DANMARKS NATIONALBANK

24 JANUARY 2023 — NO. 1

Housing wealth and consumption during Covid-19

Per Andersen
Senior Data Scientist
FINANCIAL STABILITY

pea@natinoalbanken.dk

Pernille Valentin Borgensgaard
Economist
ECONOMICS AND MONETARY POLICY

pvb@nationalbanken.dk

Andreas Kuchler
Principal Economist
ECONOMICS AND MONETARY POLICY

aku@nationalbanken.dk

Bjørn Bjørnsson Meyer
Data Scientist
FINANCIAL STABILITY

bbm@nationalbanken.dk

Housing wealth and consumption during Covid-19

Abstract

House prices increased considerably in Denmark during the pandemic, while consumption remained moderate. This memo explores the nature of the muted consumption response based on a unique data set with estimates of the value of almost all single- and multi-family houses in Denmark obtained from a machine learning model.

We find that the marginal propensity to consume (MPC) out of changes in housing wealth was modest in the years leading up to the pandemic. In addition, the largest housing wealth gains during the pandemic occurred among household segments with relatively low MPCs, such as households living in large cities with sizeable liquid assets and low loan-to-value ratios.

The Danish housing market has started to slow down. House prices have fallen and the number of transactions has decreased, following the recent surge in inflation and rising interest rates.[1] The slower pace in the housing market comes, as in other countries, after two years with record numbers of transactions and substantial house price increases. Contrary to expectations among most forecasters at the onset of the Covid-19 pandemic, house prices increased almost 20 percent from 2019 to 2022, see chart 1.[2] Rising house prices are usually expected to stimulate consumption through the increase in household wealth and/or through a relaxation of borrowing constraints. However, the aggregate consumption ratio has been relatively low since the Global Financial Crisis. It seems that the historical relationship between the wealth ratio and the consumption ratio has weakened, and the disconnect was particularly pronounced during the pandemic. In particular, the credit channel has been relatively muted over the past decade compared to previous years.

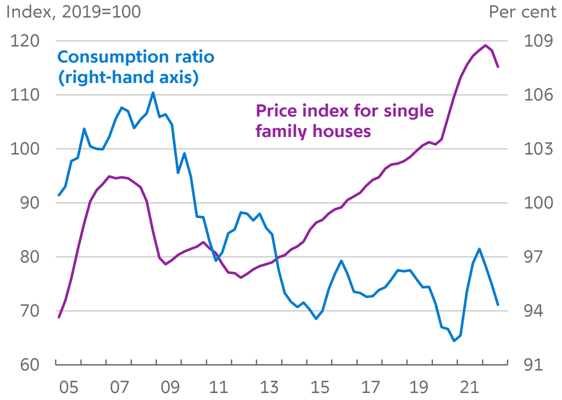
This memo explores whether the distribution of housing wealth gains across household segments could help explain the apparent disconnect between house prices and consumption during the Covid-19 pandemic. It develops and exploits a new data set with estimates of the value of almost all Danish single- and multi-family houses, obtained using a machine learning model based on rich micro data. This data set is unique, because it is available with a remarkably short lag, i.e. within a few months, and can be linked to information from the Danish administrative registers.

Combining house value estimates with administrative data from Statistics Denmark, the memo presents

1 See Danmarks Nationalbank (2022a).
2 Following the outbreak of the Covid-19 pandemic, both Danmarks Nationalbank, the Danish Ministry of Finance, and the Economic Council adjusted their outlooks for the Danish economy to include a fall in house prices in 2020, see Danmarks Nationalbank (2020), Finansministeriet (2020), and De Økonomiske Råd (2020).

House prices increased substantially, but the consumption ratio remained low throughout the Covid-19 pandemic

Chart 1



Note: Countrywide nominal house prices at the quarterly frequency and a 4-quarter moving average of the aggregate consumption ratio.
Source: Statistics Denmark, Danmarks Nationalbank and own calculations.

two key results. First, it reiterates the well-known finding that the marginal propensity to consume (MPC) out of changes in housing wealth is generally low – and has likely been lower in the years leading up to the pandemic than in previous periods. In particular, we estimate an average MPC out of changes in housing wealth of 0.4 per cent in the years from 2015 to 2019. This estimate is lower than levels seen for Danish households in previous periods, which are around 5 per cent on average.[3]

Second, we show that the distribution of house value increases across household segments may also have contributed to the moderate development in consumption during the pandemic. For instance, across household segments of varying ages, loan-to-value ratios, debt-to-income ratios, and liquid asset holdings, more significant housing wealth gains during the pandemic appear to coincide with lower MPCs in the period leading up to the pandemic. Using our MPC estimates from the pre-pandemic years, this heterogeneity implies that the expected average consumption response from 2019 to 2021 is about 30 per cent lower than it would have been if gains were – hypothetically – equally distributed across Danish house owners.

Our results generally indicate that the consumption response to changes in house prices has been relatively modest over the past years. If households react in a symmetric way to increases and decreases in house prices, the direct response of consumption to the recent decrease in house prices will also be limited.

The remainder of this memo is organised as follows. The first section presents our machine learning model of house values. The second section describes the size and distributions of housing wealth gains across households during the pandemic. Sections three and four explain our procedure for estimating MPCs and calculating the expected response to the Covid-19 housing wealth increase, and also relate the results to housing wealth decreases.

Estimating house values with machine learning

A major challenge for estimation of housing wealth is that most houses are rarely traded. Hence, for most of the housing stock, the actual value is unknown. Therefore, we use a machine learning model to estimate house values as counterfactual sales prices (Fagereng et al., 2020). We estimate the value of all Danish single-family and terraced houses.[4]

The model uses as input a wide range of features, including information on housing transactions, house characteristics, geolocation and local amenities. These features are used to explain sales prices for traded houses without strong a priori assumptions about the type of association. For example, the impact of a highway on sales prices in a local area may depend on several other characteristics in a complex and non-parametric way. The model can then be used to predict prices for all the houses that

[3] See Hviid and Kuchler (2017) and Andersen and Leth-Petersen (2021) for evidence for Denmark.

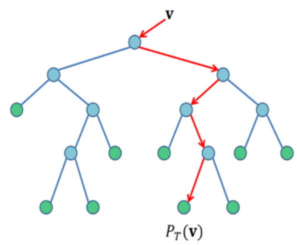
[4] The model does currently not cover owner-occupied apartment units and cooperative housing units of any type.

Machine learning model

Box 1

The data used to train the model comes from BBR, SVUR, and SDI
Data on the house characteristics come from the Danish Building and Housing Register (BBR) and include information such as the size of the property, the size of the main building, the number of bathrooms and the source(s) of heating. The data from BBR are enriched with geodata made available by the Danish Agency for Data Supply and Infrastructure (SDI). The geodata include height above sea level and the coordinates of the house, which we exploit to also compute distance from the sea and distance to the nearest highway. Finally, the sales prices used to train the model come from the Danish Sales and Valuation Register (SVUR). Only sales of single housing units on the open market are included, i.e. not sales to family members or sales of real estate portfolios, and only houses with sales prices in the range from kr. 300,000 to kr. 30 million are included in the model.

