rijksuniversiteit groningen / faculteit ruimtelijke wetenschappen

EXAM Real Estate Research
2022 / 2023
April 6, 9h00 – 12h00

ANSWERFILE

Student ID: [important to fill in]

READ CAREFULLY BEFORE STARTING!

This exam has been designed by Arno van der Vlist and Mark van Duijn.
This exam has been cross-reviewed by Sarah Mawhorter, Arno van der Vlist and Mark van Duijn.

Total credits in the exam: **60 (60% of your final grade)**
Credits needed to pass: **33**

Step 1: PLEASE READ ALL QUESTIONS FIRST! TABLES FOR CRITICAL VALUES ARE IN THE APPENDIX. FOR EACH QUESTION EXPLAIN YOUR ANSWERS IN WRITING BRIEFLY (just a number or a yes-no answer is not sufficient).

This is a closed book exam. Use this word document to formulate your answers. Do not forget to fill in your student ID above in red. Also, use your student ID in the title of the word document. You are encouraged to use the programs **STATA**, **Word**, and **Excel** to come to your answers. Include STATA commands in the word file where necessary to prove what you have done. For calculation questions, you need to show the calculation steps, and you do not earn points if the calculation steps are missing even if the result is correct.

Step 2: When finished, upload the relevant word document including your student ID on Brightspace > Real Estate Research > Exam2023.

Students can lodge an appeal against the results of an examination with the Central Portal for the Legal Protection of Student Rights (CLRS) within 6 weeks of the date on which the result was announced.

MAKE SURE YOU UPLOAD YOUR DOCUMENTS ON BRIGHTSPACE > RER > EXAM2023!

The grades will be made available no later than 10 working days after the exam. Upon completion of the grading we will return to you with information on the procedure how to examine your exam and the answers.

QUESTION 1 – CLIMATE CHANGE-RELATED MODELING (12)

A recent article in the *Review of Financial Studies* entitled, "Does climate change affect real estate prices? Only if you believe in it," examines whether housing prices reflect belief differences about climate change.

a) The authors examine two groups of agents, believers (agent *1*) and deniers (agent *2*) of sea-level rise (*S*) in the future. The model's theoretical prediction is stated in Equation (1). Explain and interpret the theoretical prediction regarding housing price (*P*):

$$P_1(S) - P_1(S') \geq P_2(S) - P_2(S'), \quad \text{for } S' > S. \tag{1}$$

(5 points)

The papers examines whether housing prices reflect belief differences about climate change. The theoretical prediction is stated in Eq. (1).

Explanation must include:

1) for any sea level rise S' > S
2) the <u>price difference</u> for agent 1 believers is larger or equal than the <u>price difference</u> for agent 2 deniers

Interpretation: The prediction is that the price effect depends on <u>sea level rise</u> (S' > S) and on whether agents are believers or deniers. The price effect for believers is larger or equal than for deniers. Now suppose that for a given S we have identical properties and P1 = P2, then this equation can be interpreted as P1(S') <= P2(S'), or house price for believers are lower or equal then the house price for deniers. Thus in neighborhoods with a great share of believers house prices are lower, ceteris paribus.

Grading: correct explanation(3 points) correct interpretation (2 points). If price levels are discussed rather than price differences are missing (-3). If S' > S is not discussed or incorrect (-1). Note that S or S' do not refer to predicted or future values but to different situations or sea level rise. Totally incorrect or inconsistent (max 1 given apparent complexity of the question). no answer (0)

b) **Formulate a statistical model to measure the effect of local measures of climate change beliefs on the relationship between flooding projections of homes and housing prices, define the measurements, and list the statistical properties. (5 points)**

The empirical model relates sales price of property i at time t, to a set of observatbles and an error term, or:

$$\begin{aligned} \ln P_{it} &= \alpha_{zd} + \alpha_y + \zeta \, \text{UnderWater}_i \times \ln H_c + \gamma' X_i \\ &\quad + \lambda'(\text{UnderWater}_i \times X_i) + \omega'(\text{UnderWater}_i \times X_z) + \varepsilon_{it}, \end{aligned} \tag{14}$$

The measurements can be defined in various ways. P is price and is in log transformation because the price distribution is typically right-skewed. H is defined as the number of believers in a neighborhood and is in log transformation, UnderWater is indicator for property in flood-prone area (1=yes), X is a set of property controls (like number of rooms, size in sq feet, parcel in sq feet), and

Statistical properties is e ~ iid N(mu, sigma^2) so we assume (Chapter 6.2 CUP)

(1) $E(u_t) = 0$

(2) $\text{var}(u_t) = \sigma^2 < \infty$

(3) $\text{cov}(u_i, u_j) = 0$ for $i \neq j$

(4) $\text{cov}(u_t, x_t) = 0$

(5) $u_t \sim N(0, \sigma^2)$

What is essential is that the model includes an interaction term between climate change beliefs (here defined as lnH) and flooding projection (here defined as UnderWater).

