of 42.5 percentage points on average above or below the actual fraction of the bonus-zoned units that was permitted. In contrast, the predictions from Step 2 yielded a mean absolute error of **0.121**, i.e. an error of just 12.1 percentage points on average.[27] Once again, there is no strict answer to the question "is that good?", however it does imply that Step 2 of the model has substantially better predictive power than a random guess.

4.4. Model prediction:

After the model's estimation, it is applied at the individual parcel level to the 2020 Sample in order to make *forward-looking* predictions.

- **In total:**

  The model predicts that, within the bonus-zoned unit cap, the parcels included in the 2020 Sample would see permitting for **29,505** new units within 5 years. Adjusting that number for an 8-year period yields **47,208** new units.[28]

  If that number is further adjusted for the rate of past permitting *above and beyond* the bonus-zoned unit cap--i.e. owing to changes in the applicable land use policy and/or exceptions made with respect to it--the parcels in the 2020 Sample would yield **89,741** new units over 8 years.[29]

  And if the number were even further adjusted to account for past levels of permitting in the city *outside* of the 2020 Sample--on parcels that were excluded from the sample ostensibly because they did not allow residential use as of 2010 or were otherwise constrained--it would be **119,447** over 8 years. However, due to requirements of State Housing Element law, the Sites Inventory cannot consider development potential that cannot be achieved without the need for legislative action.

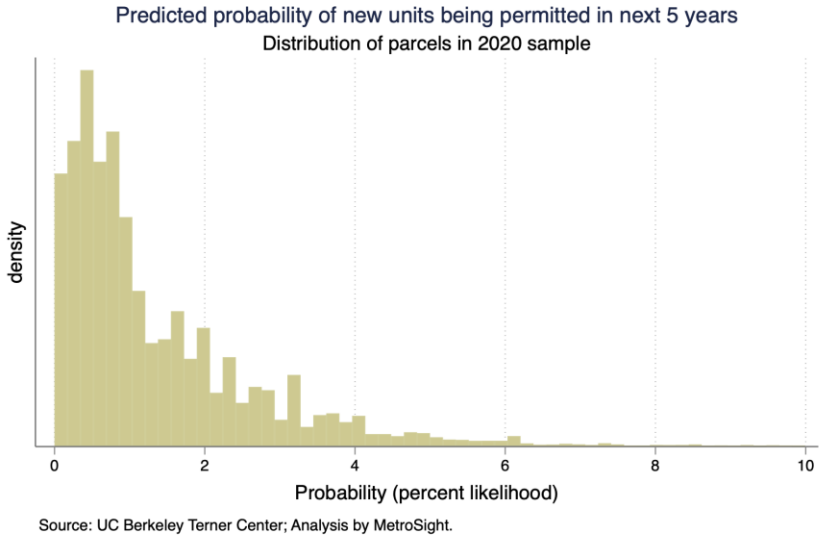
- **The parcel-level likelihood of having new units permitted:**

[27] The mean absolute error reported was obtained by way of five-fold cross-validation within the 2010 Sample.

[28] The adjustment is simply a multiplication by 8/5.

[29] The adjustment involves multiplying by the ratio of *total* new units permitted in the 2020 Sample during the 5 years from 2015 to 2019 (57,137), and new units permitted in that sample during those years *within the bonus-zoned unit cap* (30,057).

The overwhelming majority of parcels have a very low probability of seeing new units permitted during the next 5 years, however there is a long thin right tail of parcels with higher likelihood of development.



