Answering Causal Questions Using Observational Data – Achievements of the 2021 Nobel Laureates in Economics*

Zoltán Hermann – Hedvig Horváth – Attila Lindner

In the 1980s and 1990s, the field of labour economics was at the forefront of combining economic theory, high-level econometric methods and new data sources. The 2021 Nobel Prize in Economics was awarded to David Card, Joshua Angrist and Guido Imbens, who played key roles in this research, for their "empirical contributions to labour economics" (Card) and "methodological contributions to the analysis of causality relationships" (Angrist and Imbens), according to the citation. These methodological innovations are now applied in all fields of economics and in many other social science disciplines. Credible empirical analysis has transformed economics from a rather theoretical discipline into a discipline dominated by empirical results, where even the most fundamental theories can be rejected on the basis of empirical results. In this paper, we review the main methodological achievements of this period, also known as the credibility revolution, illustrated by some economic applications.

Journal of Economic Literature (JEL) codes: J2, J6, I21, I26, C2, C26

Keywords: natural experiment, minimum wage, returns to education, instrumental variables, difference-in-differences, regression discontinuity design

1. Introduction

Between 1985 and 2001, the editor-in-chief of one of the most prestigious journals in economics, the American Economic Review (AER), was Orley Ashenfelter, a professor at Princeton University, PhD supervisor and frequent co-author of two of the three 2021 Nobel laureates in Economics. At the beginning of his tenure, Prof Ashenfelter was perplexed about why a study published in the New England Journal of Medicine (NEJM) (Hearst et al. 1986), analysing the impact of

* The papers in this issue contain the views of the authors which are not necessarily the same as the official views of the Magyar Nemzeti Bank.

Zoltán Hermann is a Research Associate at the Institute of Economics of the Centre for Economic and Regional Research and Associate Professor at Corvinus University of Budapest. Email: hermann.zoltan@krtk.hu
Hedvig Horváth is Assistant Professor at University College London. Email: h.horvath@ucl.ac.uk
Attila Lindner is Associate Professor at University College London. Email: a.lindner@ucl.ac.uk

This essay greatly benefitted from the popular and scientific background materials for the 2021 Nobel Prize in Economics, published by the *Royal Swedish Academy of Sciences* (*2021a,b*).

The Hungarian manuscript was received on 19 February 2022.

DOI: https://doi.org/10.33893/FER.21.1.141

military service on subsequent mortality, comparing drafted and non-drafted age groups during the Vietnam War, had a greater impact among economists than papers in the AER. Around that time, a PhD student of Prof. Ashenfelter, Joshua Angrist, began looking at the labour market prospects of Vietnam veterans, while a colleague at Princeton University, Alan Krueger, subscribed to the NEJM to explore the reasons for the success of its articles. Working with David Card, Prof Krueger found that all NEJM articles had a clear methodological framework, a so-called 'research design', which renders the analysis transparent and the results credible (*Card 2021*).[1] Inspired by this idea they started to work on applying transparent statistical methods in economic research, in order to have a greater impact on public policy and consequently raise the scientific quality of the discipline. The concept of '*research design*' was thus introduced in economics and thanks to the revolutionary work of the laureates, their intellectual mentors and co-authors, it has become the new standard of professional rigour.

But what does this concept, previously unknown in economics, mean? We demonstrate this below with a selection of examples from the laureates' works, which illustrate both the methodological revolution taking place in the discipline and how this methodological revolution has advanced our knowledge of long-debated economic issues.

2. Answering causal questions using observational data: Natural experiments and research designs

The fundamental aim of most applied research is to explore causal links. In life sciences, the most common method is randomised controlled trials (RCT). For example, when testing new drugs, patients are randomly assigned to a treatment group and a control group, and the effect is measured by comparing the outcomes of the two groups. On average, the various characteristics of the members of the two randomly selected groups are the same on average – therefore, the difference between the two groups after treatment cannot be explained by anything else than the effect of the drug. The opportunity for this type of experiment is very limited in the social sciences. Still, in many cases we can find comparable treatment and control groups, in which – although the individuals observed are not randomly assigned to groups – the resulting situation is similar. During the Vietnam War, conscription was performed by way of draft lotteries: therefore, it was randomly decided who was drafted and who stayed at home. Public policy reforms can also create similar situations. People born in a given month have a different compulsory

[1] Zipperer, B.: *Equitable Growth in Conversation: An interview with David Card and Alan Krueger*. April, 2016. https://davidcard.berkeley.edu/interviews/interview%20with%20Card%20and%20Krueger.pdf. Downloaded: 22 February 2022.

schooling or pension age than those born in the next month. The minimum wage for people living in a particular US municipality is different from that in a similar municipality a few kilometres away, just because it belongs to another state. These situations are called natural experiments or quasi-experiments. If we can identify and leverage them in the analysis, we can convincingly identify causal effects without randomised controlled trials (RCTs). David Card, Joshua Angrist and Guido Imbens were crucial in the development and dissemination of this approach. Their work involves numerous applications, as well as the theoretical development of analytical methods.[2] This meant a radical change of approach from the previous practice in the social sciences, which tried to verify causal effects based on the statistical correlation of two phenomena. In the following, three examples that exploit various quasi-experimental situations and research designs are presented. All three cases led to results with important theoretical implications as well.

2.1. What we have learned about how the labour market works

In the 1980s and 1990s, labour economics was at the forefront of combining economic theory, high-level econometric methods and new data sources.[3] One excellent example is the economic debate on the minimum wage, to which a number of famous economists contributed (e.g. George Stigler and Milton Friedman, who was later awarded the Nobel Prize).

The minimum wage dispute was basically about what constituted the right theory of labour markets. According to the standard theory, the market for low-paid workers is competitive, and therefore wage increases lead to a loss in employment (*Stigler 1946*). In a (perfectly) competitive market, supply and demand are in equilibrium. With the introduction of a minimum wage, this equilibrium breaks down. Higher wages lead to a lower demand for labour, as labour becomes more expensive, and a higher supply of labour, as more people want to work for higher wages. As a result, minimum wages lead to a significant employment loss and unemployment.

In contrast to the standard theory, many have argued that the minimum wage does not lead to significant employment losses. For example, *Richard Lester* (*1947*) argues that the standard theory incorrectly assumes that managers apply the mathematical models of profit-maximising firms. This was backed up by surveys of company managers asking about what determines the number of people they employ. Interestingly, the most important factor for managers was not the cost of labour, but rather demand for their products.

[2] The methods applied are illustrated by examples; therefore, we rely in the following on the articles of Card and Angrist. Imbens was instrumental in developing the methodology, which is not discussed in detail here.
[3] *Interview with David Card*. Federal Reserve Bank of Minneapolis, 1 December 2006 https://www.minneapolisfed.org/article/2006/interview-with-david-card. Downloaded: 22 February 2022.