The pricing model is a gradient boosted decision tree model
In a decision tree, output is decided upon given a series of questions branching out from the top. Each question can be of the form "is the house smaller or larger than 100 square metres" or "is the house more than 500 metres from a highway". Based on the answers to these questions a path will be defined through the decision tree, eventually arriving at an estimated price of the house. A single decision tree is a weak learner and has poor performance when predicting house prices. Therefore, a series of decision trees are trained, each learning from the mistakes of the previous tree to improve performance. This series of trees makes up our model.



Geographic location and the latest sales prices of similar houses nearby are the most important model inputs
The trained model relies primarily on the geographic location of the home and on the square metre prices of similar houses nearby that have been sold recently to form its forecasts. To account for geographic location in the price model, a time-constant geographic indicator, denoted the price area, is trained on a hold-out sample. The price area indicator is based on the coordinates of the house and acts as a proxy for amenities and other unobserved local drivers of price differences. An average square metre price of similar houses nearby that have been sold within the last year is calculated and given to the model as the main feature taking into account the latest sales prices (Martinello and Møller, 2022).

are not traded. Box 1 provides further details on the data and model structure.[5]

Our sample period is from 2015 to 2021. In each year the model can predict the price of all houses on any given date based on the most recent sales. Throughout this memo, 31 December is chosen as the valuation date. An advantage of the model is that the estimates can be updated as soon as the latest sales data are available, which usually is within a few months.

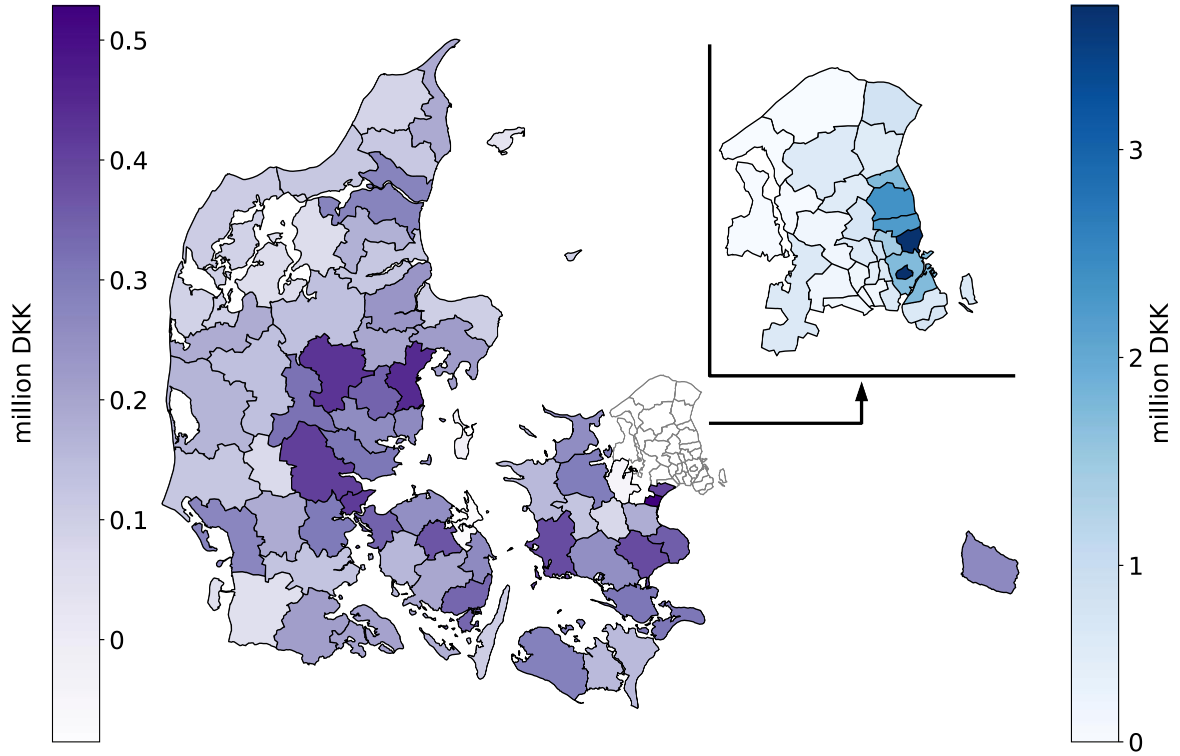
To benchmark the machine learning model, we also estimate a simple index model. An index model is a commonly used method, where either the latest sales price or the latest public valuation is updated with a local house price index. We use a subsample of houses that have been sold twice and estimate the second sales price by adding the increase or decrease in average price per square metre at the postal code level.

In a 'hold-out'- sample of sales that the machine learning model has not seen before, the median error of the estimated sales price (i.e. the difference between the estimated and the actual sales price) is 27 per cent lower in the machine learning model compared to the index model. See appendix A for more details about the model's performance.

[5] A more detailed discussion of the features and performance of the model can be found in appendix A.

Absolute house value increases by municipality, 2019-2021

Chart 2



Note: Based on average values of all single- and multi-family houses based on model price predictions. Greater Copenhagen area is singled out due to higher increases than the rest of the country.
Source: Output from machine learning model.

Different growth of house values across geography and households during Covid-19

To assess the impact of the Covid-19 housing shock, we consider growth in housing wealth from 31 December 2019 to 31 December 2021, both across municipalities and household segments. For the latter exercise, we link house ownership to income, wealth and debt registers from Statistics Denmark. Since we are interested in the impact on consumption of the distribution of housing wealth gains, we primarily consider absolute wealth gains (i.e. wealth gains measured in kr.) in the following.

The largest increases in housing values occurred in large cities

Significant geographical heterogeneity was seen in the distribution of housing wealth gains during the pandemic, see the map in chart 2.[6] Municipalities in rural areas experienced only minor growth in house prices from 2019 to 2021. In many municipalities houses in large cities increased in value by an average amount of between kr. 250,000 and kr. 500,000. However, the municipalities in the Greater Copenhagen area, especially the City of Copenhagen/the City of Frederiksberg and along the coast, saw average house values increase by well over kr. 1 million.

[6] Our model predictions serve as value measures for all houses and not just the traded houses, which provide the basis for the price index in chart 1. Hence, statistics on house value growth here do not necessarily correspond to the aggregate house price indices.

