VRIJE UNIVERSITEIT AMSTERDAM

Identification and Estimation in Spatial Economics[*]

Hans R.A. Koster[†] and Jos van Ommeren[‡]

November 3, 2022

1 INTRODUCTION

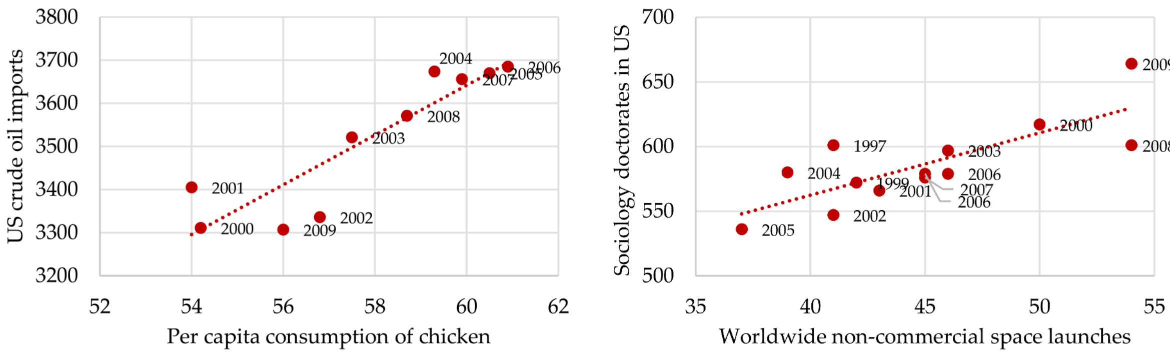
Most of you will undertake empirical research, for example when writing your MSc Thesis or, later, when working. This syllabus is concerned with setting up a research project and measuring (causal) effects of interest.

In Section 2 we introduce a step-by-step guide to design and execute a research project. You first formulate your hypotheses – which are your expectations on how reality should be understood, based on *e.g.* economic theory. Then you bring the hypotheses to the data by determining what variables *cause* each other. The effect of a treatment variable on outcome variable is then referred to as the *treatment* effect. Measuring this effect is often not straightforward for several reasons. It might be that the treatment variable is correlated with

[*]We are indebted to Angrist and Pischke (2008), who provide a very useful introduction into applied econometrics. The interested reader is strongly advised to read this book. We thank Thomas de Graaff for useful comments on previous versions of this syllabus.

[†]Department of Spatial Economics, Vrije Universiteit Amsterdam, De Boelelaan 1105 1081HV Amsterdam, email: h.koster@vu.nl. Hans is also research fellow at the Higher School of Economics in St. Petersburg, Tinbergen Institute (Gustav Mahlerplein 117 1082MS Amsterdam) and affiliated with the Centre for Economic Policy Research.

[‡]Department of Spatial Economics, Vrije Universiteit Amsterdam, De Boelelaan 1105 1081 HV Amsterdam, email: jos.van.ommeren@vu.nl. Jos is also research fellow at the Tinbergen Institute.



(a) CHICKEN AND CRUDE OIL (b) SPACE LAUNCHES AND SOCIOLOGY

Figure 1.1 – EXAMPLES OF SPURIOUS CORRELATIONS

Notes: Correlations are obtained from `spuriouscorrelations.com`. Data on total U.S. crude oil imports are from the *U.S. Department of Energy*, while data on per capita consumption of chicken are from the *U.S. Department of Agriculture*. Data on worldwide space launches are from the *Federal Aviation Administration* and awarded sociology doctorates in the U.S. are from the *National Science Foundation*.

other (unobserved) variables, or that individuals may self-select into treatment. Another possibility is that there is reverse causation. This is why you have to think of a strategy to isolate the treatment effect.

Because most of empirical economic research is concerned with *identification of causal effects*, we focus attention on a central feature of most applied research projects, which is the *identification strategy*. A causal effect implies that a certain treatment or explanatory variable *causes* a change in an outcome variable. This is different from a *correlation*, where there is a statistical association between two variables. However, even although correlations may be high, they may be entirely *spurious*, which implies that a statistical relationship is explained by another factor or factors, so-called *confounders*.

Why is then economics so much obsessed with causal effects? First of all, when looking at real-life data one may observe many statistical correlations. Let's give two examples. Using data between 2000 and 2009, there is a correlation of 0.90 between the per capita consumption of chicken and U.S. crude oil imports (see Figure 1.1a). While the correlation is very high, it seems hard to argue for a causal effect of eating more chicken on crude oil imports (or vice versa). Similarly, in Figure 1.1b we find a correlation of 0.79 between worldwide non-commercial space launches and awarded sociology doctorates in the U.S. Here again, it is very hard to argue that there is a causal effect of space launches on doctorate awards in *sociology*. So, although plain correlations are sometimes good starting points for your research, they are of limited interest when understanding how the world really works. Second, research questions in (spatial) economics are often specifically targeted at evaluating policies. A correlation between a certain policy variable and an outcome variable is then of limited interest; however a causal effect may inform us on the effectiveness of the policy.

The focus on identification of causal effects is not specific to the field of (spatial) economics. In many fields where statistical data is used to test hypotheses, one is predominantly interested in causal effects. For example, in medicine research it is very common to randomly assign participants to a treatment group and a control group, where the former receives the actual treatment and the latter a placebo. By comparing the average measure of health between the two groups afterwards, we identify a average causal effect of the medicine. But why is this the case? In Section 3 we will explain why randomized experiments do identify a causal effect of a treatment.

In spatial economics, we often cannot set up randomised experiments. For example, let's say you are interested in the effect of local air pollution on life expectancy. It would be hard to think of a way to *randomly* select people into a treatment group and/or control group. We therefore have to come up with imperfect alternatives using statistical/econometric methods. In Section 4 we consider various alternative strategies. The usefulness of these strategies depend on the specific context of the research question at hand.

When using econometric methods and data, which is almost always a *sample* of the full population, there is uncertainty in the estimated effect. The *standard error* of an estimate indicates how precise an estimate is. If the standard error is small relative to the effect, we say that the effect is *statistically significant*. It is therefore important to also correctly estimate standard errors. In Section 5 we discuss a couple of issues to bear in mind when calculating your standard errors.

In Section 6 we summarise what we think are the most important insights from this syllabus.

2 RESEARCH DESIGN: A STEP-BY-STEP GUIDE

When aiming to undertake empirical research, we consider eight (main) steps:

1. Formulate your hypotheses;
2. Determine the 'treatment' variable(s) and the 'outcome' variable(s);
3. Think of an identification strategy to identify causal effects;
4. Select samples, discuss measurement error and provide descriptives;
5. Determine functional form of variables of interest;
6. Think of different issues in estimating standard errors;
7. Estimate the model and interpret the results;
8. Provide robustness checks of the results.

These steps may seem to describe a linear process; however, when undertaking research it may be that there is feedback between different steps in the process. For example, when you obtain results in Step (7), it appears that the results are imprecise, which may imply that you need more data or need to make other selections in Step (4). Or: when undertaking a robustness analysis in Step (8) and your results appear not to be very robust, you may have to go back and improve on the identification strategy (Step (3)).

Below we will briefly discuss each step, but this syllabus will put particular emphasis on Steps (3) and (6) – defining your identification strategy and obtaining the standard errors.

STEP 1: FORMULATE YOUR HYPOTHESES

Formulating your hypotheses may seem trivial but is probably the most important step of your research. These hypotheses are not necessarily strict statistical hypothesis (*i.e.* a null hypothesis that a coefficient does not have an effect and the alternative hypothesis that a coefficient has an effect), but instead are more general descriptions of reality on what your expectations are and how reality should be understood. You should clearly state the problem that you are trying to solve or research question that you are trying to answer. Make sure that the hypotheses clearly define the topic and are not too broad.

How to form reasonable expectations on how reality may be understood? At least one useful source of information is economic theory. Many papers first develop a theory that leads to testable empirical implications. Or even more interesting, it is usually the case that economic theory indicates that a certain hypothetical policy is welfare improving (*e.g.* a tax on pollution), but is silent on the extent of the welfare gain. In such cases, one may estimate demand and supply functions, in order to determine the magnitude of welfare effects of such a policy. Another important piece of information comes from existing policies. For example, in many countries one has to pay a transaction tax when buying a new house. According to theory, residential mobility will fall because of the transaction tax. By examining the effect of current transaction taxes on residential mobility, you do not only test the hypothesis which is interesting according to economic theory, you also estimate the effect of the policy (*i.e.* the level of the transaction tax on mobility).

Another way to form expectations, although slightly more risky, is to apply so-called *reverse causal inference*. This implies that you look at some interesting outcome variable and you ask yourself what may explain this outcome. For example, you may observe a tremendous house price increase between 2013 and 2020 and you ask the question what could be determinants of this house price growth. Alternatively, you might observe a strong increase in telecommuting since March 2020 and you wonder what possibly could be a factor that explains this increase.[1] Why is reverse causal inference somewhat risky? Reverse causal inference is not necessarily firmly grounded in economic theory and related to existing policy. Hence, potential findings may be of limited interest to academics or policy makers.

[1] Reverse causal inference is to be contrasted to so-called *forward causal inference*, which supplies answers to research questions. For example, forward causal inference will inform you on how large an effect is of certain variable x on an outcome variable y.

STEP 2: DETERMINE TREATMENT AND CONTROL VARIABLES

Please look around for data and define what variables are available. Think about what your explanatory variables (*i.e.* your x variables) and your outcome or dependent variables (*i.e.* your y variable(s)) are. Make sure that y and x are actually good proxies for the things you want to measure, so that measurement error is low (we will talk more about measurement error later. Finally, please check whether y and x are continuous or discrete variables. Usually, it also makes sense to think about the order of magnitude of the effect you want to examine (maybe based on previous research) also in relation to the size of your dataset. For example, if you wish to know the effect of a new tax on fuel on car ownership, it makes sense to check how much car owners spent annually on fuel. Let's suppose this is 10% of the annual car ownership costs. Let's suppose the new tax doubles the fuel price faced by consumers. In this case, even with relatively few observations, you may be able to detect the price of the tax. But if, alternatively, the tax increases the price only marginally, you may be able to detect the effect of the price only given millions of observations.

We emphasise that although we will talk a lot about different type of (regression) models, in applied econometrics one is often interested in a specific effect of a certain variable. Regression models frequently contain many variables. Nevertheless, one is strongly advised to focus the attention on a single (or a few) x variables (while possible controlling for a host of other variable) and a single (or a few) y variables. It is much simpler to properly measure the impact of one x variable, rather than a large set. Moreover, as we will see later, an identification strategy is often valid in the context of only one or maybe a few variables. In what follows we therefore assume that there is one outcome variable y_i that is affected by one treatment variable x_i (and possibly a bunch of control variables, c_i).

STEP 3: DEVELOP AN IDENTIFICATION STRATEGY TO IDENTIFY CAUSAL EFFECTS

Once having defined your x and y, you have to think on how to measure a causal effect of x on y, rather than just a correlation. This essentially means that you have to define the 'treatment' group and a 'control' group. The latter should be otherwise identical to the people that receive treatment, except that the people in the control group do not receive treatment.

Hence, you have to carefully discuss whether there might be endogeneity issues that thwart a causal interpretation of x_i on y_i. We consider four main endogeneity issues:

1. There could be a selection bias. For example, let's assume you are interested in the effects of providing social housing on the well-being of inhabitants. Entry into social housing is, however, not exactly random (as there often exists a maximum income to be eligible for social housing and there are waiting-list.) Hence, comparing well-being of people in social housing to other people will not lead to a meaningful causal effect of providing social housing on well-being. The selection bias implies that you are comparing apples with oranges, rather than apples with apples.
2. One could worry about an omitted variable bias. Omitted variable bias arises when

a potential independent variable – which is related to both the dependent variable and an included independent variable – is omitted from the model. The result is a biased estimate of the coefficient of the included variable. An omitted variable bias is essentially the same as a selection bias (Angrist and Pischke, 2008) as omitted variables also imply that one compares apples to oranges. We will show in Section 3.2 that addressing omitted variable bias is the same as addressing selection bias.

3. Reverse causality is an issue once y affects x, rather than x influencing y.[2]

4. Fourth, once x is measured with error, the estimated effect will be incorrect. We discuss this in the next step in more detail.

In Sections 3 and 4 we will provide different strategies on what valid identification strategies could be to address endogeneity concerns. Here we want to emphasise that it is important to discuss endogeneity issues in length. More specifically:

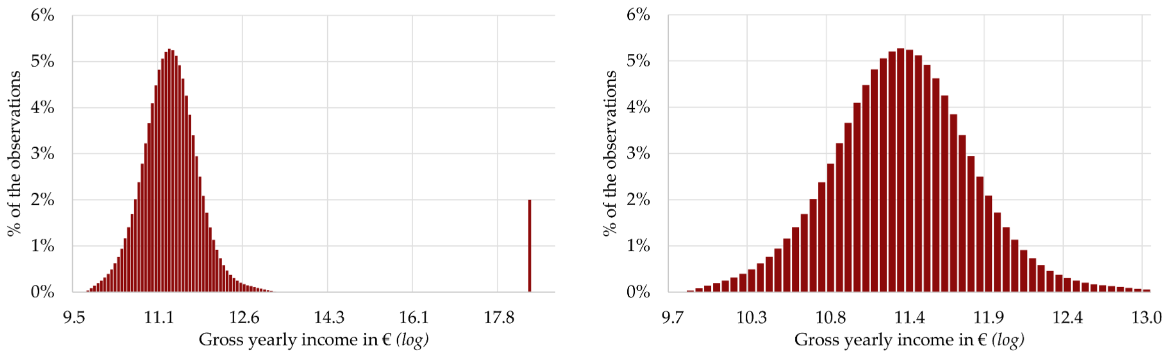
- Wonder whether there could be a selection effect;
- Carefully think about what could be potential unobserved factors that are correlated to the treatment variable and give examples;
- Make sure that reverse causality is not an issue in your context.

STEP 4: SELECT SAMPLES, DISCUSS MEASUREMENT ERROR AND DESCRIPTIVES

Much of applied economic research relies on readily available datasets. It is important to think carefully whether it is appropriate to use the full dataset. Say you are interested in the effects of school meals on health outcomes and you have information on the full population. It is then not so wise to focus on health outcomes of the whole population. Instead, it makes more sense to only include children at primary schools. However, when making selections you have to think carefully. Say you are interested in the effectiveness of a policy combating traffic congestion on highways, then it makes sense to include other roads (because motorists frequently have the option to choose between highways or other roads). To make proper selections is important as to mitigate the selection bias, which arises when treatment and control groups are not the same.

Data cleaning is the process of detecting and correcting/removing inaccurate observations from a dataset. It is then key to properly identify incomplete or incorrect observations in the data. To get a first insight into your data, please make histograms and scatterplots for the variables of interest. Let's consider the histograms of the log of income in Figure 2.1. In the left panel, you observe a weird spike at a log income of 18.42. If you take the exponent of this number you arrive at an income of €99,999,999. This seems unlikely and probably refers to missing values that are coded this way. Removing these observations from the data then

[2] Sometimes, there is a distinction made between *reverse causality* and *simultaneity*. In the former x is only influenced by y, while with simultaneity x influences y, but y also influences x. In this syllabus we will just refer to reverse causality as the phenomenon where x may be influenced by y, irrespective of whether x influences y.



(a) INCOME, RAW DATA

(b) INCOME, CLEANED

Figure 2.1 – EXAMPLES OF INCOME HISTOGRAMS

Notes: The figures are based on adapted data from *Statistics Netherlands*.

seems the right way to do in order to obtain Figure 2.1b.[3] Also check for other unrealistic values, such as house prices or incomes being negative, variables expressed in percentage terms that are negative or exceed 100, etc.