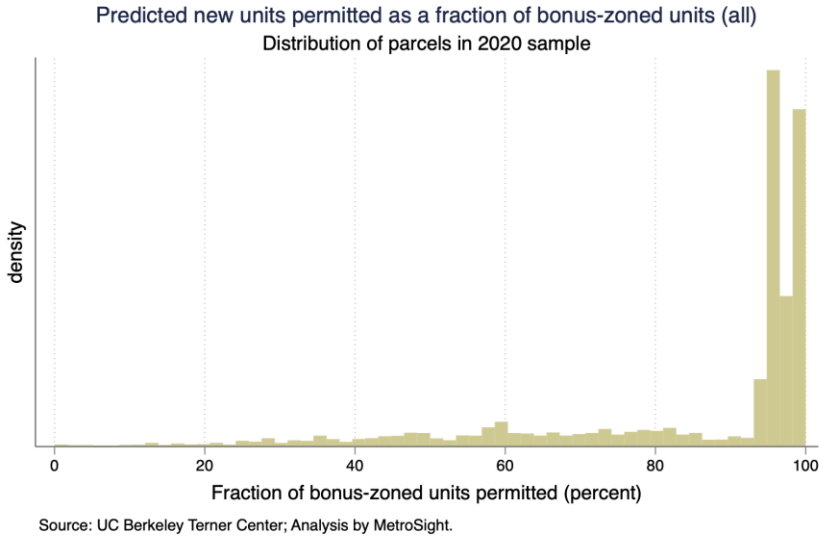
Percentile	Predicted probability of having new units permitted in next 5 years			
	All parcels	Base-zoned capacity		
		1-4 units	5-50 units	>50 units
1st	0.04%	0.04%	0.03%	0.09%
5th	0.11%	0.13%	0.08%	0.17%
10th	0.19%	0.24%	0.11%	0.21%
25th	0.45%	0.52%	0.23%	0.39%
50th	0.89%	1.01%	0.46%	0.72%
75th	1.95%	1.95%	1.22%	1.47%
90th	3.26%	3.26%	3.50%	2.84%
95th	4.30%	4.30%	4.40%	4.82%

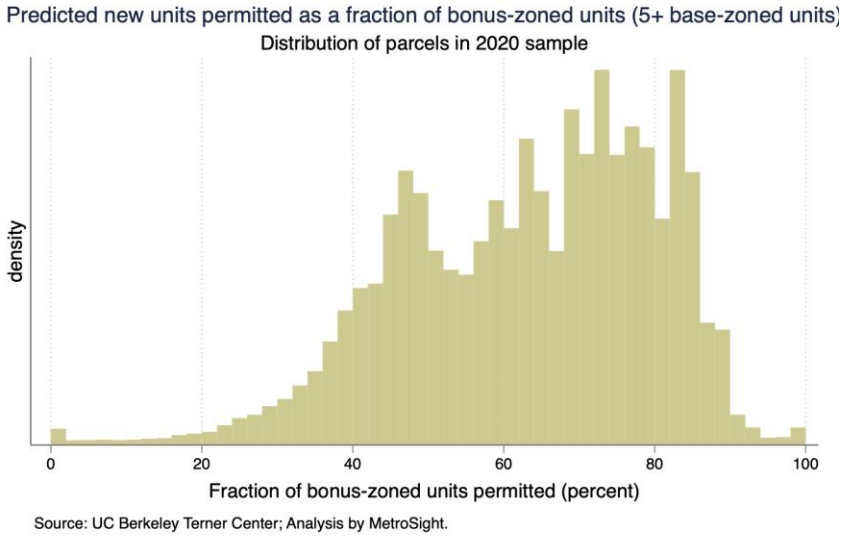
99th	8.58%	8.64%	7.31%	12.04%
Mean	**1.49%**	**1.56%**	**1.15%**	**1.44%**
Parcels	**725,541**	**602,011**	**118,915**	**4,615**

- **The fraction of bonus-zoned units permitted per parcel, conditional on new units being permitted:**

  Conditional on seeing new units permitted, a substantial share of parcels are predicted to have new unit counts ranging from 10 to 95 percent of the bonus-zoned capacity. However, more than half of parcels are predicted to yield new unit counts between 95 and 100% of the bonus-zoned capacity. Those parcels concentrated between 95-100% of bonus-zoned capacity consist primarily of parcels with base-zoning in the 1-4 unit range.