However, Lester's questionnaire approach was subject to serious criticism. Nobel laureate *Milton Friedman* (*1953*) argued that the laws of competition enforce profit maximisation even if managers do not literally solve mathematical equations. According to the argument, only managers who can figure out how to run a company efficiently and maximise profits will be successful because of competition, even if they cannot express it as clearly as mathematical equations. Milton Friedman used the example of a pool player to illustrate the argument: a good pool player cannot describe exactly where to hit a billiard ball using mathematical equations, yet if we want to understand the path of the billiard ball, we can only do so using mathematical equations. According to the argument, economic models should be tested on the basis of their predictions (e.g. whether the minimum wage reduces employment) and not on the basis of their assumptions (whether firms' demand for labour is determined by profit maximisation).

The debate on the minimum wage has therefore focused on testing the prediction of the standard theory. Early empirical results showed that the minimum wage significantly reduced employment (*Brown et al. 1982*). For this reason, by the 1980s and 1990s, the vast majority of economists believed that the minimum wage was doing more harm than good. A revolutionary study by *David Card and Alan Krueger* (*1994*) broke this consensus.

The early empirical results were mostly based on time-series analysis: the correlation between the minimum wage rate and unemployment or employment over time. The main problem with this approach is that the minimum wage rate is not simply the result of chance, but of a complex political decision, which can depend on many other factors. For example, if the minimum wage is raised more frequently under a left-wing political administration or during recessions, it is possible that the correlation of the minimum wage with employment reflects the impact of other factors, not just that of the minimum wage increase alone.

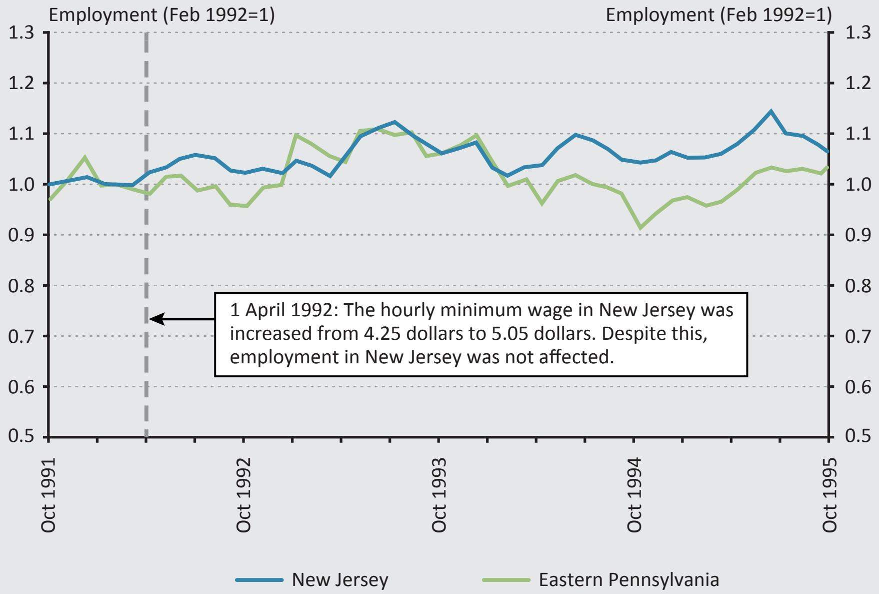
To solve such problems, *Card and Krueger* (*1994*) applied the so-called '*difference-in-differences'* estimation method to empirically analyse the effects of the minimum wage. The authors took advantage of the fact that the minimum wage in the United States is (partly) set at the state level. The state of New Jersey raised the minimum wage in April 1992, while neighbouring Pennsylvania did not. As economic conditions in New Jersey and in the eastern counties of Pennsylvania are very similar, in many respects a quasi-experimental situation arose where the minimum wage is raised in one area (treatment group), while there is no increase in another very similar area (control group).

The difference-in-differences estimation method compares the change in employment of the treatment group (New Jersey) with the change in employment of the control group (Eastern Pennsylvania). Since the economy of Eastern Pennsylvania is very similar, using the control group allows us to eliminate the effects of economic factors that would have occurred in New Jersey without the minimum wage increase, and thus provide a more accurate estimate of the causal effect of the minimum wage.

Card and Krueger collected data on employment in fast food restaurants in New Jersey and Eastern Pennsylvania both before and after the New Jersey minimum wage increase. They showed that employment in the treatment group increased relative to the control group, i.e. the minimum wage actually increased employment rather than decreased it. The new analysis, based on a more credible research design, contradicted the previous estimates in the literature that was largely based on time-series analyses. Furthermore, the finding of higher employment was hard to reconcile with the prevailing economic theory.

The difference-in-differences method has become one of the most widely used research designs in applied economics. The advantage of the method is that treatment and control group situations often arise because public policies do not affect everyone to the same extent. The empirical method is based on the assumption that the change in the control and treatment groups would be the same if the treatment group had not received the treatment. This is the so-called '*parallel trends*' assumption, which is often tested by comparing trends before the reform was introduced. *Card and Krueger* (*2000*) is a good example of this. The study uses administrative data to analyse employment trends in New Jersey and Eastern Pennsylvania between October 1991 (6 months before the minimum wage increase) and October 1995 (42 months after the minimum wage increase). The results are illustrated in *Figure 1*.

Figure 1
Impact of the New Jersey minimum wage increase



Source: Royal Swedish Academy of Sciences (2021)

It is clear that employment trends in New Jersey and Eastern Pennsylvania were very similar in the six months prior to the minimum wage increase, while employment in New Jersey increased slightly compared to Eastern Pennsylvania after 1 April 1992. The graph also illustrates that the initial positive effects do not turn negative in the long run: Employment in New Jersey was still higher than in Eastern Pennsylvania three years after the minimum wage increase.

The difference-in-differences method is now often applied in a more developed form, where there is no single control group that resembled the treatment group before the reform, but there are several control observations (country, member state, company, school, etc.), whose appropriately weighted combination fits well ('in parallel') with the pre-reform trend of the treatment group. Guido Imbens played a pioneering role in the development and statistical fine-tuning of this '*synthetic control difference in differences*' method (*Arkhangelsky et al. 2021; Athey – Imbens 2006, 2022*).