More generally, when you are worried about measurement error in your dataset, it may be relevant to know more about the consequence of measurement error for your identification. As a general rule, random measurement error in a dependent continuous variable will *not* affect the causal effect of interest (although it will lead to higher standard errors). We show this in Figure 2.2a where random measurement error (between the open dots) lead to exactly the same relationship.[4] However, random measurement error in the explanatory variable of interest causes your estimates to be biased towards zero.[5] We illustrate this in Figure 2.2b, where the fitted regression line is flatter once there is measurement error in x_i. We provide a formal explanation in Appendix A.1.

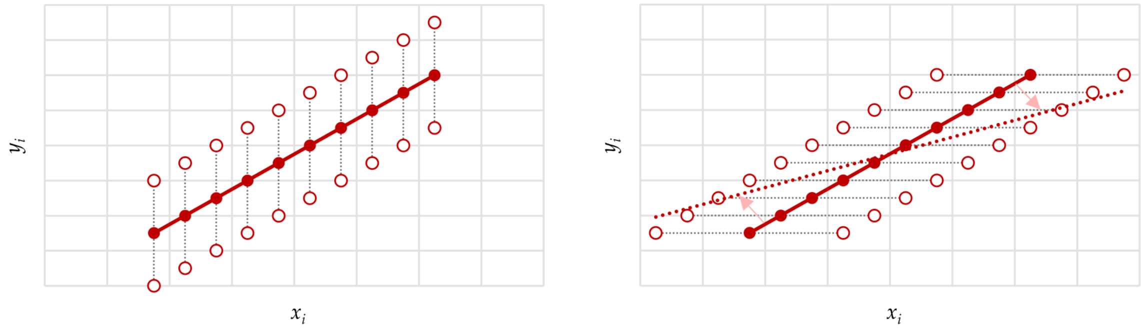
How to deal with errors in your data is a bit subjective, but here are some important rules to keep in mind:

1. Always describe the exact way you have cleaned the data. Good science implies that you, or anyone else, can replicate your work;
2. Discuss and possibly remove unrealistic values. Frequently it is subjective which values are unrealistic, but you have to make a decision anyway. With a sensitivity analysis, you can also always check whether these decisions impact your results.

[3]Removing outliers is tricky. Whether this is correct depends on the context because outliers may be based on some underlying process. Therefore, carefully think whether observations are really incorrect.

[4]For example, you wish to estimate the effect of air pollution on house prices – which is an important method to estimate the household marginal willingness of households to avoid pollution. In this case, you do not mind so much to receive the dataset with provides information about the asking house price (although you would have preferred the transaction price), at least, given the reasonable assumption that the difference between the asking and the transaction price is random (or more appropriately, is not related to air pollution).

[5]Hence, for the example given, if you would have substantial measurement error in the observations of pollution, you have to worry about downward estimates.



(a) MEASUREMENT ERROR IN y_i (b) MEASUREMENT ERROR IN x_i

Figure 2.2 – MEASUREMENT ERROR IN y_i AND x_i

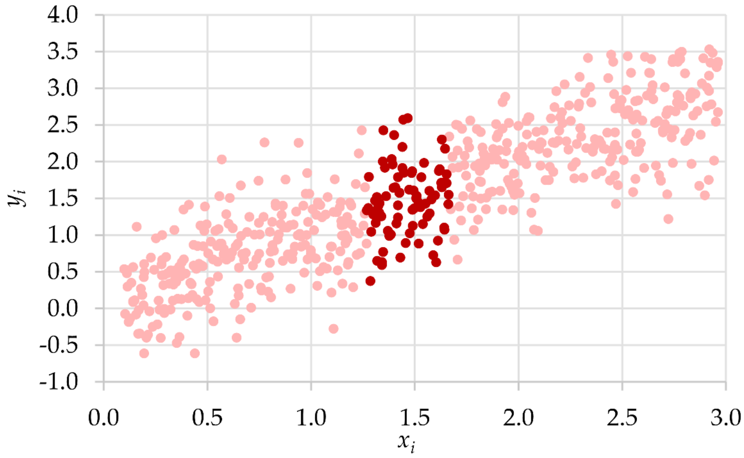


Figure 2.3 – VARIANCE IN x_i IS NECESSARY

3. Distinguish between dependent and independent variables. Sometimes it makes sense to constrain the independent variable x to a certain range.[6] However, note that variation in x_i is necessary to identify any effect. We illustrate this in Figure 2.3. If we would only focus on $x \in \{1.3, 1.7\}$ it will be hard to detect any effect of x_i on y_i, while the positive relationship between x_i and y_i is apparent when considering the full range of x_i. As a general rule of thumb, it is better not to constrain the range of the dependent variable, y_i.

4. Measurement error is context specific and therefore should be discussed in detail.[7]

[6]For example, you wish to know the effect of number of children on commuting behaviour. Now suppose there is one person in your dataset with 16 children, whereas almost all your observations have less than 4 children. It is then perfectly fine, to examine the effect of children, for a subsample of households with less than 4 children, as long as you understand that the estimated effect only holds for the selected subsample.

[7]For example, in datasets on commuting behaviour, there are always a few people with commuting distances of 200km. It is extremely unlikely that workers commute more than 150km per day, at least, if they go regularly to their work. If you have a dataset of workers who state that they go 5 times a week, to their work, 200 km must be unrealistic; but if they go only once a week, it is very realistic.

When you have made the appropriate selections, cleaned the data and discussed any remaining measurement error, please provide univariate descriptive statistics. At least provide a table with the mean, standard deviation, minimum and maximum values for all relevant variables. Think carefully about the unit of measurement of your variables, and whether or not it makes sense. Sometimes this is easy, sometimes it is not. For example, if you have received a dataset, with the age of the person, it is clear that if you are average in the dataset is 400, it is likely that age is measured in months, and not in the usual years. But for most people, if they have received a measurement of $PM_{2.5}$ pollution, they have not an immediate idea of whether an average of 30 makes sense or not.

In addition, please make histograms of continuous variables. You may also compare the means between the treatment and control groups. It is also useful to think about bivariate statistics, such as scatterplots. What is important is that you and the reader of your report receive a clear understanding on the data that are used.

STEP 5: DETERMINE FUNCTIONAL FORM OF REGRESSION EQUATION

The regressing equation to be estimated looks something like this:

$$y_i = f(x_i, c_i, \epsilon_i), \tag{2.1}$$

where y_i is the dependent variable measure for an individual i; $f(\cdot)$ is some unspecified function of x_i, control variables c_i, and the *unobserved* error term ϵ_i. The specific relationship between the outcome variable and the explanatory variables is referred to as the *functional form*.

It is very complicated to estimate (2.1), so we often assume a specific relationship relationship between c_i, x_i and y_i. Often, it makes sense to start with a linear-in-parameters relationship :

$$y_i = \beta x_i + \gamma c_i + \epsilon_i, \tag{2.2}$$

where β and γ are parameters to be estimated by a regression, usually referred to as Ordinary Least Squares (OLS). The estimates our obtained by minimising the sum of squared residuals. In case that the explanatory variables x_i and c_i are not correlated to the unobserved error term, then the estimated values of β and γ, denoted $\hat{\beta}$ and $\hat{\gamma}$ respectively, can be interpreted as causal effects and are equal to the β and γ.[8]

Many students assume (wrongly!) that the residual must be normally distributed for proper estimation of causal effects – that is, an unbiased estimator of the parameter of interest. This is generally *not* the case. When the number of observations is large (as is mostly the case in spatial economics), OLS works fine also when the residuals are not normally distributed. The main assumption here is that the residual ϵ_i is not correlated to the independent variables in the model.

[8] The residual is the difference between the observed y_i and the predicted value of y_i. In contrast to the error term, the residual is observed. By construction, the residual is not correlated to the explanatory variables.

Furthermore, the assumption of a *linear-in-variables* relationship is usually the best starting point, but may not always be appropriate.[9] Hence, in your research you should discuss why linearising the regression equation is not much of a problem. Note that you may include *non-linear effects of variables* and still use OLS with a linear-in-parameter specification for example by including quadratic terms of x:

$$y_i = \beta_1 x_i + \beta_2 x_i^2 + \gamma c_i + \epsilon_i, \tag{2.3}$$

If $\{\beta_1, \beta_2\} > 0$. This would imply that the effect of x_i gets stronger for higher values of x_i. However, if $\beta_1 > 0, \beta_2 < 0$, the effect may be positive for low values, but at a decreasing rate for higher values of x_i. Adding higher-order terms (x_i^2 – quadratic, x_i^3 – cubic, x_i^4 – quartic) may be feasible if you allow for non-linearity in one variable, but you can imagine that the number of variables to include in the regression becomes infeasibly large once you allow for flexible interactions between x_i and the control variables c_i. Moreover, higher-order terms are highly correlated with each other. The effects may therefore be hard to estimate and predictions may be very imprecise outside of the domain of x_i.[10]

We therefore also may estimate *semi-* or *non-parametric* regressions that allow for very flexible functional forms. We refer to Yatchew (2003) for a proper introduction into semi- and non-parametric regression techniques.[11]

Economists are often interested in elasticities, which indicates the percentage change of x on y. More formally, it holds that:

$$\frac{\partial y_i}{\partial x_i}\frac{x_i}{y_i} = \frac{\partial \log y_i}{\partial \log x_i} \tag{2.4}$$

We show in Appendix A.2 that the above holds. Let's consider the following *log-linear* regression equation:

$$\log y_i = \beta \log x_i + \gamma c_i + \epsilon_i, \tag{2.5}$$

where log refers to the natural log. If we take the derivative of $\log y_i$ with respect to $\log x_i$ (so $\frac{\partial \log y_i}{\partial \log x_i}$) we obtain, β. Hence, the regression coefficient in a log-linear equation captures an *elasticity*.

Apart from regression coefficients having a very convenient interpretation, there are also other reasons to use logs. Here are a couple of cases when the use of logs is preferred over a

[9] The main reason is that the linear-in-variables specification give the best linear approximation even if the correct specification is non-linear (see for example Angrist and Pischke, 2014, p. 85).

[10] It is useful to note that the relationship between the dependent and independent variable may not only be non-linear, but may also be non-monotonic. Non-monotonicity implies means that the sign of the effect is not the same for the range of the independent variable examined. In this situation, it is important to include higher order terms that allow for non-monotonicity. For example, the effect of age on commuting distance is usually found to be increasing up to 30 years, and decreasing afterwards.

[11] We also will briefly discuss semi-parametric regressions in the course *Urban Economic Challenges and Policies*. Functional form issues will explicitly come back in assignments in the course *Empirical Transport Economics*.

standard linear regression:

- Taking logs may be useful to linearise a non-linear relationship, which would be otherwise more difficult to estimate.
- When economic theory indicates that a log-linear relationship is preferable. Economists often work with Cobb-Douglas utility/production functions, which naturally lead to log-linear equations.
- The residuals have a skewed distribution. A log-transformation may then be useful to obtain residuals that are approximately symmetrically distributed. The starting point is that you examine the shape of the dependent variable y_i, because the residual usually has the shame shape as the dependent variable, particularly one explanation power of the model is low (please note that the residual is equal to the dependent variable when you do not include any independent variable).
- The spread of the residuals may vary systematically with the values of the dependent variable (which we refer to as *heteroscedasticity*). The purpose of the transformation is to remove that systematic change in spread, implying *homoscedasticity*. We come back to this issue in Section 5.1.
- One frequently estimates models using data on regions (areas). These regions differ in size (*e.g.* Amsterdam is larger than Utrecht) (Briant et al., 2010). When you estimate the effect of a variable on the change in an aggregate variable measured for this region (*e.g.* the effect of a new tax policy on the change in the number of firms in a region), it is almost always good practice to focus on the (change in the) logarithm of the aggregate variable.

STEP 6: THINK OF DIFFERENT ISSUES IN ESTIMATING STANDARD ERRORS

Applied researchers are often interested in whether the estimated coefficient, $\hat{\beta}$, is *statistically significant*. Statistical significance is the likelihood that your estimated effect is not caused by chance. Your estimate is statistically significant, given the significance level, which reflects your risk tolerance and confidence level. One often use confidence levels of 10%, 5% and 1%. If your estimate is significant at the 5% level, there is a 5% chance that there may be actually no effect.

Whether or not an estimate is statistically significant not only depends on the size of β, but also on its standard error. If the standard error is large relative to the coefficient the estimate will be less statistically significant.[12]

It is therefore important to obtain the correct standard errors; otherwise, one may be wrongfully conclude that there is an effect of the treatment variable on the outcome variable.

[12] When you have estimated $\hat{\beta}$ and the standard error $\hat{\sigma}_\beta$, you can calculate the corresponding T-statistic as $T_\beta = \hat{\beta}/\hat{\sigma}_\beta$. Given a large dataset, when $T > 2.576$, the estimate is said to be statistically significant at the 1% level. Similarly, when $T > 1.960$ ($T > 1.640$), the estimate is statistically significant at the 5% (10%) level. Hence, the higher the T the more likely to reject the null-hypothesis.

In Section 5 we therefore discuss a few issues that you should keep in mind when estimating standard errors:

- Heteroscedasticity;
- Clustering; and
- Serial correlation.

STEP 7: ESTIMATE THE MODEL AND INTERPRET THE RESULTS

Now you are ready to estimate the model you specified in Step (5). Of course, you can use standard statistical packages such as `Stata`, `R`, `MatLab` or, when you want to estimate discrete choice models, you may use `Biogeme`.

This will mean that you will obtain an estimate of β and a corresponding standard error. It is then very important to *properly interpret the estimated effect.* For that it is important to exactly know in what units y_i and x_i are measured. For example, say that y_i is the current yearly income in euro and x_i are the years of schooling. When $\hat{\beta} = 1000$ with a standard error of 200, this implies that one year of additional schooling leads to a €1000 increase in yearly income. This effect is statistically significant at the 1% level ($T = 1000/200 = 5 \gg 2.576$). However, when years of schooling would be measured in months, the estimated effect ($\beta = 1000$) would economically be much larger.

Related to this, we should distinguish between *economic significance* and *statistical significance* (see McCloskey and Ziliak, 1996; Ziliak and McCloskey, 2008). A coefficient could be statistically significant, but of limited economic impact. For example, say that in the example of the effect of schooling in income, $\hat{\beta} = 10$ with a standard error of 5. We would conclude that the estimate is statistically significant at the 5% level, but we would agree that a €10 increase income for one year of schooling would not be an awfully large effect. By contrast, if $\hat{\beta} = 10,000$ with a standard error of $7,500$ we would conclude that the estimate is not statistically significant at a conventional significance level; however, a €10,000 increase in yearly income seems to be a large and economically meaningful. Hence, when interpreting your results, do not only pay attention to statistical significance but also clearly indicate whether your result is economically meaningful and significant.

Given the prominence of log-linear regressions, we would like to say something on how to interpret coefficients in log-linear regressions. In Table 2.1 we display several options when either the outcome or treatment variables are log-transformed.

