DREAM
Danish Research Institute for
Economic Analysis and Modelling

MAKRO

Land and Non-land: Substitutes in Danish Housing Production

Christian Vikkelsø

Working paper 2024
30 August 2024

www.dreamgruppen.dk

Land and Non-land: Substitutes in Danish Housing Production

Christian Vikkelsø

Working paper
August 2024

www.dreamgruppen.dk

Abstract

This paper aims to estimate the elasticity of substitution between land and non-land factors in Danish housing production, which is a parameter utilized in MAKRO. By using the latest property and land assessments from Vurderingsstyrelsen (the Danish Property Assessment Agency), a CES-production function is estimated simultaneously with a house producing agent's first order conditions, resulting in an estimate of 1.14. This aligns closely with the currently employed estimate in MAKRO, 1.16, derived from a study by Epple et al. [2010], who rely on American urban area data. Given the Danish macroeconomic context, we expect the estimate proposed in this paper to be more appropriate for use in MAKRO.

1 Introduction

As in most economies, the Danish housing market is of major significance for the financial stability and consumer behavior in the country. A distinctive characteristic of the production of housing is its dependence on land, as land is a limited but strictly necessary input factor. In MAKRO, the housing production is modeled with a CES-production function with land and non-land as inputs. Hence, a reliable estimate of the elasticity of substitution between these two factors is needed. The currently used estimate in MAKRO is taken from Epple et al. [2010], who use American urban data to estimate the micro-elasticity for a particular area. They obtain a value of 1.16 with a non-parametric estimation approach. This paper seeks to estimate the corresponding macro-parameter for Denmark, by estimating two equations simultaneously. One of the estimation equations is a CES-production function, and the other is a relative formulation of the first-order conditions of a house-producing agent's problem. As estimation data, the most recent property assessments from Vurderingsstyrelsen (the Danish Property Assessment Agency) are utilized. This approach yields an elasticity of substitution of 1.14.

The structure of this paper is as follows: in section 2 the general assumptions made in the estimation are presented. The estimation approach is described in section 3, and the data used is presented and discussed in section 4. A simulation study related to the discussion of the data is found in section 5, before the results are presented in section 6 and concluding remarks are made in section 7.

2 Assumptions on Housing Production

Like in Epple et al. [2010], we assume housing Y consists of two components: non-land factors M, and land factors L. Non-land factors includes all mobile factors used in the production of housing, e.g. materials, energy, labor, etc. When producing housing, these factors are essentially collected and packed into what we call a "building" (which will be used synonymously with the word "non-land" in this paper). The land factor is simply the geographical area the building is placed on. Assuming housing is produced by a representative agent who buys land and non-land, and sells housing, the profit is given by:

$$\pi = P_Y Y - P_M M - P_L L \tag{1}$$

P_Y, P_M, and P_L represent the price of housing, non-land, and land, respectively. For simplification, we assume that the house-producing agent is operating in a perfect competition environment. Moreover, for each year of house construction, we normalize the price of housing so that $P_Y = 1$, as well as the price of building, $P_M = 1$.

The $P_M = 1$ assumption is made due to our use of the value of building, $V_{building}$, as a measure of M (see section 4). Hence, to ensure the relationship, $V_{building} = P_M M$, we require $P_M = 1$.[1] The implicit assumption in $P_M = 1$ is, that the construction and material costs of buildings are the same all over Denmark. Even though this is not exactly true, as e.g. construction work and material supply on small islands is more expensive due to transportation costs, we don't expect huge differences in construction costs across Denmark.[2]

Under the assumptions listed above, i.e. $\pi = 0$, $P_Y = 1$, $P_M = 1$, the resulting expression for Y is:

$$Y = M + P_L L \tag{2}$$

Apart from the identity above, ensuring the value of inputs equals the value of output, we also have an expression for Y from an assumption of production. We assume the representative agent produces housing with a CES-function:

$$Y = \gamma \left[\mu M^{\frac{\sigma-1}{\sigma}} + (1-\mu) L^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}} \tag{3}$$

in which μ is the share-parameter for M, γ is the total factor productivity, and σ is the elasticity of substitution between the two production inputs M and L that we want to estimate.

