Chapter 17
The Use of Econometrics in Labour Research

Marc K. Chan and Lili Márk

Abstract Labour economics has long been a proving ground for econometric method. The early availability of microdata, and questions that turn on choice, selection, dynamics, and policy, repeatedly generated problems that existing tools could not answer, and the responses reshaped econometric practice well beyond the field. This chapter traces that history through five themes: sample and treatment selection, duration and transition models, quasi-experimental methods, behavioural and structural modelling, and earnings dynamics. For each we set out what the field asked, how the methods developed, and which ideas have endured. The recurring pattern is method drawn from data and question.

17.1 Introduction

Labour economics plays a special role in the history of econometrics, as it was one of the first fields in which microdata became widely available. Empirical analysis, especially of microdata, was used more heavily in labour economics than in other fields of economics (Angrist & Krueger, 1999; Moffitt, 1999). Angrist and Krueger (1999) claim that, *"outside of time-series econometrics, many and perhaps most innovations in econometric technique and style since the 1970s were motivated largely by research on labor-related topics"*. Starting from the 1960s, theory, econometric methods, and microdata collection efforts evolved hand in hand (Stafford, 1986). An example of data collection motivated by economic theories is the collection of work history information to estimate earnings functions.

Marc K. Chan ✉
Faculty of Business and Economics, University of Melbourne, Melbourne, Australia, e-mail: marc.chan@unimelb.edu.au

Lili Márk
Central European University, Vienna, Austria and RMIT University, Melbourne, Australia, e-mail: Mark_Lili@alumni.ceu.edu

Three developments organise the history that follows. The first is the structural-versus-reduced-form distinction: whether to estimate the parameters of an explicit economic model or to estimate a relationship of interest more directly, and what assumptions each approach requires. The second is the arrival of new data. The household panels of the 1970s made fixed-effects and duration methods feasible, and the later spread of administrative and linked employer–employee records made possible questions that survey data could not support; often the data came first and the methods followed. The third is the shift in emphasis in applied work towards research designs whose identifying assumptions are more closely scrutinised.

Our aim is a structured overview of the main econometric approaches used in labour research: their purposes, their assumptions, the settings that called them forth, and how they have fared. The account is organised around conceptual themes rather than chronology, and is selective rather than exhaustive, concentrating on methods that remain central to the field. Section 17.2 takes up sample selection and the treatment-selection problem. Section 17.3 turns to duration and transition models. Section 17.4 surveys the quasi-experimental methods—instrumental variables, regression discontinuity, and difference-in-differences—that are widely used in applied work. Section 17.5 sets out the behavioural and structural tradition, and Section 17.6 the modelling of earnings dynamics and heterogeneity. Each section closes with a short assessment of what has aged well and what has aged less well.

17.2 Selection and Participation Modelling

Selection problems are inherent in many fundamental questions in labour economics. They occur when an outcome of interest is observed only in a non-random, self-selected subset of the population. In sample selection problems, being in the sample is itself the result of people making decisions and choosing what is best for them: wages are only observed for workers; survey outcomes are missing for those who attrit. In treatment selection problems, being in a treatment is the result of people making decisions: those who become union members or participate in programmes do so if they think this will benefit them. While we see the wages of both union and non-union members, for union members—a self-selected subset of the population—we do not observe the wages they would have earned if they had not been in the union. This missing counterfactual outcome is crucial for identifying the treatment effect.

Gronau (1974) conceptualised the *"selectivity bias"*—biased estimates of wage equations—that arises when observed wages reflect individuals' search and participation decisions. Heckman (1974) formalised it in the setting of married women's labour supply, where participation was low and strongly patterned by age, marital status, and unobservables (Killingsworth & Heckman, 1986). Heckman's canonical two-step estimator (see Heckman, 1976, 1979) addresses the missing-outcome issue in the following model: an individual works, and their wage is observed, when a latent index crosses a threshold. A participation equation is estimated by probit on the full sample; the wage equation is then estimated on workers only, with the inverse

Mills ratio—constructed from the first stage, reflecting unobservables affecting both participation and wages—added as a regressor to absorb the selection bias that would otherwise contaminate ordinary least squares. L.-F. Lee (1979) provided a closely related censored-selection formulation. Before this work, *'first-generation'* studies (Killingsworth & Heckman, 1986) had simply run ordinary least squares on samples of workers, ignoring the non-participants whose wages are unobserved.