Although *Card and Krueger's 1994* analysis has been strongly criticised, subsequent research has supported the main findings of the original study.[4] Examining the effects of 138 large minimum wage increases, *Cengiz et al* (*2019*) found that minimum wage increases have no negative impact on the employment of low earners. After summarising the results of 37 studies on the minimum wage, *Wolfson and Belman* (*2019*) concluded that the impact of the minimum wage on employment is very small and statistically indistinguishable from zero. Nevertheless, it is worth highlighting that these results should always be interpreted in the given context, with the given minimum wage level in mind. Neither *Card and Krueger* (*1994*) nor subsequent studies claim that the minimum wage can be raised beyond all limits without any employment losses.

It is important to note that, with its new, credible methodology, empirical analysis was able to reject previously highly consensual economic theories. The result itself launched an important new direction in research, leading to a more realistic description of the labour market. These new models take into account that most companies not only passively accept market wages, but have some wage-setting power. As a consequence, a certain level of minimum wage can increase employment (*Burdett – Mortensen 1998*; *Manning 2003*). Moreover, Card and Krueger's research on the minimum wage paved the way for the so-called '*credibility revolution*', where fundamental economic and social policy issues should be decided through 'credible' empirical analyses, rather than theoretical debates. As a consequence, the discipline of economics has increasingly shifted in an empirical direction, increasing its scientific validity and impact on public policy.

2.2. Credibility revolution in education research

In educational research, the methodological shift in economics linked to the names of the Nobel laureates and their fellow authors was particularly revolutionary. Prior to the 1980s, mostly sociologists, psychologists and researchers from other social science disciplines were doing research on education policy issues. However, the rigorous but transparent design-based statistical methods rooted in quasi-experimental situations have since attracted many economists to this topic to answer important, policy-relevant questions on individual labour markets. This methodological renewal, the *'credibility revolution'* (*Angrist – Pischke 2010*), has brought education economics to a point of inflection regarding both the number of people working in the field and their research methodology. In this subsection, we review the pioneering work of the Nobel laureates in this field, focusing on

[4] Many respected economists were outraged by *Card and Krueger's* (*1994*) analysis, as they saw it contradicting fundamental economic theories. For example, James Buchanan, who was awarded the Nobel Prize in Economics in 1986, argued that the analysis of David Card and Alan Krueger was unscientific and that such an article had no place in leading economics journals such as the American Economic Review.

two main issues: the returns to education in the labour market and the impact of schools/elite schools on student achievement.

2.2.1. The returns to education[5]

In economic models, wages are usually closely related to workers' productivity. In addition to traditional economic topics, economists turned their attention to education when they recognized that education could play a crucial role in determining workers' productivity. Education is defined as an *investment* in human capital that has a *return*. A fundamental and much debated question in economics is how large this return is and how to measure it. Since the work of *Jacob Mincer* (*1958, 1974*), estimation of the human capital earnings function from observational data, also known as the Mincer function after the pioneering labour economist, has become a standard in this field. The typical form of the equation is:

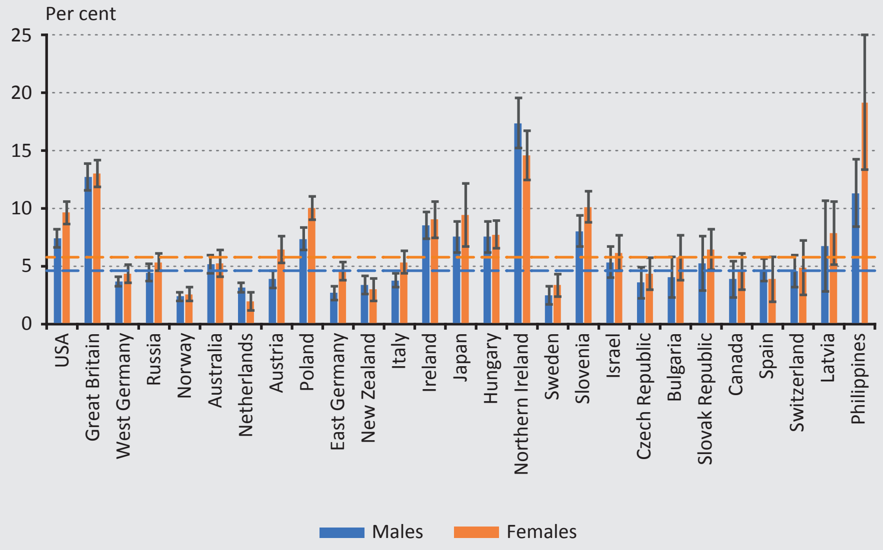
$$\log y = a + bS + cX + dX^2 + e,$$

where $\log y$ is the logarithm of earnings, S is education (measured in terms of educational attainment or years of schooling), X is labour market experience, i.e. the number of years worked,[6] e is the residual, which, in the statistically estimated form of the equation, includes the additional control variables (e.g. gender, marital status, union membership, etc.). Mincer derived this equation from an individual educational choice model and it is usually estimated by applying the ordinary least squares (OLS) method. The magnitude of the empirical relationship (b) has been measured in numerous contexts, and even though various forms of it have been fitted to the data – for example, the set of control variables or the functional form of schooling or potential labour market experience, the results are very robust: one additional year of education implies approx. 4–10 per cent higher subsequent earnings (*Figure 2*).

[5] This subsection draws heavily on *Card* (*1999*).

[6] The labour market experience is actually difficult to observe, so it is usually approximated by *potential* labour market experience: $A - c - S$, where A is the age of the individual and c is age at the start of compulsory schooling.

Figure 2
Returns to education in different countries, by gender, estimated from the Mincer equation



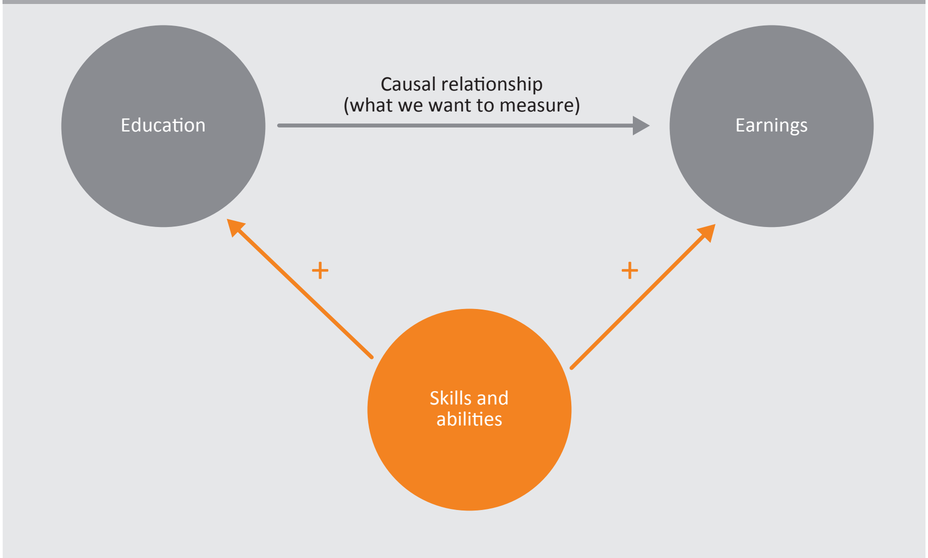
Note: The grey capped spikes represent the 95% confidence intervals generated by robust standard errors. The horizontal dashed lines indicate the estimated return for all countries pooled. In addition to education, the estimated equation also includes the following variables as controls: year dummies, union membership, marital status, age, age squared, and country dummies in the regression for all countries pooled.