The largest wealth gains were received by high income households with sizeable liquid assets

Table 1

Decile of housing wealth gain, 2019-2021	1	2	3	4	5	6	7	8	9	10
Housing wealth gain, kr. 1,000, median	-209	-19	60	130	200	270	370	520	824	1,920
Housing wealth gain, %, median	-10.60	-1.12	4.26	9.26	13.25	16.67	19.77	22.36	26.97	42.56
Housing wealth, 2019, kr. 1,000, median	2,250	1,350	1,320	1,390	1,500	1,630	1,860	2,330	3,170	4,820
Age, median	58	58	58	59	59	58	58	56	54	53
Income, kr. 1,000, median	702	610	607	619	637	663	707	779	883	1,092
Liquid assets, kr. 1,000, median	274	197	194	209	220	228	252	286	353	541
Liquid assets to income, %, median	39	32	32	34	34	34	36	37	40	48
Debt to income ratio, %, median	282	230	232	235	244	253	269	296	339	391
Ratio of total debt to house value, %, median	60	65	64	63	62	61	60	60	60	57
Total debt, kr. 1,000, median	1,296	939	939	972	1,028	1,108	1,244	1,504	1,931	2,762
Share of families with retired members	31.8	33.9	34.1	34.6	34.8	34.2	32.9	29.7	25.3	20.6
Share of families with higher education	49.9	40.2	40.1	41.4	43.4	45.5	49.3	55.9	65.2	77.5
Share of families in Aarhus and Greater Copenhagen, %	9.2	5.1	4.6	4.7	5.3	6.1	7.9	11.5	20.5	44.3

Note: The table shows descriptive statistics for households by decile of absolute housing wealth gain from end of 2019 to end of 2021. Only households living in single-family houses with an estimated sales price between kr. 300,000 kr. and kr. 30,000,000 have been included. All characteristics are measured as of end of 2019 unless otherwise noted. House values are linked to owners though Statistics Denmark's EJER register, and total housing wealth is calculated by summing owner shares of houses (excluding vacation homes) within a household (family) defined by the BEF register.
Source: Own calculations based on data described in box 1, and data from Statistics Denmark.

The largest increases in housing wealth occurred among high-income households with sizeable liquid assets

Zooming in at the household level, table 1 shows descriptive statistics for households in our sample across deciles of housing wealth gains during the Covid-19 pandemic. We see that households in the highest deciles of wealth gains are, generally, somewhat younger households with high incomes, less leverage (LTV ratio) and large stocks of liquid wealth. We also find that they are to a larger extent located in areas where households typically have larger debt relative to income (DTI), such as Aarhus and Greater Copenhagen.

High-income households with sizeable liquid assets and lower LTV ratios generally have lower marginal propensities to consume (see results in the next section, as well as e.g. Crawley and Kuchler, 2023; Andersen and Leth-Petersen, 2021). Together, this indicates that the expected consumption response of the house price increases during the pandemic would be relatively small.

Table B1 in appendix B presents the same set of descriptive statistics but by deciles of *relative* housing wealth gains (i.e. wealth gains in per cent), instead of *absolute* housing wealth gains. It shows that most of the differences that were present across households with different degrees of absolute wealth gains are not present when splitting by relative wealth gains. There are still differences along the dimensions of income, liquid assets and geography, but much smaller than was the case when considering absolute wealth gains. This indicates that the degree to which households experienced high relative wealth gains was largely explained by the fact that house prices increased to a larger degree in some areas than in others. The combination of higher growth in house prices and a higher ex ante level of house prices in areas around the larger cities seems to have been the main reason behind the heterogeneity seen in table 1.

Estimating consumption responses to changes in housing wealth

We utilise a regression framework in the years 2015-2019 to quantify the relationship between changes in housing wealth and household consumption before the pandemic. We then use the estimated model to predict the consumption response to housing wealth increases during the pandemic. Our key measure of the consumption response is the fraction of a one-krone increase in housing wealth, that a household chooses to add to consumption in the subsequent year. This fraction is usually denoted the MPC (Marginal Propensity to Consume).

Previous evidence suggests that several channels could play a role in transmitting changes in housing wealth to consumption. One way in which, say, an increase in housing wealth could impact consumption is through households becoming richer when their housing wealth increases. Therefore, they may choose to consume a part of this wealth increase even if the wealth is tied up in illiquid housing assets. This channel is called the *wealth effect*. Previous evidence, e.g. Andersen and Leth-Petersen (2021) and Hviid and Kuchler (2017), indicates that the pure wealth effect is relatively small, and that the *collateral effect* may be more important. The collateral effect refers to the fact than when house values increase, the value of houses as collateral also increases. Creditworthy households could therefore increase their borrowing and thereby also their consumption – or, for example, housing investments.

In addition to these two channels, expectations of e.g. income and productivity growth play a role. Expectations of future income are correlated with both house prices and consumption.

If, as the literature suggests, the collateral effect is the dominant channel, we would expect to see larger consumption responses among households that are more likely to be credit-constrained a priori, i.e. for example households with low liquidity and high debt levels. To take this possibility into account, our empirical setup allows for differential consumption responses to house price changes across these dimensions. We do not directly observe households' expectations, but inclusion of household fixed effects in our econometric specification takes account of persistent differences in expectations across households.

Low MPCs out of changes in housing wealth in the period from 2015 to 2019

The econometric model aims to recover the MPC out of changes in housing wealth in the years 2015 to 2019. The model provides a baseline for assessing the sensitivity of consumption to changes in housing wealth over the more recent years for which consumption data are not yet available and where consumption may have been impacted by short-run disturbances such as lockdowns and travel restrictions. Details of the model can be found in box 2.

Empirical approach to estimating MPCs out of changes in housing wealth

Box 2

Data

As a starting point for the analysis, we link the estimated house values obtained from the machine learning model to administrative register data on household wealth, debt and income from Statistics Denmark.

We impute household consumption by subtracting from disposable income the value of net savings and pension contributions, in line with Browning and Leth-Petersen (2003) and Abildgren et al. (2020).[1] This measure of consumption includes spending on e.g. durables and home improvements.

Sample restrictions

The model is estimated based on data from 2015 to 2019. It focuses on households who live in single-family houses with an estimated value of kr. 300,000 to kr. 30,000,000 and have yearly incomes of at least kr. 25,000. It excludes self-employed persons and households with members that are not fully liable for taxation to Denmark, as the registers only contain taxable income and wealth and do not separate private wealth and business wealth. It also omits observations for households involved in real-estate transactions in both the year in which the sale took place and in the previous year. Finally, to reduce the influence of outliers, we censor all non-categorical variables year-by-year at the 1st and 99th percentiles of their distributions.

Specification

We estimate heterogenous MPCs out of changes in housing wealth using the following specification for household *i* in year *t*:

$$\frac{\Delta C_{it}}{Y_{it-1}} = \beta \frac{\Delta H_{it-1}}{Y_{it-1}} + \gamma D_{it-1} + \delta D_{it-1} \times \frac{\Delta H_{it-1}}{Y_{it-1}} + \kappa \frac{\Delta \mathrm{Y_{it}}}{Y_{it-1}} + \mu_i + \varepsilon_{it},$$

where ΔC_{it}, ΔH_{it-1}, and $\Delta \mathrm{Y_{it}}$ are the nominal changes in consumption, housing wealth and disposable income, respectively. In line with Hviid and Kuchler (2017), we scale these variables by disposable income in year *t-1*.[2] D_{it-1} is a vector of dummy variables, taking the value 1 if a household belongs to a specific quartile group of the distributions for liquid assets, LTV ratios, and DTI ratios in year *t-1,* and age in year *t*, respectively. The quartiles are calculated for each year, and only for households in the estimation sample. μ_i is a household level fixed effect. ε_{it} is an error term.