Grading: model with interactions (4) statistical properties (1). No controls (-1) No properties (-1) No interaction (-2). No error term (-1) No coefficients (max 2) Inconsistent max (1) no answer (0)

c) A standard problem in modeling effect of natural hazards (like sea-level rise floodings) on housing prices is the need to distinguish between the effect of flooding and the effect of amenities. Explain why so, and how you deal with it in the empirical model. (2 points)

> The main challenge is that, all else being equal, coastal homes are also more valuable than other homes. Thus, controlling for distance from the coast is particularly important if, on the one hand, homes closer to the coast tend to have higher values, on the other hand, homes closer to the coast are also more likely to be flooded if sea levels rise. Additionally, if wealthy people live near the coast these homes may also have different characteristics. For example, these homes may be larger, or newer. This presents an omitted variable problem which is solved by controlling for distance from the coast.
>
> **Grading**: correct (2) good discussion but no solution (1.5) discussion no answer to specific problem (max 1) answer not specific to give points, or inconsistencies (max 0.5) no answer (0)

QUESTION 2 – INTERPRETATION AND CALCULATION (10)

Researchers are interested in the magnitude of economic depreciation in commercial real estate. They regress the (logarithmic transformation of) sales price on age and a set of controls. Results are reported in Table 1.

Table 1: Estimation results

	(1) Pooled	(2) Office	(3) Retail	(4) Industrial
Age	-0.0104***	-0.0173***	-0.00953**	-0.0212***
	(0.00233)	(0.00261)	(0.00455)	(0.00441)
Age^2	5.16e-05**	8.47e-05***	4.00e-05	1.05e-04
	(1.20e-05)	(1.34e-05)	(2.39e-05)	(6.16e-05)
Construction cohorts	yes	yes	yes	yes
Property characteristics	yes	yes	yes	yes
Seller type	yes	yes	yes	yes
Buyer type	yes	yes	yes	yes
Year fixed effects	yes	yes	yes	yes
Country fixed effects	yes	yes	yes	yes
Number of observations	9,596	5,592	2,601	1,403
Number of regressors	39	39	39	39
RSS	4713	2074	1159	357
Adjusted R-squared	0.58	0.69	0.61	0.62

Dependent variable is ln Price. Property characteristics include ln Size (sq.m), CBD(1=yes). Robust standard errors.

a) **Interpret the coefficients stated in column (1): indicate the marginal effect of Age on ln Price, and explain the outcome. (5 points)**

> Indicate: The marginal effect provides information on what the effect on the left hand side variable is when we change the level of a right hand side variable by one unit. Mathematically it is equivalent to the partial derivative. In this case the marginal effect is -0.0104 + 2* 5.16e-05 * Age
>
> Explanation: So marginal effect of age depends on age, or the depreciation rate depends on the age of the property, ceteris paribus
>
> Grading: If marginal effect includes only Age but otherwise correct (-1.5). Age^2 missing and interpretation incorrect (max 2.5) marginal effects mentioned for Age and Age^2 but not combined (max 4.5) no answer (0)

b) Use the Chow F test to calculate stability of coefficients across the various segments: offices, retail, and industrial. Critical F-values can be found in the Appendix. (5 points)

The Chow F test statistic is used to test stability of coefficient.

The Chow test is stated in Ch 6 CUP for the situation of 2 groups:

Box 6.7 Conducting a Chow test

(1) Split the data into two sub-periods. Estimate the regression over the whole period and then for the two sub-periods separately (three regressions). Obtain the RSS for each regression.