We should make one remark regarding the lower-left quadrant. While the logarithmic scale approximates percent changes, it is not entirely correct for large changes in either x_i or β. This particularly holds when x_i is a dummy variable. Halvorsen and Palmquist (1980) and Angrist and Pischke (2014, p. 94), show that the percent change in y_i is given by:

$$(e^{\hat{\beta}} - 1) \times 100\%. \tag{2.6}$$

Table 2.1 – INTERPRETATION IN LOG-LINEAR REGRESSIONS

	x_i	$\log x_i$
y_i	$y_i = \beta x_i + \epsilon_i$ $\hat{\beta} = \frac{\partial y_i}{\partial x_i}$ $x_i \uparrow 1 \rightarrow y_i \uparrow \hat{\beta}$	$y_i = \beta \log x_i + \epsilon_i$ $\hat{\beta} = \frac{\partial y_i}{\partial \log x_i}$ $x_i \uparrow 1\% \rightarrow y_i \uparrow \hat{\beta}/100$
$\log y_i$	$\log y_i = \beta x_i + \epsilon_i$ $\hat{\beta} = \frac{\partial \log y_i}{\partial x_i}$ $x_i \uparrow 1 \rightsquigarrow y_i \uparrow (\hat{\beta} \times 100)\%$ (for marginal changes in x)	$\log y_i = \beta \log x_i + \epsilon_i$ $\hat{\beta} = \frac{\partial \log y_i}{\partial \log x_i}$ $x_i \uparrow 1\% \rightarrow y_i \uparrow \hat{\beta}\%$

Let's say that you are interested to know the log house price difference between houses with a garden and without a garden and you find $\hat{\beta} = 0.012$. Halvorsen and Palmquist's formula shows that the house price increase would be 1.207%, which is essentially identical to the coefficient. However, if $\hat{\beta} = 0.28$, the percent increase in house prices would be 32.3%. Hence, for large changes in x_i or when $\hat{\beta}$ is large, please use Halvorsen and Palmquist's formula to calculate the percent change in the dependent variable.

STEP 8: PROVIDE ROBUSTNESS CHECKS OF THE RESULTS

The final, but an important, step is to show that your results are robust, because in the research process you frequently have to make choices of which some are somewhat arbitrary. For example, when cleaning the data, you have to make a decision whether or not an observation is valid or not. The robustness analysis consists of a set of alternative regressions where you test the sensitivity of β with respect to many decisions including definitions of variables, functional form, alternative identifying assumptions, sample selections and data corrections. What relevant robustness checks are highly depends on the context.

For example, you wish to estimate the effect of the presence of a nearby coffee-shop on the logarithm of house prices. But what is nearby? Let us suppose you define nearby as within 50m, and you find that prices decrease by 2%. As a minimum, you have to check then that if you would have chosen another definition, *e.g.* 40m, or 60m, you would get similar estimates (or that the estimates based on 40m are slightly stronger, because then the coffee-shop is more nearby, whereas with 60m the effect might be slightly weaker).

3 RANDOMISED EXPERIMENTS

3.1 THE SELECTION BIAS AND RANDOMISATION

According to Angrist and Pischke (2008, 2014) the most credible and influential research designs use random assignment to treatment and control groups. We will explain here why randomised experiments are considered to be the ideal. It will also be clear why you see them seldom in spatial economics. It is however key to understand the advantages of random assignment to treatment and control groups, because this helps you to understand how close other approaches that are common in spatial economics come close to this ideal.

Let's consider a research project evaluating the effects of a place-based policy, such as a housing subsidy, on well-being of beneficiaries of the policy.[13]

We define y_i to be the outcome variable measuring self-reported well-being of inhabitant i and x_i is a dummy treatment variable indicating whether the individual received treatment, *i.e.* whether the individual received a subsidy to upgrade her/his property. Then:

$$y_i = \begin{cases} y_{1i}, & \text{if } x_i = 1; \\ y_{0i}, & \text{if } x_i = 0, \end{cases} \tag{3.1}$$

so y_{1i} is the outcome if individual i received treatment and y_{0i} if this individual did not receive treatment. Then:

$$y_i = y_{0i} + (y_{1i} - y_{0i}) \times x_i, \tag{3.2}$$

where $(y_{1i} - y_{0i})$ denotes the causal effect of the treatment. Note that the causal effect may be different for different individuals. For example, inhabitants differ in terms of age, income and education, and it is therefore unlikely that the effect is identical for each individual.

Note that we do not observe the same individual at the same moment with and without treatment – we only observe outcomes of individuals that are treated or not treated. If we naively compare the expected well-being of the individuals which received a place-based policy to the control group of individuals that did not, we have:

$$\begin{aligned} \underbrace{\mathbb{E}[y_i|x_i = 1] - \mathbb{E}[y_i|x_i = 0]}_{\text{Observed difference in well-being}} = & \underbrace{\mathbb{E}[y_{1i}|x_i = 1] - \mathbb{E}[y_{0i}|x_i = 1]}_{\text{Average treatment effect on the treated}} \\ & + \underbrace{\mathbb{E}[y_{0i}|x_i = 1] - \mathbb{E}[y_{0i}|x_i = 0]}_{\text{Selection bias}}, \end{aligned} \tag{3.3}$$

where $\mathbb{E}[\cdot]$ denote expectations, and where the last step follows because we added and subtracted the term $\mathbb{E}[y_{0i}|x_i = 1]$. The average treatment effect is the difference between the well-being of those who received treatment ($\mathbb{E}[y_{1i}|x_i = 1]$) and the well-being of the

[13]Place-based policies refer to policies, as the name suggests, which differentiate between locations, and are very common. For example, in the Netherlands, the 83 poorest neighbourhoods of the entire country, have received substantial subsidies to upgrade the stock of public housing.Another example of a place-based policy is to build highways in poor regions within the EU.

individuals who received treatment *if they would not have received treatment* ($\mathbb{E}[y_{0i}|x_i = 1]$). However the observed difference in well-being consists of an additional term, which is the difference between the base well-being (y_{0i}) between those who received the place-based policy subsidy and those who are not. This is referred to as the *selection bias*, because the individuals which are treated differ from those which are not treated with respect to certain (unobserved and observed) variables.

In the example where you want to measure the difference in well-being between individuals who received a housing subsidy and those who did not receive the subsidy, this selection bias is likely huge because individuals with higher incomes generally report higher levels of well-being but at the same time are unlikely to receive a housing subsidy ($\mathbb{E}[y_{0i}|x_i = 1] \ll \mathbb{E}[y_{0i}|x_i = 0]$). Hence, if you do not take into account the selection bias you would wrongfully conclude that a housing subsidy decreases well-being of beneficiaries. Also in other contexts, the selection bias may be considerable and so large that it masks the average treatment effect on the treated (Angrist and Pischke, 2008).

How then to overcome the selection bias? Let's consider randomisation of the treatment. Hence, in this example this would mean that you randomly select that receive treatment, while another randomly selected group other individuals do not. This would imply that:

$$\mathbb{E}[y_{0i}|x_i = 1] = \mathbb{E}[y_{0i}|x_i = 0]). \tag{3.4}$$

Hence, by randomising treatment assignment, the expected base well-being is the same. Using (3.4), equation (3.3) simplifies to:

$$\underbrace{\mathbb{E}[y_i|x_i = 1] - \mathbb{E}[y_i|x_i = 0]}_{\text{Observed difference in well-being}} = \underbrace{\mathbb{E}[y_{1i}|x_i = 1] - \mathbb{E}[y_{0i}|x_i = 0]}_{\text{Average treatment effect}}. \tag{3.5}$$

The latter implies that if we have a randomised experiments, by comparing means between individuals in treated and control groups, we identify the *average treatment effect*.[14]

One should notice that in some settings the *average treatment effect* may not be of particular interest. For the place-based policy example, we are not interested per se in the overall increase in well-being of arbitrary chosen individuals; we are particularly interested in the increase in well-being of people that are poor. But by including rich individuals (with higher initial levels of well-being and who are less likely to see an increase in well-being because of the place-based policy), the average treatment effect is likely an underestimate of the treatment effect of poor individuals in which we are interested. This implies a paradox: *the average treatment effect may apply to no one*. Hence, in cases where heterogeneity in treatment is large, the average treatment effect may be of limited interest. One may then look at *e.g.* the median treatment effect, or the share of individuals that respond positively to a treatment (although calculating these heterogeneous effects may not be straightforward, see *e.g.* Lee, 2000).

[14] In Appendix A.3 we explain why randomisation also implies that we identify the overall *average treatment effect*, rather than just the *average treatment effect on the treated*.

3.2 ANALOGY TO REGRESSIONS

Regression is useful to analyse causal questions. Can we than incorporate the above discussion into a regression framework? Let's first assume that the treatment effect $\beta = y_{1i} - y_{0i}$ is a constant (so it does not vary with i). We then can rewrite (3.2) to:

$$y_i = \underbrace{\alpha}_{\mathbb{E}[y_{0i}]} + \underbrace{\beta}_{y_{1i}-y_{0i}} x_i \quad + \underbrace{\epsilon_i}_{y_{0i}-\mathbb{E}[y_{0i}]} \quad , \tag{3.6}$$

where ϵ_i is the random part of y_{0i}. If we evaluate the above equation *with* and *without* treatment, we have:

$$\begin{aligned} \underbrace{\mathbb{E}[y_i|x_i=1]-\mathbb{E}[y_i|x_i=0]}_{\text{Observed difference in well-being}} &= \underbrace{\beta}_{\text{Treatment effect}} \\ &\quad + \underbrace{\mathbb{E}[\epsilon_i|x_i=1]-\mathbb{E}[\epsilon_i|x_i=0]}_{\text{Selection bias}}. \end{aligned} \tag{3.7}$$

Hence, if there is a selection bias, this implies the expected value of the error term differs for the group which is treated from the group which is not treated, hence there is a correlation between the error term ϵ_i and the treatment variable x_i.

To explore this point further, let's now consider to include additional control variables c_i to (3.6). Note that if group assignment is really random, then control variables are uncorrelated to x_i and adding them should not change the estimated β. However, adding these controls may lead to more *precise* results (so smaller standard errors) if these controls explain variation in the outcome variable and therefore reduce residual variance (Angrist and Pischke, 2008). Because the standard error of β is a function of the residual variance, this leads to a lower standard error.[15]

The above assumption that the treatment effect is a constant is very restrictive. In general, the treatment effect $\beta_i = y_{1i} - y_{0i}$ is not constant, but varies with i. In this case, one aims to estimate the average value of $\mathbb{E}[\beta] = \mathbb{E}[y_{1i} - y_{0i}] = \mathbb{E}[y_{1i}] - \mathbb{E}[y_{0i}]$. Given a regression, the estimate of β, so $\hat{\beta}$, can still be interpreted as the average treatment effect in the population (ATE).

For example, assume that β takes two values which depend on an unobserved random dummy variable S, which is not correlated to any of the independent variables. Hence, $\beta_1 = y_{1i} - y_{0i}$ for $S_i = 1$ and $\beta_2 = y_{1i} - y_{0i}$ for $S_i = 0$. Consequently, the average treatment effect is equal to $\beta_1 \mathbb{E}S_i + \beta_2(1 - \mathbb{E}[S_i])$, where $\mathbb{E}[S_i]$ is equal to the share of observations for

[15] More specifically, the formula for the standard error of β is (Angrist et al., 2017, p. 95): $SE(\hat{\beta}) = \frac{\sigma_\epsilon}{\sqrt{N}} \frac{1}{\sigma_x} = \sqrt{\left(\sum_{i=1}^N \epsilon_i^2\right) / \left(N \times \sum_{i=1}^N (x_i - \bar{x})^2\right)}$, where N is the number of observations, σ_ϵ refers to the standard deviation of the regression residuals and where σ_x refers to the standard deviation of x. If more controls are added that are not correlated to x, you may observe $SE(\hat{\beta})$ becomes smaller because ϵ_i becomes smaller. Instead, if the controls are (positively or negatively) correlated to x, then the standard deviation of x, when controlling for other variables, becomes smaller, so the denominator falls and the $SE(\hat{\beta})$ may become larger when more controls are added.

Table 3.1 – HEMOGLOBIN AND ANEMIA RESULTS

	Hemoglobin	Anemia
	(1)	(2)
Sales village	0.033	-0.006
	(0.029)	(0.009)
Number of observations	11,503	11,503

Notes: We report the coefficients on the sales treatment variable from a regression with either hemoglobin concentration or has anemia as the outcome variable. Standard errors in parentheses; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

which holds that the causal effect is equal to β_1 and where $1 - \mathbb{E}[S_i]$ is equal to the share of observations for which holds that the causal effect is equal to β_2. When you estimate this model with OLS, $\hat{\beta} = \beta_1 \mathbb{E} S_i + \beta_2 \mathbb{E}[1 - S_i]$.

Application 1 — the impact of double fortified salt on anaemia. Banerjee et al. (2018) study the effects of a randomised experiment reducing 'iron deficiency anaemia', which is a frequent health issue among the poor worldwide. They focus on a particular treatment, which is salt fortified with iron and iodine. They distribute Double Fortified Salt (DFS) very cheaply (₹9, or €0.012, per kg) in 200 *randomly* selected villages, while having 200 control villages that are not aware of the existence of DFS. They then run the following regression:

$$y_{ig} = \alpha + \beta x_{ig} + \gamma c_{ig} + \mu_{g \in b} + \epsilon_{ig}, \tag{3.8}$$

where i denotes the individual, g is the village, and $\mu_{g \in b}$ are so-called Block b fixed effects, where Blocks refer to an administrative unit smaller than the District. In in each of eight Blocks they randomly select 29 villages and in each of six Blocks randomly select 28 villages.

Table 3.1 reports the results. They find essentially no effects of the treatment on the intensity. For haemoglobin, given a control mean of 12.056, the effect is economically small and statistically insignificant. The same holds for anaemia. Banerjee et al. (2018) show that the absence of both economically and statistically significant effects holds across a wide array of different demographic sub-groups.

This finding is in contrast to previous experiments showing positive effects of DFS. Banerjee et al. (2018) argue that these somewhat disappointing results are explained by *(i)* relatively modest purchases and *(ii)* a low impact of DFS for the majority of the population, even when consumed somewhat regularly.

3.3 ADVANTAGES AND DISADVANTAGES OF RANDOMISED EXPERIMENTS

Randomised experiments have several advantages, of which the most important are (see Nickson, 2015):

- It is the only type of study that establish causation. Although other identification strategies may come close to a randomised setting, they will not rely on exact randomisation.
- An experiment forces the researcher to think clearly on what the treatment is and how this is administered to respondents.
- Randomised experiments, when large enough, allow for sub-group analysis and therefore provide evidence in heterogeneity in the treatment; beyond just identifying the average treatment effect.

There are also several disadvantages of randomised experiments. We list a few here (Nickson, 2015):

- For many economic questions (such as related to economic history) random experiments are simply not possible (Pearl, 2009).[16]
- Randomisation of treatment may be unethical in the sense that withholding treatment (*e.g.* going to the hospital, education) is ethically unacceptable.
- Randomised experiments are very expensive to organise, especially when you aim to have large group sizes.
- Experiments can be logistically challenging to organise if treatment occurs at multiple sites and locations.
- Long-term effects (such as the effects of years of schooling on wages) are hard to identify.
- The setting of a randomised experiments may not always mimic real life situations, which lead to limited external validity and biases.[17]
- One may question whether full randomisation really occurs. In many settings, although the treatment is random, the selection of applicants into the experiment is not random.[18] Hence, the estimated treatment effect only applies to the group of people that choose to participate in the experiment.

A more philosophical critique on randomised experiments is that, although we may measure a causal effect, this does not mean that we know *why* there is an effect of x_i on y_i; and hence the estimate may yield little information. Deaton (2010) is very critical towards randomised experiments and argues that:

> "*without guidance from an understanding of underlying mechanisms, [randomised experiments are] unlikely to lead to scientific progress in the understanding of economic*

[16]This does not only applies to economics, but also to evolutionary biology, history, computer science, etc.

[17]These are biases that also are discussed when discussing Stated Preference data. See the syllabus on discrete choice.

[18]Think of many psychological experiments where applicants receive a small amount of money to participate. Only potential participants with a high marginal utility of money and a lower value of time will then participate (*e.g.* students).

> *development.*"

Deaton emphasises the role of economic theory in producing insights in viable mechanisms. Econometric methods, including randomised experiments, then should be used to explicitly evaluate these theoretical mechanisms.