Treatment selection problems, and the treatment-effect models used to address them, have partly separate origins. In statistics, Rubin (1974) formalised causal effects through potential outcomes[1]—the comparison, for each unit, of outcomes with and without treatment—building on earlier randomization ideas. In economics, Heckman and Robb (1985) brought non-random selection into programme evaluation, presenting methods for estimating the impact of training on earnings when enrolment is non-random, and shifting the econometric focus from the missing-data problem of sample truncation toward the estimation of causal treatment effects (see also Heckman, 1978). L.-F. Lee (1978) formulated an early switching-regression application to study union wage premiums, its separate union and non-union wage equations implicitly allowing the union effect to vary across workers.

Moffitt (1999) documents a fall in the use of sample selection-bias methods, and gives three reasons. First, the inverse Mills ratio is often highly collinear with the other regressors in the outcome equation, producing unstable estimates. Second, identification relies on strong distributional assumptions, such as bivariate normality, of the disturbances. Third, in female labour supply, some studies found that correcting for selection changed the results little (Newey, Powell & Walker, 1990), although he also remarked that the applications thus far did not explore in depth the issue of exclusion restrictions for identification, which come to the fore when distributional assumptions are relaxed. A substantial semi- and nonparametric literature was developed to relax the distributional and functional-form assumptions (see Vella, 1998 for a survey), and it remains active at the theoretical frontier (Honoré & Hu, 2020), but Heckman's parametric two-step approach remains central in applied work.

Treatment-effect models[2], meanwhile, moved to centre stage in labour economics. The key difficulty lies in the missing counterfactual: no unit is observed both with and without treatment. Much emphasis has been placed on the role of exclusion restrictions and design-based variation for identification, using instrumental variables and related strategies to isolate exogenous variation for causal inference. These developments are the subject of Section 17.4.

17.2.1 What Aged Well and Less Well

What aged well is the conceptualisation of sample selection and treatment selection: recognising that they are sources of bias and finding credible ways to address

[1] For a more detailed discussion, see section 6.3 The Statistical Roots of the Potential Outcomes Framework (1920s-1980s) of the Treatment Effects chapter of this book.

[2] For a more detailed discussion, see Chapter 6 Treatment Effects of this book.

them remains central to empirical labour economics. What aged less well were some tools, particularly those relying on strong distributional and functional-form assumptions—though the parametric two-step procedure remains common in applied work. The field's centre of gravity shifted towards treatment-effect models, where design-based variation and exclusion restrictions are emphasised for more credible identification, developments taken up in Section 17.4. Yet the underlying framework remains connected: Vytlacil (2002) showed that the latent-index selection model, which traces its roots back to the origins of the sample selection model, and the monotonicity assumption behind the LATE framework of Imbens and Angrist (1994) are equivalent, so that the treatment-effect and sample selection literatures rest on a common structure.

17.3 Duration and Transition Models

Much of labour economics is concerned not with the level of a variable but with the timing of a transition: how long a worker remains unemployed before finding a job, how long a spell on welfare lasts before exit, how quickly a displaced worker is re-employed. The natural object of study is the hazard rate—the probability that a spell ends in the next short interval, given that it has lasted until now—related to elapsed duration and to covariates such as benefit levels, age, and local labour-market conditions.

The statistical machinery for this was imported from biostatistics and reliability engineering, where the analogous problem is time to death or component failure. The nonparametric survivor-function estimator of Kaplan and Meier (1958) and the proportional-hazards regression of Cox (1972) were developed to model failure times subject to censoring, and they became the backbone of duration analysis. Lancaster (1979) brought the apparatus into economics, modelling unemployment durations and confronting the two issues that have preoccupied the field since. The first is unobserved heterogeneity: workers differ in ways the analyst cannot see, and because the most employable leave unemployment first, the composition of survivors changes as a spell lengthens, so the measured hazard falls with duration even when no individual's prospects worsen—a spurious negative duration dependence. The second is censoring: many spells are still in progress when the data end, and cannot simply be discarded. The framework handles both, through mixing distributions for heterogeneity and likelihoods that use censored spells without bias (see van den Berg, 2001 for a survey).

