FACULTY OF SOCIAL SCIENCES
UNIVERSITY OF COPENHAGEN

Bachelor thesis

Nikolaj Toepfer Melchiorsen, Frederik Ingenman Olsen & Christian Birk Gustafson

The effects of COVID-19 on the housing market

- A co-integration analysis of the Danish housing market

Submitted on: 29/04/2022

Advisor: Michael Bergman

ECTS: 15

Keystrokes: 84.000

Abstract

This paper investigates COVID-19's effect on the Danish housing market using co-integration. The methodology is inspired by Dam et Al. (2011), whereas a few ideas stem from Guglielminetti's (2021).

In our ADL/ECM estimations, we test in the periods 1992Q1-2021Q2, 2005Q1-2021Q2, and 2015Q1-2021Q2 to analyse if certain variables influence the three housing prices. After removing insignificant variables from our ADL model, we find that COVID-19 has a significant effect in all of our ADL models. The co-integration results of the full periods show no relationship between variables, whilst the opposite is true for the sub periods we have created. Additionally, a long-run solution estimation tests which variables have a more permanent effect on the housing prices. We conclude that COVID-19 has a long-run effect on some housing types, depending on the periods. Specifically, we examine that COVID-19 has a long-run effect on condominiums for the full period and weekend cottages for the sub period. However, COVID-19 has no significant long-run effect on single-family houses.

Our empirical analysis finds a correlation between the COVID-19 dummy variable and the three housing prices. We observe similar results when comparing to other studies, such as Dam et Al. (2011) and Guglielminetti's (2021).

Contribution

Name	KUid	Contribution
Nikolaj Toepfer Melchiorsen	gwv540	2.1.1, 2.1.4, 2.2.2, 2.3, 2.4.3, 3.2, 4.2, 4.3.3, 4.5.1, 4.6.2, 4.7.3
Frederik Ingemann Olsen	wxz394	2.1.2, 2.1.5, 2.2.3, 2.4.1, 2.4.4, 3.3, 4.3.1, 4.4.1, 4.5.2, 4.7.1
Christian Birk Gustafson	ldg790	2.1.3, 2.2.1, 2.2.4, 2.4.2, 3.1, 4.1, 4.3.2, 4.4.2, 4.6.1, 4.7.2

Together: Introduction, discussion and conclusion.

Contents

1 Introduction

In early 2020, the world was shocked by the COVID-19 pandemic and countries were forced to shut down, resulting in economic turbulence. These economic turbulences have led researchers like Guglielminetti to analyse COVID-19's effect on the Italian housing market. Our paper takes inspiration from these ideas. Their study concludes that more people are willing to invest in larger houses located in the suburbs. Our paper aims to find COVID-19's impact on the Danish housing market for single-family houses, condominiums, and weekend cottages. We use Guglielminetti's (2021) ideas of how COVID-19 impacted the housing market and Dam et al.'s. (2011) ideas for modelling housing prices in Denmark. In combination with our general economic studies and appropriate methods, these ideas are used to write this paper.

The COVID-19 pandemic hit Denmark in early 2020, which forced the nation into an unfamiliar situation. The fear of contagion rose dramatically due to COVID-19's damage to world health. COVID-19 created new obstacles, which resulted in unfamiliar political restrictions. These restrictions had consequences, and the Danish housing market was one of many things affected by those.

As Denmark was one of the first European countries to introduce a lockdown starting on the 13 of March 2020, people could not spend their money as they usually would have. These consequences impacted the housing demand and thereby impacted housing prices. During the first lockdown, people in non-essential functions in the public sector had to stay at home. The private sector was also affected, as employers sent their employees home. On the 18 of March, the government implemented further restrictions, resulting in even fewer social activities, Politi (2020).

To find COVID-19's significance, we derive an error correction model using an autoregressive distributed lag model for three housing types. This model tests for no co-integration between the variables, and we divide them into a full period and a sub period. Using the long-run solution, we can estimate which variables have a lasting effect on the prices for each housing type.

We divide this paper into six sections. The first section introduces the paper. The second section explains the relevant theories used in the empirical analysis and discussion, divided into an economic and an econometric theory. The third section describes the variables and the equations used in the empirical analysis. The fourth section examines the empirical results based on our data and models. In the fifth part, we debate the estimates of our empirical results and compare them to relevant studies such as Dam et Al. The sixth section concludes our paper and its findings.

2 Theory

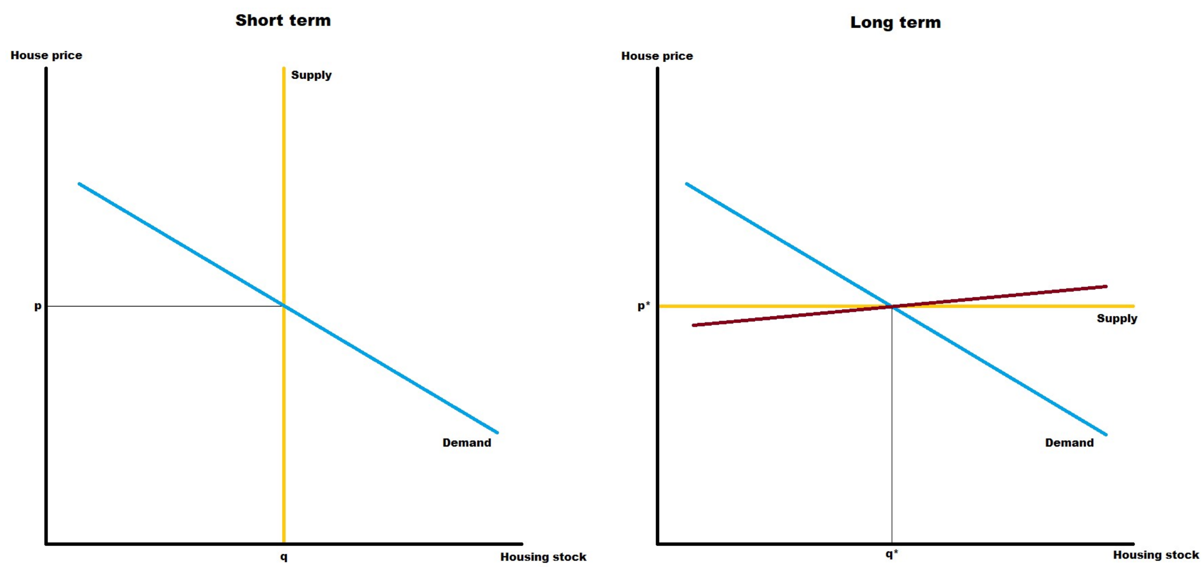
2.1 Economic theory

In this section, we reflect upon the theoretical economic aspect of our paper. Furthermore, we look into different theories that could help describe our empirical analysis results and conclude our paper.

2.1.1 General theory of supply & demand in the housing market

The demand and supply determine the housing market prices, but since it is impossible to build a house overnight, the supply is constant in the short term. The following two graphs show the short-and long term relation to price adjustment in the housing market.

Figure 1: The short & long term price adjustment in the housing market



Source: Peter Sørensen and Hans Whitte-Jacobsen 2010 ”Introducing Advanced Macroeconomics: Growth and Business Cycles”

On the left-hand side of figure 1, we observe the constant supply in the short term, which results in an endless number of houses q. Prices p can therefore only fluctuate along the supply curve, depending on the demand curve. Therefore, the prices for owner-occupied housing are determined by the demand side as there is no short term supply to ease the fluctuations that can occur in the short term. On the supply side, constructing new houses is continually active as long as it pays to build new houses instead of purchasing from the existing stock. Shifts in the demand curve to the upper right result in higher prices, boosting residential construction in the long term. But as the construction continues to increase, prices will adjust and gradually decrease towards the cost of new construction, which determines the prices in the long run.

The right-hand side is the long term market relation, where the horizontal supply curve is the real cost p^* of housing. We can imagine two scenarios. In scenario A (yellow line), if the supply is perfectly elastic, the real house prices will be constant in the long run. Therefore, an increase in the nominal house prices will equal the general inflation. In scenario B (red line), if the supply curve is an increasing function, the long term real marginal cost of the housing stock will rise. This curve is indicated in red. Therefore, an increase in economic growth will result in higher demand causing higher housing prices in the long run.

2.1.2 Tobin's Q theory

In 1969 Tobin introduced the Q ratio, which determines the investment rate by evaluating the market value ratio over replacement costs for the same asset. This theory also applies to other markets, according to Tobin.

We assume perfect market theory, meaning that the houses can easily be substituted, and the seller must accept the market price. This assumption means that the buying decision is based on construction prices versus the price of existing houses. The formula for Tobin's Q is thereby relatively simple and is given as:

$$q_t \equiv \frac{V_t^\alpha}{K_t^\alpha}$$

where V_t^α is the market price of existing homes, and K_t^α is the cost of acquiring new land and constructing new homes. Market equilibrium exists when $q = 1$. If $q > 1$, firms have an incentive to build new houses to meet consumer demand. Vice versa, if $q < 1$, the construction price is greater than the market price of existing houses, at which point it is cheaper to buy already existing houses.

2.1.3 Housing market supply

Following the general theory of housing market supply, the production function of constructing new housing I^H is:

$$I^H = AX^\beta, \quad 0 < \beta < 1,$$

The function is determined by the composite input factor X, and the constant A depends on the productive capacity of the construction sector. The parameter β indicates the decrease in marginal output of the production process over time. The construction firms' cost of building new homes includes labour L and building materials Q. For simplicity, the firms select these inputs in fixed proportions, as such:

$$L = aX, \quad Q = bX$$

where a is the amount of labour and b is the number of materials given the composite input factor X. If p is the price of a unit and X the composite input, the 'construction cost index' P can be derived as:

$$P = aW + bp^Q$$

The wage rate is given as W and the price of materials p^Q. Now the profit of the firm can be defined as:

$$\Pi = p^H I^H - PX$$

where p^H is the house prices, and $p^H I^H$ is the sales revenue of the representative construction firm. If construction firms choose the level of activity I^H, then with partial differentiation and isolation of I^H, the supply curve for housing is given as:

$$I^H = k\left(\frac{p^H}{P}\right)^{\frac{\beta}{1-\beta}}, \quad k \equiv \beta^{\frac{\beta}{1-\beta}} A^{\frac{1}{1-\beta}}$$

Firms that seek to maximise profits will increase the construction of new housing until the market is in equilibrium, which means the marginal cost of constructing a house is equal to the market price of a house. Since the price variable $\frac{p^H}{P}$ is equivalent to Tobin's Q, an increase in the market price of existing houses means it becomes more profitable for firms to construct new houses, thus, increasing housing investment.

2.1.4 Housing market demand

The consumer's budget constraint is considered to find the housing market demand function:

$$C + (\hat{r} + d)p^H s^H = Y$$

where s^H is the housing supply and C is consumer consumption. Lastly, d is the housing depreciation, meaning $dp^H H$ is the amount consumers spend on repairing and maintaining their houses. Further, $\hat{r}$ is the interest rate on mortgage debt. Thus, the consumer's total cost of housing is $(\hat{r} + d)p^H H$, which is also denoted as the simple user cost in this section. Section 3 defines a more complex user cost that better suits our data and model. To maximise the consumer's utility U, the following Cobb-Douglas function is:

$$U = H^\eta C^{1-\eta} \quad 0 < \eta < 1$$

The demand for housing H^d is determined by solving the maximisation problem:

$$H^d = \frac{\eta Y}{(\hat{r} + d)p^H}$$

The equation shows that housing demand depends positively on household income Y and negatively on the housing market price p^H.

2.1.5 Market equilibrium

In the short run, the housing supply is perfectly inelastic; hence the housing market price has to adjust depending on the demand. The market equilibrium price for the short run can be determined by setting $H^d = H$ and then solving for p^H:

$$H = \frac{\eta Y}{(\hat{r} + d)p^H} \Leftrightarrow p^H = \frac{\eta Y}{(\hat{r} + d)H}$$

The market equilibrium shows that housing price is negatively affected by the simple user cost $(\hat{r} + d)H$, housing supply H, and is positively affected by the level of income Y. According to the demand and supply section, a short-term increase in income will increase housing demand. This increase raises housing prices and creates interest from the suppliers to construct new homes. In theory, the same behaviour would occur if the real interest rate dropped.

2.2 Econometric theory

This section presents the relevant econometric theory used in the empirical analysis to test COVID-19's impact on the housing market. We estimate all regressions by OLS, and then we seek to find COVID-19's impact using co-integration. This is done by estimating an auto-distributed lag (ADL) and deriving the error correction model (ECM).