(2) The restricted regression is now the regression for the whole period, while the 'unrestricted regression' comes in two parts: one for each of the sub-samples. It is thus possible to form an F-test, which is based on the difference between the RSSs. The statistic is

$$\text{test statistic} = \frac{RSS - (RSS_1 + RSS_2)}{RSS_1 + RSS_2} \times \frac{T - 2k}{k} \qquad (6.56)$$

where RSS = residual sum of squares for the whole sample;
RSS_1 = residual sum of squares for sub-sample 1;
RSS_2 = residual sum of squares for sub-sample 2;
T = number of observations;
$2k$ = number of regressors in the 'unrestricted' regression (as it comes in two parts), each including a constant; and
k = number of regressors in (each) 'unrestricted' regression, including a constant.

The unrestricted regression is the one in which the restriction has not been imposed on the model. Since the restriction is that the coefficients are equal across the sub-samples, the restricted regression will be the single regression for the whole sample. Thus the test is one of how much the residual sum of squares for the whole sample (RSS) is bigger than the sum of the residual sums of squares for the two sub-samples ($RSS_1 + RSS_2$). If the coefficients do not change much between the samples, the residual sum of squares will not rise much upon imposing the restriction.

The test statistic in (6.56) can therefore be considered a straightforward application of the standard F-test formula discussed in chapter 5. The restricted residual sum of squares in (6.56) is RSS, while the unrestricted residual sum of squares is ($RSS_1 + RSS_2$). The number of restrictions is equal to the number of coefficients that are estimated for each of the regressions – i.e. k. The number of regressors in the unrestricted regression (including the constants) is $2k$, since the unrestricted regression comes in two parts, each with k regressors.

(3) Perform the test. If the value of the test statistic is greater than the critical value from the F-distribution, which is an $F(k, T - 2k)$, then reject the null hypothesis that the parameters are stable over time.

Here we consider a situation with <u>3 groups.</u>

Test outcome is F test statistic=

$$\frac{4713 - (2074 + 1159 + 357)}{(2074+1159+357)} \times \frac{(9596 - 3*39)}{(2 * 39)} = 37.67$$

The F critical value for a situation with (2*39) restrictions (df1) and (9596 – 3*39) degrees of freedom (df2) can be obtained from the Table and between 1.3180 and 1.2214. So let's say 1.3

F = 37.67 > F critical value (0.05) = 1.3

Conclusion: H0 is rejected

Grading: F-test outcome correct (3) critical value correct (1) correct conclusion (1). #df incorrect (-1) #restrictions incorrect (-1), critical value incorrect (-1) no Chow test or chow test incorrect (max 1). no answer (0)

Question 3 – FORECASTS (8)

Brooks & Tsolacos (2010) discuss model outcomes as shown below.

Data

	RRg	ΔVAC	$OFSg$
2002	−12.37	6.3	0.225
2003	−18.01	5.7	0.056
2004	−13.30	3.4	0.618
2005	−3.64	0.1	0.893
2006	−4.24	−0.2	2.378
2007	3.48	−2.3	2.593

Regression results

	1982–2002	
Independent variables	Coefficient	t-ratio
C	−6.81	−1.8
ΔVAC_{t-1}	−3.13	−2.5
$OFSg_t$	4.72	3.2
Adjusted R^2	0.53	
Durbin–Watson statistic	1.94	

Notes: The dependent variable is RRg, which is real rent growth; ΔVAC is the change in vacancy; $OFSg$ is services output growth in Frankfurt.

a) **Formulate the statistical model based on the regression results reported in the right panel. (2 points)**

RRg_t = C + b_1 ΔVAC_{t-1} + b_2 $OFSg_t$ + e_t

Grading:

-note that in a statistical model we do not report estimates. (When estimates are included not error term is included)

-and also correct: RRg_t = f(ΔVAC_{t-1} , $OFSg_t$) + e_t

Correct but no error term (max 1.5) specification incorrect, or coefficients included and error term (max 1) LHS incorrect or otherwise inconsistent (max 0.5) no answer (0)

c) calculate the real rental growth (RRg) out-of-sample forecast for the years 2006 and y… 2007. (4 points)

Calculation of forecast
2006 = -6.81 – 3.13 * 0.1 + 4.72 * 2.378 = 4.10
2007 = -6.81 – 3.13 * -0.2 + 4.72 * 2.593 = 6.05 (A = 3.48) so (A-F) = -2.57