The methods found their first major application in the study of unemployment insurance (UI). A central question was whether more generous benefits prolong joblessness by lowering the hazard of leaving unemployment. Using administrative spell data, Moffitt (1985) and Meyer (1990) found that the hazard of leaving unemployment rises sharply as benefits approach exhaustion, evidence that UI shapes search behaviour; Katz and Meyer (1990) separated workers expecting recall from those seeking new jobs, showing that the response differs by the nature of the spell. Atkinson and Micklewright (1991) reviewed this first wave of evidence, and Le Barbanchon, Schmieder and Weber (2024) survey the modern literature on job search and UI.

The behavioural theory behind these hazards is job-search theory (McCall, 1970; Mortensen, 1970), in which an unemployed worker samples wage offers and accepts the first above a reservation wage; this implies a hazard of leaving unemployment and links it to benefits and the arrival rate of offers. Related theories of job matching interpret separations as the outcome of gradually learning match quality, generating a turnover hazard that falls with tenure (Jovanovic, 1979), while equilibrium search models extended the idea to the determination of wages themselves (Burdett & Mortensen, 1998).

The single-spell hazard generalises naturally to transitions among several states. Rather than model only the exit from unemployment, one can treat the labour market as a set of states—typically employment, unemployment, and non-participation—with a hazard governing each possible move between them. When these transition rates are constant, the process is a continuous-time Markov chain; when the relevant states or worker types are not directly observed, it becomes a hidden Markov model, in which observed labour-force status is a noisy signal of an underlying latent state that must be inferred jointly with the transition rates. This multi-state apparatus inherits the same concerns with unobserved heterogeneity and censoring, now for a system of hazards rather than one. Castro, Lange and Poschke (2025) survey this framework and document substantial heterogeneity in transition rates across workers. These behavioural and transition-modelling frameworks, capturing the dynamics that link wages, durations, and mobility, are the foundation on which the structural models of Sections 17.5 and 17.6 are built.

17.3.1 What Aged Well and Less Well

What aged well is the conceptual apparatus: the hazard rate as the object of interest, and the explicit treatment of censoring and unobserved heterogeneity that duration data demand. These are now standard wherever the timing of an event matters, far beyond unemployment. What aged less well was the early reliance on strong functional-form assumptions to separate duration dependence from heterogeneity, a separation that is difficult to identify nonparametrically. The field's centre of gravity later shifted towards research designs that exploit policy variation in benefits or eligibility to identify behavioural responses more credibly, and towards models of transitions among multiple labour-force states.

17.4 Quasi-Experimental Methods in Labour Research

Quasi-experimental methods exploit exogenous variation generated by laws, events, cut-off values, or natural characteristics, such as geographic borders, gender, or date of birth, to mimic random variation from experiments in variables that would otherwise be endogenously related to the outcome of interest. In the context of social insurance benefits, for example, it is difficult to determine whether any labour market impact is due to the benefits themselves or to past labour supply and earnings, which determine both benefit receipt and benefit amounts. Cut-off values in eligibility rules (e.g., minimum working history requirements to be eligible for a benefit) or changes in benefit durations and amounts can provide variation in the benefit generosity that can be used to estimate their impact. The survey by Meyer (1995) on quasi-experimental econometric designs made many economists aware of these techniques, which were already in use in other social sciences (Angrist & Pischke, 2010) and includes many applications from labour economics, such as the impact of social insurance programmes on labour supply, the effects of minimum wage laws, the effects of immigration on the labour market and the impact of taxes and tax reforms on labour supply.

Quasi-experimental methods, or ‘natural experiments’ in economics gained momentum at the beginning of the 1990s, with many seminal applications originating from the field of labour economics (e.g., Card & Krueger, 1994). Although the econometric methods had been used in economics for some time, they were interpreted in the treatment effects framework using language from experimental studies during this period (e.g., instrumental variables were given a new interpretation in the LATE framework and fixed effects estimation was used for difference-in-differences models). The emphasis in empirical work shifted towards finding credible control groups, understanding sources of variation, as well as sensitivity analyses, and design-based approaches to ensure that results could be interpreted causally.