Source: Trostel et al. (2002): Table 2

This finding is often interpreted – misleadingly – as ‘an extra year of education *increases* later earnings by about 4–10 per cent’. The word ‘increases’ may give us the impression that higher education causes higher earnings. Partly, perhaps, because there is a good chance that this causal relationship is what we are interested in. For example, we want to know whether our child should go to university. Or, as public policy advisors, we need to answer the question of whether it is worthwhile to support the expansion of secondary/higher education. However, the results of Mincer type regressions estimated by OLS are not suitable for drawing such conclusions on causation, for several reasons.

First, because the skills and abilities of individuals are difficult to observe: consequently, we usually don’t have data on this. However, children with higher abilities are more likely to have better academic outcomes, and thus are more likely to study longer (e.g. go to university). Better skills/abilities, on the other hand, have a direct impact on earnings, even without higher education.[7] Therefore, when we find a positive relationship between education and earnings without observing skills/abilities, we cannot be sure whether we are really seeing an effect of education, or rather of better skills. This is illustrated in *Figure 3*.

Figure 3
Omitted variables (abilities)



The coefficient estimated from the Mincer equation, $\hat{b}$, combines the direct causal effect of education and the indirect effect of better ability (*ability bias*). Since we expect both to be positive, we say $\hat{b}$ is likely to overestimate the pure causal effect of education, which we are investigating.

The second problem with the causal interpretation of the Mincerian education coefficient is that if there are large differences between individuals in terms of their return to an extra year of education. It is possible that individuals for whom this return is particularly high may choose ('self-select') to complete a higher level of education. In other words, reverse causality is conceivable, which would also strengthen the positive correlation between education and earnings. This again suggests that $\hat{b}$ estimated by OLS might be biased upwards.

[7] Such variables, for which we either have no data or 'forgot' to include them in our regression model, but which are related to both our main independent variable (education) and our outcome variable (earnings), are called *omitted variables*.

Some of the main contributions of the 2021 Nobel laureates in economics were to highlight that correlation does not necessarily imply causation and to develop methods to establish causal relationships using observational data. To do this, they exploited – using the phrase coined by the laureates and their co-authors – 'quasi experimental' situations similar to random experiments, but occurring naturally in real life. In addition to the 'difference-in-differences' method described above, the other and perhaps best-known of these quasi-experimental research designs is the so-called *instrumental variables (IV)* method, which has been frequently used since the 1990s to measure the returns to education.

The essence of the IV framework is illustrated by the blue shapes in *Figure 4*, using the example of *Angrist and Krueger* (*1992*). The research design of the study exploits that during the Vietnam War, those who enrolled in college in the United States could avoid being drafted. However, conscription was performed by assigning a *random draft lottery* number to each military-age man based on their date of birth, who then were drafted in ascending order of their lottery numbers according to the military's personnel requirements.[8] In this particular regulatory environment, the lottery number used for conscription can be considered as an instrumental variable which

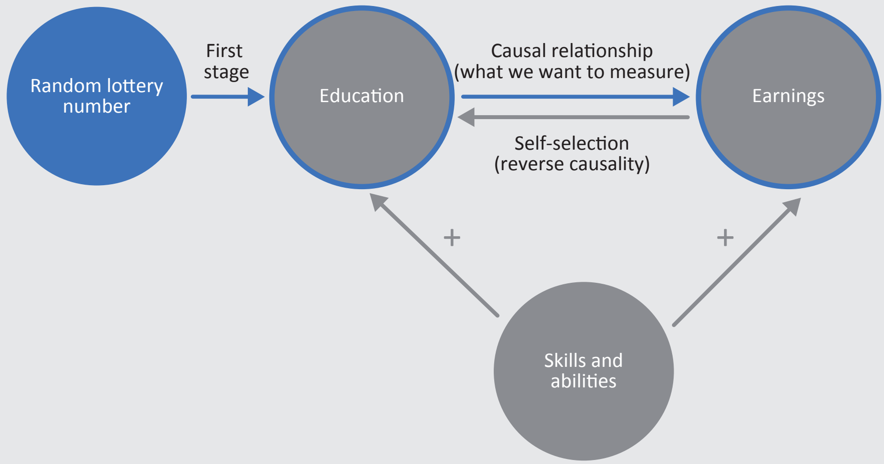
1. may determine educational attainment – as university enrolment was a way of avoiding conscription, those with a lower lottery number (due to being drafted earlier) were more likely to go to university by pure chance,

2. does not affect earnings through any channel other than education, e.g. ability – again, because the draft lottery number is randomly assigned (based on date of birth, regardless of abilities).

In other words, a change in the instrument (a lower lottery number) increases education without changing abilities. As a result, the potential differences in earnings between individuals can only be attributed to differences in the lottery number and hence in education, and not to differences in abilities. In this case, we can be certain that the difference in earnings is *caused* by differences in education, since abilities do not differ as a result of the lottery number.

[8] The same conscription scheme was used by *Hearst et al.* (*1986*), mentioned at the beginning of this paper, and by *Angrist* (*1990*), inspired by them.

Figure 4
Instrumental variables (draft lottery number) mechanism



In general, one major advantage of the IV method is that a 'good' instrument can purge the OLS estimate from both omitted variable bias (stemming from skills in the case of the returns to education) and bias from reverse causality (self-selection).[9] The drawback of the method, however, is that it is very difficult to find a 'good' instrument, because it would have to meet the following two conditions, to put it a bit more rigorously:

1. *Relevance*: the instrument should strongly correlate with the endogenous variable (in our case, education) whose causal effect we are interested in but applying OLS estimation would give a biased estimate due to the omitted variable/reversed causality problems described above.

2. *Exclusion restriction*: the instrument should not be related through any other channel to the outcome variable (in our example, earnings).