Under the given assumptions, the equations (1) and (3) can be used to obtain the following first order conditions for the market equilibrium:

$$1 = \mu \left(\frac{Y}{M} \right)^{\frac{1}{\sigma}} \gamma^{\frac{\sigma-1}{\sigma}} \tag{4}$$

$$P_L = (1-\mu) \left(\frac{Y}{L} \right)^{\frac{1}{\sigma}} \gamma^{\frac{\sigma-1}{\sigma}} \tag{5}$$

It is these first order conditions, as well as the CES-function, we utilize in the estimation of σ.

3 Estimation Strategy

The method of estimation used in this paper is the Feasible Generalized Non-linear Least Squares (FGNLS) applied on a system of two equations. In the FGNLS procedure, the estimation equations are initially estimated one by one using non-linear least squares. The residuals from these estimations are then used to calculate a covariance matrix to form the FGLS-estimator, that is used to estimate both equations together. The formulation and estimation of the

[1] The level of this P_M is of no importance, because a higher level of P_M would just lead to a proportionally smaller M when we have $V_{building} = P_M M$.

[2] We could (and have tried, see Vikkelsø [2024]) come around this problem of splitting value into price and quantity by using an entirely different measure of quantity, e.g. the living area of the dwelling. However, we don't see living area as a good representation of the quantity of a building, as characteristics such as quality, energy efficiency, etc. are not accounted for. Hence, we have decided to use the building value as the measure of quantity, based on the idea that all these characteristics are satisfactory described by the value.

equations in this working paper is mainly inspired by León-Ledesma et al. [2010] who argues, that estimating the CES-production function together with the first order conditions of the market equilibrium, improves identification. León-Ledesma et al. [2010] also suggest to normalize the variables with respect to the mean observations, hence ensuring the estimation results are invariant to which units are used to measure the variables.

In contrast to León-Ledesma et al. [2010], this working paper formulates the two first order conditions from (4) and (5) into a single estimation equation by setting them relative to each other. It is shown in the DREAM working paper Kastrup and Vikkelsø [2023] that this formulation removes bias towards unity in the estimation results, arising from unsystematic measurement errors in prices. However, Kastrup and Vikkelsø [2023] utilize two price measures, one for each factor input, while this paper due to the $P_M = 1$ assumption only have one, P_L. For this reason, the results from Kastrup and Vikkelsø [2023] are not transferable to this setting, as the measurement errors in the prices don't "cancel out" in the two-equation formulation when P_M is fixed. Nonetheless, formulating the estimation in two instead of three equations makes the estimation process simpler, hence it is the method we use in this paper, while keeping in mind that a bias towards unity in the estimate can be expected.

In conclusion, two estimation equations are estimated simultaneously: the first estimation equation is the CES-production function normalized with respect to the mean observations and set in log, and the second equation is the two first order conditions normalized with respect to the mean observations, set relative to each other, and put in log. That is:

$$\log \frac{Y_i}{\overline{Y}} = \log \xi + \frac{\sigma}{\sigma - 1} \log \left(\overline{\mu} \left(\frac{M_i}{\overline{M}} \right)^{\frac{\sigma-1}{\sigma}} + (1 - \overline{\mu}) \left(\frac{L_i}{\overline{L}} \right)^{\frac{\sigma-1}{\sigma}} \right) \tag{6}$$

$$\log \frac{P_M}{P_{L,i}} = \log \frac{\overline{\mu}}{1 - \overline{\mu}} + \log \frac{\overline{L}}{\overline{M}} + \frac{1}{\sigma} \left(\log \frac{L_i}{\overline{L}} - \log \frac{M_i}{\overline{M}} \right) \tag{7}$$

in which i is the index of the observations, $\overline{Y}$, $\overline{M}$ and $\overline{L}$ represent the geometric means of the respective variables, and $\overline{\mu}$ is the arithmetic mean of the share parameters.