2.2.1 Unit root testing

In order to properly use our data, we have to identify if the data shows stationarity. Stationarity occurs when the mean is constant, the volatility is constant, and the time series does not show seasonality. Though it is often very intuitive to observe stationarity based on graphical analysis, there are situations where it can be misleading. We do a unit root test called the Augmented Dickey-Fuller test (ADF) to ensure stationarity. The Dickey-Fuller test (DF) estimates if an AR(1) model is stationary. We will explain the theory behind a DF test, which is then used to understand the ADF for an AR(p) model. The overall method for the two tests is the same. The DF test starts with an AR(1) model written as:

$$y_t = \theta y_{t-1} + \epsilon_t$$

The time series is a unit root if θ=1. This is then tested by rewriting the model where we subtract y_{t-1} on both sides:

$$\Delta y_t = \varpi y_{t-1} + \epsilon_t \text{ where } \Delta y_t = y_t - t_t - 1 \text{ and } \varpi = \theta - 1$$

This creates a model where the left-hand side is stationary under our null hypothesis, written as:

$$H_0 : \varpi = 0 \text{ or } H_A : \varpi < 0$$

This is then tested by comparing the t-statistic of ϖ and the Dickey-fuller distribution, calculated by OxMetrics or found in a DF table for critical values. The t-distribution for ϖ is written as:

$$t_\varpi = \frac{\hat{\varpi}}{se(\hat{\varpi}}$$

If $t_\varpi <$ DF$_{criticalvalue}$, then we reject the H_0, meaning that the time series has a unit root, indicating that it is stationary. If $t_\varpi >$ DF$_{criticalvalue}$, then we do not reject the H_0, thus cannot reject the presence of unit root.

We will now extend the test to an AR(p) model, where we consider the case of $p = 3$ lags:

$$y_t = \theta_1 y_{t-1} + \theta_2 y_{t-2} + \theta_3 y_{t-3} + \epsilon_t$$

We note that a unit root in $\theta(z) = 1 - \theta_1 z - \theta_2 z^2 - \theta_3 z^3$ corresponds to the equation:

$$\theta(1) = 1 - \theta_1 - \theta_2 - \theta_3 = 0$$

The test is the very similar to the DF test statistics and can be extended to :

$$\xi_{DF} = \frac{\sum_{i=1}^{3} \hat{\theta}_i - 1}{se(\sum_{i=1}^{3} \hat{\theta}_i)}$$

We want to avoid testing a restriction on $1 - \theta_1 - \theta_2 - \theta_3$, that involves all p=3 parameters, so we will rewrite the models as:

$$y_t - y_{t-1} = (\theta_1 - 1)y_{t-1} + \theta_2 y_{t-2} + \theta_3 y_{t-3} + \epsilon_t \Leftrightarrow$$

$$y_t - y_{t-1} = (\theta_1 - 1)y_{t-1} + (\theta_2 + \theta_3)y_{t-2} + \theta_3(y_{t-3} - y_{t-2}) + \epsilon_t \Leftrightarrow$$

$$\Delta y_t = (\theta_1 + \theta_2 + \theta_3 - 1)y_{t-1} + (\theta_2 + \theta_3)(y_{t-2} - y_{t-1}) + \theta_3(y_{t-3} - y_{t-2}) + \epsilon_t \quad (1)$$

where

$$\pi = \theta_1 + \theta_2 + \theta_3 - 1 = -\theta(1)$$

$$c_1 = -(\theta_2 + \theta_3)$$

$$c_2 = -\theta_3$$

From equation (1) the hypothesis that $\theta(1) = 0$ can be simplified to the null hypothesis that

$$H_0 : \pi = 0 \text{ vs. } H_A : -2 < \pi < 0$$

Here the test statistics for H_0 are denoted as the ADF test, where the asymptotic distribution is the same as the DF test in the AR(1) described above. It should be noted that the test for $\pi = 0$ is the only one that follows the DF distribution, whereas the test related to c_1 and c_2 have standard asymptotics. The difference is that the hypothesis, $c_1 = 0$, does not introduce any unit-roots. This means that the unit root test is just performed as a test for $\pi = 0$ for our regression with p lags. When testing for an AR(p) model, it is important to include sufficient lags to ensure that the errors are i.i.d. When we determine the number of lags, we use the general-to-specific method. This method also ensures that our model is well-specified before the unit root test is applied.

2.2.2 Co-integration

Co-integration is used to analyse how much the different variables affect housing prices. We use the ADL/ECM approach to derive the co-integration with the Engle & Granger theorem. The theorem states that if two or more variables have the same underlying stochastic trend, such that a linear combination exists between the variables, the variables are said to co-integrate. In other words, when a linear combination of non-stationary variables becomes stationary, it is denoted as co-integration. Therefore, in general you can have p variables $X_t = (x_{1t}, x_{2t}, ..., x_{pt})'$, that are all non-stationary, but using a particular linear combination e.g.

$$\alpha' x_t = \alpha_1 x_{1t} + ... + \alpha_p x_{pt}$$

may create a stationary process, where $\alpha = (\alpha_1, ..., \alpha_p)$ is a vector of parameters. The ECM is our way to create a linear combination, which we use to test for co-integration.

Consider our vector of variables for single-family houses for the full period:

$$X_t = \begin{pmatrix} \log p_t^H \\ \log Y_t^d \\ u_t \\ y_t \\ \sigma_t \\ \upsilon_t \\ 1 \end{pmatrix}, \hat{\beta} = \begin{pmatrix} 1 \\ -\beta_2 \\ -\beta_3 \\ -\beta_4 \\ -\beta_5 \\ -\beta_6 \\ -\mu \end{pmatrix}$$

If the variable for real house prices error corrects (as shown later in figure 4, for the ECM), then there is co-integration. This only applies if we assume that the real house prices error correct. Using the above-defined vector, we proceed to write the vector which the standard t-test will be based upon:

$$X_t' = \left(\log\, p_t^H \;\log\, Y_t^d \; y_t \; u_t \; \sigma_t \; \upsilon_t \; 1\right) \;\wedge \gamma' = \left(1 - \frac{\psi_2}{\psi_1} - \frac{\psi_3}{\psi_1} - \frac{\psi_4}{\psi_1} - \frac{\psi_5}{\psi_1} - \frac{\psi_6}{\psi_1} - \frac{\delta}{\psi_1}\right)$$

where the speed of adjustment is ψ_1. The null hypothesis is that the speed of adjustment is equal to 0 and the alternative hypothesis is that the speed of adjustment must be less than zero:

$$H_0 : \psi_1 = 0 \wedge H_A : \psi_1 < 0$$

The simple t-test is:

$$\hat{\tau}_{\psi_1} = \frac{\hat{\psi}_1}{se\left(\hat{\psi}_1\right)} \tag{2}$$

If we accept the null hypothesis, the error-correction model might be non-stationary. Therefore, assuming that the real house prices error corrects on all the variables included, ψ_1 will be negative. If the real house prices do not error correct, $\psi_1 = 0$. In our paper $\psi = \alpha$. The test for no error-correction is based on the assumption that only housing prices error corrects the different co-integration relations.

2.2.3 Determine number of lags

There is a wide range of methods to determine the optimal number of lags, q. Using the ADF-test as a test for the optimal lag number of differences, we choose the general-to-specific approach due to its simplicity. We start with $q = 12$ lags and go down until the last lag is still significant and we reach a well-specified model. Simultaneously, we test whether the model suffers from autocorrelation in the residuals to make sure residuals will not be inconsistent. If autocorrelation is present, dummy variables will be included to control for significant deviations in the data. If the model continues to suffer from misspecification, insignificant lags will be removed, which will result in holes in the lag structure. It is preferable to avoid holes in the lag structure, it is considered a last resort.

2.2.4 ADL & ECM

This paper proceeds with an (autoregressive distributed lag) ADL model. The reason for choosing an ADL model is that we will look at the dynamic response to an intervention in a given variable (COVID-19). Furthermore, it produces fewer misspecifications. Therefore, a univariate model will not suffice for this purpose. The object of interest is the conditional mean of y_t given $y_{t-1}, y_{t-2}, ..., x_t, x_{t-1}, x_{t-2}, ...$ An ADL model is denoted as ADL(q_1, q_2), where q_1 is the number of lags on the dependent variable and q_2 is the number of lags on the independent variables.

Assuming linearity of the conditional mean, we obtain the ADL model, which can be written in the general form:

$$y_t = \delta + \theta_1 y_{t-1} + \phi_0 x_t + \phi_1 x_1 + \epsilon_t \tag{3}$$

for $t = 1, 2, ..., T$, with $E(\epsilon | y_{t-1}, y_{t-2}, ..., x_t, x_{t-1}, x_{t-2}...) = 0$. In order to rule out heteroskedasticity we assume that the error correction term ϵ is independent and identically distributed $i.i.d.(0, \sigma^2)$.

Using the general ADL model, we can formulate an (Error Correction Model) ECM, which is used to test for no co-integration between variables. We derive the ECM step-by-step due to its essential part in the empirical analysis.
Starting from (11):

$$y_t = \delta + \theta_1 y_{t-1} + \phi_0 x_t + \phi_1 x_{t-1} + \epsilon_t$$

Next we use recursive substitution to obtain:

$$y_t - y_{t-1} = \delta + (\theta_1 - 1) y_{t-1} + \phi_0 x_t + \phi_1 x_{t-1} + \epsilon_t$$

Rewriting this we obtain:

$$y_t - y_{t-1} = \delta + (\theta_1 - 1) y_{t-1} + \phi_0 (x_t - x_{t-1}) + (\phi_0 + \phi_1) x_{t-1} + \epsilon_t$$

Now we can rewrite and insert Δ to symbolize the difference between the variable from period t-1 to t:

$$\Delta y_t = \delta + (\theta_1 - 1) y_{t-1} + \phi_0 \Delta x_t + (\phi_0 + \phi_1) x_{t-1} + \epsilon_t$$

We then rewrite the equation so the level only appears once:

$$\Delta y_t = \theta_0 \Delta x_t - (1-\theta_1)(y_{t-1} - \frac{\delta}{1-\theta_1} - \frac{\phi_0+\phi_1}{1-\phi_1} x_{t-1}) + \epsilon_t \quad (4)$$

where

$$\delta = \frac{\delta}{1-\theta_1} \text{ and } \alpha = \frac{\phi_0+\phi_1}{1-\phi_1}$$

Model (4) is the so-called ECM. We use the Engle & Granger two-step procedure for the co-integration analysis.

2.3 Long-run solution

The long-run solution describes the deviation from the steady-state in the previous period. To reach a steady state, the variables need to eliminate deviations and move towards the long-run solution. The long-run solution for the full period of single-family houses, can be derived by defining the steady states as $\log p_t^H = \log p_{t-1}^H$, $\log Y_t^d$=$\log Y_{t-1}^d$, $y_t = y_{t-1}$, $u_t = u_{t-1}$, $\sigma_t = \sigma_{t-1}$, and $\upsilon_t = \upsilon_{t-1}$.

The equation for long-run solution is given as:

$$\alpha_1 \log p_t^H + \delta + \alpha_2 \log Y_t^d + \alpha_3 u_t + \alpha_4 y_t + \alpha_5 \sigma_t + \alpha_6 \upsilon_t = 0$$

$$\Leftrightarrow$$

$$\log p_t^H = -\frac{\delta}{\alpha_1} - \frac{\alpha_2}{\alpha_1} \log Y_t^d - \frac{\alpha_3}{\alpha_1} y_t - \frac{\alpha_4}{\alpha_1} u_t - \frac{\alpha_5}{\alpha_1} \sigma_t - \frac{\alpha_6}{\alpha_1} \upsilon_t \quad (5)$$

Where the long-run coefficients are given as:

$$\mu = -\frac{\delta}{\alpha_1} = \frac{\delta}{1-\theta_1-\theta_2}$$

$$\beta_2 = -\frac{\alpha_2}{\alpha_1} = \frac{\alpha_2}{1-\theta_1-\theta_2}$$

$$\beta_3 = -\frac{\alpha_3}{\alpha_1} = \frac{\alpha_3}{1-\theta_1-\theta_2}$$

$$\beta_4 = -\frac{\alpha_4}{\alpha_1} = \frac{\alpha_4}{1-\theta_1-\theta_2}$$

$$\beta_5 = -\frac{\alpha_5}{\alpha_1} = \frac{\alpha_5}{1-\theta_1-\theta_2}$$

$$\beta_6 = -\frac{\alpha_6}{\alpha_1} = \frac{\alpha_6}{1-\theta_1-\theta_2}$$

It should be noted that the long-run solution will be significant for the models with co-integration. If the model has no sign of co-integration, the long-run solution could be misleading.

2.4 Misspecification

The general-to-specific principle determines the lag structure for the preferred model, after which we test for misspecifications. More formally, we test for the four asymptotics to obtain a well-specified model; autocorrelation, heteroskedasticity, normality and ARCH-effects.

The first test is for no autocorrelation in the residuals. If autocorrelation is present, the estimator might be inconsistent. We use the Breusch-Godfrey (LM) test for no first-order autocorrelation, which tests the omitted variable ϵ_t. Due to our lag structure we need to test for higher-order autocorrelation, meaning the auxiliary regression must include more lags, $\tilde{\epsilon}_{t-1}, \tilde{\epsilon}_{t-2}, ..., \tilde{\epsilon}_{t-m}$.