  Limiting the set of parcels to those with base-zoning of 5 or more units leads to a more even distribution, with the bulk of parcels having predicted numbers of new units typically ranging between 40 and 90 percent of bonus-zoned capacity.





- **The number of new units permitted per parcel:**

  More than half of the parcels in the 2020 Sample are predicted--conditional on having new units permitted--to yield just one new unit (in practice, this often shows up as a predicted conditional number of new units permitted below 1). Indeed, the 99th percentile parcel in this respect is predicted to see 37.2 new units permitted.

  However, a disproportionate share of predicted new units are concentrated among parcels with a high predicted unit count: Half of new units are predicted to emerge on parcels with 12.4 or more units permitted. Similarly, as much as a quarter of new units are predicted to emerge on parcels with 32.1 or more units, and ten percent of new units are predicted to emerge on parcels predicted to have at least 88.5 units.

Predicted number of new units permitted per parcel[30]					
Percentile of all parcels	**Conditional**	**Unconditional**	**Percentile of total predicted new units**	**Conditional**	**Unconditional**
1st	0.93	0.0005	1st	0.95	0.0045
5th	0.95	0.0017	5th	0.96	0.0095
10th	0.95	0.0032	10th	0.98	0.0160
25th	0.96	0.0064	25th	2.58	0.0398
50th	0.99	0.0135	50th	12.4	0.127
75th	2.32	0.0321	75th	32.1	0.507
90th	8.54	0.0724	90th	88.5	2.27
95th	14.3	0.1300	95th	224	6.79
99th	37.2	0.4290	99th	1,265	31.4
Mean	**3.65**	**0.041**	**Mean**	**62.0**	**2.90**

4.5. Limitations and cautions:

A number of limitations and cautions ought to be highlighted:

- As noted in the overview, the predictive methods used in the study amount ultimately capture correlation, not causation.
- A potentially important limitation of the study is that it does not fully account for lot assembly. In cases in which the permitting of new units follows lot assembly, and in which such lot assembly results in city parcel identification numbers *changing* between the time at which 2010 Sample was observed and the time of

[30] "Conditional" or "unconditional" refers to conditioning on new units being permitted (whereas the conditional case is conditioned on new units being permitted, in the unconditional case the predicted number of units also reflects the likelihood of seeing new units permitted). The figures correspond to the 2020 Sample.

permitting, the permitting of new units on such parcels will fail to be observed. Such cases would likely bias estimates in both steps of the model downwards.

- Another limitation of the analysis (due to time constraints) is that it does not distinguish between new units permitted through addition or alteration from those permitted through the construction of a new building.

- Both the set of control variables (covariates) included in Steps 1 and 2 and the regression methods used in those steps affect the predictions. The execution of this study included a fair amount of specification testing and trial of different alternative methods, and those presented were chosen because they are--in order of importance--feasible, valid, compelling and simple. However, alternative specifications may lead to different results.

5. TOC exercise:

The implications of the Transit-Oriented Communities (TOC) program with respect to bonus-zoned capacity are accounted for in the model. Thanks to the program, sites within certain distances of transit that place them within the program's four tiers now have greater bonus-zoned capacity than they did before, and this increase is reflected in the 2020 Sample data, and therefore in the model's predictions.

However, the implications of TOC on other aspects of entitlement are absent from the model. For example, as noted in the overview, the relaxation of minimum parking requirements in developments subject to TOC bonuses may very well raise the likelihood of seeing new units permitted on a parcel beyond the impact of a higher bonus-zoned capacity on its own.

To gauge the magnitude of the TOC program's influence on the likelihood of seeing new units permitted on a parcel and their number, a difference-in-differences exercise was conducted. The exercise involves estimating the outcomes during the periods before and after the introduction of TOC, and doing so separately for parcels inside and outside of the areas subject to the program (those subsets of parcels are referred to as the "treatment" and "control" groups). Suppose that the difference between the outcome in the before and after periods *outside* of the TOC areas reflects the general trend in the outcome over time, whereas that difference *inside* the TOC areas reflects that general trend *as well as* the influence of the TOC program. In that case, the difference between the before and after differences inside and outside of the TOC areas--the so-called difference in differences--isolates the influence of the TOC program.