This estimation approach is suitable for both time series and cross sectional data. In this paper, we have cross sectional data, as all the property assessments used in the estimation (see section 4) are made for all Danish properties at a single point in time. I.e. we have data on (almost) all existing dwelling-properties in Denmark in 2023. However, these dwellings were constructed at different points in time which should be addressed when estimating. First of all, the preferences of the house buyers, and thereby the house-producing agent, may have changed over time. It is not certain that the general preference of the land to building ratio of a dwelling was the same in 1950 as it is today. Hence, including properties with buildings from many decades ago may yield an outdated estimate. Therefore, we decide to only use properties with buildings constructed in the time span from 2010 to 2021 in the estimation.[3]

Furthermore, an issue arises from the fact that the residual between assessed property value and assessed land value is used as the measure for the quantity of building, M (described in further detail in next section). This residual assessment is the value of a given building in 2023-prices, which we consider as a good measure of "how much" building a property consists of. However, as we have $P_M = 1$, and $P_M M$ should express the cost the house-producing agent had when he constructed the building (cf. equation (1)), the fact that the assessments are all in 2023-prices makes this interpretation problematic. One could imagine, that if the building was constructed in e.g. 2010, the general relation between construction costs and land costs might have been different than in 2021. We could utilize general land price and construction cost indices, however adjusting e.g. P_M with a construction year dependent index (which we

[3] 2010 is chosen as the first year in the time span to limit the attention to buildings constructed in the past decade, however expanding the time span to begin in 2000 and thereby including another decade, doesn't change the obtained estimate substantially.

can assume to be increasing in time), would yield inaccurate measures for either Y or M. If P_M is adjusted dependent on the year of construction without adjusting the Ms in the same year, older properties Y would be "smaller" than newer in general, as the quantity of housing cf. (2) is given by $Y = P_M M + P_L L$ (so, smaller P_M would lead to smaller Y). If instead the Ms are adjusted as well to still maintain the $V_{building} = P_M M$ relation, older buildings would be systematically measured larger than newer.

To avoid these issues, we decide to treat each year in the time span 2010 to 2021 by itself, meaning for all $t \in [2010, 2021]$, we estimate a t-specific elasticity of substitution by only estimating on properties with buildings constructed in year t. In this way, it is plausible to assume the relation between the general level of P_M and the general level of P_L, the house-producing agent faced when he produced, is equal for all the observations used in the same estimation. This implies the respective measures of M, L, and Y are comparable for all observations within the same estimation procedure.

4 Data

The data used to set up the estimation variables are land area and the year of construction for owner-occupied dwellings gathered from the Danish BBR-register, as well as property and land assesments from Vurderingsstyrelsen (Danish Property Assessment Agency) of November 2023. The BBR-register contains data on all buildings in Denmark, and Vurderingsstyrelsen assesses the values of all owner-occupied dwellings, including both a land valuation and a property valuation. The assessments from Vurderingsstyrelsen are found on vurderingsportalen.dk and has for this project been web-scraped during December 2023. The webscraped assessments includes an adress-ID that can be linked with adress-IDs in the BBR-register, making it possible to construct a data set for all owner-occupied dwellings in Denmark containing land area, year of construction, and the public valuations of both the land and the property in total.

4.1 Property and Land Assessments

It has been possible to web-scrape and match data between vurderingsportalen.dk and BBR for 1 879 472 properties. Though, only a small subset of these are used in the estimation, as it will be explained below. Before that, a small discussion of the measures constructed for M, L, P_L, and Y is needed.