The second test is for heteroskedasticity, which is done through a Breusch-Pagan test with the same LM statistic LM_T. If the variance of the time series shifts over a given period, it signifies heteroskedasticity, meaning the explanatory variables are correlated with the error term. Homoscedasticity is the opposite of heteroskedasticity, ensuring that the coefficient converges towards a normal distribution.

A test for normality is also performed. Since we use continuously observed variables, i.e. (income), the normality test checks for skewness and kurtosis. Non-normality of ϵ_t does not invalidate the consistency of the OLS estimator or its asymptotic normality, yet it is still a relevant test to complete to ensure that the estimates are consistent. Because if ϵ_t has a severely skewed distribution, it might be helpful to transform the dependent prior to estimation (e.g. using log income instead of income itself). We use the Jarque-Bera test, which is an LM test for normality.

Lastly, we want to secure no ARCH effects because it would otherwise make the OLS estimation of the ADL estimations inefficient. We use the standard Breusch Pagan LM test for no heteroskedasticity and apply it to this particular form of heteroskedasticity.

We use Akaike Information Criterion AIC as a robust check to determine the optimal number of lags. AIC tests when the model is most specified and is given as:

$$AIC = \log \hat{\sigma}^2 + \frac{2k}{T}$$

As the AIC test statistic becomes smaller, the model becomes better specified.

2.4.1 Formal test for no autocorrelation

Using the Breusch-Godfrey (LM) test, we base it on the auxiliary regression of the estimated residual of the ADL model.

$$\hat{\epsilon}_t = x_t^{'}\delta + \alpha\hat{\epsilon_{t-1}} + u_t \tag{6}$$

where $\hat{\epsilon}_t$ is the estimated residuals from the ADL model and u_t is the new error term. x_t is the original explanatory variable and it is included in (6) to allow for the fact that x_t may not be strictly exogenous and thus may be correlated with $\hat{\epsilon}_{t-1}$. The null hypothesis of no autocorrelation is:

$$H_0 : \alpha = 0 \text{ vs. } H_A : \alpha \neq 0$$

We compute the LM statistic, $\xi_{AR} = T \cdot R^2$, where T denotes the number of observations and R^2 is the coefficient of determination in the auxiliary regression (6). Under the null hypothesis, the statistic is asymptotically distributed and will converge towards an χ^2-distribution with one degree of freedom, which corresponds to the number of lags used in the residuals in the auxiliary regression.

2.4.2 Formal test for heteroskedasticity

There are several ways of performing a heteroskedasticity test, but we proceed with the Breusch-Pagan test. Using the general ADL model as an example. The test uses auxiliary regression of the squared residuals on the original regressors and their squares:

$$\hat{\epsilon_t}^2 = \alpha_0 + x_{1t}\alpha_1 + ... + x_{kt}\alpha_k + x_{1t}^2\delta_1 + ... + x_{kt}^2\delta_k + u_t \tag{7}$$

The hypothesis for the test is that all coefficients, except the constant, are equal to zero. This proves that none of the parameters or the parameters squared depends on the variance of the residuals:

$$H_0 : \alpha_1 = ... = \alpha_k = 0 = \delta_1 = ... = \delta_k = 0 \text{ vs. } H_A : \text{ At least one parameter} \neq 0$$

The LM statistic $\xi_{HET} = T \cdot R^2$, is distributed as $\chi^2(2k)$ under the null, where 'k' is the number of variables in the ADL model.

2.4.3 Formal test for normality

We use the Jaque-Bera test for normality. The skewness measures the asymmetry of the distribution, and the kurtosis measures how much of the probability mass is placed in the tail of the distribution. Skewness and kurtosis can be written as:

$$S = E\left[\left(\frac{\epsilon}{\sigma}\right)^3\right] \text{ and } K = E\left[\left(\frac{\epsilon}{\sigma}\right)^4\right]$$

with the sample counterparts:

$$S = T^{-1}\sum_{t-1}^{T}\left(\frac{\hat{\epsilon}_t - \bar{\epsilon}_t}{\hat{\sigma}}\right)^3, \; K = T^{-1}\sum_{t-1}^{T}\left(\frac{\hat{\epsilon}_t - \bar{\epsilon}_t}{\hat{\sigma}}\right)^4$$

where $\bar{\epsilon}_t = T^{-1}\sum_{t-1}^{T}\hat{\epsilon}_t$ is the estimated mean and $\sigma^2 = \sum_{t-1}^{T}(\hat{\epsilon}_t - \bar{\epsilon}_t)^2$ is the estimated variance of the residuals. The standardised residuals are given by $\frac{\epsilon_t - \bar{\epsilon}_t}{\sigma}$. The null hypothesis H_0 of the residuals implies that S and K is asymptotically normal. The Jarque-Bera test statistic is given by:

$$\xi_{JB} = \xi_S + \xi_K$$

where $\xi_S = \frac{T}{6}S_T^2$ and $\xi_K = \frac{T}{24}(K_T - 3)^2$, which is true under the assumption of normality, H_0. Both S and K converge to a χ^2-distribution with one degree of freedom. This is due to the Gaussian distribution, where the (S=0) and the (K=3), meaning that the joint test to execute is $S = (K-3) = 0$. The H_0 of normality of S and K corresponds to ξ_{JB} converges towards a χ^2-distribution with two degrees of freedom. The H_0 can be written in the form:

$$H_0 : \xi_{JB} \sim \chi^2(2) \text{ vs. } H_A : \; H_0 \text{ not true}$$

If we reject the null hypothesis, it means that we have normality misspecification and vice versa.

2.4.4 Formal test for ARCH-effects

We will consider the hypothesis of no ARCH-effects up to order q for the following auxiliary regression model:

$$\hat{\epsilon}_t^2 = \gamma_0 + \gamma_1\hat{\epsilon}_{t-1}^2 + \gamma_2\hat{\epsilon}_{t-2}^2 + ... + +\gamma_p\hat{\epsilon}_{t-p}^2 + error$$

where $\hat{\epsilon}_t$ is the estimated residual from the regression used in this paper. The null hypothesis of no ARCH-effects is:

$$H_0 : \gamma_1 = \gamma_2 = ... = \gamma_p = 0$$

which means that the expected value of $\hat{\epsilon}_{t-1}^2$ is a constant, γ_0 for all t. The alternative hypothesis is that at least one γ_i is different from zero, $i = 1, 2, ..., p$. If we accept the null hypothesis it means we have no ARCH-effects. We will use the LM statistic given as:

$$\xi_{ARCH} = T \cdot R^2$$

where R^2 is the coefficient of determination from the auxiliary regression described above. The test statistic is asymptotically distributed as a $\chi^2(p)$ if we accept the null. It should be noted that we first need to test no autocorrelation. Because if the residuals are not autocorrelated, but the squared residuals are, that is an indication of ARCH effects.

3 Data

3.1 The variables

Our paper seeks to estimate a full period model for single-family houses, condominiums, and weekend cottages in the Danish housing market. This section defines and discusses the data sets and variables for the empirical analysis. We use quarterly data primarily from Danmarks Nationalbank's database "MONA". Some macro variables are taken from Statistics Denmark since they were not included in the MONA database. All variables are defined below:

Table 1: Variables

$\log p_t^{EN}$	Real housing price for single-family houses
$\log p_t^{EJ}$	Real housing price for condominiums
$\log p_t^{SOM}$	Real housing price for weekend cottages
$\log Y_t^d$	Real disposable income
y_t	Minimal first year payment
u_t	User cost
$\log s_t^{EN}$	Supply of single-family houses
$\log s_t^{EJ}$	Supply of condominiums
$\log s_t^{SOM}$	Supply of weekend cottages
σ_t	Dummy variable for the financial crisis
υ_t	Dummy variable for COVID-19

Source: Denmarks Nationalbank Mona database, Statistics Denmark and own calculations.
Note: t is time for the given period.

Real housing price is deflated by dividing it with the consumer price index. We took the logarithm of the real housing price to make it possible to compare the values over time. The housing price is determined for each of the three housing types: single-family houses, condominiums, and weekend cottages. Where the price index is in 2010-prices.

Real disposable income is calculated by taking the logarithm of disposable income divided by the consumer price deflator. This is done because the disposable income is in current prices with 2010 = 100.

User cost is defined as

$$u = ((1-t)r^{30} - \pi) + \bar{s} + d - \pi^H$$

where we use the definition from Dam et al. (2011, pp.15). Where t is the tax rate for housing-related interest cost, r^{30} is the 30-year bond yield, π denotes the expected inflation rate, $\bar{s}$ is the total housing stock calculated at market price, d is the depreciation rate. Lastly, π is the expected inflation rate of households, which we have approximated through an autoregressive process AR(1):

$$\pi = (1-\gamma)\pi_1 + \gamma \log \frac{P}{P_4}$$

The equation above uses the exponential smoother estimate from Cogleys (2002), assuming adaptive expectations. We continue with the Dam et al. (2011) approach, where we set $\gamma = 0.125$ and denote P as the consumer price index (CPI). Using quarterly data, P_4 is CPI lagged four quarters before the given period. Furthermore, changes in real house prices are denoted as π^H. However, as Dam et al. (2011) states, π^H is difficult to estimate, and the depreciation rate is somewhat constant over time, which is why we assume $d = 0$ and $\pi^H = 0$. This means that both variables are excluded from the equation, leading to the equation of the user cost applied in this paper:

$$u = ((1-t)r^{30} - \pi) + \bar{s}$$

Minimal first year payment is defined as:

$$y = (1-t)r^{min} + \hat{s} + repay$$

where real interest rates r^{min} is the minimum bond of ten years, $\hat{s}$ denotes the total housing taxes as a ratio of the total housing stock calculated at market price. *repay* is the repayment rate for a fully leveraged house with full utilisation of adjustable-rate loans, and it is based on different loans throughout the period, Dam et al. (2011, p. 15 & 72).

Supply for housing is from Realkreditrådet, and the data is from 2004(1) to 2021(2). It measures the number of houses for sale in all of Denmark. The number of sales is counted as the number of online listings. Due to the lack of internet use until the mid-2000s, the data started in 2004. This applies to the three housing types used in the paper.

The Dummy variables are based on two events, the financial crisis in 2008 and the COVID-19 pandemic. The financial crisis is equal to one in 2008(1)-2009(4) and zero otherwise. The COVID-19 dummy variable represents the period 2020(2)-2021(2), equal to one and zero otherwise. We can control for large outliers and estimate their impact on housing prices with these dummy variables. Both variables are exogenous. This means that other variables in our model do not determine the variables. Instead, factors outside of the model determine the value of the exogenous variables.

3.2 The full period models

Using the defined variables, we can estimate a full period housing price model for single-family houses:

$$\log p_t^{EN} = \alpha_0 + \alpha_1 y_t + \alpha_2 \log Y_t^d + \alpha_4 u_t + \alpha_5 \upsilon_t + \alpha_6 \sigma_t + \epsilon_t$$

note that housing supply for single-family houses is not included since the model starts in 1992(1) and ends in 2021(2).

Our analysis also includes condominiums and weekend cottages, here it is important to note that our full period model starts in 2005(1) and ends in 2021(2). The start period is effected by the four lags included in the model. Furthermore, these models also include housing supply for condominium and weekend cottages. The full period housing price model for condominiums is given as:

$$\log p_t^{EJ} = \alpha_0 + \alpha_1 y_t + \alpha_2 \log Y_t^d + \alpha_4 u_t + \alpha_5 \log s_t^{EJ} + \alpha_6 \upsilon_t + \alpha_7 \sigma_t + \epsilon_t$$

and for weekend cottages:

$$\log p_t^{SOM} = \alpha_0 + \alpha_1 y_t + \alpha_2 \log Y_t^d + \alpha_4 u_t + \alpha_5 \log s_t^{SOM} + \alpha_6 \upsilon_t + \alpha_7 \sigma_t + \epsilon_t$$

3.3 The sub period models

We split our data into sub period models starting from 2015(1)-2021(2) to test the conclusion's validity of no co-integration. In the sub period, we include the variable $\log(s_t^{EN})$ for single-family houses and exclude the dummy variable σ as the financial crisis is estimated to end in 2009(4) and why it is no longer a concern. The same is done for condominiums and weekend cottages, where we also exclude the dummy variable σ. The sub period housing price model for single-family houses is given as:

$$\log p_t^{EN} = \alpha_0 + \alpha_1 y_t + \alpha_2 \log Y_t^d + \alpha_4 u_t + \alpha_5 \log s_t^{EN} + \alpha_6 \upsilon_t + \epsilon_t$$

whereas for condominiums:

$$\log p_t^{EJ} = \alpha_0 + \alpha_1 y_t + \alpha_2 \log Y_t^d + \alpha_4 u_t + \alpha_5 \log s_t^{EJ} + \alpha_6 \upsilon_t + \epsilon_t$$

and for weekend cottages:

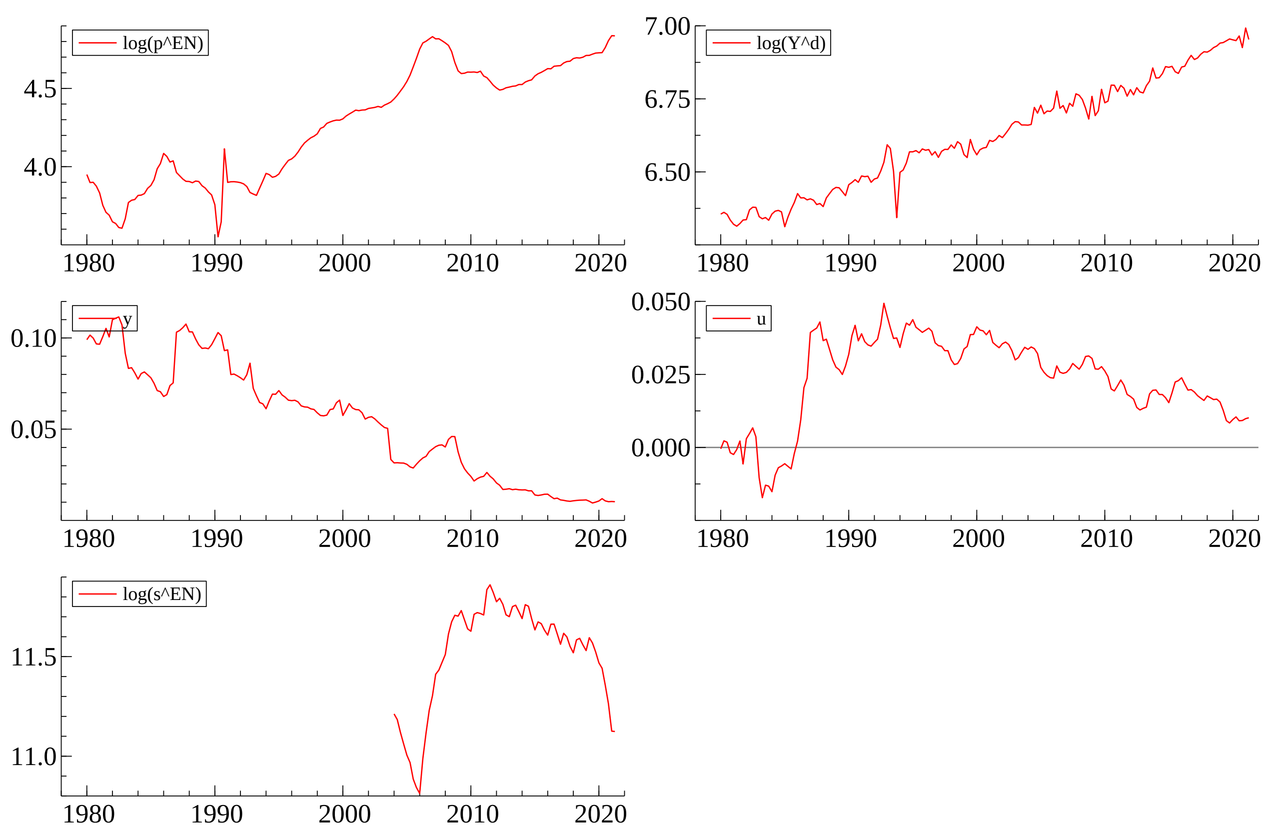
$$\log p_t^{SOM} = \alpha_0 + \alpha_1 y_t + \alpha_2 \log Y_t^d + \alpha_4 u_t + \alpha_5 \log s_t^{SOM} + \alpha_6 \upsilon_t + \epsilon_t$$

4 Empirical analysis

4.1 Graphical analysis

Figure 2, shows five different time series. We have included the real housing price for single-family houses $\log(p_t^{EN})$, real disposable income $\log(Y_t^d)$, minimal first year payment y_t, user cost u, and the supply of single-family houses for sale $\log(s_t^{EN})$.

Figure 2: Time series plots



Note: The figures are based on quarterly data. Here u and y are in %, log (s^{EN}) is the logarithmic of the number of houses for sale, log (Y^d) is the logarithmic to the disposable income in billion DKK, and log p^{EN} is the logarithmic to the single-family house price index with 2010(1)=100.
Source: Denmarks Nationalbank MONA database and Statistics Denmark

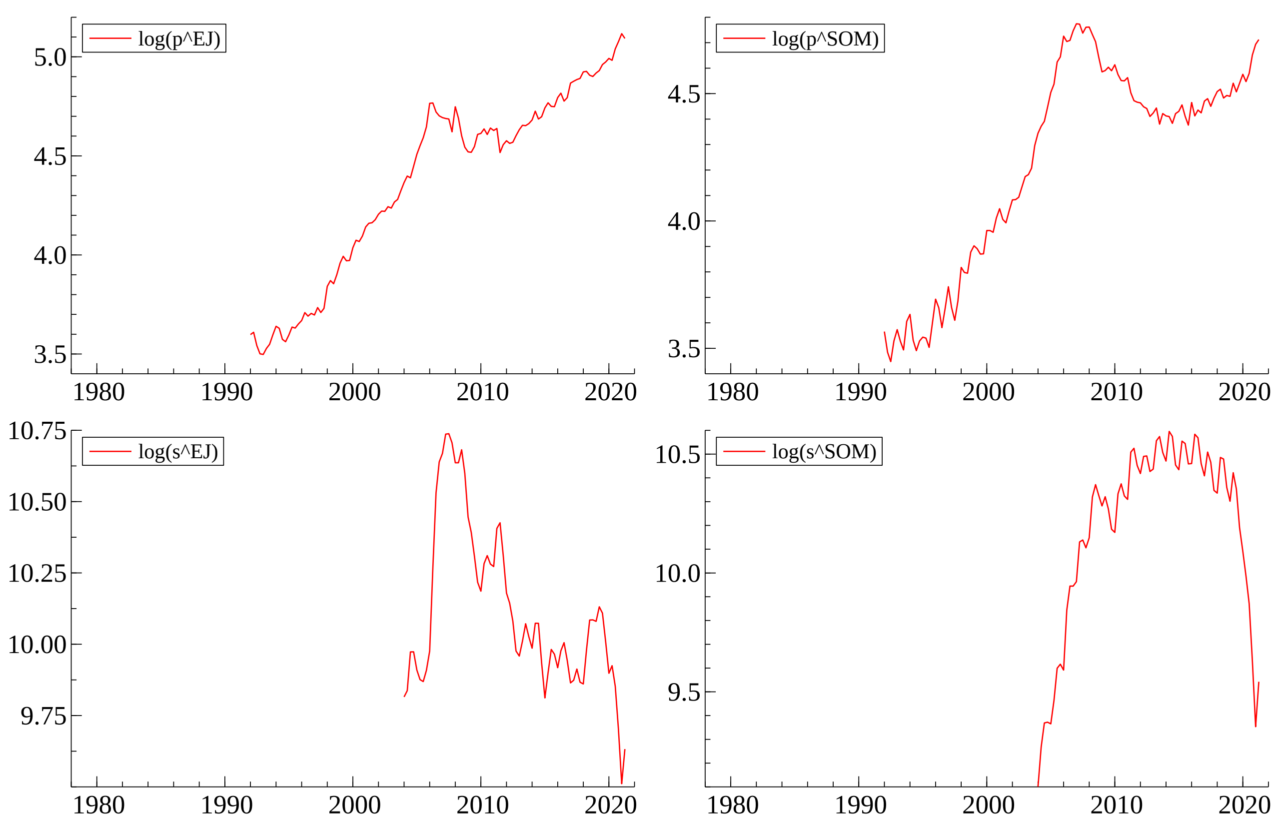
The graphs show no indications of stationarity since none of them fluctuates around a constant mean with a constant variance. The logarithmic scaled housing and log to the real disposable income seems to have an upward trend, which could be seen as trend-stationarity. The minimal first year payment graph seems to have a downward trend without a constant variance. This indicates non-stationarity and, therefore, might be a unit root process.

The graph for the user cost has experienced an overall increase throughout the period 1980(1)-

2021(2) and thus is non-stationary. From 1980-1985 we analyse that the user cost has become negative which means that buying a house was accompanied by negative expenses. This is probably because the repayments of loans were disregarded in the period.

There seems to be a level shift starting in 1987. This is probably a result of the fiscal intervention under Schlüter's government "Kartoffelkuren". A level shift might bias the formal ADF test towards non-stationarity. According to Bergman et al. (2014, pp. 171), the level shift is primarily due to three main drivers. The increased nominal interest rate where the inflation came down significantly after 1982 due to Denmark's shift to a hard currency peg. Second, in 1987 a new tax reform took effect, which led to a decrease in the mortgage interest expenses. This explains the increase in user costs in the period 1987. Third, a decrease in real international interest rates in 1982-1987 tends to reduce the user cost.

Figure 3: Time series plots



Note: The figures are based on quarterly data.
Source: Statistics Denmark

Figure 3, shows the change in real housing prices for weekend cottages $\log(p^{SOM})$, condominiums $\log(p^{EJ})$ for the period 1992(1)-2021(2). The supply for weekend cottages $\log(s^{SOM})$ and condominiums $\log(s^{EJ})$ start in the period 2004(1)-2021(2).

In 2020 we saw a drastic increase in prices for single-family housing. We assume this is due to the COVID-19 pandemic, which is further explored in section 4. The same trend can be found in weekend cottages and condominium prices (see figure 3). The supply of the three different housing types seems to follow the same trend. In the period 2004-2007, the supply increased, and in 2020 we study a remarkable decrease in the supply, which we assume to be caused by the COVID-19 pandemic. However, the supply of condominiums started to decrease from 2009 to 2018, unlike single-family houses and weekend cottages.

With further inspection of the graphical analysis, we provide a table with descriptive statistics using a simple AR(1) for each time series to test for stationarity. In table 2 and 3, we estimate that the t-values are significant for all estimates. Next, we examine that all of the coefficients for the variables are smaller than one, indicating stationarity. This is not in line with our graphical assumption. It should be noted that the models suffer from misspecification. However, the AR-test for $\log(p_t^{EN})$ and $\log(p_t^{EJ})$ indicates that we can accept the hypothesis of no autocorrelation at a 5% critical level. This indicates that this model might be more reliable than the rest.

We continue the analysis with the assumption that all of the presented variables are non-stationary. Due to the misspecification errors in our AR models, the graphical inspections will be a better at approximating for stationarity.

Table 2: AR(1) descriptive statistic for single-family houses

	$\log p^{EN}$	y	u	$\log Y^d$
Sample period	1980(1)-2021(2)	1980(1)-2021(2)	1980(1)-2021(2)	1980(1)-2021(2)
No. of obs.	165	165	165	165
Mean	4.276	0.054	0.024	6.624
Std.Error	0.366	0.031	0.015	0.190
AR(1) estimation with constant term				
Coefficient	0.995	0.991	0.971	0.991
t-value	91.0	98.9	57.6	80.7
AR 1-5 test	[0.1720]	[0.0740]	[0.0050]	[0.0000]
Normality test	[0.0000]	[0.0000]	[0.0000]	[0.0000]
Hetero test	[0.0035]	[0.1051]	[0.2109]	[0.7146]

Note: p-values in [·]

Note: Data from 1980(1)-2021(2).

Source: Denmarks Nationalbank MONA database and Statistics Denmark.

Table 3: AR(1) descriptive statistic for condominiums and weekend cottages

	log p^{SOM}	log p^{EJ}	log s^{SOM}	log s^{SOM}
Sample period	1992(1)-2021(1)	1992(1)-2021(2)	2004(1)-2021(2)	2004(1)-2021(2)
No. of obs.	117	117	69	69
Mean	4.238	4.376	10.205	10.122
Std.Error	0.405	0.471	0.371	0.286
AR(1) estimation with constant term				
Coefficient	0.988	0.995	0.926	0.962
t-value	95.7	135	29.6	24.8
AR 1-5 test	[0.0000]	[0.0820]	[0.0000]	[0.0000]
Normality test	[0.9984]	[0.0003]	[0.0004]	[0.0442]
Hetero test	[0.0000]	[0.8343]	[0.0122]	[0.8287]

Note: p-values in [·]

Note: Data from 1992(1)-2021(2).

Source: Denmarks Nationalbank MONA Database and own calculations.

4.2 Unit root test

Table 4 shows the unit root results. We estimate five AR(p) models for the variables $\log(p_t^{EN})$, y, u, and $\log(Y^d)$:

Table 4: Unit root results

	log p^{EN}	y	u	log Y^d	log s^{EN}
ADF test statistic	-1.552	-1.274	-0.2636	-0.1589	1.583
P-value	[0.5456]	[0.6667]	[0.1514]	[0.0141]	[0.3548]
Number of lags in AR(p)	2	2	5	5	9
Determinisstic terms	δ	δ	δ	δ	δ
Conclusion	I(1)	I(1)	I(1)	I(1)	I(1)

δ indicates a constant term is being used.

Note: Data from 1992(1)-2021(2).

Source: Denmarks Nationalbank MONA Database, Denmarks statistics and own calculations.