4 ALTERNATIVE IDENTIFICATION STRATEGIES

4.1 EXHAUSTIVE SET OF CONTROLS

Because randomised experiments often cannot be applied to economic settings we have to think of alternative ways to measure a *causal effect* of x_i on y_i. The strategy to deal with any bias including the selection bias (see Section 3.2) is called an *identification strategy*.

The most obvious first step to deal with selection bias is to include a host of control variables, c_i. Let's write:

$$y_i = \alpha + \beta x_i + \gamma c_i + \epsilon_i. \tag{4.1}$$

By controlling for additional variables, the omitted variable bias might be reduced, because now the omitted variable bias is equal to $\mathbb{E}[\epsilon_i|x_i = 1, c_i] - \mathbb{E}[\epsilon_i|x_i = 0, c_i]$. Hence, if the control variables are correlated with the main variable of interest, x_i and with the error term, ϵ_i, then the omitted variable bias is smaller.[19] Although it may be theoretically possible that we observe all relevant controls in a certain application, in most applications one cannot be certain to observe all relevant control variables. This holds particularly for many research projects in spatial economics, as it close to impossible to have data on all relevant determinants of, say, house prices, air pollution or travel destination choices, and it is frequently likely that the variable you are interested in is correlated to one of those determinants. Please note that this gives rise to an omitted variable bias of which the size x_i depends on the strength of the correlation with the omitted variables.

A standard way is then to either use *first-differenced data* or include *fixed effects*. Let's now consider that you have temporal data, so you observe individuals in different periods t (*e.g.* data for several years) where you observe some individuals at least twice. We further make a difference between controls that vary over time, c_{it} and controls that do not, denoted by d_i. Hence, we generalise (4.1), and we write:

$$y_{it} = \alpha_t + \beta x_{it} + \gamma c_{it} + \delta d_i + \epsilon_{it}. \tag{4.2}$$

Please notice that we have allowed here for time fixed effects: α_t. In other words, we have

[19] One can reduce this bias not only by estimating regressions with controls, but also by estimating so-called *matching estimators*, which, in practice, give almost the same result as OLS. In case of matching estimators, one matches each observation with another observation (or a combination of observations) that is similar in terms of control variables, but differs in terms of x_i. The average difference in the dependent variable, y_i, between the two samples (one sample for which $x_i = 1$, another sample for which $x_i = 0$), gives the average effect of x_i on y_i.

added a dummy variable for each t (*e.g.* a dummy for each year of observation). Why would you do that? One reason is that there are often trends in the variables. For example, housing prices usually increase over time because of inflation. Similarly, in most countries, air pollution usually has a strong trend (because of technological developments).

Let us now denote differences over time by Δ. So, for example, $\Delta y_{it} = y_{it} - y_{it-1}$. We can write:

$$\begin{aligned} y_{it} - y_{it-1} &= (\alpha_t - \alpha_{t-1}) + \beta(x_{it} - x_{it-1}) + \gamma(c_{it} - c_{it-1}) + \delta(d_{it} - d_{it-1}) + \epsilon_t - \epsilon_{t-1}. \\ \Delta y_{it} &= \Delta\alpha_t + \beta\Delta x_{it} + \gamma\Delta c_{it} + \Delta\epsilon_{it}. \end{aligned} \tag{4.3}$$

Because $d_{it} = d_{it-1}$, the term cancels from the regression. Hence, first-differencing the data controls for all *time-invariant* characteristics of an individual i. Many factors are actually (almost) time-invariant (think of: birthplace, gender, cultural factors, mother tongue, geographical features of the landscape, etc.), so using first-differencing is a powerful step to mitigate omitted variable bias problems, although it will not completely address it when many factors (c_{it}) may change over time.

Note that an important prerequisite for first-differencing the data is that there is variation in the treatment over time; otherwise the effect cannot be identified will also be cancelled. Hence what do we learn from this? In general, focusing on treatments that change over time are more convincing. For example, if we have data for several years, and at a certain moment, some individuals receive an improvement of housing, we are more likely to get a causal effect. If you're interested in the effect of coffeeshops on house prices, it is more convincing to analyse data when coffeeshops are forced to close because of a certain policy, then to analyse the effect of coffee shops on house prices when there is no variation in coffee shops within your dataset.

A similar approach to mitigate omitted variable bias is to include group fixed effects. Let's say you can attribute each individual i to groups g. These groups can be regions, municipalities, neighbourhoods, or even individuals themselves if you observe individuals over time. Hence, we write:

$$y_{ig} = \alpha + \beta x_{ig} + \gamma c_{ig} + \mu_g + \epsilon_{ig}, \tag{4.4}$$

where μ_g denotes the group fixed effect. Including many fixed effects is computationally difficult (for example, if you have data for the full population, you may have to include millions of fixed effects). Fortunately, for linear models, one can get rid of the fixed effects, with a trick using the averages. This trick is not only extremely useful, it is also very insightful. The fixed effects estimator implies that you subtract the group average, denoted by $\bar{y}_g$, from (4.4):

$$\begin{aligned} y_{ig} - \bar{y}_g &= \beta(x_{ig} - \bar{x}_g) + \gamma(c_{ig} - \bar{c}_g) + \mu_g - \bar{\mu}_g + \epsilon_{ig} - \bar{\epsilon}_g. \\ \Delta y_{ig} &= \beta\Delta x_{ig} + \gamma\Delta c_{ig} + \Delta\epsilon_{ig}. \end{aligned} \tag{4.5}$$

where we now redefine Δ as the difference of i with the average of the group g to which i

belongs to. So, for example, $\Delta y_{ig} = y_{ig} - \bar{y}_g$. Note that when estimating (4.4), one does not explicitly estimate the group fixed effect.

Here we assume that we can make a distinction between control variables that vary within the group c_{ig} and only between groups d_g. By including fixed effects at the group level, the latter controls cancel from the regression equations (as $d_g = \bar{d}_g$) in the same way as the μ_g cancel out. Hence, by including fixed effects we control for all factors that vary only between groups. These can be *e.g.* municipal or national taxes, demographic compositions of neighbourhoods, country-wide environmental policies.

Generally speaking, the more refined the definition of the groups the more convincing the fixed effects estimator is.[20] For example, let us suppose you have cross-sectional data, and you wish to know the effect of an attractive energy label on house prices (in some countries, energy labels are required by law to inform buyers of the energy efficiency of the house), as an attractive energy labels may provide an incentive to homeowners to improve insulation of their house as they know they can get a higher price if they sell the house. The difficulty here is that the presence of an attractive energy label is likely correlated with unobserved characteristics that are related to the spatial environment. Hence, it is a good idea to include neighbourhood fixed effects, but it would even be better to include street fixed effects, or even postcode fixed effects, if streets contain several postcodes.

However, the downside is that the treatment variable always needs to vary *within* the group; otherwise we cannot identify the effect of interest. For example, if all houses within the same postcode have the same energy label, then including postcode fixed effects does not work. Moreover, even if there is variation within the groups, the fixed effects may absorb part of the effect where one is interested in (Abbott and Klaiber, 2011). For example, say you are interested in the effects of pollution on health outcomes, while you include local fixed effects (say neighbourhood fixed effects). Although strictly speaking pollution varies within the neighbourhood, it will be hard to detect differences.

When you have a panel data of individuals, you have to decide whether or not to use fixed effects estimators or first-differences. Note that if you have (maximally) 2 observations per individual, both estimators generate identical results, but the results may differ if you have more observations per individual. It is frequently the case that the fixed effect estimator is (much) more efficient than by taking first-differences, so you obtain smaller standard errors.[21] However, the fixed effects estimator have the drawback that it imposes slightly stronger exogeneity conditions, as the first difference estimator combines information about individuals that are one period apart, whereas the fixed effects estimator uses information from individuals over longer periods.

[20] You may also include multiple fixed effects (say year fixed effects and region fixed effects). Recent packages in `Stata` allow for the inclusion of many types of fixed effects. Please consider the command `reghdfe`.

[21] The intuition is that with fixed effects you subtract the group-averaged variables which, generally, has less random variation than the variables one time period lagged, which you use when calculating first differences; however, given strong serial correlation in the residuals, even this result is not general.

Application 2 — Footfall and retail vacancies. For retail firms, the number of potential shoppers in a certain location is likely one of the most important determinants of location choice. This is investigated in Koster et al. (2019) where, among other things, the impact of number of pedestrians passing by a shop, referred to as *footfall*, on the vacancy probability is investigated. They use data on footfall counts from *Locatus* in 1,253 shopping streets and data on locations of shops in those streets between 2003 and 2015.

Ideally, we would like to compare retail firms in identical streets with randomly assigned levels of footfall. They aim to estimate the following regression:

$$y_{it} = \alpha + \beta \log x_{it} + \mu_t + \epsilon_{it}, \tag{4.6}$$

where y_{it} denotes a dummy variable that equals one when the property i is vacant in a given year t. x_{it} is the number of pedestrians per hour on a regular Saturday in year t and μ_t are year fixed effects.

One may be concerned that endogeneity thwarts a causal interpretation of β as locations with more pedestrians may be more attractive for retailers for other reasons. Koster et al. (2019) therefore only include observations close to shopping street intersections and include *intersection fixed effects.* The design is illustrated for the city centre of Rotterdam in Figure 4.1. The idea is to compare shops are around the corner with each other that have very different levels of footfall. However, by including intersection fixed effects we control for all differences between shopping streets.

Table 4.1 reports the results based on Linear Probability Models. The coefficient in column (1) implies that doubling of footfall leads to an decrease in the vacancy rate of $(\log 2 - \log 1) \cdot -0.0304 = 2.1$ percentage point. Given an average vacancy rate of about 6%, this effect is substantial.

One may concerned that the results are inconsistent because of reverse causality; a vacant shop may also imply that fewer pedestrians are attracted; in other words, v_{it} may influence x_{it}. To address this concern two instruments are considered: the number of cinemas in 1930 within 200m and the number of shops in 1832 within 200m. Because of historic reasons, past concentrations of shops/cinemas affect the footfall positively today. However, these are not impacted by contemporary vacancy rates. In column (2) and (3) it is shown that these instruments are reasonably strong (especially for cinemas in 1930) and the results are somewhat stronger. However, given the larger standard errors they are not statistically significantly different from the baseline OLS estimates.

4.2 INSTRUMENTAL VARIABLES

The strategy to include fixed effects or first-difference the data is often combined with the use of *instrumental variables* (IV). IV is first described in Wright (1928) and applied in Reiersol (1941). It is now a very important tool within the realm of applied economics

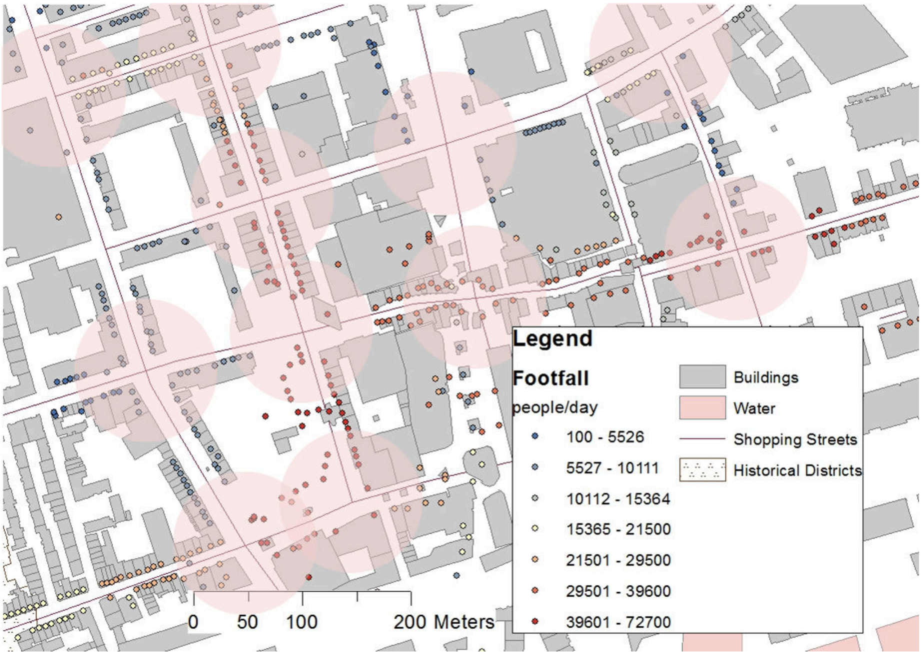


Figure 4.1 – FOOTFALL AND SHOPS IN ROTTERDAM

Table 4.1 – FOOTFALL AND RETAIL VACANCIES
(Dependent variable: shop is vacant)

	OLS	2SLS	2SLS
		Instrument: Cinemas in 1930	Instrument: Shops in 1832
	(1)	(2)	(3)
Footfall *(log)*	-0.0304***	-0.06161***	-0.0500*
	(0.0016)	(0.0176)	(0.0258)
Shop and location characteristics	Yes	Yes	Yes
Intersection fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Number of observations	394,389	394,389	207,242
R^2	0.0448		
Kleibergen-Paap F-statistic		37.37	9.777

Notes: We only keep observations within 250m of an intersection. In column (2) we instrument for footfall with the number of cinemas in 1930 within 200m and in column (3) the number of shops in 1832 within 200m. Standard errors in parentheses; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

(Stock and Trebbi, 2003). The first Nobel prize winner in Economics, Ragnar Frisch, received his prize for his contribution to IV. Arguably, its most important use is to address omitted variable bias (as well as reverse causation and measurement error bias), and it is considered to be the main contribution of economists to statistical theory.

To keep the exposition as simple as possible, suppose you are interested to estimate the following regression (without control variables) and you have one instrument:

$$y_i = \alpha + \beta x_i + \epsilon_i, \tag{4.7}$$

where x_i is (strongly) correlated to ϵ_i. For example, if you are interested in estimating a demand function, then y_i would be the quantity of the good and x_i would be the price of the good, which is correlated to ϵ_i.

We can use two-stage least squares (2SLS) to estimate the causal effect of x_i on y_i:

$$\begin{aligned} x_i &= \zeta + \eta z_i + \xi_i, && (1^{\text{st}} \text{ stage}) \\ y_i &= \alpha + \beta \hat{x}_i + \epsilon_i, && (2^{\text{nd}} \text{ stage}) \end{aligned} \tag{4.8}$$

where z_i is called an instrument, and $\hat{x}_i$ is the predicted value from the first stage using the estimated coefficient $\hat{\eta}$. We emphasise the two conditions for valid instruments:

- *The instrument should be relevant*, implying that $\text{cov}[z_i, x_i] \neq 0$. In other words, $\hat{\eta}$ should be a strong instrument. As a rule of thumb, most researchers interpret this that the instrument should be statistically significant with an F-statistic exceeding at least 10.[22] When having multiple endogenous variables, please look at the Kleibergen-Paap F-statistic to investigate whether the instrument is strong.
- *The instrument should be exogenous*, implying that $\text{cov}[z_i, \epsilon_i] = 0$. This assumption is not testable, because, by definition, we do not observe unobservable variables implying that we cannot assess $\text{cov}[z_i, \epsilon_i]$. Hence, the exogeneity of an instrument should be based on sound economic reasoning.[23] Note that this assumption also implies that the instrument z_i *only* influences y_i *via* x_i.[24]

Now suppose again as above that $\beta_i = y_{1i} - y_{0i}$ is not constant, but varies with i. We have seen above that OLS then provides the average treatment effect for the population. Unfortunately, this is not the case when you use instrumental variables. For example, suppose that the instrument z_i is only observed when $S_i = 1$, but there is no information about this variable otherwise. For example, if the instrument is the gender of couples with (at least) two children (Angrist and Evans, 1998), then this instrument is not observed for individuals with less than two children.[25] Consequently, the instrumental variable approach provide the average

[22] The F-statistic is the square of the t-statistic in case of one variable, so the t-statistic should exceed 3.16 when you have one instrument.