Interpretation

The model yields estimates of the marginal propensity to consume (MPC) out of changes in housing wealth of $\beta + \delta D_{it-1}$, being heterogenous across 256 household segments with different combinations of quartile groups for liquid assets, LTV ratios, DTI ratios and age. Coefficient estimates are presented in table B2 in appendix B.

1. Browning and Leth-Petersen (2003) show that the register-based measure of imputed consumption is fairly consistent with self-reported household consumption from the Danish Expenditure Survey. Abildgren et al. (2020) update the measure of imputed consumption using more recent data and confirm the consistency with survey data on average as well as with national accounts data.
2. This ensures that the effect of control variables is modelled in terms of relative changes in consumption rather than absolute changes. This is preferable since the latter are heavily influenced by the level of consumption.

MPCs out of changes in housing wealth are generally limited. We estimate the average MPC to be 0.4 per cent. This is at the lower end compared to previous estimates for Denmark in earlier years, but it corresponds with the relatively low aggregate consumption ratio in the years prior to the pandemic.[7]

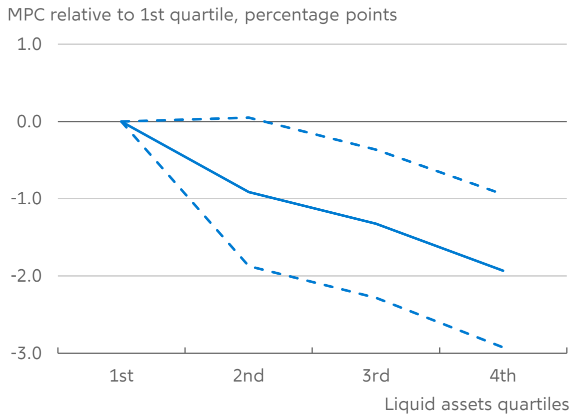
MPCs decrease with household liquidity and increase with leverage

As explained in box 2 and discussed above, we allow MPCs to vary in the econometric model across different dimensions that the literature has identified as important. Chart 3 shows the differences across groups of households with different degrees of liquidity, leverage (income or value-denominated) and age. Note that the figure shows the marginal differences, i.e. holding other characteristics fixed. In

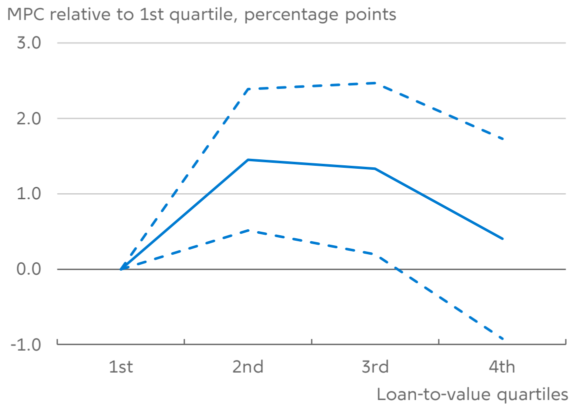
7 Some of the more rigorous previous studies have aimed at recovering the MPC out of unexpected changes in housing wealth, which from a behavioral point of view is the interesting metric, but which also requires additional information to estimate – such as surveys or natural experiments. Our estimate captures the response to both expected and unexpected changes in housing wealth.

The marginal propensity to consume out of changes in housing wealth differs across household segments, being highest among households with... Chart 3

... 1) few liquid assets

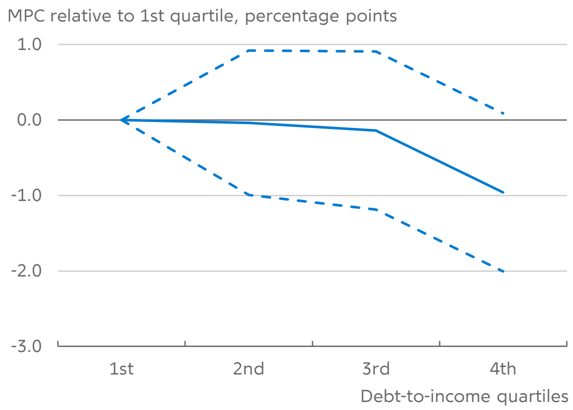


... 2) LTV ratios in the middle of the distribution

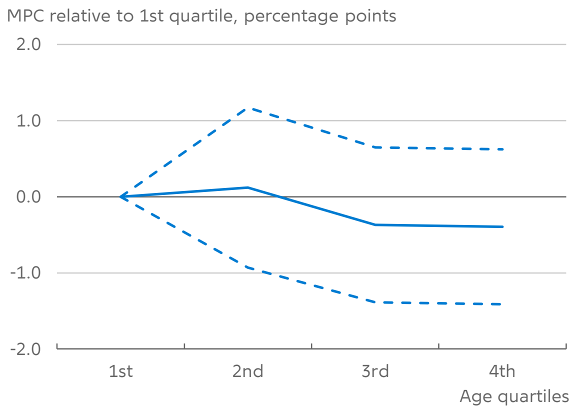


It does not vary significantly across...

... a) DTI ratios



... b) age groups



Note: The figure shows estimates (solid) of the additional effect of increasing housing wealth on consumption for households belonging to each quartile group for liquid assets, LTV ratios, DTI ratios and age, respectively, relative to the first quartile group, with corresponding 90% confidence bands (dashed). These estimates compare to an average MPC of 0.4 per cent. Estimates and standard errors are presented in table B2 in appendix B.

Source: Own calculations based on data described in box 1, and Statistics Denmark.

practice, these variables are correlated. For example, households with high levels of debt tend to have lower liquidity than other households.

It appears that households that are more likely to be borrowing-constrained, i.e. households with few liquid assets and relatively high LTV ratios, tend to have higher MPCs. This is consistent with the collateral effect discussed above. However, in contrast to households in the 2nd and 3rd quartiles for LTV, households in the 4th quartile do not, on average, have MPCs that are significantly higher than households in the 1st quartile. This may be because they are so highly leveraged that they cannot borrow fully against an increase in housing wealth due to, e.g., regulation and credit policies. MPCs do not vary significantly across DTI ratios and age groups.