We choose two lags for $\log(p_t^{EN})$ as reducing the lag length further will cause autocorrelation. For disposable income, the optimal lag length is five. For user cost and minimal first year payment, it is two. We use the ADF test based on a critical value of -2.89 on a 5% critical level. The results show that for $\log(p_t^{EN})$, we cannot reject the null hypothesis that the variable is a unit root. This is shown by the ADF-test in table 4, where $\log(p_t^{EN})$ has a test statistic of -1.55, which is higher than the critical value of -2.89. Since the test statistic is greater than the critical value, we cannot reject the null hypothesis of the variable being a unit root process including a constant term. Continuing the same strategy with the rest of the variables, we cannot reject the null hypothesis for any of the variables. All the variables are non-stationary $I(1)$.

Next, we will examine the variables for condominiums and weekend cottages. The unit root test for the variable $\log(s_t^{EN})$ starts in the period 2015(1), as the variable is only included in the sub period. The four other variables start in the period 1992(1) and end in 2021(2). Table 5 represents the unit root results for condominiums and weekend cottages. Thus, four more AR(p) models are estimated:

Table 5: Unit root results for condominiums and weekend cottages

	log p^{SOM}	log p^{EJ}	log s^{SOM}	log s^{EJ}
ADF test statistic	-2.130	-0.6360	-0.7063	-1.020
P-value	[0.1603]	[0.3986]	[0.6983]	[0.6276]
Number of lags in AR(p)	9	4	10	6
Deterministic terms	δ	δ	δ	δ
Conclusion	I(1)	I(1)	I(1)	I(1)

δ indicates a constant term is being used.

Note: Data from 1992(1)-2021(2).

Source: Denmark Statistic and own calculations.

We first observe that all variables are greater than the critical value of -2.89. Therefore, we cannot reject the null hypothesis of the variable being a unit root process. This implies that none of the variables are stationary, which is also in line with our graphical assumption. The variables for supply start in 2005(1), and the price variables start in 1992 until 1995 due to the different lag structures.

4.3 Co-integration results

4.3.1 Single-family houses

For the co-integration test we use the variables $\log(p_t^{EN})$, $\log(Y_t^d)$, y_t, u_t, $\log(s_t^{EN})$, and the two dummy variables σ_t and υ_t. We perform two different co-integration tests, one for the main sample 1992(1)-2021(2) and one for the sub period 2015(1)-2021(2). This co-integration setup is also applied to the two other housing types, condominiums and weekend cottages. Our vector of variables and $\hat{\beta}$ can be written as:

$$X_t = \begin{pmatrix} \log p_t^{EN} \\ \log Y_t^d \\ u_t \\ y_t \\ \upsilon_t \\ \sigma_t \\ 1 \end{pmatrix}, \hat{\beta} = \begin{pmatrix} 1 \\ -\beta_2 \\ -\beta_3 \\ -\beta_4 \\ -\beta_5 \\ -\beta_6 \\ -\mu \end{pmatrix}$$

If there exists a vector, $\hat{\beta}$ such that X_t, defined in the theory section, is a stationary process then it will be $I(0)$. This property is denoted co-integration. We expect the following signs of the variables in the later estimated co-integration vector to be, $\beta_2 > 0$, $\beta_3 < 0$, $\beta_4 < 0$, $\beta_5 > 0$, $\beta_6 < 0$ and $\mu > 0$.

In the theory section, we estimated the long-run solution, which can be used to calculate the direct impact of the co-integration vector on the variable $\log(p_t^{EN})$ in equilibrium. The disposable income is expected to impact the housing prices positively, $\beta_2 > 0$ since an increase in disposable income would make it more affordable to buy a house.

The user cost is expected to have a negative impact on housing prices. An increase in user cost implies an increase in real interest rates and house-related taxes. Therefore, it becomes more expensive to own a house when user cost increases, resulting in lower house prices.

We expect the β_4 coefficient to be negative due to an increase in the minimal first year payment, which is expected to decrease the disposable household income. Thus, loans become more difficult to obtain, thereby decreasing the prices.

We expect the COVID-19 dummy to positively impact the housing prices, as the pandemic restricted the population to stay at home, $\beta_5 > 0$. This should increase the importance of the home, thereby increasing people's willingness to invest in the right home. Furthermore, it makes the households spend less money on vacations and other leisure activities, which would otherwise decrease the household's disposable income. This should, in combination, create an increase in housing prices.

The dummy variable for the financial crisis is expected to have a negative impact on housing prices, $\beta_6 < 0$. There are many effects related to the financial crisis, but the most important assumption is that the housing market was in a bubble, stabilising housing prices when people started to default on their loans.

It should be noted that the dummy variable σ is not included in the sub period for all of the housing types. Also, the supply variable for single-family houses is only included in the sub period, which is estimated to impact housing prices negatively. This is because a supply surplus would decrease prices, as described thoroughly in section 2.1.1 for housing demand and supply.

4.3.2 Condominiums

In the co-integration test for condominiums we use the variables $\log(p^{EJ})$, $\log(Y^d)$, y, $\log(s^{EJ})$, σ and υ. Our vector of variables and β can be written as:

$$X_t = \begin{pmatrix} \log p_t^{EJ} \\ \log Y_t^d \\ y_t \\ \log s_t^{EJ} \\ \upsilon \\ \sigma \\ 1 \end{pmatrix}, \hat{\beta} = \begin{pmatrix} 1 \\ -\beta_2 \\ -\beta_3 \\ -\beta_4 \\ -\beta_5 \\ -\beta_6 \\ -\mu \end{pmatrix}$$

We expect the following signs of the variables to be, $\beta_2 > 0$, $\beta_3 < 0$, $\beta_4 < 0$, $\beta_5 > 0$, $\beta_6 < 0$ and $\mu > 0$.

The supply of condominiums is expected to have a negative impact on the condominium prices, $\beta_4 < 0$. As our supply and demand theory proves, an increase in supply decreases condominium prices.

4.3.3 Weekend cottages

In the co-integration test for weekend cottages, we use the variables $\log(p_t^{SOM})$, $\log(Y_t^d)$, y_t, $\log(s_t^{SOM})$, υ, and σ. Our vector of variables and β can be written as:

$$X_t = \begin{pmatrix} \log p_t^{SOM} \\ \log Y_t^d \\ y_t \\ \log s_t^{SOM} \\ \upsilon \\ \sigma \\ 1 \end{pmatrix}, \hat{\beta} = \begin{pmatrix} 1 \\ -\beta_2 \\ -\beta_3 \\ -\beta_4 \\ -\beta_5 \\ -\beta_6 \\ -\mu \end{pmatrix}$$

We expect the following signs of the variables to be, $\beta_2 > 0$, $\beta_3 < 0$, $\beta_4 < 0$, $\beta_5 > 0$, $\beta_6 < 0$ and $\mu > 0$.

The supply of weekend cottages is also expected to have a negative impact on the weekend cottage prices, $\beta_3 < 0$, using the same argument as earlier.

4.4 ADL & ECM estimates for single-family houses

4.4.1 ADL

We estimate an ADL(4,4) model for the main sample 1992(1)-2021(2) without the variable $\log(s_t^{EN})$, which is only included in the sub period:

$$\begin{aligned}
\log p_t^{EN} = \; & \underset{(0.0776)}{1.606} \log p_{t-1}^{EN} - \underset{(0.113)}{0.5277} \log p_{t-2}^{EN} - \underset{(0.0472)}{0.08803} \log p_{t-4}^{EN} \\
& - \underset{(0.0351)}{0.06702} \log {Y^d}_{t-1} + \underset{(0.0401)}{0.1174} \log Y_{t-2}^d + \underset{(0.0406)}{0.04264} \log Y_{t-3}^d \\
& - \underset{(0.035)}{0.1119} \log Y_{t-4}^d - \underset{(0.54)}{1.271} \, \mathrm{u}_t + \underset{(0.523)}{1.41} \, \mathrm{u}_{t-1} \\
& - \underset{(0.261)}{1.058} \, \mathrm{y}_t + \underset{(0.279)}{0.8115} \, \mathrm{y}_{t-3} + \underset{(0.00839)}{0.0307} \, \upsilon_{t-1} \\
& - \underset{(0.0133)}{0.02876} \, \upsilon_{t-3} - \underset{(0.0158)}{0.01986} \, \upsilon_{t-4} - \underset{(0.00855)}{0.02719} \, \sigma_{t-3} \\
& + \underset{(0.00843)}{0.02307} \, \sigma_{t-4} + \underset{(0.227)}{0.1744}
\end{aligned}$$

where the standard errors are in the parentheses. Table 6 shows the estimates of the preferred ADL model for the main sample 1992(1)-2021(2). The estimated ADL model is well specified. This is due to the p-value [0.49] for no autocorrelation, which is greater than the 5% critical level [0.05]. Therefore, we cannot reject the null hypothesis of no autocorrelation. Next, the p-value for no ARCH effects is [0.97], meaning we cannot reject the H_0 of no ARCH effects. The same is true for no heteroskedasticity [0.12] and normality [0.13]. The most important misspecification test is the one which ensures no autocorrelation, though we obviously aim to avoid misspecification in all our tests.

The optimal number of lags to include is four. However, some of the variables are insignificant, which is solved with the general-to-specific approach, where we exclude one variable after another based on their 10% critical level. This implies that our ADL model almost only contains significant variables. The constant is not significant, but we do not remove it as we assume it is important to include a constant for the no co-integration test.

We choose to split our data into two samples to test if the results are robust or if the final conclusion will change. The preferred sub period starts in 2015(1) and ends in 2021(2). We choose this sub period as the test for no autocorrelation cannot be rejected for the ADL models we estimate.

Table 6: The table shows estimates of the ADL model for single-family houses in the long- and short run with various restrictions imposed. Standard errors in (·) and p-values in [·] for misspecification tests. SFH stands for single-family houses.

	(SFH full period)	(SFH sub period)
Constant	0.174 (0.227)	3.133 (0.916)
COVID-19	.	−0.015 (0.004)
COVID-19$_1$	0.031 (0.008)	0.017 (0.006)
COVID-19$_2$	.	0.030 (0.006)
COVID-19$_3$	−0.029 (0.013)	.
COVID-19$_4$	−0.020 (0.016)	−0.023 (0.006)
Financialcrisis$_3$	−0.027 (0.009)	.
Financialcrisis$_4$	0.023 (0.008)	.
log s^{EN}	.	−0.086 (0.025)
log s_2^{EN}	.	−0.111 (0.033)
log Y_1^d	−0.067 (0.035)	.
log Y_2^d	0.117 (0.040)	0.162 (0.054)
log Y_3^d	0.043 (0.041)	.
log Y_4^d	−0.112 (0.035)	.
log p_1^{EN}	1.606 (0.078)	0.111 (0.125)
log p_2^{EN}	−0.528 (0.113)	.
log p_3^{EN}	.	0.472 (0.084)
log p_4^{EN}	−0.088 (0.047)	.
u	−1.271 (0.540)	.
u$_1$	1.410 (0.523)	.
y	−1.058 (0.261)	.
y$_3$	0.811 (0.279)	.
$\hat{\sigma}$	0.011	0.003
Log-lik.	376.830	118.493
AIC	-6.099	-8.346
HQ	-5.937	-8.206
SC/BIC	-5.700	-7.862
No autocorr. 1-5	[0.49]	[0.51]
No ARCH 1-2	[0.93]	[0.81]
No hetero.	[0.46]	[0.91]
Normality	[0.13]	[0.31]
T	118	26
Sample start	1992(1)	2015(1)
Sample end	2021(2)	2021(2)

Source: Denmarks Nationalbank MONA database, Statistics Denmark and own calculations.

4.4.2 ECM

Based on the ADL model for full period, we estimate the ECM:

$$\begin{aligned}
\Delta\log p^{EN} \quad = \quad & \underset{(0.113)}{0.5277}\ \Delta\log p^{EN}_{t-1}\ +\ \underset{(0.227)}{0.1744}\ -\ \underset{(0.0424)}{0.04815}\ \Delta\log Y^{d}_{t-1} \\
& + \underset{(0.0393)}{0.06924}\ \Delta\log Y^{d}_{t-2}\ +\ \underset{(0.035)}{0.1119}\ \Delta\log Y^{d}_{t-3}\ -\ \underset{(0.54)}{1.271}\ \Delta u_t \\
& + \underset{(0.0464)}{0.07843}\ \log p^{EN}_{t-1}\ -\ \underset{(0.0472)}{0.08803}\ \log p^{EN}_{t-4}\ -\ \underset{(0.0343)}{0.01888}\ \log Y^{d}_{t-1} \\
& - \underset{(0.261)}{1.058}\ \mathrm{y}_t\ +\ \underset{(0.279)}{0.8115}\ \mathrm{y}_{t-3}\ +\ \underset{(0.324)}{0.1388}\ \mathrm{u}_{t-1} \\
& + \underset{(0.00839)}{0.0307}\ \upsilon_{t-1}\ -\ \underset{(0.0133)}{0.02876}\ \upsilon_{t-3}\ -\ \underset{(0.0158)}{0.01986}\ \upsilon_{t-4} \\
& - \underset{(0.00855)}{0.02719}\ \sigma_{t-3}\ +\ \underset{(0.00843)}{0.02307}\ \sigma_{t-4}
\end{aligned}$$

The ECM for full period and sub period can be found in table 7. Using the equation for t-statistic, we test for no co-integration:

$$1992Q2 - 2021Q2 : \hat{\tau}_{\psi_1} = \frac{\hat{\psi}_1}{se\left(\hat{\psi}_1\right)} \Leftrightarrow \frac{-0.00961}{0.00759} = -1.27$$

The t-statistic is higher than our 5% critical value -4.38, the null hypothesis cannot be rejected and thus there is no co-integration relationship in this model. The critical values can be found in table 9. An explanatory factor could be that the full period model includes the financial crisis period, which might interfere with the results of co-integration between log(p^{EN}) and the COVID-19 dummy. This is also the case for condominiums and weekend cottages.