To ensure relevance, we use the so-called *first stage* regression, which regresses the endogenous variable (education) on the instrument(s) and other control variables (e.g. potential labour market experience). In this regression, we can determine the relevance of the instrument(s), i.e. whether the instrument is *strong,* in light of the *F*-statistic testing for the joint significance of the coefficients on the instrument(s).[10]

[9] Moreover, it may also address a third problem: the bias arising from the measurement error in schooling/education. See *Card* (*1999*) for more details.
[10] In the article by *Angrist and Krueger* (*1992*) mentioned earlier, it turned out that the draft lottery number as an instrument for education was weak – a low number was not strongly related to whether or not one went to university (*Card 1999*).

Although there are rules of thumb about the *F*-value above which the instrument is deemed sufficiently strong/relevant (*Staiger – Stock 1997; Stock – Yogo 2005*), it is in fact not possible to establish this without any doubt if someone is not fully convinced of the exclusion restriction (validity) of the instrument. A weak instrument that violates the exclusion restriction even if only a tiny bit (affecting earnings also through other channels, not just education) may even increase the bias of the OLS estimate (see details below). The exclusion restriction, however, is particularly difficult to verify, as there is no formal statistical test for it. Thus, the only option for researchers is to argue for it in a detailed and rigorous way, transparently laying out the supposed mechanism of the instrument, i.e. 'where the identification comes from.'

Card's (*1995*) perhaps autobiographically inspired article[11] uses the proximity of one's birthplace to the nearest university as an instrument for education (university degree). This can be regarded as a natural experiment, since the place of birth of an individual can be considered random, but it influences university attendance: it costs less for the individual to go to university if they were born close to one, than if they were born far away. For this reason, those born nearby are more likely to enrol and graduate from university than those born further away, as the argument for the relevance of the instrument goes. At the same time, it is not related to one's earnings whether the individual was born near a university through anything else than the person's degree, says the argument for the exclusion restriction. However, as it turns out, neither of the conditions fully hold: the instrument is only strong among those with lower-educated parents, and places close to universities tend to offer better job opportunities, meaning that proximity to university also affects earnings in other ways, through better local labour market conditions. As in many other IV studies measuring the returns to education, the numerical results in this paper show that the IV estimate far exceeds the OLS estimate. This remains somewhat of an unsolved puzzle, since both the presence of omitted variables (e.g. abilities) and the reverse causality (due to self-selection) would imply that the OLS is biased upwards, so an unbiased IV estimate should imply a lower return. What are the possible explanations for this apparent contradiction?

On the one hand, it is possible that the instrument is both invalid (i.e. does not meet the exclusion restriction) and weak, a combination that can inflate the bias of the OLS estimate. Whether this entirely explains the observation that the IV exceeds the OLS is not considered likely/convincing by the researchers (*Card 1999*). In the case of *Card* (1995), for example, the IV is still about 30 per cent higher than the

[11] David Card grew up on a farm in Canada, but there was a university nearby. He partly attributes his educational attainment to this (*Interview with David Card*. Federal Reserve Bank of Minneapolis, 1 December https://www.minneapolisfed.org/article/2006/interview-with-david-card).

OLS when controlling for an individual's family background, when the exclusion restriction and relevance of the IV are far more convincing.

On the other hand, it is possible that the IV is larger because the instrument also removes the downward bias in the OLS due to measurement error (see *Footnote* 9).

Moreover, many, including *Card* (*1999*), believe that IV studies are most likely to estimate the treatment effect (in our example, the returns to education) for specific groups for whom the return is larger than the population average. To shed more light on this, consider *Angrist and Krueger's* (*1991*) estimate of the returns to education. In this paper, the authors exploit the feature of the US public education system that in most states, primary school starts on 1 September of a given year for those who turned 6 years old before 1 January of that school year, while those who turn 6 on or after 1 January, only start school in the following September. Because of this, a person born on 1 January (in the first quarter of the year) typically starts school at the age of 6 and ¾, while someone born on 31 December (the fourth quarter) starts a year younger at 5 and ¾. Therefore, by their 16th birthday, when they reach the upper compulsory schooling age, those born in the first quarter of the year have been in school for one year less (for just over 10 years) than those born in the fourth quarter. If the quarter of birth is independent of other factors that determine earnings, we can use it as an instrument for education, and the difference in earnings between those born in the first and fourth quarters identifies the return to the difference in their education. However, as Angrist and Krueger point out, we only see education differences between the two groups (those born in the first vs. fourth quarter) among early school leavers, but not among those who go to university or post-graduate education. The differences in earnings are also concentrated in this group. However, this is a special group: they are the ones who are kept in school for a longer period of time exclusively due to the compulsory schooling laws. It is for this specific group that this study measures the returns to education, not for the general population. Yet the returns to education in this group can be very different from those of the general population.

This idea is formalised by *Imbens and Angrist* (*1994*) and *Angrist et al.* (*1996*) in the concept of the *local average treatment effects*. Still using the example of *Angrist and Krueger* (*1991*), the population can be divided into 4 segments according to how much they attend/would attend school if born in the first and fourth quarters:[12]

[12] Note that it is not obvious from the data to which group an individual belongs, since the allocation is based not only on the schooling chosen for the actual date of birth, but also on the so-called counterfactual choice, i.e. what the individual would have chosen if he or she had been born at a different date. The latter cannot be observed. Yet, using the method developed by Imbens and Angrist, we can describe these groups with the characteristics observed in the data.

1. The *always takers*: whether born in the first or fourth quarter, they go to school for 'longer';
2. the *never takers*: whether born in the first or fourth quarter, they go to school for 'shorter';
3. *compliers*: they attend school longer if they were born in the fourth quarter, than they do if they were born in the first quarter;
4. *defiers*: they attend school longer if they were born in the first quarter than they do if they were born in the fourth quarter.

Imbens, Angrist *et al.* show that under certain conditions, for example when there are no defiers, the IV estimate estimates the effect of treatment (in our example, education) on the group of compliers.[13] However, the group of compliers may be different for different instruments, for whom the returns to education may also be different. *Angrist and Krueger* (*1991*), for example, find that in the group held in high school by the compulsory schooling laws the return to one extra year of schooling is about 7.5 per cent, which is barely different from the OLS estimate. By contrast, for the group of compliers (those who only go to university if there is one nearby), *Card* (*1995*) finds the return to education – in their case university – to be well above those estimated by OLS.