[23] An attentive reader may remark that if one has multiple instruments one may use *overidentification* tests to assess the validity of the instruments. However, with overidentification tests, one can only investigate the validity of one instrument, *conditional on the validity of the other instrument(s)*. Furthermore, even if both instruments are valid, this does not imply that they give the same results, because each instrument provides a local average treatment effect. Hence, while overidentification tests are informative on the robustness of your results, they provide limited information on the overall exogeneity of the instruments.

[24] Let's say you are interested in the effect of development aid on economic growth. A problem is that economic growth may also impact development aid, because if countries get richer they will receive less aid. Deaton (2010) discusses that many studies have used *country dummies* as instruments for aid. While these are likely strong instruments, they are unlikely to meet the exogeneity condition, as country dummies may impact economic growth in a myriad of ways other than via aid (*e.g.* via geography, cultural factors, FDI, etc.).

[25] This instrument is useful in a number of contexts. In most countries, the gender of children can be interpreted as random, but still has influence on households' decisions. For example, households with children of different gender are more likely to occupy larger houses further from the city centre, because in most cultures, parents prefer children not to share rooms with children of another gender. It is also the case that

treatment effect, where the average is estimated for a subsample of individuals with $S_i = 1$. This average is known as the 'local average treatment effect' (LATE) (see Imbens and Angrist, 1994). Hence, in case treatment heterogeneity is important (so different individuals having different β's), different instruments may lead to different β's, because different individuals are used in the estimation procedure to identify the effect.

The use of IV is closely related to the estimation of (inverse) *demand and supply functions*, and therefore closely related to economic theory. Instruments then can be used to shift the supply function to identify the demand function or vice versa. We explain this in more detail in Appendix A.4 and provide an application to estimating supply and demand for parking.

In economics, there is an ongoing discussion on whether the local average treatment effect is of interest. Proponents of the (quasi-)experimental approach (see Angrist and Krueger, 2001; Angrist and Pischke, 2008; Imbens, 2010; Gibbons and Overman, 2012) have argued that a causal estimate is almost always better than not addressing selection bias. Others (like Deaton, 2010; Heckman, 2010) are more critical and advocate the use of 'structural' models to properly learn about the exact mechanisms of how x_i impacts y_i.

While the conditions for valid instruments are straightforward, in practice it is not so easy to find instruments that are both strong and uncorrelated to the error term ϵ_i. For example, the effect of gender of children on decision-making by households is presumably very small, so you will need a very large dataset in order to estimate such an effect. Consequently, this instrument might work if you have the data of the full population for several years of a specific country, but is less likely to work if you have 'only' a few hundred thousand observations.

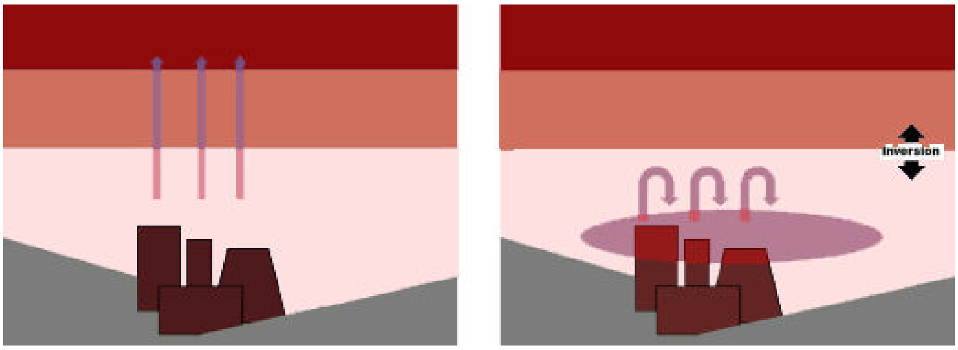
Whether an instrument is valid depends highly on the context and specific research questions. Below we will give some examples of likely valid instruments. Frequently, economic theory can be helpful to determine whether or not an instrument is potentially valid.

Application 3 — Air pollution and road safety. Sager (2019) aim to answer the question whether air pollution affects road safety. Recent evidence indicates that short-term fluctuations in air pollution may impair productivity and human behaviour in ways that could affect road safety (see *e.g.* Graff Zivin and Neidell, 2012). For example, pollution may drivers render more risk-taking and aggressive, and reduce cognitive performance.

Ideally, we would like to compare locations with randomly varying levels of pollution. A problem is that air pollution is a result of traffic and therefore is not randomly assigned across locations. Hence, the number of accidents per day may be affected by unobserved confounders, such as population density of regions, speed limits, road designs and types, etc.

Sager (2019) proposes an instrumental variable approach where temperature inversions are used as an instrument for variation in pollution levels. In Figure 4.2 we illustrate that temperature inversions are deviations from the norm that temperatures decrease

households where children are of the same gender are more likely to have another child.



(a) NORMAL CONDITIONS (b) INVERSION EPISODE

Figure 4.2 – EXAMPLES OF SPURIOUS CORRELATIONS

Notes: In Figure 4.2a temperature is lower in higher altitudes so that pollutants rise and disperse. In Figure 4.2b a warmer layer at higher altitudes prevents pollutants to rise.

with higher altitudes. During inversion episodes, warmer air at higher altitudes implies that pollutants are 'trapped' close to the ground. Does this instrument meet the two conditions for instrument validity? Most likely, because the instrument probably has a strong impact on air pollution, while temperature inversions are random and are unlikely to be correlated to economic conditions of locations.

The two-stage least squares estimation yields:

$$\begin{aligned} x_{it} &= \zeta + \eta z_{it} + \theta c_{it} + \nu_i + \nu_t + \xi_{it}, && (1^{\text{st}} \text{ stage}) \\ \log y_{it} &= \alpha + \beta \hat{x}_{it} + \gamma c_{it} + \mu_i + \mu_t + \epsilon_{it}, && (2^{\text{nd}} \text{ stage}) \end{aligned} \tag{4.9}$$

where x_{it} is the concentration of particulate matter ($PM_{2.5}$) in region i on day t and y_{it} is the number of vehicles involved in accidents. z_{it} is a measure of night-time inversion magnitude, c_{it} are control variables capturing weather conditions, and $\{\nu_i, \mu_i\}$ and $\{\nu_t, \mu_t\}$ are region and day fixed effects respectively.

The traffic and weather data are from the UK between 2009 and 2014. Table 4.2 reports the main results. Columns (1)-(3) show the first-stage results. It is shown that inversion strength has a strong and significant impact on the concentration of particulate matter. When fixed effects and weather controls are included, the coefficient in column (3) indicates that an additional degree in inversion strength leads to an increase of $1.395\mu g/m^3$ in daily average $PM_{2.5}$ concentrations. The first-stage F-statistic is well above 10 in all specifications. Columns (4)-(6) of Table 4.2 show the second-stage results. The most convincing specification in column (6) with fixed effects and controls imply that a $1\mu g/m^3$ increase in $PM_{2.5}$ concentration leads to an increase in accidents of 0.4%. This effect is statistically significant at the 1% level, and although small, not economically negligible.

Table 4.2 – POLLUTION AND ACCIDENTS

	First-stage results			Second-stage results		
	Dep. var.: PM$_{2.5}$			*Dep. var.: the log of number vehicles in accidents*		
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	OLS	2SLS	2SLS	2SLS
Inversion strength (°C)	1.830***	1.725***	1.395***			
	(0.0967)	0.0831)	(0.0771)			
PM$_{2.5}$				0.008***	0.004***	0.004***
				(0.0022)	(0.0008)	0.0010
Region-year fixed effects	No	Yes	Yes	No	Yes	Yes
Day-of-week fixed effects	No	Yes	Yes	No	Yes	Yes
Month fixed effects	No	Yes	Yes	No	Yes	Yes
Weather controls	No	No	Yes	No	No	Yes
Number of observations	265,723	265,723	247,106	265,723	265,723	247,106
Kleibergen-Paap F-statistic	358.5	430.8	327.3	358.5	430.8	327.3

Notes: Cluster-robust standard errors are in parentheses, allowing for two-clustering over NUTS3 regions and days. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

4.3 QUASI-EXPERIMENTAL METHODS

4.3.1 THE MAIN IDEA

There is a trend in applied spatial economics to use quasi-experimental variation in variables of interest to pin down a causal effect of x_i on y_i (Gibbons and Overman, 2012). These shocks can relate to national policy changes, (arbitrary) policy rules, historic events such as earthquakes and bombings, etc. These shocks cannot be influenced by individual decision makers and are therefore more likely to provide exogenous variation in x_i. If the shock is really random, this is good news as the selection bias will be zero (recall equation (3.5)) A concern with using these kind of shocks is that shocks may not only impact the variable x_i but also other (omitted) variables. For example, say you are interested in the effect of pollution on accidents and you consider to use the Covid-19 crisis as a shock to pollution, the problem is that Covid-19 did not only affect pollution, but also many other factors. Whether a quasi-experimental shock is really random depends highly on the research context and question at hand.

A general concern that applies to quasi-experimental identification strategies is that results may not be generalisable (Gibbons and Overman, 2012). Like in randomised experiments and IV, if effects of x_i are heterogeneous, it may be that the estimated β cannot be extrapolated to the general population. Most researchers agree, however, that it is better to have plausibly causal estimates for a specific group in the population, than to estimate non-causal parameters without particular meaning for the whole population (Imbens, 2010; Gibbons and Overman, 2012).

Application 4 — The Berlin Wall and agglomeration economies. Ahlfeldt et al. (2015) is interested in measuring the magnitude of agglomeration economies, which are cost savings arising from urban agglomeration, *i.e.* higher density, due to reductions in transport costs of goods, people and ideas (Glaeser, 2008). However, it is generally hard to distinguish between agglomeration economies and attractive features of a location that attracts firms (*e.g.* geographic features like rivers, local policies, etc.) (Koster et al., 2014).

Ahlfeldt et al. (2015) aim to exploit quasi-experimental variation in Berlin's density of economic activities to identify the importance of agglomeration economies. More specifically, Berlin has been divided in East and West-Berlin from 1961 until 1989 (see Figure 4.3). Especially in later years it was essentially impossible to travel from East to West Berlin and therefore locations on both sides of the Berlin Wall could not benefit from each other's economic activities (*i.e.* they could not benefit from each other's density). The building of the wall (the division), as well as its fall in 1989 (the reunification), can be interpreted as a shock to the economy of the city.

Here we consider the effects two outcome variables: the log of floor space prices and the log of employment. Let's consider the following specification:

$$\Delta y_i = \alpha + \sum_{k=1}^{K} \beta_k \mathbb{I}_{ik} + \mu_{i\in g} + \epsilon_i, \tag{4.10}$$

where Δy_i is the change in the outcome variable in block i after one of the two shocks, $\mathbb{I}_i k$ denotes indicator variables for whether block i lies within a 500m intervals, denoted by k, from the pre-war central business district (CBD) (*i.e.* 'Mitte'), while β_k are coefficients to be estimated. $\mu_{i\in g}$ are district g fixed effects.

Table 4.3 reports the results. In column (1) we observe that the rent has strongly dropped after the division for locations close to the CBD (relative to those far away). For example, the decrease in rents within 500m of the CBD due to division is $(e^{-0.567} - 1) \times 100 = 43\%$ (see equation (2.6)) as compared to locations that are further than 3km away (which is the reference category). The decrease in rents become smaller once a location is further away from the CBD. A likely explanation is provided in column (2) where it is shown that the decrease in employment density also seems to be greater close to the CBD (although the results are somewhat imprecise). The effects are statistically and economically significant.

In columns (3) and (4) we show similar results but now for reunification. The effects are the opposite as the CBD became relatively more central after reunification. For example, the first coefficient in column (3) indicates that rents within 500m of 'Mitte' increased by 50%, while employment increased by an astonishing 384%. However the effects are more localised and dissipate within 2km.

Ahlfeldt et al. (2015) then set-up a structural general equilibrium to use this quasi-

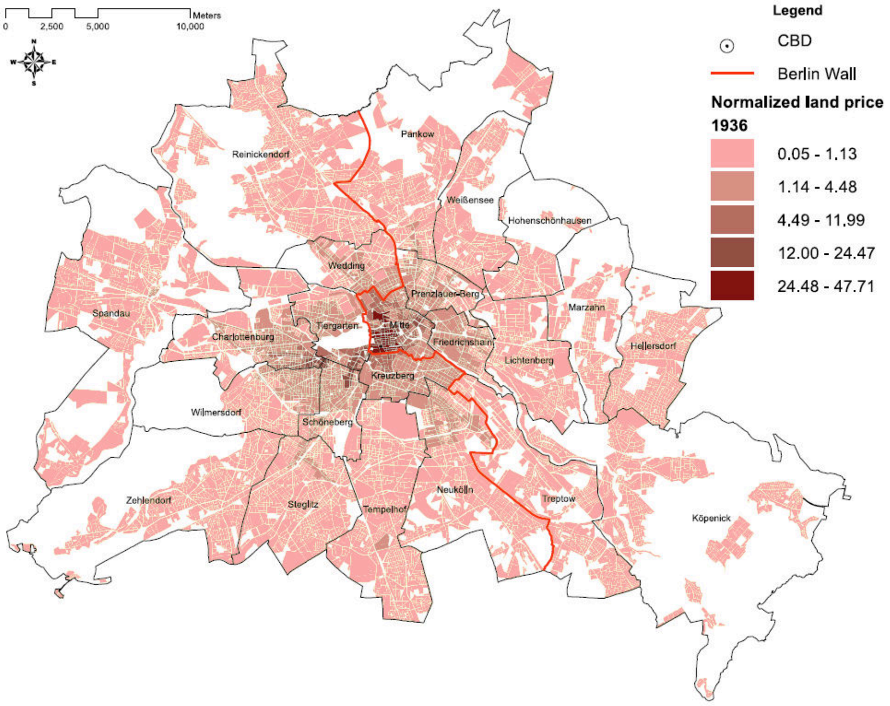


Figure 4.3 – THE BERLIN WALL

experimental variation in rents and density to estimate the magnitude and spatial extent of agglomeration economies, showing that these are very important within cities.

4.3.2 REGRESSION-DISCONTINUITY DESIGN

A specific example of a quasi-experimental research method is the so-called *Regression-Discontinuity Design* (RDD). Let's assume we have a treatment variable x_i that is discrete and dependent on a continuous variable r_i:

$$x_i = \begin{cases} 1, & \text{if } r_i \geq r_0; \\ 0, & \text{if } r_i < r_0. \end{cases} \tag{4.11}$$

We are interested in the effects of the treatment on an outcome variable y_i:

$$y_i = \alpha + \beta x_i + \gamma r_i + \epsilon_i. \tag{4.12}$$

Note that x_i here is a *fully deterministic function* of the variable r_i, which we refer to it as the *running variable*. In other words, if you know the value for r_i you also know whether someone received treatment. In case of a fully deterministic function this is known as a *sharp* RDD.