The sub period model for single-family housing starts in 2015(1) and ends in 2021(2) with the new variables given in table 7. Using the equation for t-statistic again, the test for no co-integration gives:

$$2015Q1 - 2021Q2 : \hat{\tau}_{\psi_1} = \frac{\hat{\psi}_1}{se\left(\hat{\psi}_1\right)} \Leftrightarrow \frac{-0.417}{0.064} = -6.52$$

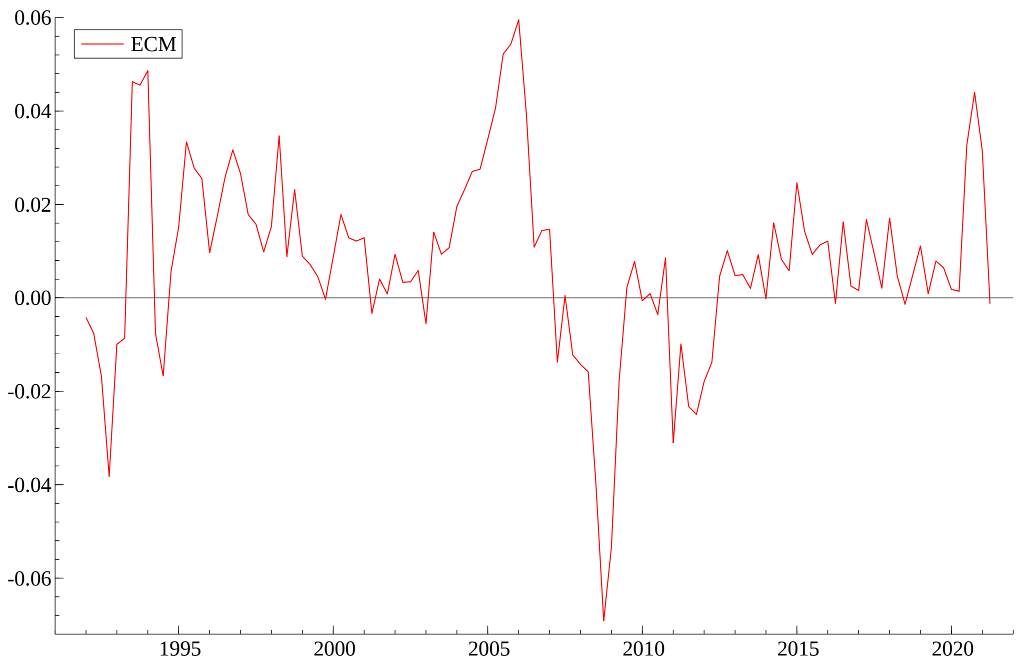
Here we examine that the t-statistic is greater than our critical value and thus the null hypothesis is rejected. It is important to note that in our sub period model we include the variable log s^{EN}.

Table 7: The table shows estimates of the ECM for single-family houses with various restrictions imposed. Standard errors in (·) and p-values in [·] for misspecification tests.

	(SFH ECM full period)	(SFH ECM sub period)
Constant	0.174 (0.227)	3.133 (0.916)
COVID-19	.	-0.015 (0.004)
COVID-19_1	0.031 (0.008)	0.017 (0.006)
COVID-19_2	.	0.030 (0.006)
COVID-19_3	-0.029 (0.013)	.
COVID-19_4	-0.020 (0.016)	-0.023 (0.006)
Δlog Y_1^d	-0.048 (0.042)	.
Δlog Y_2^d	0.069 (0.039)	.
Δlog Y_3^d	0.112 (0.035)	.
Δlog p_1^{EN}	0.528 (0.113)	.
Δu	-1.271 (0.540)	.
Financialcrisis_3	-0.027 (0.009)	.
Financialcrisis_4	0.023 (0.008)	.
log s^{EN}	.	-0.086 (0.025)
log s_2^{EN}	.	-0.111 (0.033)
log Y_1^d	-0.019 (0.034)	.
log Y_2^d	.	0.162 (0.054)
log p_1^{EN}	0.078 (0.046)	-0.889 (0.125)
log p_3^{EN}	.	0.472 (0.084)
log p_4^{EN}	-0.088 (0.047)	.
u_1	0.139 (0.324)	.
y	-1.058 (0.261)	.
y_3	0.811 (0.279)	.
$\hat{\sigma}$	0.011	0.003
Log-lik.	376.830	118.493
AIC	-6.099	-8.346
HQ	-5.937	-8.206
SC/BIC	-5.700	-7.862
No autocorr. 1-5	[0.49]	[0.51]
No ARCH 1-2	[0.96]	[0.81]
No hetero.	[0.25]	[0.91]
Normality	[0.13]	[0.31]
T	118	26
Sample start	1992(1)	2015(1)
Sample end	2021(2)	2021(2)

Source: Denmarks Nationalbank MONA database, Statistics Denmark and own calculations.

Figure 4: ECM for single-family houses



Note: Data from 1992(1)-2021(2)
Source: Denmarks Nationalbank MONA Database and own calculations

Figure 4 visualises the ECM single-family houses in the full period. We observe that the figure seems to be stationary. The fluctuations describe the different periods where the house prices are different from the equilibrium. A positive fluctuation indicates that the housing prices are above their structural level, and vice versa when negative. This is also in line with the assumption about the effect of the financial crisis, where we observe that the fluctuation is negative in the period 2008-2010. Furthermore, the fluctuation in the period 2020-2021(2) is positive, which is in line with our COVID-19 assumption.

4.5 ADL & ECM estimates for Condominiums

4.5.1 ADL estimates

We estimate an ADL(4,4) model for condominiums in the period 2005(1)-2021(2), which is our full period. Table 10 shows the estimates for the preferred models, which include both the full period and the sub period for condominiums. Again, we choose to include significant lags on a 10% critical level. We first observe that the variable, user cost, has no significant value, so we exclude that variable. This could be because the user cost is estimated for single-family houses and therefore has no real meaning in describing condominiums. However, the rest of the variables are still significant and are therefore included in the model.

Next, we observe that both models for the full period and the sub period have no problems with misspecifications. The test statistic for no autocorrelation for the full period is [0.17], thus greater than our critical value, so we cannot reject the null hypothesis of no autocorrelation. The estimate for normality is [0.09], where we again cannot reject the null hypothesis of no normality at a 5% critical level. The estimate for no heteroskedasticity is [0.70], meaning we do not have any problems with heteroskedasticity. Lastly, we do not have any ARCH effects with an estimate of [0.21].

We continue with the sub period, 2015(1)-2021(2) and again find that the user cost does not have any significant value for condominiums. We also choose to exclude the dummy variable σ as it is no longer relevant for the estimated period. Therefore, we end up with an ADL(4,4) model for the sub period. We observe that the estimate for no autocorrelation is [0.37], which is higher than the full period. Next, we observe that the estimates for normality, no heteroskedasticity, and ARCH-effects are [0.79], [0.86], [0.92]. This is closer to zero than the full period, indicating that the sub period model is more specified than the full period.

We now estimate two ECM based on the ADL models, which we will use for the co-integration tests.

4.5.2 ECM estimates

The table for the ECM for full period is in table 11. Using the ECM for t-statistic, we can test for no co-integration. We start with the full period:

$$2005Q1 - 2021Q2 : \hat{\tau}_{\psi_1} = \frac{\hat{\psi}_1}{se\left(\hat{\psi}_1\right)} \Leftrightarrow \frac{0.2231}{0.0699} = 3.19$$

As the t-statistic is higher than our 5% critical value of -4.19, the null hypothesis cannot be rejected, and thus there is no co-integration relationship in this model.

The sub period model for condominiums starts in 2015(1) and ends in 2021(2) with the new variables which were given in table 11. Using the equation for ECM again, the test for no co-integration gives:

$$2015Q1 - 2021Q2 : \hat{\tau}_{\psi_1} = \frac{\hat{\psi}_1}{se\left(\hat{\psi}_1\right)} \Leftrightarrow \frac{-0.7701}{0.0707} = -10.89$$

Since the t-statistic is lower than our critical value, the null hypothesis can be rejected and thus, there is a sign of a co-integration relationship in this model.

4.6 ADL & ECM estimates for weekend cottages

4.6.1 ADL

The last housing type we choose to include is weekend cottages. The preferred ADL model is an ADL(2,3) for the full period 2005(1)-2021(2), which can be found in table 12. First, we observe that the variable for user cost is significant on a 10% critical level. Next, we estimate that the model has no problems with misspecification. The test statistics for no autocorrelation, normality, heteroskedasticity, and ARCH-effects are $0.76, 0.68, 0.60$, and 0.87. All the test statistics are greater than the 5% critical level.

Next, looking at the sub period from 2015(1)-2021(2), the variable for user cost and σ dummy variable are excluded from the ADL(1,4) model. This model, too, has no problems with misspecifications. The test statistics for no autocorrelation, normality, no heteroskedasticity and no ARCH-effects are $[0.39], [0.71], [0.23]$, and $[0.47]$. All the test statistics are greater than the 5% critical level. Compared to the full period model, the sub period does not seem more specified.

4.6.2 ECM estimates

The ECM for full period and the sub period can be found in table 13. Using the estimated ECM, we can test for no co-integration. We start with the full period:

$$2005Q1 - 2021Q2 : \hat{\tau}_{\psi_1} = \frac{\hat{\psi}_1}{se\left(\hat{\psi}_1\right)} \Leftrightarrow \frac{-0.092}{0.036} = -2.56$$

As the t-statistic is greater than our 5% critical value of -4.38, the null hypothesis cannot be rejected, and thus there is no co-integration relationship in this model.

The sub period model for weekend cottages starts in 2015(1) and ends in 2021(2) with the new variables given in table 13. Using the equation for the ECM once again, the test for no co-integration gives:

$$2005Q1 - 2021Q2 : \hat{\tau}_{\psi_1} = \frac{\hat{\psi}_1}{se\left(\hat{\psi}_1\right)} \Leftrightarrow \frac{-1.52}{0.102} = -14.90$$

Here, the t-statistic is lower than our critical value, the null hypothesis can be rejected, and thus there is a sign of a co-integration relationship in this model.

4.7 The long-run solution

4.7.1 Single-family houses

Using the ECM for single-family houses, we estimate the static long-run solution to determine the impact of the different variables' long run impact on housing prices. We find the long-run solution for the ECM by estimating the co-integration vector, $\hat{\beta}$, as shown in section 2.3 equation (4). We will start with the long period and next estimate the long-run solution for the sub period:

$$\hat{\beta}' = \begin{bmatrix} 1 & \frac{-0.0189}{-0.00961} & \frac{0.139}{-0.00961} & \frac{-0.247}{-0.00961} & \frac{-0.00412}{-0.00961} & \frac{0.174}{-0.00961} & \frac{-0.0179}{-0.00961} \end{bmatrix}$$

$$\Leftrightarrow$$

$$\hat{\beta}' = \begin{bmatrix} 1 & 1.96528 & -14.44558 & 25.6816 & 1.86635 & 0.42918 & -18.161 \end{bmatrix}$$

We can rewrite this as the long-run solution:

$$X_t\hat{\beta} = \begin{pmatrix} \log p_t^{EN} \\ \log Y_t^d \\ u_t \\ y_t \\ \upsilon_t \\ \sigma_t \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ -\beta_2 \\ -\beta_3 \\ -\beta_4 \\ -\beta_5 \\ -\beta_6 \\ -\mu \end{pmatrix} = \begin{pmatrix} 1 \\ 1.97 \\ -14.45 \\ 25.68 \\ 1.87 \\ 0.43 \\ -18.16 \end{pmatrix}$$

The long-run solution is where $X_t\hat{\beta} = 0$:

$$\log p_t^{EN} + \mu - \beta_2 \log Y_t^d - \beta_3 u_t - \beta_4 y_t - \beta_5 \upsilon - \beta_6 \sigma = 0 \Leftrightarrow$$

$$\log p_t^{EN} + 18.16 + 1.97 \log Y_t^d - 14.45 u_t + 25.68 y_t + 1.87 \upsilon + 0.43\sigma = 0 \Leftrightarrow$$

$$\begin{aligned} \log p_t^{EN} = \; & \underset{(0.5524)}{18.16} - \underset{(0.6561)}{1.97} \log Y_t^d + \underset{(0.6741)}{14.45} \; \mathrm{u}_t \\ & - \underset{(0.3660)}{25.68} \; \mathrm{y}_t - \underset{(0.3264)}{1.87} \; \upsilon - \underset{(0.5686)}{0.43} \; \sigma \end{aligned} \tag{8}$$

When analysing the long-run solution for the full period of single-family houses, we achieve five variables besides the constant term. Though their signs are contrary to our economic understanding, it does not matter since all the variables are insignificant. This means that in the long term, we reject the hypothesis that any of the variables matter to the price of single-family housing.