The analysis by *Angrist and Krueger* (*1991*) has been widely replicated. *Bound et al.* (*1995*) showed that the quarter of birth as an instrument for education is weak, and therefore the IV estimates, even in a large sample, may be inconsistent. This article triggered a whole wave of methodological research, developing practical advice for cases where researchers have many weak *instruments* at their disposal (e.g. *Staiger – Stock 1997*). New solutions to this problem are also emerging today, with machine learning algorithms becoming more widespread.[14]

2.2.2. The effect of elite schools

For education researchers, education policymakers and parents alike, the question of which schools are good and what makes them work better than others is a fundamental one. At first glance, this seems like a trivial question: just look at which schools are at the top of the school rankings available everywhere. These rankings are based on data that show very clearly the achievement of students attending these schools: graduation rates, post-secondary attainment rates, standardised test scores. In most countries there are significant differences between

[13] In addition, the authors have also worked out the details of how to characterise the group of compliers using their characteristics observed in the data.
[14] See, for instance, the application of *Belloni et al.* (*2011*) in *Derenoncourt*'s (*2022*) paper.

schools with regard to these indicators, and usually a few elite schools stand out from the rest.

However, it is also clear that the composition of students in these schools is also very different. As they are very popular, they select the students who can perform the best from a large number of applicants, who then go on to perform really well. But what is the role of selection on the one hand and the school and the higher standards of education on the other? This question is particularly important if the government wants to encourage schools to improve the quality of education, as it requires measuring quality. In the US, several states and metropolitan school districts introduced this type of accountability reforms in the 1990s, and in the early 2000s it was introduced as a federal programme (No Child Left Behind programme).

Researchers have long tried to separate the effects of student characteristics and school quality using various statistical methods. On the one hand, we can try to directly identify the impact of student characteristics by including them as control variables in the analysis. On the other hand, we can look at the change in achievement (the '*value added*') of individual students over time. This approach is based on the assumption that the effect of students' individual characteristics is summed up in their previous test scores, and thus the change in their test scores is more or less attributable to the school. These two approaches, and various combinations of them, are very often used to measure the performance of individual schools or teachers. While everyone agrees that these estimates are generally much closer to the actual quality of schools compared to simple averages of outcome indicators, it is an open question how accurate they are. There is good reason to believe that in some cases, such as in prominent, elite schools, the role of unobserved student characteristics, which the traditional approach above cannot take into account, can be very important. Students who apply to these schools are likely to be more motivated and more diligent, and their academic performance is also a priority for their parents, who invest more in their development in a broader sense. These students presumably progress faster, so that the increase in achievement compared to their previous results is greater, and the value added calculation does not fully eliminate the effect of such selection.

How can we measure the quality of schools more accurately? If we consider school quality as the impact of schools on students' knowledge and skills, we can apply the methods of causal analysis, the logic of natural experiments. Angrist's research has been seminal in this area.

As he pointed out even in his Nobel Prize lecture, one of the issues that has long been a concern of his, and that he and his co-authors explored in several studies, is the effectiveness of so-called charter schools. These are non-public schools that receive public funding under a contract (a '*charter*') with the government (in the

US, unlike in Hungary, this is not usually the case for private and parochial schools), but may follow a curriculum and pedagogical practices different from those in traditional public schools. They often operate in segregated metropolitan areas and are successful in educating poor students who live there. Many see these schools as an opportunity to reform the US education system and expect them to reduce the huge achievement gap between minority and white students. While others believe that selection is behind the success of charter schools, where students and parents are more motivated and engaged than the average student with similar socio-economic backgrounds.

Angrist and his co-authors studied charter schools in Massachusetts, where in the case of over-subscription, lotteries are used to decide who gets into the school (*Angrist et al. 2010*; *Angrist et al. 2012*). Using lottery draw as a natural experiment, the subsequent results of the lucky ones admitted and those who rejected were compared. The IV method was applied, using the result of the random lottery numbers as an instrument for admission. In the case of the schools studied, a significant positive effect was found: the test score gains of admitted students at one school (Lynn), for example, were more than a third of a standard deviation higher in maths and a tenth of a standard deviation higher in reading and writing after one year than their peers' who were not admitted as a result of the random lottery (*Angrist et al. 2012*).[15]

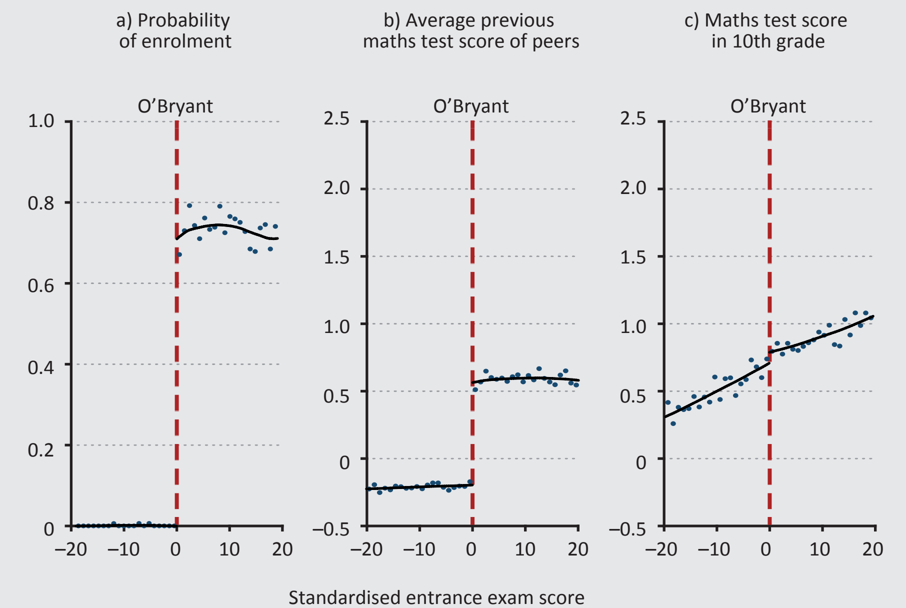
Another issue where Angrist and his fellow researchers have made significant progress is the impact of elite schools (*Abdulkadiroğlu et al. 2014*). In their ground-breaking study, they examined the impact of the three most prestigious elite high schools in both Boston and New York. Students are admitted to these schools solely on the basis of an admission test. A direct comparison of admitted and rejected applicants does not show the effect of schools, as the two groups are very different: admitted students are not randomly selected: they have achieved much better academic results in the past. However, students on either side of the admission threshold, who have just got in and those who have just missed the cut-off, are very similar; we can assume that they do not differ in any of their unobserved characteristics. Of course, the past achievement of students below the threshold is somewhat poorer, and this should be taken into account in the analysis. In such situations, when the probability of being treated suddenly and significantly jumps when a given variable passes some threshold, we can use the *regression discontinuity design*, as Angrist and colleagues did in their analysis of elite schools. Their study has not only contributed to the analysis of the impact of schools, but has also become an exemplary application of the regression discontinuity design.

[15] By comparison, these are large effects; the raw test score difference between black and white students in third grade is almost 1 standard deviation (*Fryer – Levitt 2006*).