This period has been termed the ‘credibility revolution’ in empirical economics by Angrist and Pischke (2010), who focuses on the origin story in empirical labour economics. Labour economists began to question the reliability of evaluations of government training programmes leading to a series of large scale randomized experiments (see Card, Kluve & Weber, 2018 for an overview). A landmark paper by LaLonde (1986) compared econometric evaluations of the National Supported Work Demonstration, a temporary employment programme, with experimental results, revealing substantial differences. Another early example is that of Solon (1985) who conducted an early difference-in-differences analysis of taxing unemployment benefits that lead to a decrease in net benefit levels.

In this section, we summarise key developments in the use of key methods: instrumental variable estimation, regression discontinuity, and difference-in-differences. Since becoming part of the standard econometric toolkit in labour economics, these methods have been the subject of active research in econometrics and have undergone significant development.

17.4.1 Instrumental Variable Estimation

The method of instrumental variable (IV) estimation has been available since the 1920s.[3] During the 1950s-1970s it was mainly a tool for estimating structural parameters in simultaneous equation models (SEMs) in the framework developed by the Cowles Commission.[4] The canonical example is supply-demand models, where the outcomes of interests are determined jointly and causality is assumed both ways. Identification meant that the structural parameters of the model could be recovered from the joint distribution of observed variables. Technically, instrumental variable estimation was the tool to identify impacts of endogenous variables, through including certain variables—instruments—in some equations only. IV estimation focused less on explaining why the selected instruments are correlated with the key endogenous variable but not directly with the outcome variable, and were viewed more like a technical step in the algebraic identification problem of structural models. Several estimation techniques emerged during this era (limited information estimation (LIML), full information estimation (FIML), two- and three-stage-least squares (2SLS, 3SLS)) and 2SLS became the most common way of estimating IV models.

The new wave of IV applications in the 1990s primarily addressed omitted variable biases through 'natural experiments' broadening the areas where IV estimation is used. This new approach became a widespread and standard tool in the econometrics toolkit after leading examples from labour economics (e.g., Angrist & Krueger, 1991) and the reinterpretation of IV in the treatment effects framework (Imbens & Angrist, 1994), which was motivated by examples from labour economics such as the returns to education.

One of the canonical examples of the use of IVs is the estimation of the returns to human capital accumulation. Since the 1950s, human capital theory and the estimation of returns to education have been central questions in labour economics (see, for methodological reviews and surveys of empirical literature Angrist & Krueger, 1999, Card, 1999, and Deming & Silliman, 2025). Mincer (1974) developed a model framework that captured the relationship between schooling, experience, and wages, summarised in an equation known to labour economists as the 'Mincer equation'. When estimating this equation empirically, the key identification problem is that people with higher unobserved ability tend to have more years of education and higher wages. However, ability is typically not observed. Furthermore, it is unclear how innate ability should be measured. This leads to an omitted variable problem, which makes the standard OLS estimation biased. This phenomenon is called the 'ability bias' shortly (Deming & Silliman, 2025). The 'ability bias' could be resolved by randomly assigning years of education to individuals; however, this would be a

[3] In his seminal work Wright (1928) estimated the supply and demand elasticities of flaxseed using 'curve shifters' that we would now call instrumental variables. 'Curve shifters' are factors that either influence demand but leave the cost of production unaffected (e.g., the price changes of a substitute good) or influence supply while leaving demand unaffected. This was the first use of IV estimation (Angrist & Krueger, 2001; Stock & Trebbi, 2003).

[4] For an overview about the emergence of SEMs see section 6.2 The Structural Equation Paradigm (1940s-1980s) in the Treatment Effects chapter of this book.

highly unrealistic experiment. Some papers from the 1970s attempted to address this issue by controlling for ability using observed test scores. However, this approach had significant limitations (Griliches, 1977). Consequently, IV estimation gained prominence.

Finding an instrument—a variable correlated with years of education, but does not directly affect wages—helps us to identify the causal relationship between education and wages (Card, 1999; Angrist & Krueger, 2001). In their influential paper, Angrist and Krueger (1991) were the first to use a 'natural experiment'—a term coined by Rosenzweig and Wolpin (2000)—based on school starting rules to identify the impact of education on wage. They exploit a true random source of variation that is the birth date of individuals, which is plausibly not correlated with ability. More specifically, they exploit the rule that children must start school in the calendar year in which they turn six. Coupled with compulsory schooling laws requiring students to stay in education until their sixteenth birthday, this rule generates exogenous variation in years of education for students for whom the compulsory schooling cut-off age is binding. Students born in the fourth quarter of the year start school at a younger age (before their sixth birthday, as school starts in the autumn) than those born in the first quarter, resulting in more years of education by the age of 16. Subsequently, numerous other papers have employed IV methods to identify this relationship. For instance Duflo (2001) and Khanna (2023) use expansions in public schooling to define instruments for a fuzzy regression discontinuity design, while Harmon and Walker (1995) and Oreopoulos (2006) exploit changes in compulsory schooling age. Duflo (2001) and Khanna (2023) use a fuzzy regression discontinuity design.