Let's provide an example. Consider the case where you study the effects of study grants on

Table 4.3 – RENTS, EMPLOYMENT AND THE BERLIN WALL

	Division		Reunification	
	Rents *(log)*	Empl. *(log)*	Rents *(log)*	Empl. *(log)*
	(1)	(2)	(3)	(4)
CBD 0-500m	-0.567***	-0.691*	0.408***	1.574***
	(0.071)	(0.408)	(0.090)	(0.479)
CBD 500-1000m	-0.422***	-1.253***	0.289***	0.684**
	(0.047)	(0.293)	(0.096)	(0.326)
CBD 1000-1500m	-0.306***	-0.341	0.120***	0.326
	(0.039)	(0.241)	(0.033)	(0.216)
CBD 1500-2000m	-0.207***	-0.512***	-0.031	0.336**
	(0.033)	(0.199)	(0.023)	(0.161)
CBD 2000-2500m	-0.139***	-0.436***	0.018	0.114
	(0.024)	(0.151)	(0.015)	(0.118)
CBD 2500-3000m	-0.125***	-0.280***	-0.000	0.049
	(0.019)	(0.130)	(0.012)	(0.095)
District fixed effects	Yes	Yes	Yes	Yes
Number of observations	6,260	2,844	7,050	5,602
Kleibergen-Paap F-statistic	0.51	0.12	0.32	0.03

Notes: Data on pre-division is from 1936, during the division it is from 1986 and from reunification it is from 2006. Standard errors adjusted for spatial correlation are in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

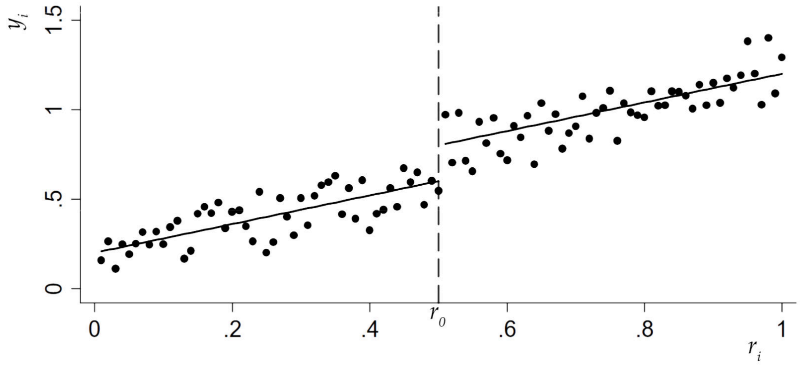
wages 10 years later, denoted by y_i. If students receive study grants, then students are more likely to spend more time on their studies, which would make them more productive later. Now suppose that study grants are given to students based on their grades (*e.g.* observed for their first year of their study). Let's assume that the running variable are grades, denoted by r_i. Further, assume that a student always receives a grant if her grades exceed r_0. Figure 4.4a shows an example of regression (4.12).

We observe a positive linear relationship between grades r_i and wages y_i, which makes sense as students with higher grades (who tend to be students who are brighter and study harder) are expected to earn more later. However, at r_0 we observe a *jump* in y_i. This is exactly the treatment effect of the grant. Note that controlling for the running variable is very important, as comparing someone with $r_i = 1$ to $r_i = 0$ would not only capture the treatment effect but also a selection effect because students with higher grades receive higher wages 10 years later.

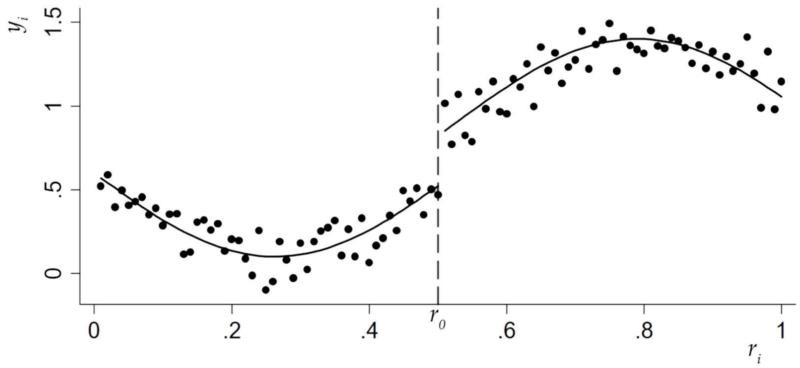
In this example, we assume a linear relationship between r_i and y_i. However, in real-life data there is not much reason to assume that this linear relationship exactly holds true. More specifically, we would prefer:

$$y_i = \alpha + \beta x_i + f(r_i) + \epsilon_i, \tag{4.13}$$

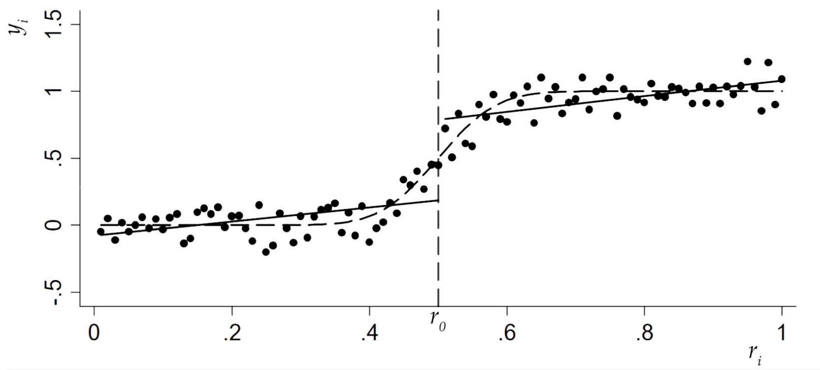
where $f(\cdot)$ is some unspecified function of the running variable. We may for example use polynomials to estimate $f(\cdot)$, implying that we include r_i, r_i^2, r_i^3 etc. Figure 4.4b shows an



(a) LINEAR $\mathbb{E}[y_i|r_i]$



(b) NON-LINEAR $\mathbb{E}[y_i|r_i]$



(c) NON-LINEARITY MISTAKEN FOR A DISCONTINUITY

Figure 4.4 – THE SHARP REGRESSION-DISCONTINUITY DESIGN

Notes: These figures are reproduced from Angrist and Pischke (2008).

example. Here again we observe a jump in wages at r_0 at the moment students receive the grant.

That it is important to control for the running variable is illustrated in Figure 4.4c. With a linear function of r_i, we mistakenly would conclude that the grant has a positive effect on wages. However, if we control more flexibly for the running variable, we do not see a jump in the regression function and $\beta \approx 0$. Hence, in any RDD one should show sensitivity of the results to different specifications of the running variable.

To circumvent this issue, one may also focus on a small area $(2 \cdot \delta)$ close to the cut-off r_0. Let's consider:

$$\begin{aligned} &\mathbb{E}[y_i | r_0 - \delta < r_i < r_0] \approx \mathbb{E}[y_{1i} | r_i = r_0], \\ &\mathbb{E}[y_i | r_0 \leq r_i < r_0 + \delta] \approx \mathbb{E}[y_{0i} | r_i = r_0], \end{aligned} \tag{4.14}$$

so that:

$$\lim_{\delta \to 0} \mathbb{E}[y_i | r_0 - \delta < r_i < r_0] - \mathbb{E}[y_i | r_i = r_0 \leq r_i < r_0 + \delta] = \mathbb{E}[y_{1i} - y_{0i} | r_i = r_0]. \tag{4.15}$$

The above equations imply that if we compare students that *just* got the grant with students that *just did not* got the grant (so when $\delta \to 0$) we identify the treatment effect of the grant. In practice there is a trade-off because if we only include students that are essentially at the trade-off we have too few observations to estimate the effect of interest. Hence, we re-emphasise that when adopting an RDD one should carefully investigate whether the treatment effect is robust to various values of δ, implying how many observations you include on both sides of the threshold. The value of δ is referred to as the *bandwidth*.

We have seen above that OLS will provide you with the average treatment effect, if the control variables (including fixed effects) nullify the omitted variable bias. In contrast, the IV approach will provide you with the local average treatment effect (LATE). How does this hold for the Regression Discontinuity Design? It is straightforward to see that this design provides you with the LATE. Specifically, suppose that β varies for different levels of the running variable r. For example, it is very plausible that the effect of a grant on wages differs with the skills and ambitions level of students. This implies that the effect estimated using the Regression Discontinuity Design depends on the level of r_0, so the estimated effect will be a LATE, defined by the sample close to r_0.

In many real-life applications assignment is not fully deterministic. In the example of students receiving a grant when there grades exceed r_0 it is more likely that not all students receive a grant when $r_i > r_0$, but that this also depends on other factors (say family background, past achievements, etc.). It may also be the case, that some students who have grades just below r_0 will get the grant, for example, if they can show that where ill during the exams. Nevertheless, there is still a large jump in the *probability* to receive the grant when $r_i > r_0$, denoted by $\mathbb{P}[x_i = 1 | r_i]$. When there is a jump in the probability of receiving treatment at some point, we refer to a *fuzzy* RDD. More formally, we have:

$$\mathbb{P}[x_i = 1 | r_i] = \begin{cases} g_1(r_i), & \text{if } r_i \geq r_0; \\ g_0(r_i), & \text{if } r_i < r_0, \qquad \text{where } g_1(r_i) \neq g_0(r_i). \end{cases} \tag{4.16}$$

Let's assume here that treatment is more likely for higher values of the running variable, so

$g_1(r_i) > g_0(r_i)$.[26] The relationship between the probability of treatment and the running variable then can be expressed as:

$$\mathbb{P}[x_i = 1|r_i] = g_0(r_i) + \Big(g_1(r_i) - g_1(r_i)\Big)z_i, \tag{4.17}$$

where $z_i = \mathbb{I}(r_i \geq r_0)$. Hence, the dummy variable z_i indicates the point where the probability on receiving the treatment is expected to jump.

Angrist and Pischke (2008, pp. 259-263) show that a fuzzy RDD leads to a standard two-stage least square estimation, where the discontinuity in the running variable is used as an instrument:

$$\begin{aligned} x_i &= \zeta + \eta z_i + g(r_i) + \xi_i, && (1^{\text{st}} \text{ stage}) \\ y_i &= \alpha + \beta \hat{x}_i + f(r_i) + \epsilon_i, && (2^{\text{nd}} \text{ stage}) \end{aligned} \tag{4.18}$$

So in the first stage we regress the treatment variable on a dummy indicating, which we will call the instrument, whether r_i exceeds r_0 (*i.e.* so whether the grade is high enough to possibly obtain a grant.), as well as a flexibly function of the running variable r_i. In the second stage the fitted value of x_i is used to estimate the causal effect of x_i on y_i.

Because the fuzzy design can be interpreted as an IV approach, one has to consider whether z_i satisfies the two above-mentioned conditions for instrument exogeneity (relevance and exogeneity). Instrument relevance is usually not so much an issue: if a minimum grade of r_0 is really necessary to obtain the grant, there is usually a strong increase in the probability to obtain the grant and $\text{cov}[x_i, z_i \neq 0]$. Instrument exogeneity $\text{cov}[x_i, \epsilon_i = 0]$ is also met as long as one controls flexibly for the running variable. Consider comparing students that just have a too low grade ($r_i < r_0$) with students that just have high enough grades, the only difference in wages at r_0 is attributable to jump in the probability to receive the grant.

When interested in applying RDDs in practice, we strongly advise you to have a careful look at Imbens and Lemieux (2008) and Lee and Lemieux (2010), which are excellent introductions in all the dos and don'ts of RDDs.

Application 5 – Urban renewal and neighbourhood attractiveness. There are many large-scale urban renewal projects aiming to improve built environment and reduce social inequality. Koster and Van Ommeren (2019) evaluate whether such policies are effective in making neighbourhoods more attractive by looking at house price differences before and after implementation of a large-scale programme in the Netherlands. This so-called *Krachtwijken*-scheme invested about a billion euro in 83 deprived neighbourhoods from 2008 onwards mainly by improving public housing. These neighbourhoods where ranked according to a so-called deprivation score, which we refer to as z_0. We show examples of targeted neighbourhoods for Amsterdam in Figure 4.5. We gather data on 434,033 housing transactions *of privately-owned properties* in 3,138 neighbourhoods covering whole of the Netherlands between 2000 and 2014.

[26] Note that if students never get the grant with grades below r_0, then $g_0(r_i) = 0$.

In an ideal setting, one would assign randomly treatment to neighbourhoods. A main issue in identifying the causal effect of the programme is that particularly deprived neighbourhoods with low house prices have been selected. Hence, a naive regression of log house prices on a dummy indicating whether a neighbourhood is part of the programme would lead to a strong *negative* coefficient.

To address endogeneity concerns we first-difference the data, so that we have:

$$\Delta y_{it} = \Delta\alpha + \beta\Delta x_{it} + \gamma\Delta c_{it} + \Delta\mu_t + \Delta\epsilon_{it} \tag{4.19}$$

where Δy_{it} is the log difference of house price between year t_1 and t_0 and x_{it} is a dummy indicating whether the neighbourhood has been treated. While first-differencing may mitigate the selection effect, one may still be concerned that more deprived neighbourhoods have different price trend, *e.g.* due to gentrification or trends in crime rates.

We therefore apply an RDD. In Figure 4.6 we plot assignment as a function of the deprivation z-score. There is a clear discrete jump in the probability of treatment at $r_i = 7.3$. However, some neighbourhoods were not treated despite having a sufficiently high score. Moreover, two neighbourhoods with scores lower than r_i were eventually treated. This implies that we have a *fuzzy* RDD: there is a jump of the probability to receive treatment at $r_i = 7.3$. Hence, we have a two-stage least squares estimation procedure:

$$\begin{aligned} \Delta x_{it} &= \Delta\zeta + \eta\Delta z_{it} + \theta\Delta c_{it} + \Delta\nu_t + \Delta\xi_{it}, && \text{if } r_0 - \delta < r_i < r_0 + \delta, && (1^{\text{st}} \text{ stage}) \\ \Delta y_{it} &= \Delta\alpha + \beta\Delta\hat{x}_{it} + \gamma\Delta c_{it} + \Delta\mu_t + \Delta\epsilon_{it}, && \text{if } r_0 - \delta < r_i < r_0 + \delta, && (2^{\text{nd}} \text{ stage}) \end{aligned} \tag{4.20}$$

Hence, we keep changes in prices and the treatment status with neighbourhoods that have a z-score that is sufficiently close to r_i, *i.e.* with a difference that is smaller than δ. The results do not very much for the chosen bandwidth. Here we discuss results using a recommended bandwidth δ using a procedure proposed by Imbens and Kalyanaraman (2012). Table 4.4 reports the main results. In column (1) we show that when first-differencing the data we find a positive effect of the KW investment programme of $(e^{0.0441} - 1) \times 100 = 4.5\%$. The effect is statistically significant at the 1% level. When employing the fuzzy RDD we find an effect of 3.3%. Although the effect is slightly lower, the effect is not statistically significantly different from the estimate in column (1). Hence, although price trends are somewhat more positive in deprived neighbourhoods, the bias is rather small.

Are the estimates economically meaningful? Yes, a back-of-the-envelop calculation in Koster and Van Ommeren (2019) shows that the total increase in house prices, capturing the increased attractiveness due to the programme, is likely larger than the total investment costs. This means that the return on investment of a place-based policy which makes poor neighbourhoods more attractive is high.