Grading:
When the method is correct but for dVAC t is used rather than t-1, also when done incorrect twice (3 points)
when the method is correct but data misunderstood and used incorrect for dVAC and for OFS, or C missing (max 2 points) when the method is incorrect and data misunderstood (max 1 point).
inconsistent (max 1) no answer (0)

d) calculate the mean squared error (MSE) over these two years by comparing Actual and Forecasted values. (2 points)

for 2006: (A = -4.24) so (A-F) = -8.34
for 2007: (A = 3.48) so (A-F) = -2.57
thus
MSE = 1/n sum (A-F)^2 = ½ ((-8.34^2) + (-2.57^2)) = (69.55 + 6.60) /2 = 38.08%

Grading:
correct, or when 3b is incorrect but used correctly (2 points)
MSE calculation correct, per year but not over years (max 1.5 points)
MSE incorrect but A and F given (1 point)
only A mentioned (0.5)
inconsistent (max 0.5)
no answer (0)

Question 4
Dow…

Question 4 – Analysis in STATA (30 credits)

Download rer-exam-data.dta from Brightspace > Real Estate Research > Exam2023 > Data, save it in a separate map, and open the data in STATA. For the following questions, use rer-exam-data.dta as a starting point. **Make sure you include the appropriate syntax code in your answers to show exactly what you have done.** If you want to show each and every step you made in Stata, you may want to create a 'log' file in Stata. Include: "log using rer-exam.log, replace" in your *.do file. You can use the "log close" command to close the log file.

The data contain information about condominium and executive condominium home sales in a particular large district in Singapore. You will be taking sales prices "T_Price" (in Singaporean dollars) as the dependent variable, distance to the central business district "CBD" (in km) as your key independent variable, and you will need to include all relevant control variables. Only keep data from 2010 onwards.

A. (5 points). Adjust and transform data if appropriate, and include a table with summary statistics of the net sample (incl. information on all relevant variables you will use for analysis).

3 point cleaning the data: 1 for only keeping data from 2010 onwards. 1 for discussing outliers. 1 for dropping outliers in terms of sales prices and/or floor area.

1 point for transforming relevant variables (e.g. ln(price) instead of using price)

1 point for showing summary statistics (incl. all variables) with each variable having the same number of observations. No points if the number of observations are not similar for each variable.

B. (15 points). Run a hedonic regression model where you include your key independent variable and all other relevant control variables in the dataset.

i. Show the regression results.

ii. Interpret the coefficient of your key independent variable. Be explicit on significance, sign, and size (interpret the economic effect size even if the coefficient is not significant).

iii. A critical reader notes that you assume a specific functional form between your dependent and key independent variable. First, describe the functional form you assumed when looking at the regression results in Question B.i. Second, adjust your model specification and show regression results where you allow for a more flexible functional form between your dependent and key independent variable. Finally, provide an argument for using a more flexible function form.

2 points for showing the regression results (-1 for including year, month and/or neighborhood variables as continuous variables)

4 points for correctly interpreting the coefficient: 1 for sign, 1 for significance, 2 for exact interpretation of the coefficient.

3 points for describing the functional form in Question B.i: One assumes a linear functional form between the dependent (lnP) and independent variable (CBD).

3 points for adjusting the model specification by including polynomials or splines for the key independent variable and showing the regression results

3 points for providing arguments for using a more flexible functional form. One argument is that the data point to a non-linear functional form between distance to the CBD and the natural logarithm of sales price of condominiums and executive condominiums (-points for inconsistencies or invalid arguments)

C. (10 points) Theory indicates that the effect of floor area "FL_AREA" (in m2) on the sales price depends on the type of property "PROPERTY_TYPE2" (condominium and executive condominium). Run a hedonic regression model to test the above mentioned theory.

i. Show the regression results.

ii. Conclude, based on your regression results, whether there is a significant difference in the effect of floor area on sales prices between the different types of properties.

iii. Describe, based on your regression results, how much the effect of floor area on sales prices differs between the two property types (interpret the effect even if the coefficient is not significant).

2 points for showing the regression results including interaction variables (and the independent variables on their own) in the regression model. Note that you need to interact a continuous variable (FL_AREA) with a dummy variable (PROPERTY_TYPE2). (-1 for not including control variables in the estimated specification).

2 points for concluding whether or not you find a significant coefficient for the interaction variable (or significant difference between 2 coefficients when splitting the data)

6 points for showing how much the effect of floor area on sales prices differ between condominiums and executive condominiums. Sign and effect size are important here. Basically you only need to interpret the coefficient of the interaction variable. Possible conclusion: The effect of floor area on sales prices differs 0.032% between executive condominiums and condominiums. The effect is smaller for executive condominiums than for condominiums.

THE END!