Consumption and house prices during the Covid-19 house price boom and beyond

As discussed earlier, consumption growth has been relatively muted since the beginning of the pandemic. During the lockdowns, consumption possibilities were legally restricted, and consumption was generally reoriented towards durable goods and online trade (Andersen et al., 2020). However, the consumption of services such as travel and restaurant visits was also restricted at other times during the pandemic, and some households may also have increased their savings rate due to the generally higher uncertainty.[8]

At the same time, house prices increased considerably. One likely contributing factor was that many households may have increased their preference for housing (Hetland et al., 2021).[9]

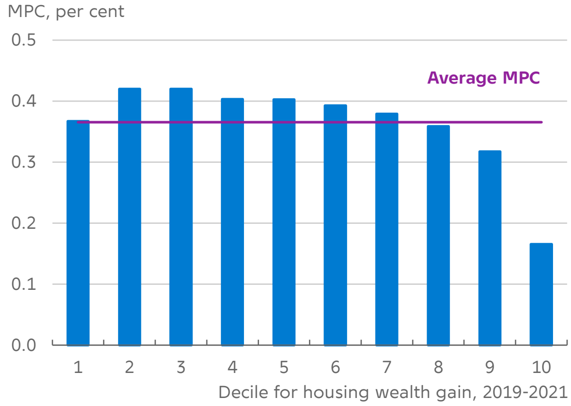
Our results demonstrate that the propensity to transmit housing wealth increases into consumption was relatively limited before the pandemic. In the absence of major behavioural changes during the pandemic, our results indicate that we should not expect households to have increased their consumption a lot as a consequence of the house price increases from 2019 to 2021 – in line with the muted development in the aggregate consumption ratio.

On top of the generally low MPCs estimated in our model, an additional factor has contributed to reducing the consumption response to higher house prices: the fact that the highest housing wealth gains have accrued to households with a lower-than-average propensity to consume out of housing wealth gains. Combining the estimates of housing wealth gains from our machine learning model with our estimates of the marginal propensity to consume, we see exactly this pattern, see chart 4.

We do not have evidence indicating whether this pattern is different from previous periods with booming house prices. However, our results do suggest that the heterogeneity in housing wealth gains – in addition to the generally low propensities to consume out of house price increases – has been an explanatory factor behind the relatively muted consumption response during the pandemic.

The largest house value increases occurred among households with relatively low MPCs

Chart 4



Note: The figure shows predicted MPCs for each household averaged by decile of (absolute) housing wealth gain from 2019 to 2021. Estimates are based on regression results from table B2 in appendix B.

Source: Own calculations based on model output for house values and registers from Statistics Denmark.

In chart 5 we illustrate the potential importance of the distribution of the housing wealth gains for the aggregate consumption response. The chart compares the estimated average consumption response based on the model in two scenarios; one in which housing wealth gains follows the actual distribution, and a counterfactual scenario in which the same aggregate housing wealth gains were

8 Andersen et al. (2022) document that savings rates were higher in 2020 than in the previous years, in particular among younger households.

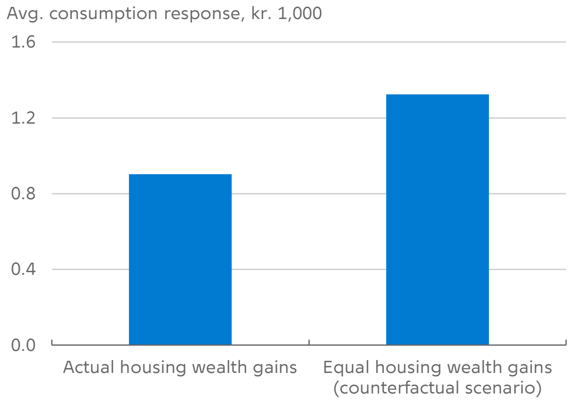
9 Here, the preference for housing refers to the marginal product of housing. The factors limiting consumption may have contributed to increasing house prices, which could give rise to concerns about reverse causality. However, in this analysis, we only consider the behavior of existing homeowners, and use the estimates to predict the consumption response of homeowners that owned their house before the onset of the pandemic. Therefore, this type of reverse causality is not problematic for our results.

distributed equally among all house owners.[10] The estimated consumption response is 30 per cent lower than it would be in the counterfactual scenario where housing wealth gains were distributed equally among house owners.

The effects we estimate here are for existing homeowners only. In order to evaluate the full effect of house price increases on consumption, one should also consider the reaction of sellers and, in particular, buyers and prospective buyers. They could also be expected to change their savings behaviour in response to house price increases.

The distribution of housing wealth gains across household segments muted the aggregate consumption response

Chart 5



Note: The figure shows the aggregate consumption response of an average household to changes in housing wealth from 2019 to 2021. It compares two scenarios; one where households receive *actual* housing wealth gains, and one hypothetical scenario where every household in our sample of homeowners receive the average and hence *equal* housing wealth gains.

Source: Own calculations based on model output for house values and registers from Statistics Denmark.

Consumption responses to decreasing house prices would likely be limited

Recently, house prices have started to decrease following the surge in inflation, and house prices are projected to decrease further over the coming year.[11] Our results generally show that the consumption response to changes in house prices has been relatively modest over the past years, and if households react in a symmetric way to increases and decreases in house prices, the direct response of consumption to decreasing house prices would also be limited.

There is evidence from previous literature to suggest that households react less to decreasing house prices compared to increasing house prices (Andersen and Leth-Petersen, 2021). But there is also evidence showing that households may increase their savings when house prices drop in order for them to retain their solvency and flexibility in future housing decisions (Hviid and Kuchler, 2017). However, this result was found in the aftermath of the Global Financial Crisis, and the desire to deleverage may therefore be specific to house price decreases following credit booms.

Finally, there is evidence pointing towards a correlation between equity extraction and the attractiveness of refinancing (Andersen and Leth-Petersen, 2021; Danmarks Nationalbank, 2022b, box 6). The recent interest rate increases have given rise to considerable refinancing activity among borrowers with fixed-rate mortgages. This could imply an upwards pressure on consumption, if, for example, households extract equity in connection with refinancing. However, this effect is likely to be countered by increasing prudence due to, e.g., the projected decrease in house prices, the inflation outlook as well as increasing interest rates.

Overall, our results indicate that in the absence of major behavioural changes, the consumption response to decreasing house prices will be limited.

[10] This scenario serves a purely illustrative purpose.

[11] See Danmarks Nationalbank (2022a).

References

Abildgren, Kim, Andreas Kuchler, America Solange Lohmann Rasmussen and Henrik Sejerbo Sørensen (2020), Registers or surveys – Does the type of microdata matter for empirical analyses of consumption behavior?, *Journal of Economic and Social Measurement*, vol. 45(3-4), pp. 237-254.

Adolfsen, J., S. Gudiksen, A. T. Martinello, B. M. Mønsted, A. Schmidt, and K. Sonberg (2022), Segmentation of the housing market with internet data: Evidence from Denmark, *Danmarks Nationalbank Working Paper,* no. 188, August.

Andersen, Henrik Yde, Alina Bartscher, Thomas Krause and Andreas Kuchler (2022), Inequality and Savings, *Danmarks Nationalbank Economic Memo*, no. 11, September.

Andersen, Henrik Yde and Søren Leth-Petersen (2021), Housing wealth or collateral: How home value shocks drive home equity extraction and spending, *Journal of the European Economic Association*, vol. 19(1), pp. 403-440.

Andersen, Svend Greniman, Stine Ludvig Bech, Simon Juul Hviid and Andreas Kuchler (2020), Forbruget dykker midlertidigt på grund af corona, *Danmarks Nationalbank Economic Memo*, no. 7, June.