The starting point of the method is that receiving the treatment depends on the value of a certain variable (*running* or *forcing variable*). For example, students can be admitted to an elite school if their admission test score is above a certain threshold. Admission test performance also correlates with the subsequent results, which is demonstrated by the regression curve fitted (by some parametric or non-parametric method) to the two variables. The effect of being in the treatment group is shown by the magnitude of the 'jump' observed in the outcome variable at the threshold.

Figure 5 illustrates the results for an elite school in Boston. It is clearly visible that, at the admission threshold, the probability of enrolling in the school increases sharply (panel a), and the composition of peers also changes significantly (panel b). However, contrary to what might be expected, there is no sharp increase in students' subsequent maths test scores (panel c): there is no difference in the subsequent academic achievement of students who were admitted to an elite school and those who were just rejected. This surprising result is expressed also in the title of the study: The Elite Illusion.

Figure 5
Elite illusion



Source: Abdulkadiroğlu et al. (2014): Figures 1, 2 and 4

The zero impact of elite schools on test scores is interesting for several reasons. First, it contributes to the literature on peer effects in education. How does the composition of the class, e.g. whether the classmates are high achievers or disruptive, affect the student? The literature on the issue provides a number of sharply different results, but this effect is very difficult to identify empirically (*Angrist 2014*). The study by Angrist et al. answers this question indirectly. It shows that the average previous test scores of the classmates of students who are enrolled in elite schools are much higher than those who failed to enrol, but this has no effect on the subsequent performance of the student, i.e. in this case the achievement of classmates does not matter (*Abdulkadiroğlu et al. 2014*).[16]

However, the question remains: if they have no real effect, why are these schools so popular amongst parents and students? Would applicants be wrong to chase the illusion of elite education? Or does student achievement as measured by test scores not capture what students gain by studying in these schools? For the moment, these remain open questions, but how strong these preferences are is shown by a recent study analysing elite schools in Chicago (*Angrist et al. 2019*). Here, the authors found that enrolment in elite schools has a negative effect on student achievement, because many of the students not admitted are enrolled in charter schools that do improve their performance. Nevertheless, applicants prefer elite schools to charter schools.

Causal estimation of school effects is only possible in exceptional cases, and the results refer to a specific, often unique group of students and schools. What can we learn from this type of analyses in general? In another recent study, Angrist and co-authors (*Angrist et al. 2017*) used data from the centralised admission system in Boston public schools. When assigning sixth-grade students to high schools, a random lottery plays a significant role in the matching algorithm.[17] Thus, for a very large number of schools, they were able to estimate the causal effect of the school and compare this with traditional value-added-type indicators of school achievement.[18] They found that the school quality estimated by the value-added method does give a biased estimate, but that the bias is not too large and so such data should not be ignored in education policy decisions.

[16] A similar result was obtained by *Angrist and Lang* (*2004*) in a previous study: the achievement of middle-class students in suburban schools in a Boston integration programme was not affected by the appearance of a few poor, black students from the inner city in the classroom. *Angrist* (*2014*) argues that many analyses overestimate the effect of peers, and that it is in fact much weaker than we think.

[17] In Hungary, a similar centralised admissions algorithm matches applicants with schools at both secondary and tertiary level, but there, random elements play a negligible role.

[18] *LaLonde's* (*1986*) study pioneered this strategy in the context of job training programmes.

3. Closing remarks

The emergence of a new approach based on natural experiments has fundamentally transformed the way economics research is conducted, and *Angrist and Pischke* (*2010*) rightly referred to this transformation as the '*credibility revolution*. This required the development of new analytical methods, in which David Card, Joshua Angrist and Guido Imbens played a key role. This brought economics closer to the natural-scientific ideal of understanding and empirically verifying causality. But perhaps more importantly, this new toolkit can help us better understand the economy and society.

References

Abdulkadiroğlu, A. – Angrist, J. – Pathak, P. (2014): *The Elite Illusion: Achievement Effects at Boston and New York Exam Schools*. Econometrica, 82(1): 137–196. https://doi.org/10.3982/ECTA10266

Angrist, J.D. (1990): *Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records*. American Economic Review, 80(3): 313–336.

Angrist, J.D. (2014): *The perils of peer effects*. Labour Economics, 30(October): 98–108. https://doi.org/10.1016/j.labeco.2014.05.008

Angrist, J.D. – Dynarski, S.M. – Kane, T.J. – Pathak, P.A. – Walters, C.R. (2010): *Inputs and Impacts in Charter Schools: KIPP Lynn*. American Economic Review, 100(2): 239–243. https://doi.org/10.1257/aer.100.2.239

Angrist, J.D. – Dynarski, S.M. – Kane, T.J. – Pathak, P.A. – Walters, C.R. (2012): *Who Benefits from KIPP?* Journal of Policy Analysis and Management, 31(4): 837–860. https://doi.org/10.1002/pam.21647

Angrist, J.D. – Hull, P.D. – Pathak, P.A. – Walters, C.R. (2017): *Leveraging Lotteries for School Value-Added: Testing and Estimation*. Quarterly Journal of Economics, 132(2): 871–919. https://doi.org/10.1093/qje/qjx001

Angrist, J.D. – Imbens, G.W. – Rubin, D.B. (1996): *Identification of Causal Effects Using Instrumental Variables*. Journal of the American Statistical Association, 91(434): 444–455. https://doi.org/10.2307/2291629

Angrist, J.D. – Krueger, A.B. (1991): *Does compulsory school attendance affect schooling and earnings*. Quarterly Journal of Economics, 106(4): 979–1014. https://doi.org/10.2307/2937954

Angrist, J.D. – Krueger, A.B. (1992): *Estimating the Payoff to Schooling Using the Vietnam-Era Draft Lottery*. NBER Working Paper No. 4067. https://doi.org/10.3386/w4067

Angrist, J.D. – Lang, K. (2004): *Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program*. American Economic Review, 94(5): 1613–1634. https://doi.org/10.1257/0002828043052169

Angrist, J.D. – Pathak, P.A. – Zárate, R.A. (2019): *Choice and Consequence: Assessing Mismatch at Chicago Exam Schools*. NBER Working Paper No. 26137. https://doi.org/10.3386/w26137

Angrist, J.D. – Pischke, J.-S. (2010): *The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics*. Journal of Economic Perspectives, 24(2): 3–30. https://doi.org/10.1257/jep.24.2.3

Arkhangelsky, D. – Athey, S. – Hirshberg, D.A. – Imbens, G.W. – Wager, S. (2021): *Synthetic Difference-in-Differences*. American Economic Review, 111(12): 4088–4118. https://doi.org/10.1257/aer.20190159

Athey, S. – Imbens, G.W. (2006): *Identification and Inference in Nonlinear Difference-in-Differences Models*. Econometrica, 74(2): 431–497. https://doi.org/10.1111/j.1468-0262.2006.00668.x

Athey, S. – Imbens, G.W. (2022): *Design-based analysis in Difference-In-Differences settings with staggered adoption*. Journal of Econometrics, 226(1): 62–79. https://doi.org/10.1016/j.jeconom.2020.10.012

Belloni, A. – Chernozhukov, V. – Hansen, C. (2011): *LASSO Methods for Gaussian Instrumental Variables Models*. arXiv:1012.1297v2. https://arxiv.org/abs/1012.1297

Bound, J. – Jaeger, D.A. – Baker, R.M. (1995): *Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variables is weak*. Journal of the American Statistical Association, 90(430): 443–450. https://doi.org/10.1080/01621459.1995.10476536

Burdett, K. – Mortensen, D.T. (1998): *Wage differentials, employer size, and unemployment*. International Economic Review, 39(2): 257–273.