Vurderingsstyrelsen writes the following about their assessments on their web-page:

- Property assessment: the value of the entire property. I.e. the value of the land as it is exploited and the buildings that are placed on it (Vurderingsportalen [2024]).
- Land assessment: the value of the land if it was undeveloped. The land is assessed based on what would be the optimal utilization, that would yield the highest land value (Vurderingsportalen [2024]).

Note that these definitions in someway contradict each other when it comes to the concept of "land assessment". In the property assessment, the land is seen as a part of the property "as it is exploited", while the land assessment concerns what could be understood as the market value of the land. Land value "as it is exploited" is meaningless under the assumption of a competitive housing market, and it implies that a decomposition of the property assessment $V_{property}$ into the published land assessment V_{land} and an unpublished building assessment $V_{building}$ is not possible, according to Vurderingsstyrelsen. I.e. Vurderingsstyrelsen does not propose that $V_{property} = V_{land} + V_{building} \implies V_{building} = V_{property} - V_{land}$ is meaningful. Going down Vurderingsstyrelsens line of thought, one could imagine that the concept of land value utilized when coming up with $V_{property}$ is represented by another valuation, $V_{land}^{exploited}$, implying the decomposition $V_{property} = V_{land}^{exploited} + V_{building}$ in which both $V_{land}^{exploited}$ and $V_{bygning}$ is unknown, as they are not published by Vurderingsstyrelsen. However, it must be true that $V_{land} \geq V_{land}^{exploited}$, as it would be nonsense for the land owner to have exploited the land to more than its potential and thereby making the land more valuable than its market value. The equality will hold true, if the land owner has exploited the land exactly to its full potential. When we construct the measures for the estimation variables used in this paper, we assume the equality holds true for all dwellings in the data set. I.e. we make the assumption that $V_{property} = V_{land} + V_{building}$. In consequence, the value of the buildings will tend to be undershot, as the V_{land} variable used when calculating $V_{building} = V_{property} - V_{land}$ is potentially larger (but never smaller) than the "true decomposition variable" $V_{land}^{exploited}$. The bigger the discrepancy is between $V_{land}^{exploited}$ and V_{land}, the more undershot is the estimate $V_{building}$. In the same manner, it is also the case that $V_{property}$ is assessed too low in relation to its "true" market value, if V_{land} is far from $V_{land}^{exploited}$. To avoid the most extreme cases of this, dwellings in which $V_{land} \geq V_{property}$ are excluded from the estimation data, which also ensures that the residual $V_{building}$ is strictly positive in all cases. The effects on the elasticity estimate from $V_{building}$ and $V_{property}$ potentially being too low is explored in a simulation study presented in section 5.

4.2 The Constructed Measures

As mentioned above, we have data for 1 879 472 properties. However, a series of exclusions is imposed on this data set. Firstly, condominiums, houses on land not owned by the house-owner, and undeveloped land are excluded from the data set, which is 19.5% of the data. The concept of "land" belonging to a condo unit is abstract and will typically be shared between multiple owners, which is why the condominiums are excluded. The cases of house-owners who don't own the land their house is put on, as well as the data entries with undeveloped land, are excluded, because these don't fit into the way this paper treats "housing" as a combination of "land" and "building". Note that holiday homes are kept in the data set.

As explained in section 3, we only want to estimate on properties with buildings constructed in the time span of 2010-2021, to ensure the elasticity estimate is not affected by "old" preferences. This leaves 77 156 observations without condominiums, etc.

For properties with $V_{land} > V_{property}$, a negative building value will be obtained if we do as described in the subsection above, i.e. $V_{building} = V_{property} - V_{land}$. One could argue that a negative building value is just a result of the building being effectively worthless, and therefore a tear down cost exists, however, as house-producing agents presumably don't produce houses for tear down, the existence of negative building value doesn't make sense in the framework of this paper. Furthermore, a negative as well as 0 building value will not work when applying the log-function in the estimation. Hence, we find it necessary to exclude observations with $V_{land} \geq V_{property}$ from the estimation, which is the case for 467 of the remaining 77 156 observations. Another advantage of the $V_{land} \geq V_{property}$-exclusion is, that it might remove some of the most extreme cases in which a large discrepancy between V_{land} and $V_{land}^{exploited}$ is present (see discussion of this in the subsection above).