- **Sub period**

$$\log p^{EN} = \underset{(0.0002)}{7.51} - \underset{(0.6561)}{1.97} \log s^{EN} + \underset{(0.0028)}{0.39} \log Y^d + \underset{(0.4841)}{0.02} \upsilon$$

We get one significant variable based on a 5% critical value when compiling our coefficients based on the sub period. This is the real disposable income, which positively affects the housing prices in the long run. A one percent increase in disposable income should increase the prices for single-family houses by 0.39 percent. This increase follows basic economic theory, which says that a higher income should increase spending on normal goods.

4.7.2 Condominiums

- **Full period**

Using the ECM for condominiums shown in table 11, we estimate the static long-run solution to determine the impact of the different variables on the housing prices for condominiums. The equation for the full period is:

$$\log p^{EJ} = \underset{(0.0010)}{-15.94} + \underset{(0.0004)}{0.75} \log s^{EJ} + \underset{(0.0021)}{1.94} \log Y^d - \underset{(0.1220)}{10.79} \text{ y} + \underset{(0.0172)}{0.41} \upsilon$$

where the p-values are in (). The minimal first year payment variable is insignificant, with a p-value of (0.12) and a 5% critical value. We observe that the rest of the variables are significant on a 5% critical level. Next, we exclude the financial crisis dummy variable because it is insignificant in our ECM. Firstly, a one percentage change in supply for condominiums will increase the prices for condominiums by 0.75 percentage. This increase does not align with our supply and demand theory expectations. Second, a one percentage change in disposable income will lead to a 1.94 percent increase in prices, which is in line with our expectations. Third, the minimal first year payment variable is insignificant in our long-run solution, with a p-value of (0.12). Therefore, we will not interpret the coefficient. Fourth, one percentage change in COVID-19 will lead to a 0.41 percent increase in prices, which is also in line with our assumption.

- **Sub period**

Next, we estimate the static long-run solution for the sub period.

$$\log p^{EJ} = \underset{(0.000)}{-3.51} - \underset{(0.0063)}{0.13} \log s^{EJ} + \underset{(0.000)}{1.60} \log Y^d$$
$$- \underset{(0.0000)}{21.70} \text{ y} + \underset{(0.7850)}{0.01} \upsilon$$

The coefficient shows that a one percent change in condominium supply will decrease the prices by 0.13 percent. This decrease is different compared to the full period for condominiums. We expected the supply to have a negative impact on the prices, which it does have in the sub period. A one percent change in disposable income will increase prices by 1.60 percent, almost the same percentage change for the full period. The minimal first year payment variable will decrease the prices by 21.70 percent with a one percentage change in minimal first year payment. This sign is also expected. The last variable, COVID-19, is insignificant, with a p-value of (0.78), which means we will not interpret the results.

4.7.3 Weekend cottages

- **Full period**

Using the ECM for weekend cottages shown in table 13, we estimate the static long-run solution to determine the impact of the different variables on the housing prices for weekend cottages. The equation for the full period is:

$$\log p^{SOM} = \underset{(0.0004)}{17.20} - \underset{(0.0004)}{1.22} \log s^{SOM} - \underset{(0.0746)}{0.97} \upsilon$$
$$- \underset{(0.0910)}{0.36} \sigma$$

where the p-values are in (). The dummy variables for COVID-19 (0.0746) and the financial crisis (0.0910) are insignificant on a 5% critical value. The housing supply of weekend cottages is the only variable left in the long-run solution for the full period. One percentage change in the supply of weekend cottages will decrease the price for weekend cottages by a 0.13 percentage. This decrease aligns with the theory section about housing demand & supply.

- **Sub period**

We estimate the static long-run solution for the sub period model of weekend cottages. Note that this model does not include the financial crisis dummy.

$$\log p^{SOM} = \underset{(0.0009)}{2.65} - \underset{(0.0000)}{0.10} \log s^{SOM} - \underset{(0.0003)}{6.32} \text{ y}$$
$$+ \underset{(0.0001)}{0.42} \log Y^d - \underset{(0.0000)}{3.84} \text{ u} + \underset{0.0032}{0.05} \upsilon$$

The long-run solution shows that all p-values are below the critical value, which means all variables are significant. As expected, a one percent increase in housing supply decreases prices by 0.09 percent. A one percentage increase in minimal first year payment decreases the price by 6.31 percent, which aligns with our theory. A one percentage increase in disposable income increases the price by 0.42 percent. Whereas a one percent increase in user cost decreases price by 3.84 percent. Lastly, a one percentage increase in the COVID-19 dummy variable increases the price by 0.05 percent, confirming our theory.

5 Discussion

This section discusses and compares our data, methodology, and conclusion with other studies and research on similar topics.

5.1 Data

This paper uses variables from the Danish National Bank's database MONA for single-family houses, user cost, minimal first year payment, and disposable income. The data goes back to the first quarter of 1980 and ends in the second quarter of 2021. Furthermore, we apply data from Statics Denmark for the variables used to describe condominiums and weekend cottages. This data goes back to the first quarter of 1992 and ends in the fourth quarter of 2021. In section 4.1, Graphical analysis, the housing price variables are based on the second quarter of 1992, which experienced a significant decline in 1990-1991. This decline could interfere with our chosen model. The sudden drop in $\log(p_t^{EN})$ can also be seen in figure 2. The early 1990s recession in western countries could explain the sudden drop in housing prices. This recession particularly affected Denmark since it is a small open economy and thus more sensitive to changes from outside forces. The same could be said for 2008-2010 when the world economy was affected by the financial crisis. Therefore, we included a dummy variable to account for large outliers and start our full period model in 1992.

The second dummy variable in our model is to test for the co-integration relationship between housing prices and the presence of COVID-19. Denmark confirmed its first case of COVID-19 on 27th February 2020. Shortly after, the government induced a lockdown starting on 13th March.

The dummy variable for COVID-19 starts in the second quarter of 2020 and ends in the second quarter of 2021. Some people might argue that the dummy variable should start in the first quarter of 2020 since other countries were already affected by COVID-19, but since Denmark was not affected until the end, and housing prices are stable in the short run, we decided only to include it from the second quarter.

Autocorrelation is present when estimating an ADL model for single-family houses, starting in the first quarter of 1980. Therefore, the model would be insignificant, and the results would not suit our paper. In order to obtain a model with no misspecification problems, we choose to start in the first quarter of 1992. A limitation of this method is the lack of objectivity, meaning that the results are not a perfect reflection of the data. The same issues are seen in condominiums and weekend cottages, where the ADL model suffers from misspecifications if we start earlier than 2005. When starting in 2005, we can include a variable for housing supply for condominiums and weekend cottages. We get a more exact model by including an additional variable for these models. However, this complicates comparing the full period model for single-family houses, condominiums, and weekend cottages since the model for single-family houses do not include the housing supply variable. Since the main focus of the paper was to see if COVID-19 had an effect on housing prices, it was deemed optimal to include a variable for housing supply for condominiums and weekend cottages. This results in different starting points for the full period models since we focus on having more data. As previously explained, we split our data into sub period models to check if the conclusion changes.

To further expand on our analysis, regions could be included. This would make our estimates more precise as some regions have more specific housing types than others. Another way to elaborate on our paper could be by including more variables such as housing size, grace-period loans, expected inflation, and investment payoff. An advantage of including more variables could be that our model is better specified. However, it could result in overfitting, which means that the model contains too many variables and thus explains less. A good model can explain as much as possible whilst being as simple as possible.

5.2 Methodology

There are several approaches to determining co-integration between variables. In our case, we chose the ADL/ECM approach. To use the ADL/ECM approach, we assumed that only the housing price error corrects. We assume that user cost and disposable income do not error correct. When estimating the co-integration between the variables, we use the Engle & Granger method to assume that only the housing price error corrects. One of the cons of this method is that it does not examine the individual variable time series. Furthermore, autocorrelation could still be a problem. Nevertheless, we choose a model with no autocorrelation, fixing the issue. The Engle & Granger method is useful as long as the ADL model is well specified.

As we have more than five variables, it would be efficient to use an ADL model since it can interpret numerous variables. We can create a long-run solution based on the well-specified ECM and then interpret the significant variables. These results would help us estimate the effect of COVID-19 compared to other variables. Another approach could be to estimate a difference-in-difference estimator. However, this is difficult for COVID-19. For example, we would need a control group which is similar but is not affected by COVID-19. This is also too simple for our purpose.

5.3 Co-integration results

When testing for co-integration, we find that the three housing type models in the full period have no signs of co-integration. Therefore, the ECM would be misleading, and the results of the long-run solution could be inconsistent. However, we still estimated the long-run solution for these models to test if the variables were significant without co-integration. The first model for single-family houses in the full period has insignificant variables only. The two other housing types models show significant variables, which contradicts our expectations. The explanation for this could be that the significance level is independent of the co-integration results. Therefore, it could be argued that the long-run relationships for the full period should not be interpreted at all.

5.4 Comparison with other studies

5.4.1 Dam et Al. (2011)

We have compared our paper to Dam et al.'s "Developments in the Market for Owner-Occupied Housing in Recent Years". Using the article as inspiration for estimating the housing market, we decided to include a variable to estimate COVID-19's effect on the housing market.

Our paper has slightly different periods and data, but our graphs and values are very similar. This is seen on the graphs for user cost and minimal first year payment. We chose to add a dummy for financial crisis and COVID-19. These dummies could create differences in the conclusion compared to Dam et al. They found co-integration between the variables, which is comparable to our results. This result is likely due to the same choice of approach, namely the ADL/ECM approach. Therefore, we conclude that our estimates are reliable.

5.4.2 Guglielminetti (2021)

The study by Guglielminetti finds that the COVID-19 pandemic has led to an increase in demand for houses and a shift towards dwellings with specific characteristics. These characteristics include more outdoor space and larger surface areas. Their article uses the instrumental variable approach, which exploits structural differences across provinces. Their measurements, such as containment, work from home, and epidemiological conditions, proves that several factors lead to an increase in demand for houses in Italy.

The COVID-19 pandemic has a more significant effect on condominiums than weekend cottages & single-family houses. This effect could indicate that more people move into the city rather than moving to the suburbs, which is in contrast to Guglielminetti's studies. Furthermore, Guglielminetti uses different variables to control COVID-19, which could be why our results differ. Even though their results are different, it is essential to consider that our study takes part in Denmark whilst theirs is in Italy.

5.4.3 Cepos (2020)

According to BRØNS-PETERSEN, Cepos has estimated COVID-19 as a negative supply shock to the economy. This shock is primarily due to the lockdown, which prevented industries from constructing at their usual rate. Furthermore, Cepos claims that the negative supply shock results in a corresponding decline in household income, which creates a negative demand shock. The fall in demand will be on the same scale as the supply shock.

Compared to our results, we observe that COVID-19 has an increased effect on housing prices by decreasing the supply, which is in line with our theory. The negative supply shock will lower the construction of housing while increasing prices. Furthermore, Cepos' claim of a negative demand shock due to COVID-19 will also lower construction and housing prices. Nevertheless, we disagree with Cepos' claim that the fall in demand will be the same as the supply shock.

5.5 The impact of the pandemic on housing demand: quantitative analysis

Many variables can affect COVID-19's impact on housing prices. There are no long run effects on single-family houses based on the COVID-19 variable. However, the rest of the variables are insignificant for the long-run solution. We did not expect this insignificance, as we assumed that some of the variables, such as disposable income, would have a long-run effect. Also, one could argue that it is reasonable that the COVID-19 variable is insignificant for the long-run solution because COVID-19 might not have a long term effect but only a short term effect. When estimating the prices for condominiums and weekend cottages, we observe a significant effect in the long run based on the COVID-19 variable. However, the percentage change in prices from a one percent change in COVID-19 is minimal and does not have a remarkable effect in the long run.

In the short term, we observe an increase in prices for all three housing types in the COVID-19 period. We believe that COVID-19 created an increase in demand for houses. This is based on the assumption that more people will work from home and become sick, creating a higher demand for workhouses with more home-office space and outdoor space to do activities. This does not mean that demand could not be affected by other variables such as low-interest rates and grace period. Furthermore, the households cannot use their disposable income on vacations and experiences due to restrictions, which should increase their disposable income.