Natural experiments—just as Angrist and Krueger (1991)—often rely on truly 'natural' sources of exogenous variation such as twin births (e.g., Rosenzweig and Wolpin (1980)), birth dates, gender and weather events. In their review, Rosenzweig and Wolpin (2000) examine papers that use similar designs to answer important questions in labour economics, such as estimating the returns to human capital investments, the determinants of labour supply using weather variation, and the effect of children on female labour supply. They highlight that, even in situations where the source of variation is very plausibly exogenous, implicit modelling assumptions are needed for the estimated parameters to be correctly identified. They argue that, without a structural model, it is difficult to understand what parameters the results of an IV estimation actually estimate.

There are many other significant questions in labour economics where IVs have often been used for identification. Without attempting to list them all, we present some examples that have led to many subsequent papers using similar designs, while focusing on the earliest applications. Many of these examples attempt to identify the impact of endogenous factors on individuals' labour supply. When estimating the effect of children on parents' labour supply, particularly on mothers' employment and earnings, fertility and family size are endogenous. Angrist and Evans (1998) used the sex composition of the first two children as an instrument for having an additional child assuming parental preference for a mixed sibling-sex composition. There is substantial theoretical and empirical work on the impact of social insurance programmes on labour supply, where the receipt of disability or unemployment insurance (DI or UI) benefits is endogenous. Maestas, Mullen and Strand (2013) used examiners' allowance rates as an instrument for DI benefit receipt. Similarly, Kling (2006) used randomly assigned judge leniency as an instrument for the length of incarceration when estimating the impact of incarceration on employment and earnings. Further examples from the 1990s can be found in Angrist and Krueger (2001).

A major methodological innovation is the use of Bartik instruments, also known as shift-share instruments, to estimate the impact of changes in labour demand on local labour markets, named after Bartik (1991) and later adopted by Blanchard and Katz (1992). Unlike earlier uses of IV estimation in labour economics, where the impact of individual supply-side factors was the main area of interest, Bartik instruments allow for credible identification at the local labour market level. The challenge of identification is that employment growth in a region reflects both supply- and demand-side factors, as well as aggregate shocks that affect the entire economy. As Goldsmith-Pinkham, Sorkin and Swift (2020) explain *"these papers define the instrument as the local employment growth rate predicted by interacting local industry employment shares with national industry employment growth rates"*. Notable examples from labour economics include Bound and Holzer (2000), who study the impact of local job growth on the employment and earnings of racial and demographic groups, and Autor, Dorn and Hanson (2013), who study the effects of import competition from China on labour market outcomes for US workers. Shift-share instruments are also commonly used today in other fields of economics, such as trade, finance, public economics and regional and urban economics.

17.4.1.1 Unobserved Heterogeneity in Treatment Effect

Since the 1980s, economists have become increasingly interested in the implications of unobserved heterogeneity in treatment effects (UHTE). Mogstad and Torgovitsky (2024) provides an intuitive explanation using the example of returns to college education, showing how people's non-random choices lead to unobserved heterogeneity in treatment effects in any setting where they make choices about the variable of interest. *"Those [people] who choose to attend college tend to be those who would benefit from it"*, thus the choice of individuals is directly affected by their expected treatment effect.

In their landmark paper, Imbens and Angrist (1994) introduced the concept of the local average treatment effect (LATE). By combining the potential outcomes framework from statistics and IV estimation from economics, they showed that, under a monotonicity assumption, IV estimation provides an estimate of the average treatment effect for 'compliers', or units whose decision is impacted by the instrument. They consider the simplest case of binary treatment and binary instruments. As UHTE is a generic feature of most economic applications, especially in labour economics, and because of the intuitive appeal of the framework, the LATE framework has become the standard way of interpreting IV estimations (Mogstad & Torgovitsky, 2024).