The difference-in-differences exercise applied here proceeds essentially along the lines just described and using a linear regression, but with some nuances. First, as the TOC program was introduced in November 2017, the exercise uses the 3-years 2015-2017 as the "before" period and the year 2020 as the "after" period.[31] Second, the estimates are conditioned on all of the different covariates included in the corresponding steps of the model--Step 1 for the likelihood estimate and Step 2 for the estimate of the conditional number of new units permitted--so that the estimates from the exercise may be interpreted as the program's effect holding all of those factors fixed. Finally, in order to omit the direct influence of the TOC program via parcels' bonus-zoned capacity, the value

[31] Due to the time lag between introduction of a new zoning system and the time it takes for individual development projects to reach the building permit state, the impact of the TOC Program is unlikely to emerge in the data until sufficient time has passed from its onset. In order to best capture the effects of the TOC Program on permitting activity the model defines the "after" period as the year 2020, which is the latest full year of available data and it begins two years after the program's creation.

of the bonus-zoned capacity covariate is updated from its 2010 value to its 2020 value for observations in TOC areas in the "after" period. That modification controls for the change in bonus-zoned capacity generated directly by the TOC program, and so the estimated effect of TOC derived from the exercise should be interpreted as the effect of TOC above and beyond that direct influence via the bonus-zoned capacity.

The regression specification used for both outcomes--the likelihood of having new units permitted, and their number conditional on such permitting--is as follows:

$$y_{it} = \alpha \cdot after_t + \gamma \cdot treat_i \ + \delta \ \cdot after_t \cdot treat_i \ + X_i\beta + \epsilon_{it}$$

Where y_{it} is either an indicator for whether the observation corresponding to parcel *i* in period *t* (the "before" or "after" period) had new units permitted during that period, or the number of new units permitted conditional on such permitting; $after_t$ is an indicator for the observation corresponding to the "after" period; $treat_i$ is an indicator for parcel *i* being inside TOC areas; X_i is a vector of all the covariates included in Step 1, with bonus-zoned units updated to its 2020 (prediction set) value for observations with $after_t = 1$ and $treat_i = 1$; and ε_{it} is an error term.

The difference-in-differences estimates for the effect of the TOC program (the estimate of γ) are a negligible change in the likelihood of having new units permitted, and a 29.0 percentage point increase in the conditional number of new units permitted as a share of the parcel's bonus-zoned capacity.

Adjusting the likelihood and conditional number of units in the forward-looking predictions using those estimated TOC-driven boosts raises the 8-year prediction for new units permitted within the bonus-zoned cap from 47,208 to **61,158**.[32]

[32] In the event that the TOC boosts raised either the likelihood above 1 or the conditional number of units as a share of bonus-zoned capacity above 1, the boosted values were truncated at 1.

Endnote from City Staff:

Upon receipt of the parcel-level model predictions, additional modifications and adjustments were made to the results, which led to an overall reduction in the total anticipated development potential reflected through the model (see Chapter 4). This included adjustments to:

- Remove vacant parcels located in a Very High Fire Hazard Severity Zone (VHFHSZ), as those sites are presumed to have additional impediments to development that may not be captured in the model;
- Remove parcels containing restricted affordable units that are subject to a land use covenant or other regulatory program that would preclude their redevelopment during the 8-year period;
- Remove parcels that would not permit an overall net increase in residential units or had an unconditional predicted unit potential of zero units; and
- Remove parcels that are otherwise unlikely to redevelop based on the current use (such as public schools or active government services).

As a result of these adjustments, the model anticipates a total development potential of 42,764 units over the eight-year period.