In addition, properties with either $V_{property} = 0$, $V_{land} = 0$, living area less than 10 m^2 or land area less than 10 m^2 are removed, as these are probably observational errors. It is the case for only 8 of the remaining observations. This leaves 76 681 properties in the time span 2010-2021, ready for estimation. Table 7.0.1 in Appendix shows number of observations, distributed on the 12 years of the time span and different regions of Denmark.

The estimation variables are assigned the following measures present in the data set:

- $M \equiv V_{property} - V_{land}$
- $P_M \equiv 1$
- $L \equiv$ Land area

- $P_L \equiv \frac{V_{land}}{\text{Land area}}$
- $Y \equiv P_M M + P_L L$

Note that Y above is formulated as in (2), but it is easily seen this equals to $V_{property}$. Summary statistics of the estimation variables are found in table 7.0.2 in Appendix.

The land-related variables L and P_L are measured in m^2 and $\frac{\text{kr.}}{\text{m}^2}$, respectively. M is measured in kr., and as P_M is set to 1 for all observations, it is unitless.

5 Simulation Study: The Effect of the Observed Residual Being Too Low

As described in section 4.1, we suspect the assessed property value $V_{property}$ and the residually given building value $V_{building}$ are too low, if we take the assessments of the land value when "optimally exploited" seriously. With the notation from the previous section, the bigger the discrepancy is between $V_{land}^{exploited}$ and V_{land}, the more undershot are the assessments of property and building values.

Note that it is unclear exactly how one should think of the land assessment V_{land} and how reliable this estimate is. It is not given that the uncertainties between $V_{property}$, V_{land} and their residual $V_{building}$ are as its laid out in this paper. Though, the simulation study can still give a sense of the magnitude of the effect from systematic measurement errors on the elasticity estimate.

With these reservations laid out, the simulation study will now be described. First, a level of scaling between $V_{land}^{exploited}$ and V_{land} is selected as $\varepsilon^* \in (0, 1]$. The assumed relation between the two variables will be $V_{land}^{exploited} = \varepsilon^* V_{land}$. Selecting $\varepsilon^* = 1$ is assuming all land areas are optimally exploited, hence no measurement errors exists. Selecting $\varepsilon^* = 0.7$ is assuming the house-producing agent generally only exploits 70% of the land's potential.

After the selection of ε^*, 10 000 triplets of $\log(V_{property})$, $\log(V_{land})$ and $\log(L)$ are drawn from a multivariate normal distribution with mean vector and covariance matrix corresponding to the empirical. From these, the following variables are constructed:

- $M^{obs} \equiv V_{property} - V_{land}$ (the usual "observed" building value)
- $M^* \equiv V_{property} - \varepsilon^* V_{land}$ (the "true" building value, when ε^* is assumed)
- $P_M \equiv 1$
- $P_L \equiv \frac{V_{land}}{L}$
- $Y^{obs} \equiv M^{obs} + P_L L$
- $Y^* \equiv M^* + P_L L$

Then two estimations can be conducted and compared to measure the effect of the discrepancy between $V_{land}^{exploited}$ and V_{land}, i.e. M and Y being measured too low. The first estimation is based on the "observed" simulated data M^{obs}, L, P_L, and Y^{obs}, and the other on the "true" simulated data M^*, L, P_L, and Y^*.