Early findings from 'natural experiments' raised important conceptual questions. One such puzzle was that the IV estimate of the return to education in Angrist and Krueger (1991) was larger than the OLS estimate, despite the fact that the IV estimate is supposed to solve the 'ability bias' which leads to overestimation of the returns to education due to the unobserved and thus omitted variable, ability. These kinds of contradictions were key drivers to gain a better understanding of the role of UHTE. The puzzle was partly resolved by the realisation that, with IV estimation, one can estimate the treatment effect for a subgroup of the population on whom the instrument has an influence, and that this does not equal the average treatment effect for the population if heterogeneous treatment effects are present (Imbens & Angrist, 1994). Weak instruments—2SLS is biased towards the OLS estimate when the correlation with endogenous regressors is low—also received more attention and it became a common practice to accept first stage estimates with $F \geq 10$ (Staiger & Stock, 1997) and later with $F \geq 16.4$ (Stock & Yogo, 2005) for t-tests to give reliable results. Keane and Neal (2023) describe the "power asymmetry" for not only weak instruments in the earlier sense: if the bias is positive, the t-test of the 2SLS estimate has little power to detect true negative effects while inflated power to detect positive effects. They suggest a broader use of the Anderson-Rubin test (Anderson & Rubin, 1949).

The canonical LATE framework and subsequent discussions typically consider a scenario without covariates, in which the model yields intuitive results. However, most real-world applications incorporate control variables to relax the exogeneity constraint and employ 2SLS for estimation. Since Imbens and Angrist (1994), it has become common practice to interpret IV estimators as LATE, or at least as a positively weighted average of LATEs, however Blandhol, Bonney, Mogstad and Torgovitsky (2026) and Mogstad and Torgovitsky (2024) highlight that, in most cases, the LATE interpretation is not applicable. The linear IV estimator with covariates, in fact, is a weighted average of treatment effects for both compliers and always-takers, where always-takers are negatively weighted. For the LATE interpretation to hold, a saturated specification is needed. Blandhol et al. (2026) surveyed 99 papers that used a linear IV estimator with covariates and found that only one used this specification. Thus, they argue that other methods of estimating LATE, which have mostly remained undiscovered by applied researchers so far, may play a larger role in the future.

The literature on marginal treatment effects (MTE) asks the question of what we can learn about heterogeneous treatment effects based on IV estimation, beyond having a single LATE for compliers as in the simplest binary instrument – binary treatment case (for further discussions of MTE, see Cornelissen, Dustmann, Raute & Schönberg, 2016 and Heckman & Vytlacil, 2007). Crucially, LATE is defined by the instrument used and is often not the parameter of interest. The key idea is that when we have multivalued instruments or more binary instruments, the resulting overall IV estimate is a specific weighted average of underlying covariate-specific pairwise LATEs at different value-pairs of the instrument capturing different values of the unobserved preference for the treatment (U_D), which is hard to interpret on its own and hides interesting treatment effect heterogeneity (Imbens & Wooldridge, 2009). MTEs, however, capture the treatment effect at a particular value of U_D. Björklund and Moffitt (1987) introduced the concept of marginal treatment effects and the concept was further developed in a series of papers by Heckman and Vytlacil (1999, 2001a, 2001b, 2005) and Mogstad, Santos and Torgovitsky (2018) more recently. Applications in labour economics include the estimation of the heterogeneous returns to education (Carneiro, Heckman & Vytlacil, 2011; Carneiro, Lokshin & Umapathi, 2017) that illustrate the large differences in the returns to schooling for low- vs. high-resistance individuals as previously hypothesised in the literature; the impact of foster care on future earnings (Doyle, 2007); the labour supply response to disability insurance (French & Song, 2014; Maestas et al., 2013).

17.4.2 Difference-in-Differences and Event Studies

The concept of a difference-in-differences (DiD) model emerged early in labour economics. Obenauer and von der Nienburg (1915) analysed the introduction of a minimum wage for certain retail worker groups in Portland, Oregon. The authors compared the means of important outcome variables before and after the policy was introduced in Portland and in other cities where there was no change to the minimum wage (Lechner, 2011). This paper also provides early evidence that minimum wages do not increase unemployment. Another early example, mentioned by Angrist and Krueger (1999) is that of Lester (1946), who also analysed the impact of minimum wages, while Currie, Kleven and Zwiers (2020) mention that the first paper that used the DiD method is Ashenfelter (1978) who analysed the impact of training programmes on earnings, although he did not use DiD terminology. These early examples did not use DiD language and Ashenfelter and Card (1985) coined the term difference-in-differences (Card, 2022).