Browning, Martin and Søren Leth-Petersen (2003), Imputing Consumption from Income and Wealth Information, *The Economic Journal*, vol. 113(488), pp. F282-301.

Crawley, Edmund and Andreas Kuchler (2023), Consumption heterogeneity: Micro drivers and macro implications, *American Economic Journal: Macroeconomics*, vol. 15(1), pp. 314-341.

Danmarks Nationalbank (2020), Udsigt til gradvis genopretning af økonomien, *Danmarks Nationalbank Analyse (Udsigter for dansk økonomi)*, no. 12, June.

Danmarks Nationalbank (2022a), Presset i økonomien bør dæmpes, *Danmarks Nationalbank Analyse (Udsigter for dansk økonomi)*, no. 11, September.

Danmarks Nationalbank (2022b), Stigende inflation og russisk invasion har øget volatiliteten, *Danmarks Nationalbank Analyse (Monetære og finansielle tendenser)*, no. 4, March.

Danmarks Statistik (2021), Statistikdokumentation for husholdningernes formue i fast ejendom 2021.

De Økonomiske Råd (2020), Corona-scenarier for dansk økonomi og de offentlige finanser, *Baggrundsnotat*, April.

Fagereng, Andreas, Martin Blomhoff Holm, and Kjersti Næss Torstensen (2020), Housing wealth in Norway, 1993–2015, *Journal of Economic and Social Measurement*, vol. 45(1), pp. 65-81.

Finansministeriet (2020), Økonomisk Redegørelse, Maj.

Hetland, Simon Thinggaard, Simon Juul Hviid, Jesper Pedersen and Adrian Michael Bay Schmith (2021), Housing market robustness should be strengthened, *Danmarks Nationalbank Analyse*, no. 16, June.

Hviid, Simon Juul and Andreas Kuchler (2017), Consumption and savings in a low interest-rate environment, *Danmarks Nationalbank Working Paper*, no. 116, June.

Johannemann, Jonathan, Vitor Hadad, Susan Athey, and Stefan Wager (2019), Sufficient representations for categorical variables, arXiv preprint: 1908.09874.

Martinello, Alessandro and Niels Møller (2022), Opportunities and Risks in the Residential Sector During a Green Transition: House Prices, Energy Renovations and Rising Energy Prices, *Danmarks Nationalbank Working Paper,* no. 190, September.

Appendix A: Residential housing valuation with machine learning

The housing valuation model used in this analysis builds on data from the Danish Building and Housing Register (BBR), the Danish Sales and Valuation Register (SVUR), and geodata from the Danish Agency for Data Supply and Infrastructure (SDI), see box 1 in the main text. The main model is trained and calibrated on 75 per cent of the sales occurring after 2014, while the remaining 25 per cent are kept for evaluation purposes. The model uses input features such as distance from the closest body of water, house price indexes, and sales density.

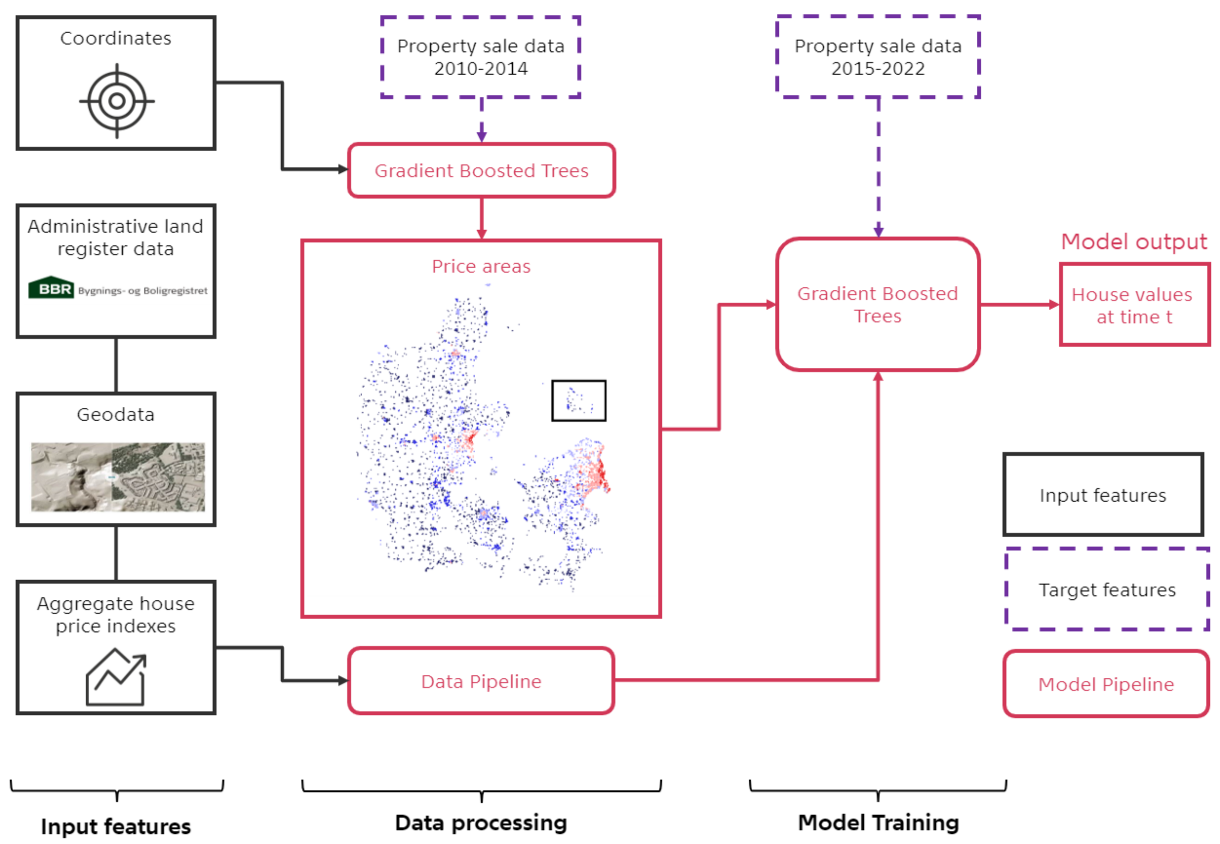
Chart A1 sketches how the model ingests this data and learns from it. First, the model pre-processes the input data. For example, geographic coordinates are processed separately into interpretable price areas. Second, we train the model using sales prices for sales occurring after 2014 as target features. Finally, the model is fed input features from non-transacted houses to estimate their market values.

Data processing

Most of the input features are pre-processed according to a flexible data pipeline which considers the type of the feature ingested. Continuous features are standardised to have a mean of zero and standard deviation of one. Rare instances for categorical features (together accounting for less

The full valuation model consists of a data pipeline, an auxiliary price area model and a final house value model

Chart A1



Note: The figure describes the property valuation model in its entirety. Input features in black (see table A1) are pre-processed before entering the final property valuation model. Property sale data (purple) before and after 31 December 2014 are used as target feature for the auxiliary price area model and the final property valuation model, respectively. Trained models and pipelines are highlighted in red. By using input data on non-transacted properties, the model is then able to impute a property value for all properties.
Source: Own representation.

than 1 per cent of the training data) are grouped together. The pipeline then one-hot encodes categorical features into a matrix of dummy variables. Finally, for each categorical, ordinal and numeric feature, the pipeline replaces null values with means, and whenever null values consist in more than 1 per cent of the feature values, it creates a dummy indicating the values for which the original feature was missing.