This process is repeated 1 000 times, and several ϵ^*s are tested, namely $\varepsilon^* \in \{1, 0.95, 0.9, 0.85, 0.8, 0.75, 0.7\}$.The results are presented in figure 5.0.1. It is seen, that the lower the ε^* is, the greater an upwards bias is seen in the estimates based on the observed variables in which M and Y are measured too low. The difference in median estimates for different ε^* are shown in table 5.0.1. The table shows, that changing ε^* from 1 to 0.7 (from perfect exploitation of land to 70% exploitation), reduces the true estimate with 0.10. We don't see this as a substantial deviation in the estimate, hence the estimation method seems satisfactorily robust against systematic measurement errors of this sort.

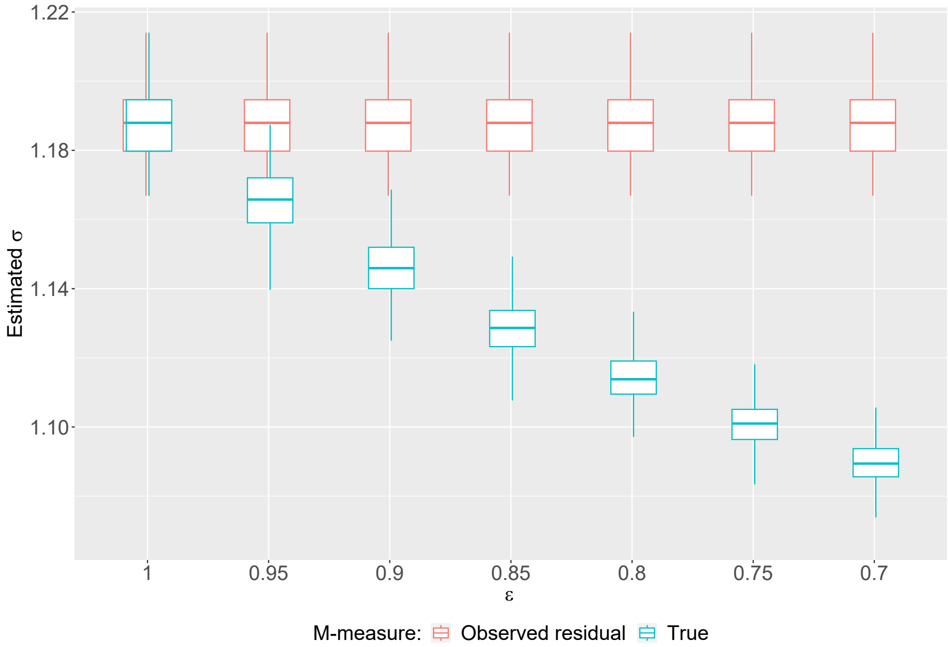

Figure 5.0.1: Simulation study. The selected ε^* is shown on the first axis, and the estimates of the elasticity of substitution between M and L is shown on the second axis. Each box represents minimum, maximum (the whiskers' endpoints) and first, second, and third quartile (the box) of 1 000 estimations with 10,000 draws of $\log(V_{property})$, $\log(V_{land})$ and $\log(L)$ in each simulation. The "true" simulations are based on $M^*$and Y^*, and the "observed" simulations are based on M^{obs} and Y^{obs}.

ε^*	median($\hat{\sigma}^{obs}$) − median($\hat{\sigma}^*$)
1.00	0.00
0.95	0.02
0.90	0.04
0.85	0.06
0.80	0.07
0.75	0.09
0.70	0.10

Table 5.0.1: Summary of the results from the simulation study. The second column presents the difference in the medians of 1 000 estimates based on M^{obs}, Y^{obs} and M^*, Y^*, respectively.