6 Conclusion

This paper examines COVID-19's effects on the housing price of single-family houses, condominiums, and weekend cottages. We start with an ADL model based on Dam et al.'s model and determine the relevant macroeconomic variables for our model. Thus, we estimate the specific ADL model and calculate the corresponding ECM using this method. We derive the co-integration test for the relevant variables and calculate the long-run solution for our full- & sub period models using the ECM.

Through the ADL model, we exclude insignificant lags. We exclude the user cost and the minimal first year payment variables as they are insignificant for the sub period model of the single-family houses. We analyse that the user cost variable is insignificant for both the full-& sub period models when looking at condominiums. For weekend cottages in the full period, we find that the variable user cost, minimal first year payment, and disposable income are insignificant. For the sub period, we observe that all variables are significant.

Our co-integration tests show that our full period models do not cointegrate since the t-statistic is below our 5% critical value. However, our sub period models show co-integration between some of the variables. Using the long-run solution on both model periods, COVID-19 is insignificant for both model periods in single-family houses. In contrast, COVID-19 is only significant for condominiums' full period model. Lastly, COVID-19 is significant in the full & sub-period for weekend cottages on a 10% critical value. However, we can reject that COVID-19 has a long-run effect on weekend cottages in the full period for a 5% critical value. None of our ECM suffers from misspecifications which implies that the models are well specified.

When comparing to other studies, we find that our results are similar to Dam et al.s' but still differ in the choice of variables. Furthermore, we find that Guglielminettis' results are very different from ours. This might be due to the cultural differences between Denmark and Italy. Lastly, we disagree with Cepos' claim that COVID-19 will have a negative demand and supply shock on the housing price. Overall, we can conclude that all housing types are affected by COVID-19, though not necessarily in the long run.

References

Bergman, M. U., Sillemann B. T. and Sørensen P. B. (2014) " house Prices in Denmark and Sweden", Reform Capacity and Macroeconomiv Perfomance in the Nortdic Countries, Oxford University Press, 01/2015, page 163-186

Brøns-Petersen Otto, "How much will the shutdown of the Danish economy during the corona crisis cost?" Accessed 26-02-2022.
cepos.dk

Davidson, R., and J. G. MacKinnon (1993): "Estimation and Inference in Econometrics". Oxford University Press, Oxford.

Dam, N., Hvolbøl, T., Pedersen, E., Sørensen, P. and Thamsborg, S. (2011), "Boligboblen der bristede: Kan boligpriserne forklares? Og kan deres udsving dæmpes?", Kvartalsoversigt 01/2011, Danmarks Nationalbank, page 47-70.

Danmark Nationalbank Database "MONA", accessed 20/03/2022
Danmarks Statistik, accessed 20/03/2022
https://www.dst.dk/da/

Dickey, D. A. and Fuller, W. A. (1981), "Likelihood Ratio Statistics For Autoregressive Time Series With A Unit Root", Econometrics, 9/04/1981, page 1057-1072.

Ericsson, N. R., and J. G. MacKinnon (2002): "Distributions of Error Correction Tests for Cointegration," The Econometrics Journal, 5(2), 285-318.

Engle & Granger, Robert F. & C. W. J. (1987). "Co-integration and Error Correction: Representation, Estimation, and Testing". In: Econometrica, Vol. 55, No. 2, page 251-276.

Guglielminetti Elisa, Loberto Michele, Zevi Giordano and Zizza Roberta. "Living on my own: The impact of the COVID-19 pandemic on housing preferences", page 8-9, accessed 21-02-2022
https://www.bancaditalia.it/pubblicazioni/qef/2021-0627/QEF_627_21.pdf

Joshua Daniel Freimark (2020) "Tobin's Q theory of investment applied to the United States housing market", page 8-9, accessed 21-02-2022
http://ses.wsu.edu/wp-content/uploads/2020/07/FREIMARK-MS-Project.pdf

Nielsen Heino Bohn, ”A Course in Time Series Econometrics”. 03-09-2021

Peter Birch Sørensen and Hans Jørgen Whitta-Jacobsen (2010), ”Introducing Advanced Macroeconomics: Growth and Business Cycles”, *Investment and asset prices*, 14.4, page 406-411

Politi. ”Til dig, der er privatansat”, accessed 26-02-2022
https://politi.dk/coronavirus-i-danmark/til-dig-der-er-privat-ansat

Verbeek Mano (2017), ”A Guide to Modern Econometrics, fifth edition”.

Appendix A

Table 8: Critical values for unit root test

Distribution	1%	2.5%	5%	10%
N(0, 1)	-2.33	-1.96	-1.64	-1.28
DF	-2.56	-2.23	-1.94	-1.62
DF_e	-3.43	-3.12	-2.86	-2.57
DF_l	-3.96	-3.66	-3.41	-3.13

Note: The figur shows the asymptottic critical values for the Dicky-Fuller unit root test. It is the one-sided test for $\pi = 0$.
Source: Repreduced from Davidson and MacKinnon (1993).

Table 9: Critical values for PcGive test for no co-integration

PcGive test for no co-integration						
Number of est. par. to I(1) var. in long-run solution	Constant in (8.38)			Constant and trend in (8.38)		
	1%	5%	10%	1%	5%	10%
1	-3.79	-3.21	-2.91	-4.25	-3.69	-3.39
2	-4.09	-3.51	-3.19	-4.50	-3.93	-3.62
3	-4.36	-3.76	-3.44	-4.72	-4.14	-3.83
4	-4.59	-3.99	-3.66	-4.93	-4.34	-4.03
5	-4.80	-4.19	-3.87	-5.11	-4.52	-4.21
6	-4.99	-4.38	-4.06	-5.29	-4.70	-4.38
7	-5.17	-4.56	-4.23	-5.46	-4.86	-4.53
8	-5.34	-4.73	-4.40	-5.61	-5.01	-4.69

Note: The figure shows the asymptotic critical values for the tests for no co-integration. The distribution relies on the amount of estimated long-run parameters to I(1) variables.
Source: Reproduced from Ericsson and MacKinnon (2002).

Apendix B

Table 10: The table shows estimates of the ADL model for condominiums with various restrictions imposed. Standard errors in (·) and p-values in [·] for misspecification tests.

	(Condominiums ADL full period)	(Condominiums ADL sub period)
Constant	3.549 (1.060)	−3.510 (0.751)
COVID-19	−0.026 (0.030)	.
COVID-19_1	0.036 (0.034)	0.040 (0.009)
COVID-19_3	−0.051 (0.035)	0.017 (0.013)
COVID-19_4	−0.050 (0.040)	−0.052 (0.015)
Financialcrisis	0.074 (0.025)	.
Financialcrisis_1	−0.129 (0.023)	.
$\log(p^{EJ})_1$	0.888 (0.061)	.
$\log(p^{EJ})_4$	0.335 (0.079)	0.230 (0.071)
$\log(s^{EJ})$	−0.167 (0.026)	.
$\log(s^{EJ})_1$	.	−0.236 (0.032)
$\log(s^{EJ})_2$	.	0.139 (0.031)
$\log Y_3^d$	−0.431 (0.175)	0.407 (0.205)
$\log Y_4^d$	.	0.819 (0.209)
y	.	−16.714 (2.285)
y_3	−6.764 (2.479)	.
y_4	9.167 (2.491)	.
$\hat{\sigma}$	0.028	0.009
Log-lik.	150.316	90.719
AIC	-4.161	-6.209
HQ	-3.991	-6.070
SC/BIC	-3.730	-5.725
No autocorr. 1-5	[0.17]	[0.37]
No ARCH 1-2	[0.37]	[0.92]
No hetero.	[0.70]	[0.86]
Normality	[0.10]	[0.79]
T	66	26
Sample start	2005(1)	2015(1)
Sample end	2021(2)	2021(2)

Source: Denmarks Nationalbank MONA database, Statistics Denmark and own calculations.

Table 11: The table shows estimates of the ECM for condominiums with various restrictions imposed. Standard errors in (·) and p-values in [·] for misspecification tests.

	(Condominiums ECM full period)	(Condominiums ECM sub period)
Constant	3.549 (1.060)	−3.510 (0.751)
COVID-19	−0.026 (0.030)	.
COVID-19_1	0.036 (0.034)	0.040 (0.009)
COVID-19_3	−0.051 (0.035)	0.017 (0.013)
COVID-19_4	−0.050 (0.040)	−0.052 (0.015)
Financialcrisis	0.074 (0.025)	.
Financialcrisis_1	−0.129 (0.023)	.
$\log(p^{EJ})_1$	−0.112 (0.061)	.
$\log(p^{EJ})_4$	0.335 (0.079)	−0.770 (0.071)
$\log(s^{EJ})$	−0.167 (0.026)	.
$\log(s^{EJ})_1$	.	−0.236 (0.032)
$\log(s^{EJ})_2$	.	0.139 (0.031)
$\log Y_3^d$	−0.431 (0.175)	0.407 (0.205)
$\log Y_4^d$	.	0.819 (0.209)
y	.	−16.714 (2.285)
y_3	−6.764 (2.479)	.
y_4	9.167 (2.491)	.
$\hat{\sigma}$	0.028	0.009
Log-lik.	150.316	90.719
AIC	-4.161	-6.209
HQ	-3.991	-6.070
SC/BIC	-3.730	-5.725
No autocorr. 1-5	[0.17]	[0.37]
No ARCH 1-5	[0.57]	[0.93]
No hetero.	[0.70]	[0.86]
Normality	[0.09]	[0.79]
T	66	26
Sample start	2005(1)	2015(1)
Sample end	2021(2)	2021(2)

Source: Denmarks Nationalbank MONA database, Statistics Denmark and own calculations.

Table 12: The table shows estimates of the ADL model for weekend cottages with various restrictions imposed. Standard errors in (·) and p-values in [·] for misspecification tests.

	(Weekend cottages ADL full period)	(Weekend cottages ADL sub period)
Constant	1.583 (0.250)	4.043 (1.068)
COVID-19	-0.078 (0.021)	.
COVID-19_2	0.049 (0.030)	0.124 (0.018)
COVID-19_3	-0.061 (0.030)	.
COVID-19_4	.	-0.046 (0.019)
Financialcrisis_3	-0.033 (0.010)	.
$\log(\text{p}^{SOM})_1$	0.439 (0.111)	-0.525 (0.102)
$\log(\text{p}^{SOM})_2$	0.469 (0.106)	.
$\log Y^d$	.	0.643 (0.140)
$\log(\text{s}^{SOM})$	-0.145 (0.038)	0.154 (0.031)
$\log(\text{s}^{SOM})_1$	-0.085 (0.039)	-0.292 (0.035)
$\log(\text{s}^{SOM})_3$	0.118 (0.030)	.
u_1	.	-5.852 (1.047)
y_2	.	-9.631 (2.306)
$\hat{\sigma}$	0.024	0.011
Log-lik.	158.097	86.673
AIC	-4.488	-5.975
HQ	-4.357	-5.849
SC/BIC	-4.156	-5.539
No autocorr. 1-5	[0.76]	[0.39]
No ARCH 1-2	[0.87]	[0.47]
No hetero.	[0.60]	[0.23]
Normality	[0.68]	[0.71]
T	66	26
Sample start	2005(1)	2015(1)
Sample end	2021(2)	2021(2)

Source: Denmarks Nationalbank MONA database, Statistics Denmark and own calculations.

Table 13: The table shows estimates of the ECM for weekend cottages with various restrictions imposed. Standard errors in (·) and p-values in [·] for misspecification tests.

	(Weekend cottages ECM full period)	(Weekend cottages ECM sub period)
Constant	1.583 (0.250)	4.043 (1.068)
COVID-19	−0.078 (0.021)	.
COVID-19_2	0.049 (0.030)	0.124 (0.018)
COVID-19_3	−0.061 (0.030)	.
COVID-19_4	.	−0.046 (0.019)
$\text{Dlog}(p^{SOM})_1$	−0.561 (0.111)	.
$\text{Dlog}(s^{SOM})$	−0.145 (0.038)	0.154 (0.031)
Financialcrisis_3	−0.033 (0.010)	.
$\log(p^{SOM})_1$	.	−1.525 (0.102)
$\log(p^{SOM})_2$	−0.092 (0.036)	.
$\log Y^d$	.	0.643 (0.140)
$\log(s^{SOM})$_1	−0.230 (0.036)	−0.138 (0.025)
$\log(s^{SOM})$_3	0.118 (0.030)	.
u_1	.	−5.852 (1.047)
y_2	.	−9.631 (2.306)
$\hat{\sigma}$	0.024	0.011
Log-lik.	158.097	86.673
AIC	-4.488	-5.975
HQ	-4.357	-5.849
SC/BIC	-4.156	-5.539
No autocorr. 1-5	[0.76]	[0.39]
No ARCH 1-2	[0.87]	[0.47]
No hetero.	[0.65]	[0.26]
Normality	[0.68]	[0.71]
T	66	26
Sample start	2005(1)	2015(1)
Sample end	2021(2)	2021(2)

Source: Denmarks Nationalbank MONA database, Statistics Denmark and own calculations.