Municipality indicators and geographic coordinates are treated separately. For municipality indicators we use the sufficient representation approach for categorical variables proposed by Johannemann et al. (2019). To transform geographic coordinates into interpretable price areas, we exploit an auxiliary regularised gradient boosted trees model trained on latitude and longitude as input features and standardised (within calendar year) sales price quantiles occurring between 2000 and 2014 as targets (Martinello and Møller, 2022; Adolfsen et al., 2022). Intuitively, the model draws static boundaries across geographic areas spanning both within and across administrative boundaries indicating in which quantile of (country-wide) square metre price a house sold in that area will belong. Technically, the model outputs the vectors of probabilities of each house belonging to each quantile as input features. The version of the model used for this analysis uses nine quantiles.

Table A1 shows an overview of the input features used in the model, their type (which determines its pre-processing), and their data source.

Model training

The full model consists of a data pipeline, an auxiliary price area model, and a final house valuation model. We highlight these components in red in chart A1. These components are trained on input features (highlighted in black) and use granular sales prices (highlighted in purple) as target features.

Our modelling sample consists of sales of single family and terraced houses occurring as private market transactions and involving a single property, and for which the sales price was between kr. 300,000 and kr. 30,000,000. We use 75 per cent of the sample for model training. Model hyperparameters are calibrated within the training set through a standard random search, across five folds within the training data.

Model performance

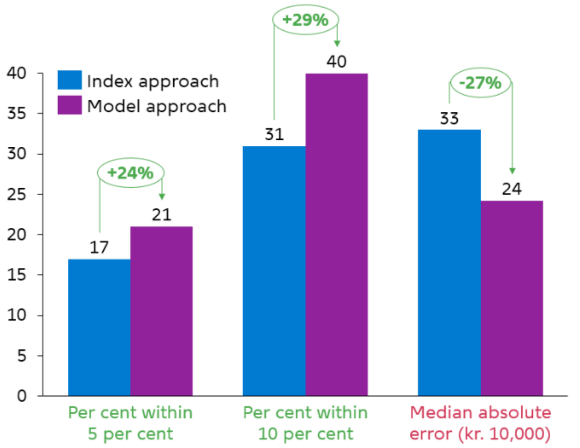
As aforementioned, we also estimate a simple index model of house values to serve as a comparison for our machine learning model. To evaluate the precision of our estimates, we are limited to the subsample of traded houses in the remaining 25 per cent of the sample not used to estimate the model. To estimate the index model, we further restrict the sample to houses that have been traded twice between 2014 and 2021. This secures an initial price that we can multiply by the square metre price index of traded houses at the postal code level. Hence, the evaluation sample is undeniably a small fraction of the full housing stock. We expect the precision of the index model to benefit from a relatively short span between the first and the second sale, but our machine learning model is equally favoured as a recent sales price is among the inputs.

We compare our machine learning model to the index model on three primary metrics: the percentages of house price estimates that are correct within 5 per cent and 10 per cent of actual sales prices for traded houses, respectively, as well as the size of the median absolute error. The result can be seen in chart A2. On all three metrics the reference model is outperformed by our machine learning model.

Due to the selected sample, our index model solely serves the purpose of a benchmark. Another potential benchmark could be the model used by Statistics Denmark. Statistics Denmark (2021) estimates house values (specifically the variable MARKEDSVAERDI in the FORMEJER-register) for the residential housing stock by projecting public valuations from 2013 to the relevant period for us by using correction factors calculated on the basis of property type, price range and postal code.[12] We cannot compare our machine learning estimates directly to MARKEDSVAERDI as it contains the actual sales price whenever a house is traded in a given year.

Performance of machine learning model compared to index model

Chart A2



Note: The sample is restricted to houses that were traded twice between 2014 to 2021. *Within 5/10 per cent* refers to the fraction of estimates that are within 5/10 per cent of the actual sales price. The *absolute error* is the absolute difference between estimate and actual sales price in kr.

Source: Own estimates based on data from the Danish Building and Housing Register (BBR), the Danish Sales and Valuation Register (SVUR), and geodata from the Danish Agency for Data Supply and Infrastructure (SDI).

[12] Due to a change in the Danish tax code, public evaluations have not been updated since 2013.

Input features of the model

Table A1

Input feature	Feature type	Data source
Habitable m^2 of the property	Numeric	BBR
Size (m^2) of main building	Numeric	BBR
Plot size in m^2	Numeric	BBR
Area of cellar	Numeric	BBR
Average price per m^2 of properties sold within a 100m/1km/10km radius within the last year (the smallest radius for which we observe at least 5 sales)	Numeric	BBR+SVUR
Last sales price of the property	Numeric	SVUR
Height above sea level	Numeric	BBR+SDI
Distance to coastline	Numeric	BBR+SDI
Distance to highway		BBR+SDI
House price index (municipality)	Numeric	SVUR
Moving average of house price index (municipality)	Numeric	SVUR
Number of rooms	Ordinal	BBR
Number of bathrooms	Ordinal	BBR
Number of floors	Ordinal	BBR
Number of buildings in the property	Ordinal	BBR
Year of sale	Ordinal	BBR
Month of sale	Ordinal	BBR
Days since last private market sale	Ordinal	BBR
Days since any sale	Ordinal	BBR
Construction year (main building)	Ordinal	BBR
Age of main building	Ordinal	BBR
Renovation year	Ordinal	BBR

Note: The table shows input features of the model. Input features are pre-processed in the data pipeline according to their type (numeric, ordinal, categorical, coordinates or fixed effects) as specified in the main text.