6 Results

Using 76 681 properties, we estimate an elasticity of substitution between the land and non-land factors in Danish housing production for each year in 2010-2021. The results are shown in table table 6.0.1. The mean elasticity of substitution across the years is 1.143, and the median estimate is 1.141. Hence, we suggest 1.14 as our preferred estimate of the elasticity of substitution. Note we expect a bias towards unity arising from unsystematic measurement errors in the price of land, due to the method of estimation. On the other hand, it is shown in the simulation study from section 5, that if the land value assessments from Vurderingsstyrelsen are taken seriously, implying the property value (and thus the building value) is assessed too low, the estimates will be slightly upwards biased. Combined with the pull from unity due to unsystematic measurement errors, we are confident, that the elasticity must be above 1, and 1.14 seems like a reasonable estimate. Though, remark that these estimates are based on property assessments as estimation data instead of real world observations, which of course is a major source of uncertainty. Not only is it possible that these assessments are systematically wrong, we also utilize them in a way they are not meant to when we calculate building values residually.

In the Appendix, a robustness check is made by doing similar estimations on the data, when properties with land values close to equal the property values have been excluded (specifically properties for which the land value is greater than 95% of the property value). The idea being, that if land value is almost equal the property value, this might be a case of an owner, who hasn't exploited his land very well. This yields slightly lower estimates, the mean being 1.11. We don't see this as a substantial change, hence it doesn't affect our confidence in the estimates presented in this section.

Year of con.	# properties	$\hat{\sigma}$
2010	4 833	1.138 (0.009)
2011	5 550	1.130 (0.008)
2012	4 741	1.056 (0.007)
2013	4 390	1.139 (0.008)
2014	4 632	1.108 (0.008)
2015	5 288	1.120 (0.007)
2016	6 238	1.197 (0.008)
2017	7 437	1.154 (0.007)
2018	7 806	1.186 (0.007)
2019	8 350	1.163 (0.006)
2020	8 545	1.143 (0.006)
2021	8 871	1.187 (0.006)

Table 6.0.1: Summary of the results from the estimation of elasticity of substitution between land and non-land in Danish housing production, distributed on 12 years of construction. The number of properties used in each estimation are shown in the middle column, and the estimates are reported in the right-most column with standard errors in parenthesis below their corresponding estimate.

7 Conclusion

In summary, this paper proposes 1.14 as a macro estimate for the substitution of elasticity between land and non-land in Danish housing production. Even though this estimate is influenced by several uncertainties, especially due to the rough assumptions on equal construction costs across Denmark, the use of assessments as estimation data instead of "real world" observable data, as well as the use of an estimation procedure with bias towards unity if unsystematic measurement errors are present, we still consider this estimate more relevant for MAKRO than the currently used estimate from Epple et al. [2010]. Epple et al. estimate on American urban data, while this paper utilizes assessments from Vurderingsstyrelsen covering Denmark, which obviously is more relevant within the context of housing modeling in MAKRO. The estimate is further backed by a robustness check presented in the Appendix, in which properties with extreme land-to-property value ratios are excluded from the estimation. Although this yields slightly lower estimates, it doesn't change the qualitative conclusion that we find land and non-land to be substitutes in Danish housing production, i.e. the the elasticity is greater than 1, and we find it plausible to assert it lies around 1.14.

References

D. Epple, B. Gordon, and Sieg H. A new approach to estimating the production function for housing. *American Economic Review*, 100(3):905–924, 2010. (document), 1, 2, 7

C. B. Kastrup and C. Vikkelsø. Overcoming bias in elasticity of substitution estimation: A novel approach with empirical application. https://dreamgruppen.dk/Media/638376480629276921/overcoming_bias_in_elasticity_of_substitution_estimation.pdf, 2023. Arbejdspapir, DREAM. 3

M. A. León-Ledesma, P. McAdam, and William A. Identifying the elasticity of substitution with biased technical change. *American Economic Review*, 100(4):1330–1357, 2010. 3

C. Vikkelsø. Bachelorprojekt, land og bygning - substitutter i dansk boligproduktion. *Københavns Universitet*, 2024. 2

Vurderingsportalen. Vurderingsstyrelsen. https://www.vurderingsportalen.dk/ejerbolig/vurdering/foreloebige-vurderinger/metode/foreloebig-vurdering-af-helaarshuse-og-sommerhuse/, 2024. Visited August 6, 2024. 4.1