Brown, C. – Gilroy, C. – Kohen, A. (1982): *The effect of the minimum wage on employment and unemployment*. Journal of Economic Literature, 20(2): 487–528.

Card, D. (1995): *Using geographic variation in college proximity to estimate the return to schooling*. In: Christofides, L.N. – Grant, E.K. – Swidinsky, R. (eds.): Aspects of labour market behaviour: Essays in honour of John Vanderkamp. University of Toronto Press, Toronto, Canada, pp. 201–222.

Card, D. (1999): *The Causal Effect of Education on Earnings*. In: Ashenfelter, O.C. – Card, D. (ed.): Handbook of Labor Economics. Elsevier, Volume 3, Part A, pp. 1801–1863.

Card, D. (2021): *Design-Based Research in Empirical Microeconomics*. Lecture at the Nobel Prize ceremony, 8 December. https://davidcard.berkeley.edu/papers/CardNobelLecture.pdf

Card, D. – Krueger, A. (1994): *Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania*. American Economic Review, 84(4): 772–793.

Card, D. – Krueger, A. (2000): *Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania: Reply*. American Economic Review, 90(5): 1397–1420.

Cengiz, D. – Dube, A. – Lindner, A. – Zipperer, B. (2019): *The effect of minimum wages on low-wage jobs.* Quarterly Journal of Economics, 134(3): 1405–1454. https://doi.org/10.1093/qje/qjz014

Derenoncourt, E. (2022): *Can You Move to Opportunity? Evidence from the Great Migration*. American Economic Review, 112(2): 369–408. https://doi.org/10.1257/aer.20200002

Friedman, M. (1953): *The Methodology of Positive Economics*. In: Essays in Positive Economics. University of Chicago Press, pp. 3–43.

Fryer, R.G. – Levitt, S.D. (2006): *The Black-White Test Score Gap Through Third Grade*. American Law and Economics Review, 8(2): 249–281. https://doi.org/10.1093/aler/ahl003

Hearst, N. – Newman, T.B. – Hulley, S.B. (1986): *Delayed effects of the military draft on mortality: A randomized natural experiment*. New England Journal of Medicine, 314(10): 620–624. https://doi.org/10.1056/NEJM198603063141005

Imbens, G.W. – Angrist, J.D. (1994): *Identification and Estimation of Local Average Treatment Effects*. Econometrica, 62(2): 467–475. https://doi.org/10.2307/2951620

LaLonde, R.J. (1986): *Evaluating the Econometric Evaluations of Training Programs with Experimental Data.* American Economic Review, 76(4): 604–620.

Lester, R.A. (1947): *Marginalism, Minimum Wages, and Labor Markets*. American Economic Review, 37(1): 135–148.

Manning, A. (2003): *Monopsony in Motion: Imperfect Competition in Labor Markets*. (Course Book ed.). Princeton: Princeton University Press.

Mincer, J. (1958): *Investment in Human Capital and Personal Income Distribution*. Journal of Political Economy, 66(4): 281–302. https://doi.org/10.1086/258055

Mincer, J. (1974): *Schooling, experience and earnings*. Columbia University Press, New York.

Royal Swedish Academy of Sciences (2021a): *Natural experiments help answer important questions*. An easy-to-understand background to the 2021 Nobel Prize in Economics. https://www.nobelprize.org/uploads/2021/10/popular-economicsciencesprize2021-3.pdf. Downloaded: 10 January 2022.

Royal Swedish Academy of Sciences (2021b): *Answering causal questions using observational data.* A scientific background to the 2021 Nobel Prize in Economics. https://www.nobelprize.org/uploads/2021/10/advanced-economicsciencesprize2021.pdf. Downloaded: 10 January 2022.

Staiger, D. – Stock, J.H. (1997): *Instrumental Variables Regression with Weak Instruments*. Econometrica, 65(3): 557–586.

Stigler, G.J. (1946): *The Economics of Minimum Wage Legislation*. American Economic Review, 36(3): 358–365.

Stock, J.H. – Yogo, M. (2005): *Testing for weak instruments in Linear IV regression.* In: Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg. Cambridge University Press, pp. 80–108. https://doi.org/10.1017/CBO9780511614491.006

Trostel, P. – Walker, I. – Woolley, P. (2002): *Estimates of the economic return to schooling for 28 countries*. Labour Economics, 9(1): 1–16. https://doi.org/10.1016/S0927-5371(01)00052-5

Wolfson, P. – Belman, D. (2019): *15 years of research on U.S. employment and the minimum wage*. Labour, 33(4): 488–506. https://doi.org/10.1111/labr.12162

Essays of the Series on the Work of Nobel Laureates in Economics

Kóczy, Á.L. – Kiss, H. J. (2017): *Hart and Holmström's Contributions to Contract Theory*. Financial and Economic Review, 16(1): 162–174. https://en-hitelintezetiszemle.mnb.hu/letoltes/laszlo-a-koczy-hubert-janos-kiss.pdf

Neszveda, G. (2018): *The Contribution of Thaler to Behavioural Economics*. Financial and Economic Review, 17(1): 153–167. https://doi.org/10.25201/FER.17.1.153167

Meyer, D. (2019): *Optimising Nobel Prize Laureates*. Financial and Economic Review, 18(1): 125–141. https://doi.org/10.33893/FER.18.1.125141

Major, K. (2020): *Studying Poverty in Economics – The Work of the 2019 Nobel Prize Laureates*. Financial and Economic Review, 19(1): 119–131. https://doi.org/10.33893/FER.19.1.119131

Biró, P. – Magyarkuti, Gy. (2021): *The Work of Milgrom and Wilson in the Theory and Practical Application of Auctions*. Financial and Economic Review, 20(1): 127–151. https://doi.org/10.33893/FER.20.1.127151