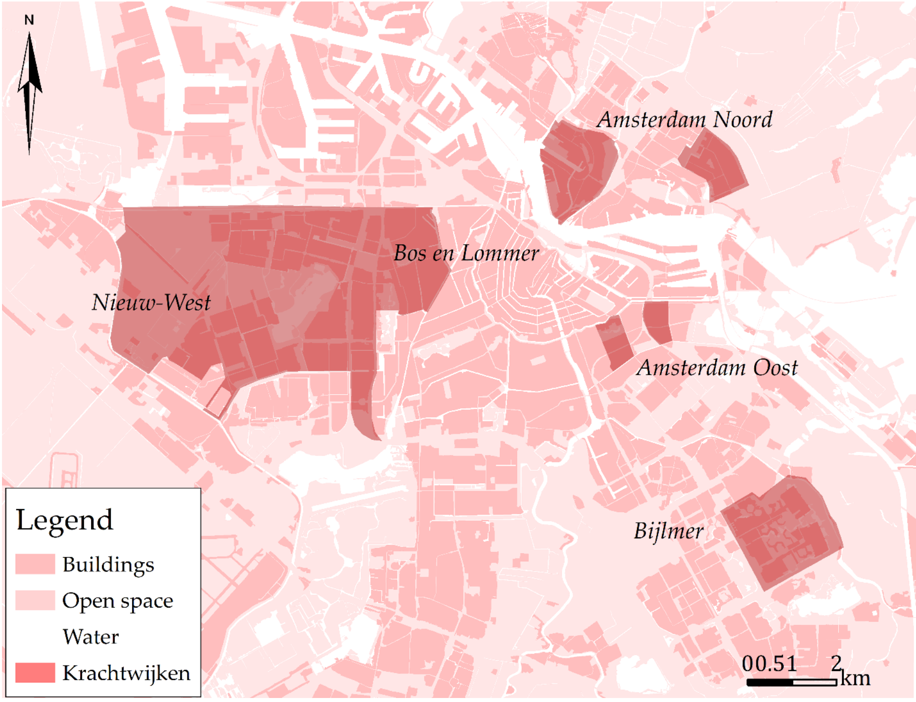


Figure 4.5 – KRACHTWIJKEN IN AMSTERDAM IN 2008

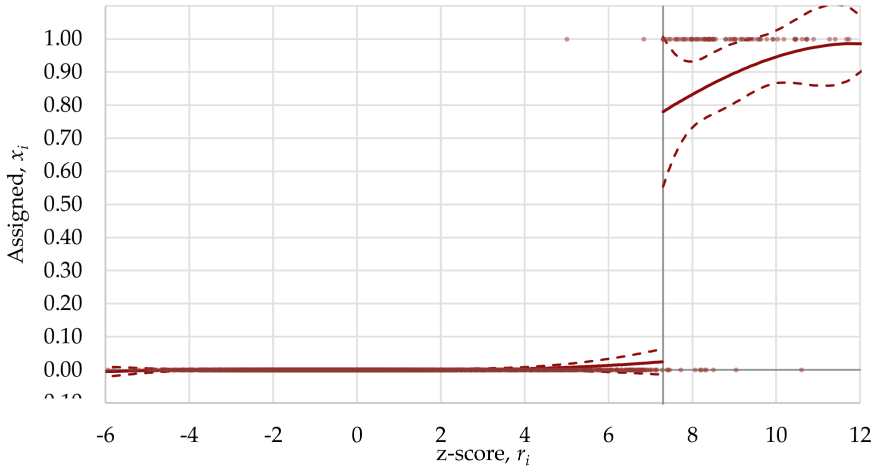


Figure 4.6 – ASSIGNMENT AND THE Z-SCORE

5 STANDARD ERRORS

To conclude whether an effect is statistically significant, it is important to obtain reasonable standard errors. In this Section we discuss three issues that you should bear in mind when estimating the standard errors of the effect: *(i)* heteroscedasticity, *(ii)* clustering and *(iii)* serial correlation.

Table 4.4 – URBAN RENEWAL AND HOUSE PRICES
(Dependent variable: the change in the log of house prices)

	First-differences	+Fuzzy RDD
	(1)	(2)
ΔKW investment	0.0441***	0.0329***
	(0.0114)	(0.0122)
Number of observations	169,664	22,589
R^2-within	0.375	
Kleibergen-Paap F-statistic		5444
Bandwidth, δ		3.383

Notes: We exclude observations within 2.5km of targeted neighbourhoods to avoid picking up spillover effects beyond the neighbourhood boundaries. In column (3) the change in the KW investment is instrumented with the change in the eligibility based on the scoring rule. Standard errors are clustered at the neighbourhood level and in parentheses; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

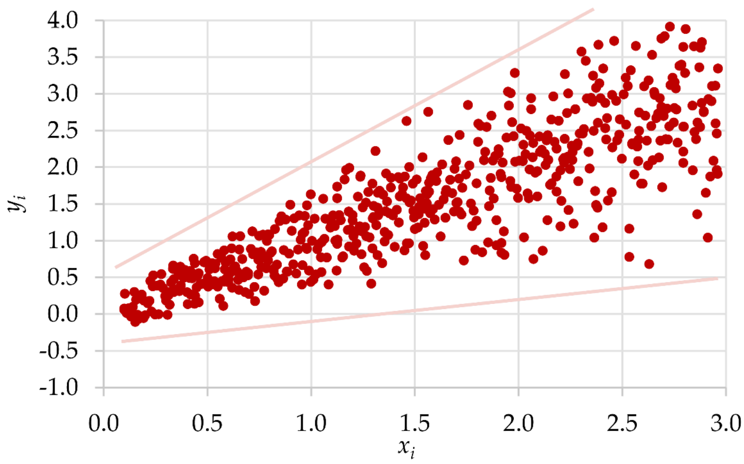


Figure 5.1 – HETEROSCEDASTICITY

5.1 HETEROSCEDASTICITY

What is heteroscedasticity? This is essentially the observation that the variance of your residuals change once the values of x_i changes, as displayed in Figure 5.1. This may be an issue as statistical packages (such as `Stata`) provide standard errors assuming that the variance of the residuals is constant (*i.e.* homoscedasticity) as the default option.

Fortunately, the issue of heteroscedasticity is not a major problem in practice. Just use so-called *robust* standard errors. These standard errors are asymptotically valid, also in the presence of heteroscedasticity (asymptotically means when the number of observations approaches infinity). Hence, as a general rule-of-thumb use standard errors that are robust to the presence of heteroscedasticity. Almost always, these standard errors somewhat exceed the standard errors assuming homoscedasticity (think about 10-20%), so by using robust standard errors, your conclusion will be conservative (i.e. you may conclude that there is no effect, while actually there is an effect.)

The only worry arises when having a a small sample (there are no very clear rules when a sample is small, but think in the order of <1000 observations). Although robust standard errors are asymptotically valid (so when the number of observations goes to infinity), these are biased for small sample (Angrist and Pischke, 2008). Hence, for small samples, the researcher should check whether heteroscedasticity is an issue and whether robust standard errors are smaller from standard errors that require homoscedasticity.

5.2 CLUSTERING

A more problematic issue – particularly for spatial economists who are interested in policies that differ by area and who use individual data – is the assumption in OLS that observations are independent from each other.

Let's consider the case that you are interested in the effects of municipal taxes on individual well-being, to see if higher taxes imply a better provision of public goods. We would then have the following bivariate regression (ignoring any endogeneity issues):

$$y_{ig} = \alpha + \beta x_g + \epsilon_{ig}, \tag{5.1}$$

where y_{ig} is the well-being of an individual i living in municipality g, and x_g is the tax rate that only varies at the municipal level.

However, because x_g only varies at the municipal level, there is likely correlation between residuals of individuals i and j within municipality g:

$$\mathbb{E}[\epsilon_{ig}\epsilon_{jg}] = \rho_\epsilon \sigma_\epsilon^2 > 0, \tag{5.2}$$

where ρ_ϵ is the residual intra-municipality correlation coefficient and σ_ϵ^2 is the residual variance. In other words, the error term has a group structure like $\epsilon_{ig} = \psi_g + \omega_{ig}$ and individual observations are not independent. Moulton (1990) shows that this may lead to much too small standard errors.

We think this is intuitive. Please recall the standard formula for the standard error of $\hat{\beta}$, which assumes that all observations are independent of each other:

$$SE(\hat{\beta}) = \frac{\sigma_\epsilon}{\sqrt{N}} \frac{1}{\sigma_x}, \tag{5.3}$$

where N is the number of observations. As we have remarked above, σ_β becomes smaller once N increases. But also notice that the standard error becomes smaller for a larger standard deviation in the x variable. We find this also intuitive. If you have lots of variation in the x variable, then there is more information in the data to estimate β more precisely (see Angrist and Pischke, 2014, p. 96, and Figure 2.3).

The standard error given intra-municipality correlation is more complicated, and less intuitive. Therefore, let us make the extreme assumption that the intra-municipality correlation is equal to 1, whereas the number of municipalities is equal to G, so where G is much smaller than N. This is extreme, of course, as it implies that everyone within the municipality has the same well-being, but this assumption proves the point. In this case, the standard error can be calculated as:

$$SE(\hat{\beta}) = \frac{\sigma_\epsilon}{\sqrt{G}} \frac{1}{\sigma_x}, \tag{5.4}$$

because there is only variation between municipalities, so the relevant number of observations refers to the number of municipalities. Hence, the standard error is $\sqrt{N/G}$ larger, when erroneously was assumed that the observations are independent. For example, in this example, if you would have 9,000 individual observations and only 90 municipalities, your standard error would be 10 times larger than you think, which most likely would affect your conclusion. Because intra-group correlation is usually non-negligible, but far less than one, it is frequently the case that the standard errors are several times larger when this issue is ignored.

More intuitively. In the context of this example, what is the number of observations to consider: 9000 or 100? We think it is the latter, because taxes vary only at the municipality level. Even if you would increase the number of observations *per municipality*, you would not increase variation in terms of levels of taxation (*i.e.* the standard deviation of taxation, σ_x, remains the same), so the standard errors remain the same. However, if you will be able to obtain more information on about 10 other municipalities, then your standard errors would approximately fall by more than 5% (*i.e.* $1 - \sqrt{100/90}) \times 100$).

Fortunately, there is a relatively straightforward method to adjust standard errors by *clustering the standard errors* at the corresponding level. In the example, one should cluster the standard errors at the municipality level. This almost always lead to *higher* standard errors, implying that you are less likely to reject the null-hypothesis that $\beta = 0$. Please note that clustering should also be applied when using other regression models, such as two-stage least squares, Poisson regression and Logit models. As long as observations are not independent, this should be addressed somehow by adjusting the standard errors.[27]

When clustering, one should keep in mind that errors that allow for correlation within groups also have the problem that for few clusters the correlation ρ_ϵ is underestimated and therefore clustered standard errors are incorrect. But what is 'few'? The answer is not entirely clear, but a general rule-of-thumb is that when the number of clusters exceed 50 one is reasonably safe (Angrist and Pischke, 2008).

5.3 SERIAL CORRELATION

A final issue that is worth mentioning is the presence of serial correlation in panel data. Like in the situation with aggregate variables there is correlation between residuals, but now

[27]What if you include a set of variables that are measured at different (spatial) levels? Recent packages in `Stata`, such as `reghdfe` allow for clustering at different levels.

not within groups, but between individuals over time: $\mathbb{E}[\epsilon_{it}\epsilon_{it-1}] \neq 0$. In case both your dependent variable and main independent variable of interest are at the same level (*e.g.* both individual, or both at the municipality level), then serial correlation will be acknowledged in the calculations of the standard error so you do not have to worry.

Issues are more complicated, when you have individual data and the independent variable of interest is an aggregated variable. Although Angrist and Pischke (2008) acknowledge that this issue is still under study, a quick-and-dirty fix is to cluster at the unit at which you expect serial correlation. Going back to the example of municipal taxes and well-being, say that you have panel data with observations about tax levels for several years, to estimate the following model:

$$y_{igt} = \alpha + \beta x_{gt} + \epsilon_{igt}. \tag{5.5}$$

Note that x_{gt} is measured at the municipality-year level (it varies between municipalities, as well as between years). One recommended solution is to cluster at the municipality level rather than at the municipal-year level to mitigate concerns related to serial correlation.

6 SUMMARY

In this syllabus we have provided a step-by-step guide to undertake a research project. In Section 2 we considered eight steps that should be considered in any project:

1. Formulate your hypotheses;
2. Determine the 'treatment' variable(s) and the 'outcome' variable(s);
3. Think of an identification strategy to identify causal effects;
4. Select samples, discuss measurement error, and provide descriptives;
5. Determine functional form of variables of interest;
6. Think of different issues in estimating standard errors;
7. Estimate the model and interpret the results;
8. Provide robustness checks of the results.

Most of applied econometrics is particularly concerned with Step (3), which is the identification of a causal effect of a treatment variable on the outcome variable. We show in Section 3.1 that causality is in principle not guaranteed because people may self-select into treatment, which we refer to as the *omitted variable bias effect*. We have discussed that randomisation addresses problem. Hence, randomised experiments are in principle the preferred method to *identify a causal effect* of a treatment.

Because randomised experiments are almost never possible in spatial economics, we consider alternative identification strategies to plausibly measure a causal effect. We consider *(i)* adding controls, with an emphasis on adding several types of fixed effects, *(ii)* use instrumental variables, and *(iii)* exploit quasi-experimental variation in the treatment variable, where we have particularly focused on the use of a regression-discontinuity design. Of course, combinations of different identification strategies are possible and preferable if that increases the belief that one can establish causality.

Finally, we paid attention to the estimation of the standard errors, which is step (6) in the step-by-step guide. Heteroscedasticity – implying that the variance changes for different levels of the dependent variable – is relatively straightforward to address in larger samples. Clustering may address, or at least, mitigate concerns related to dependence in the data. Please be aware that statistical significance does not mean that the effect is also economically meaningful. We emphasise that providing a sound economic interpretation of the results is paramount.

We end here with a quote from Angrist and Pischke (2008), which does apply to any research project you may undertake in the future:

> "*Econometrics applied to coherent causal questions, regressions and 2SLS almost always make sense. Your standard errors probably won't be quite right, but they rarely are. Avoid embarrassment by being your own best septic, and especially, DON'T PANIC!.*"

References

Abbott, J. K. and H. A. Klaiber (2011). "An Embarrassment of Riches: Confronting Omitted Variable Bias and Multiscale Capitalization in Hedonic Price Models." In: *Review of Economics and Statistics* 93, pp. 1331–1342 (cit. on p. 21).

Ahlfeldt, G. M. et al. (2015). "The Economics of Density: Evidence from the Berlin Wall". In: *Econometrica* 83.6, pp. 1217–2189 (cit. on p. 28).

Angrist, J. D. and W. N. Evans (1998). "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size". In: *American Economic Review* 88.3, pp. 450–477 (cit. on p. 24).

Angrist, J. D. and A. B. Krueger (2001). "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments". In: *The Journal of Economic Perspectives* 15.4, pp. 69–85 (cit. on p. 25).

Angrist, J. D. and J. Pischke (2014). *Mastering Metrics: The Path from Course to Effect*. Princeton: Princeton University Press (cit. on pp. 10, 12, 14, 37).

Angrist, J., S. Caldwel, and J. Hall (2017). "Uber vs. Taxi: A Driver's Eye View". In: *NBER working paper series, 23891* (cit. on p. 16).

Angrist, J. and J. Pischke (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press (cit. on pp. 1, 6, 14–16, 25, 31, 33, 37–40).

Banerjee, A., S. Barnhardt, and E. Duflo (2018). "Can Iron-Fortified Salt Control Anemia? Evidence from Two Experiments in Rural Bihar". In: *Journal of Development Economics* 133, pp. 127–146 (cit. on p. 17).

Briant, A., P. P. Combes, and M. Lafourcade (2010). "Dots to Boxes: Do the Size and Shape of Spatial Units Jeopardize Economic Geography Estimations?" In: *Journal of Urban Economics* 67.3, pp. 287–302 (cit. on p. 11).

Deaton, A. (2010). "Instruments, Randomization, and Learning about Development". In: *Journal of Economic Literature* 48.2, pp. 424–455 (cit. on pp. 18, 19, 24, 25).

Gibbons, S. and H. Overman (2012). "Mostly Pointless Spatial Econometrics?" In: *Journal of Regional Science* 52.2, pp. 172–191 (cit. on pp. 25, 27).

Glaeser, E. (2008). *Cities, Agglomeration, and Spatial Equilibrium*. Oxford: Oxford University Press (cit. on p. 28).

Graff Zivin, J. and M. Neidell (2012). "The Impact of Pollution on Worker Productivity". In: *American Economic Review* 102.7, pp. 3652–3673 (cit. on p. 25).