Appendix

Descriptive Statistics

Descriptive statistics of the data used in estimation is found in the tables below. In table 7.0.1 the number of observations used in the estimation for each year of the covered time span is shown. The distribution of these observations across different regions of Denmark is also shown. One of the stand outs is Eastern Jutland, with more than 1 000 observations in every year of the time span, and more than 2 000 in the last four years. In contrary, Bornholm stands out as the region with the fewest observations, only varying between 2 and 21 through the years.

Summary statistics of the estimation variables is found in table 7.0.2 (all years pooled together).

Region	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**	**2021**
City of Copenhagen	55	83	142	202	259	161	259	244	268	153	102	84
Gr. Copenhagen	142	284	239	203	287	357	363	522	506	514	369	444
Northern Zealand	296	388	398	387	519	540	602	699	779	673	742	863
Bornholm	7	4	10	5	4	7	5	2	5	3	21	20
Eastern Zealand	166	249	261	296	275	389	565	610	426	473	485	433
W. and S. Zealand	464	524	377	244	249	303	414	602	584	728	621	908
Funen	400	417	300	258	273	320	358	426	422	558	772	730
Southern Jutland	949	909	717	623	592	641	793	898	847	1 133	1 477	1 296
Eastern Jutland	1 119	1 434	1 280	1 097	1 150	1 421	1 633	1 962	2 255	2 217	2 067	2 279
Western Jutland	690	676	531	490	499	515	533	632	710	972	849	889
Northern Jutland	545	582	486	585	525	634	713	840	1 004	926	1 040	925
Total	4 833	5 550	4 741	4 390	4 632	5 288	6 238	7 437	7 806	8 350	8 545	8 871

Table 7.0.1: Number of observations for each year and region.

Variable	M [kr.]	$L[m^2]$	$P_L[\frac{kr.}{m^2}]$	Y [kr.]
Minimum	1 000	24	1	191 000
1st quartile	1 789 000	375	772	2 547 000
Median	2 391 000	811	1 527	3 447 000
Mean	2 784 808	1 817	3 021	4 225 800
3rd quartile	3 178 000	1 020	3 522	4 771 000
Maximum	59 020 000	446 664	87 600	75 140 000
Std. deviation	1 865 678	7 371	4 507	3 030 556
Skewness	5.06	16.09	4.69	4.06

Table 7.0.2: Summary statistics of the estimation variables.

Robustness check

In table 7.0.3, the results from doing a similar estimation as in section 6 are presented, the only change being properties for which the ratio of assessed land value to assessed property value is greater than the 99.9%-fractile (which turns out to be 0.951) are left out from the estimation. This removes 77 observations from the data set, in which one could suspect the owners haven't exploited their land very well. The mean estimate is 1.110, and the median is 1.118. We don't see this as a substantial change from what was presented in section 6.

Year of con.	# properties	$\hat{\sigma}$
2010	4 830	1.115 (0.008)
2011	5 546	1.087 (0.008)
2012	4 739	1.047 (0.007)
2013	4 387	1.121 (0.008)
2014	4 627	1.077 (0.007)
2015	5 283	1.091 (0.007)
2016	6 224	1.131 (0.007)
2017	7 432	1.135 (0.006)
2018	7 797	1.129 (0.006)
2019	8 344	1.143 (0.006)
2020	8 537	1.114 (0.006)
2021	8 858	1.127 (0.006)

Table 7.0.3: Summary of the results from the estimation of elasticity of substitution between land and non-land in Danish housing production, distributed on 12 years of construction when properties with land to property value assessment ratios above 0.951 are left out from the estimations. The number of properties used in each estimation are shown in the middle column, and the estimates are reported in the right-most column with standard errors in parenthesis below their corresponding estimate.