Input features of the model (continued)

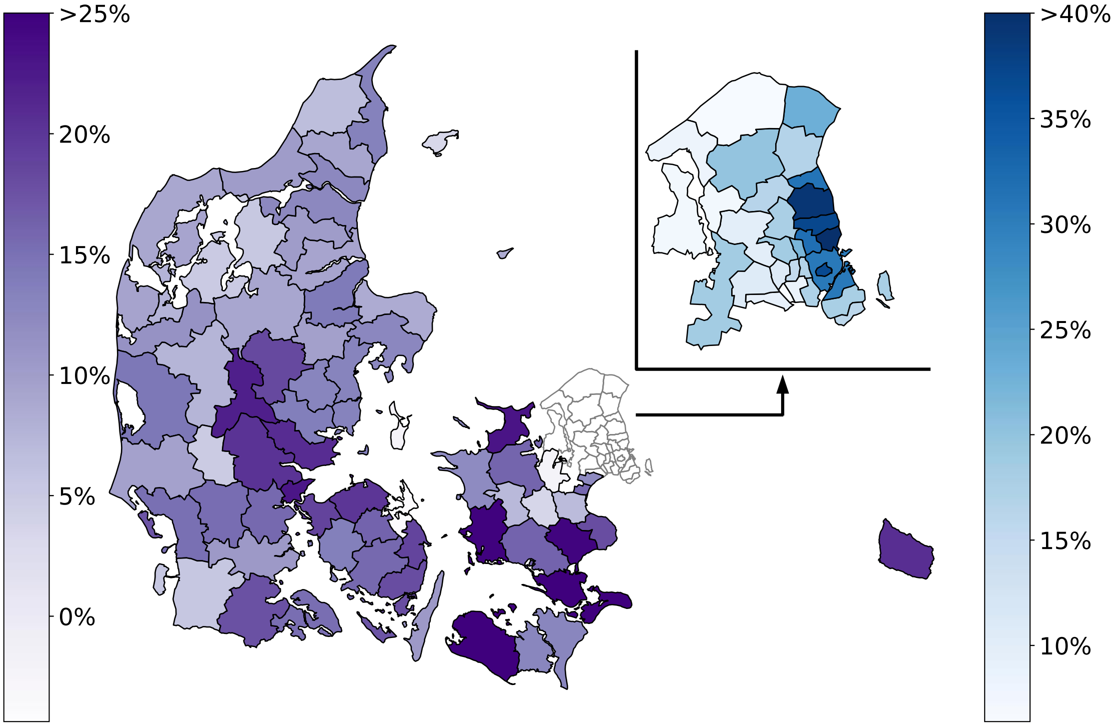
Input feature	Feature type	Data source
Property type (detached/terraced)	Categorical	BBR
Heating source	Categorical	BBR
Energy source	Categorical	BBR
Type of roof	Categorical	BBR
Has a garage (1/0)	Categorical	BBR
Has a shed (1/0)	Categorical	BBR
Last sale was either not an open-market transaction, it involved multiple properties, or its price was excessively low/high	Categorical	SVUR
Code of last sale (market transaction, family transfer, auction, other)	Categorical	SVUR
Municipality	Fixed effects	BBR
Coordinates	Coordinates	BBR

Note: The table shows input features of the model. Input features are pre-processed in the data pipeline according to their type (numeric, ordinal, categorical, coordinates or fixed effects) as specified in the main text.

Appendix B – Supplementary figures and tables

Relative house value increases by municipality, 2019-2021

Chart B1



Note: The figure shows average value growth of all single- and multi-family houses based on model price predictions from our model. The greater Copenhagen area is singled out due to extreme value increases relative to rest of the country.

Source: Own calculations.

The highest *relative* wealth gains were received by households in the greater cities

Table B1

Decile of relative housing wealth gain, 2019-2021	1	2	3	4	5	6	7	8	9	10
Housing wealth gain, kr. 1,000, median	-210	-20	60	150	230	300	380	490	690	1,260
Housing wealth gain, %, median	-11.05	-1.33	3.67	7.80	11.60	15.52	19.93	25.53	34.12	57.58
Housing wealth, 2019, kr. 1,000, median	1,950	1,800	1,910	1,960	1,960	1,940	1,920	1,930	2,020	1,695
Age, median	58	58	56	56	57	57	57	57	57	56
Income, kr. 1,000, median	654	678	707	719	722	722	723	733	746	742
Liquid assets, kr. 1,000, median	244	239	241	251	258	261	268	277	299	299
Liquid assets to income, %, median	37	35	34	35	36	36	37	37	40	39
Debt to income ratio, %, median	262	258	266	270	271	274	273	274	277	283
Ratio of total debt to house value, %, median	61	62	63	62	62	61	61	60	58	60
Total debt, kr. 1,000, median	1,127	1,150	1,229	1,267	1,272	1,280	1,277	1,295	1,320	1,353
Share of families with retired members, %	33.0	32.2	30.4	30.3	30.9	31.1	31.7	31.6	31.8	29.3
Share of families with higher education, %	45.7	46.5	48.8	50.1	50.7	51.6	51.5	53.0	55.3	54.4
Share of families in Aarhus and Greater Copenhagen, %	6.9	8.6	9.3	9.8	9.9	10.6	11.1	13.4	18.8	20.2

Note: The table shows descriptive statistics for households by decile of relative housing wealth gain from the end of 2019 to the end of 2021. Only households living in single-family houses with an estimated sales price between kr. 300,000 and kr. 30,000,000 have been included.

Source: Own calculations based on data described in box 1, and data from Statistics Denmark.

Estimated consumption responses to housing wealth changes (regression results) Table B2

Regression results: Consumption response to housing wealth changes, ***p<0.01, **p<0.05, *p<0.1

Dependent variable: Change in consumption	
Change in housing wealth	0.0107
	(0.00776)
Ch. Housing wealth * Quartile of liquid assets	
2	-0.00912
	(0.00584)
3	-0.0132**
	(0.00584)
4	-0.0193***
	(0.00603)
Ch. Housing wealth * Quartile of Loan-to-Income	
2	-0.000354
	(0.00582)
3	-0.00138
	(0.00637)
4	-0.00958
	(0.00637)
Ch. Housing wealth * Quartile of Debt-to-Value	
2	0.0145**
	(0.00570)
3	0.0134*
	(0.00691)
4	0.00405
	(0.00806)
Ch. Housing wealth *Quartile of age	
2	0.00122
	(0.00639)
3	-0.00369
	(0.00618)
4	-0.00395
	(0.00618)

PUBLICATIONS

NEWS

News offers quick and accessible insight into an Analysis, an Economic Memo, a Working Paper or a Report from Danmarks Nationalbank. News is published continuously.

ANALYSIS

Analyses from Danmarks Nationalbank focus on economic and financial matters. Some Analyses are published at regular intervals, e.g. *Outlook for the Danish economy* and *Financial stability*. Other Analyses are published continuously.

REPORT

Reports comprise recurring reports and reviews of the functioning of Danmarks Nationalbank and include, for instance, the *Annual report* and the annual publication *Danish government borrowing and debt.*

ECONOMIC MEMO

An Economic Memo is a cross between an Analysis and a Working Paper and often shows the ongoing study of the authors. The publication series is primarily aimed at professionals. Economic Memos are published continuously.

WORKING PAPER

Working Papers present research projects by economists in Danmarks Nationalbank and their associates. The series is primarily targeted at professionals and people with an interest in academia. Working Papers are published continuously.

DANMARKS NATIONALBANK
LANGELINIE ALLÉ 47
DK-2100 COPENHAGEN Ø
WWW.NATIONALBANKEN.DK

Danmarks Nationalbank's Economic Memos are published at www.nationalbanken.dk. A free electronic subscription is also available at the website. The subscriber receives an e-mail notification whenever a new Economic Memo is published.

Please direct any enquiries directly to the contributors or to Danmarks Nationalbank, Communications, *Kommunikation@nationalbanken.dk.*

DANMARKS NATIONALBANK