Halvorsen, R. and R. Palmquist (1980). "The Interpretation of Dummy Variables in Semilogarithmic Equations". In: *American Economic Review* 70.3, pp. 474–475 (cit. on pp. 12, 13).

Heckman, J. J. (2010). "Comparing IV with Structural Models: What Simple IV can and Cannot Identify". In: *Journal of Econometrics* 156.1, pp. 27–37 (cit. on p. 25).

Imbens, G. W. (2010). "Better LATE than nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)". In: *Journal of Economic Literature,* 48.2, pp. 399–423 (cit. on pp. 25, 27).

Imbens, G. W. and J. D. Angrist (1994). "Identification and Estimation of Local Average Treatment Effects". In: *Econometrica* 62.2, pp. 467–475 (cit. on p. 25).

Imbens, G. and K. Kalyanaraman (2012). “Optimal Bandwidth Choice for the Regression Discontinuity Estimator”. In: *Review of Economic Studies* 79.3, pp. 933–959 (cit. on p. 34).

Imbens, G. and T. Lemieux (2008). “Regression Discontinuity Designs: A Guide to Practice”. In: *Journal of Econometrics* 142.2, pp. 615–635 (cit. on p. 33).

Koster, H. R. A., I. Pasidis, and J. Van Ommeren (2019). “Shopping Externalities and Retail Concentration: Evidence from the Netherlands”. In: *Journal of Urban Economics* 114, p. 103194 (cit. on p. 22).

Koster, H. R. A. and J. Van Ommeren (2019). “Place-based Policies and the Housing Market”. In: *Review of Economics and Statistics* 101.3, pp. 1–15 (cit. on pp. 33, 34).

Koster, H. R. A., J. Van Ommeren, and P. Rietveld (2014). “Agglomeration Economies and Productivity: A Structural Estimation Approach using Commercial Rents”. In: *Economica* 81.321, pp. 63–85 (cit. on p. 28).

Lee, D. and T. Lemieux (2010). “Regression Discontinuity Designs in Economics”. In: *Journal of Economic Literature* 48.2, pp. 281–355 (cit. on p. 33).

Lee, M. J. (2000). “Median Treatment Effect in Randomized Trials.” In: *Journal of the Royal Statistical Society B* 62.3, pp. 595–604 (cit. on p. 15).

McCloskey, D. N. and S. T. Ziliak (1996). “The Standard Error of Regressions”. In: *Journal of Economic Literature* 34.1, pp. 97–114 (cit. on p. 12).

Morgan, S. L. and C. Winship (2015). *Counterfactuals and Causal Inference: Methods and Principles for Social Research.* 2nd Editio. New York: Cambridge University Press (cit. on p. A2).

Moulton, B. (1990). “An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units”. In: *Review of Economics and Statistics* 72.2, pp. 334–338 (cit. on p. 37).

Nickson, C. (2015). *Randomised Control Trials* (cit. on pp. 17, 18).

Pearl, J. (2009). *Causality.* 2nd. Cambridge: Cambridge University Press (cit. on p. 18).

Reiersol, O. (1941). “Confluence Analysis by Means of Lag Moments and Other Methods of Confluence Analysis”. In: *Econometrica* 9, pp. 1–24 (cit. on p. 22).

Sager, L. (2019). “Estimating the Effect of Air Pollution on Road Safety using Atmospheric Temperature Inversions”. In: *Journal of Environmental Economics and Management* 98, p. 102250 (cit. on p. 25).

Stock, J. H. and F. Trebbi (2003). “Retrospectives: Who Invented Instrumental Variable Regression?” In: *Journal of Economic Perspectives,* 17.3, pp. 177–194 (cit. on p. 23).

Van Ommeren, J. and D. Wentink (2012). “The (Hidden) Cost of Employer Parking Policies”. In: *International Economic Review* 53.3, pp. 965–978 (cit. on pp. A3, A4).

Wright, P. (1928). *The Tariff on Animal and Vegetable Oils.* New York: MacMillan (cit. on p. 22).

Yatchew, A. (2003). *Semiparametric Regression for the Applied Econometrician.* Cambridge: Cambridge University Press (cit. on p. 10).

Ziliak, S. and D. N. McCloskey (2008). *The Cult of Statistical Significance: How the Standard Error Costs us Jobs, Hustice, and Lives.* Ann Arbor MI: University of Michigan Press (cit. on p. 12).

APPENDIX

A.1 RANDOM MEASUREMENT ERROR

Let's return to measurement error in the dependent variable. Suppose you do not observe y_i, but you observe y_i^*, where $y_i^* = y_i + u_i$, where u_i denotes random measurement error. This implies that you aim to run the following regression:

$$y_i = \beta x_i + \gamma c_i + \epsilon_i, \tag{A.1}$$

but you do estimate the following regression:

$$\begin{aligned} y_i^* - u_i &= \beta x_i + \gamma c_i + \epsilon_i, \\ y_i^* &= \beta x_i + \gamma c_i + \omega_i, \end{aligned} \tag{A.2}$$

where $\omega_i = \epsilon_i + u_i$. In other words, the variance of the error term ω_i is larger (because $Var(\omega) = Var(\epsilon) + Var(u)$), but otherwise (A.2) is essentially the same as (A.1). The only consequence of a larger variance is that the predictive error will be larger.

Now suppose that there is measurement error in x_i, so that you observe $x_i^* = x_i + u_i$. In this case you aim to estimate:

$$y_i = \beta x_i + \gamma c_i + \epsilon_i \tag{A.3}$$

but you do estimate:

$$\begin{aligned} y_i &= \beta(x_i^* - u_i) + \gamma c_i + \epsilon_i. \\ &= \beta x_i^* + \gamma c_i + \epsilon_i - \beta u_i. \end{aligned} \tag{A.4}$$

Now, by assumption, u_i is positively correlated to x_i^*, so $\epsilon_i - \beta u_i$ is correlated to x_i^*, which is not in line with the assumption of OLS.

More intuitively, consider the case that u_i is very large compared to x_i. Because x_i^* is then a random variable, $\beta \to 0$. Hence, in general, random measurement error in x_i leads to a bias towards zero of β.

A.2 LOG-LINEAR EQUATIONS AND ELASTICITIES

In Section 2.5 we claim that the elasticity is given by:

$$\frac{\partial y_i}{\partial x_i}\frac{x_i}{y_i} = \frac{\partial \log y_i}{\partial \log x_i}. \tag{A.5}$$

To show this, let's assume a standard log-linear regression equation:

$$\log y_i = \beta \log x_i + \gamma c_i + \epsilon_i, \tag{A.6}$$

where y_i is the dependent variable measures for agent i,x_i is a continuous variable of interest, and c_i are control variables, while ϵ_i is an unobserved effect. Note that $\frac{\partial \log y_i}{\partial \log x_i} = \beta$.

By taking the exponent of (A.6), we can write:

$$y_i = e^{\beta \log x_i + \gamma c_i + \epsilon_i}. \tag{A.7}$$

Let's now take the derivative of this function with respect to x_i and use the chain rule of differentiation:

$$\frac{\partial y_i}{\partial x_i} = e^{\beta \log x_i + \gamma c_i + \epsilon_i} \times \frac{1}{x_i}\beta. \tag{A.8}$$

From (A.7) we know that $y_i = e^{\beta \log x_i + \gamma c_i + \epsilon_i}$, so that (A.8) can be simplified to:

$$\frac{\partial y_i}{\partial x_i} = y_i \times \beta \frac{1}{x_i}. \tag{A.9}$$

By re-arranging and replacing β by $\frac{\partial \log y_i}{\partial \log x_i}$, we have:

$$\frac{\partial y_i}{\partial x_i}\frac{x_i}{y_i} = \frac{\partial \log y_i}{\partial \log x_i}. \tag{A.10}$$

Hence, the regression coefficient β in a log-linear equation captures an *elasticity*.

A.3 AVERAGE TREATMENT EFFECTS

Please note the difference between the *average treatment effect* and the *average treatment effect on the treated*. With randomisation, the average treatment effect is the same as the average treatment effect on the treated because there is a randomised selection of people based on the whole population.

Let's consider β be the individual-causal effect of the treatment and π be the proportion of population that takes the treatment (Morgan and Winship, 2015). Let's make a distinction between the average treatment effect on the treated ($\mathbb{E}[\beta|x_i = 1]$) and the average treatment effect $\mathbb{E}[\beta]$. The *average treatment effect on the treated* is defined as:

$$\begin{aligned}\mathbb{E}[\beta|x_i = 1] &= \mathbb{E}[y_{1i} - y_{0i}|x_i = 1] \\ &= \mathbb{E}[y_{1i}|x_i = 1] - \mathbb{E}[y_{0i}|x_i = 1],\end{aligned} \tag{A.11}$$

while the *average treatment effect* is defined as:

$$\begin{aligned}\mathbb{E}[\beta] = \pi\mathbb{E}[y_{1i}|x_i = 1] + (1-\pi)\mathbb{E}[y_{1i}|x_i = 0] \\ - \pi\mathbb{E}[y_{0i}|x_i = 1] + (1-\pi)\mathbb{E}[y_{0i}|x_i = 0]\end{aligned} \tag{A.12}$$

Randomisation implies that $\mathbb{E}[y_{0i}|x_i = 1] = \mathbb{E}[y_{0i}|x_i = 0]$ and $\mathbb{E}[y_{1i}|x_i = 1] = \mathbb{E}[y_{1i}|x_i = 0]$, implying that:

- The baseline outcome in the treatment group equals the baseline outcome in the control group;
- People in the control group would experience the same increase in the outcome variable *if they were to be treated.*

Hence, in settings where there is no randomisation, one should notice that, even after addressing the selection bias, the average treatment effect on the population may be different from the average treatment effect on the treated.

A.4 INSTRUMENTAL VARIABLES AND SUPPLY AND DEMAND FUNCTIONS

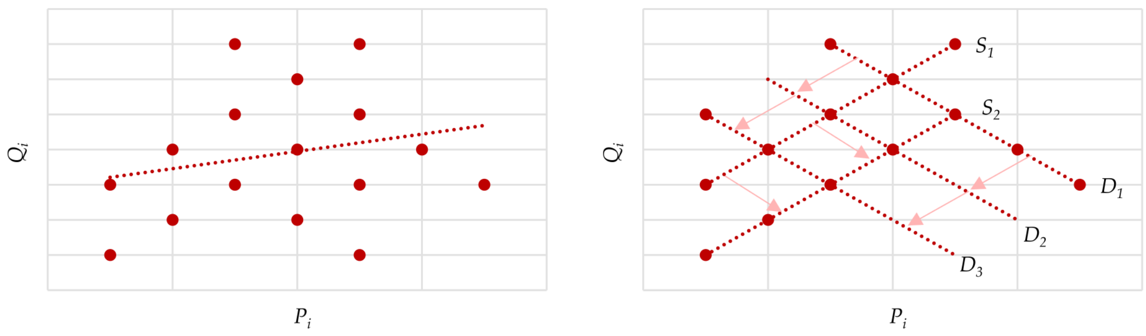
The use of IV is closely related to the estimation of (inverse) *demand and supply functions*, and therefore closely related to economic theory. Recall that an inverse demand function describes the (negative) effect of a price on the quantity demanded, whereas the inverse supply function describes the effect of the price on the quantity supplied, which is usually positive, because supplying more of a good is more costly.

We start from the notion that we do *not* observe the demand and supply function, but we only observe prices and quantities which are in the outcome of the interaction of the demand and supply functions, *i.e.* the equilibrium. Consequently, in general, if you regress the quantity of a variable on its price, you will neither get a demand nor a supply function, but a mixture of both functions. This is illustrated in Figure A1a. In this example, regression of quantity on prices gives you a positive relationship, but does not provide any information about the underlying functions.

IV approaches are very useful to disentangle demand and supply functions from each other. We illustrate this in Figure A1b. When you aim to estimate demand functions using IV, one employs an instrumental variable that shifts the supply function, but *not* the demand function. Note in the figure that when you shift the supply function downwards (from S_1 to S_2), one gets information about the slope of the demand function. Similarly, when you aim to estimate a supply function, one has to use an instrument which shifts the demand function, but not the supply function.

Note in the figure that when you shift the demand function downwards (from D_1 to D_2 and then to D_3), one obtains information on the slope of the supply functions. Please keep in mind that many variables determine both demand and supply and are therefore not useful as instruments, but only as control variables.

Application A.1 — Estimating the demand and supply for parking. Van Ommeren and Wentink (2012) are interested to estimate the employers' demand and supply function of parking that is offered to employees. Knowledge of both functions are key to determine the welfare loss of distortionary income taxation. In almost all countries around the world, employers offer parking for free to workers. One reason they do that is that workers do not pay any income tax on receiving a free parking space. In contrast, wages



(a) OBSERVATIONS ON PRICES AND QUANTITIES (b) INVERSE DEMAND AND SUPPLY FUNCTIONS

Figure A1 – DEMAND AND SUPPLY

which are taxed as income, so employers tend to offer more parking spaces to workers than they would if parking spaces would be taxed as a fringe benefit. As a consequence, one expects welfare losses in the employer parking market.

In Van Ommeren and Wentink (2012), the researchers used information about office space rented by employers, including information about the number of parking spaces rented, and the annual rent paid for each parking space. The age of the building rented by the employer was used as an instrument to estimate the parking demand function. The idea is that building age still has an effect on the costs of having a parking space (in old buildings, *e.g.* constructed before the car was invented, it is very expensive to add a parking space), so it affects the supply function, whereas the parking demand by employees should not depend on the age of the building directly.

Table A1 reports the main results for parking demand. The coefficient in column (3) shows that the price elasticity of demand, *i.e.* the effect of log price on log parking spaces, is about -0.60 with a standard error of about 0.15.

The size of the office rented by the employer was used as an instrument to estimate the parking supply function. The idea is that larger offices usually have more employees which increases the demand for parking, whereas the cost per parking space should not depend directly on number of employees. Van Ommeren and Wentink (2012) find that the parking supply function is essentially horizontal, *i.e.* fully elastic, so the price/cost of parking does not depend on the quantity supplied. For offices, mainly in the suburbs of large cities, this makes sense, as having a larger parking lot usually means only using more land, so the costs per parking space are constant; it mainly reflects the price of land.

When estimating the demand as well as the supply function, the study used postcode fixed effects as control variables. The idea is here that, presumably, the demand for parking will vary over space (*e.g.* close to railway station it will be less than in the suburbs), but so will supply costs (*e.g.* in the city centre, to convert land to parking will

Table A1 – DEMAND FOR PARKING

(Dependent variable: the log of number of parking spaces)

	2SLS Baseline specification	2SLS Add municipality f.e.	2SLS Add area f.e.
	(1)	(2)	(3)
Price *(log)*	-0.824***	-0.918***	-0.612***
	(0.146)	(0.186)	(0.150)
Controls (3)	Yes	Yes	Yes
Year fixed effects (7)	Yes	Yes	Yes
Municipality fixed effects (129)	No	Yes	Yes
Area fixed effects (516)	No	No	Yes
Number of observations	394,389	394,389	207,242
Kleibergen-Paap F-statistic	12.46	12.81	24.33

Notes: The instrument is the construction year of the property. Controls include the log of floor area, the log of distance to the nearest highway ramp and the log of the distance to the nearest railway station. Robust standard errors in parentheses; *** $p < 0.01$, ** $p < 0.5$, * $p < 0.10$.

be more expensive than in the suburbs). Adding these controls did not change the results much, but makes the instruments stronger.

It appears that the welfare losses of providing free parking to employees are roughly equal to about 10% of the parking building costs. Hence, loosely speaking, about 10% of the parking building costs are wasted from an economic point of view. This excludes any additional environmental and congestion costs because of increased driving.