Ashenfelter (1978) is very important for the entire ‘natural experiment’ movement as it highlights the importance of identifying appropriate control groups for specific analyses. In the context of training programmes, the author illustrates that participants typically experience a decline in employment and earnings immediately prior to joining a programme, which is why they end up in it. Therefore, comparing them to non-participants without considering these trends would result in upwardly biased estimates due to reversion to the mean.

The simplest form of a DiD model involves two cross-sectional units—one treated and one control—and there are two periods of observations —one before and one after the treatment or policy change. The earliest canonical papers using this simplest DiD from the ‘natural experiment’ era of the late 1980s and 1990s come from labour economics and clearly demonstrate the intuitive appeal of this method. The Mariel Boatlift study by Card (1990) analysed the sudden influx of Cuban immigrants to the Miami labour market in 1980, which led to a 7% increase in the labour force. The study compared the employment trajectories in Miami with those in selected comparison cities (see Angrist and Krueger (1999) for a discussion). Card and Krueger (1994) studied the impact of the minimum wage increase in New Jersey by surveying fast-food restaurant workers in New Jersey and Pennsylvania. They also found that it did not lead to the expected significant changes in employment.

DiD has become a very popular method in economics, particularly in labour economics. Currie et al. (2020) find that the proportion of applied microeconomic NBER Working Papers referencing DiD began to rise around the mid-1990s, reaching almost a quarter of applied microeconomic NBER papers and over 15% of papers in applied mircroconomics in the top five journals by 2020. Bertrand, Duflo and Mullainathan (2004) report that, in their survey of applied DiD papers published between 1990 and 2000 in six selected journals[5] the most common outcome variables are employment and wages.

In labour economics, DiD has been used to study a wide range of topics, including the effects of minimum wages on employment (see Dube & Lindner, 2024 for a review), the impact of training and other labour market programmes for the unemployed on labour market outcomes, and the analysis of labour supply (see Blundell, Duncan & Meghir, 1998 for a survey). It has also been used to study the effects of various types of labour market regulations, such as workers' injury compensation on injury-related absenteeism, and maternity and disability insurance regulations. Further references to DiD in labour economics can be found in Angrist and Krueger (1999) and Lechner (2011).

In the simplest canonical form of difference-in-difference (DiD) estimation, a simple comparison of the relevant sample means is equivalent to a regression with binary variables for the cross-sectional units indicating treatment and time after treatment, as well as their interaction. Estimating with ordinary least squares (OLS) is useful for making inferences, and in the simplest case it yields the same result as two-way fixed effects (TWFE) estimation. This realisation has led to TWFE becoming the standard method for estimating DiD models in more complex settings, where there are more cross-sectional units and/or time periods, as well as more events to analyse that may occur at different points in time (i.e., staggered DiD), and triple DiD models (for example Gruber, 1994). In order to identify average treatment effects, the key identifying assumption of parallel trends in the control and treated units should also generalise.

Since the 2000s, several papers have raised concerns about DiD methods that have often been ignored by researchers using them. Firstly, issues surrounding inference with serial correlation and the clustering of standard errors were discussed in several papers, both in general for fixed effects models and DiD models separately (see for example Donald and Lang (2007); Hansen (2007a, 2007b); Kézdi (2004); Wooldridge (2003)). Bertrand et al. (2004) highlight the role of serial correlation in the underestimation of standard errors in DiD models where multiple time periods are used. Using state-level female wage data from the US Current Population Survey in their simulations, they found that randomly generated placebo laws had a significant impact in 45% of cases (compared to 5% if standard errors had been correctly estimated). This realisation has led to increased interest in how to appropriately handle the serial correlation problem when estimating DiD models with TWFE (Angrist & Pischke, 2009; Lechner, 2011). The most common way to handle this

[5] The six journals that Bertrand et al. (2004) review are the American Economic Review, the Industrial Labor Relations Review, the Journal of Labor Economics, the Journal of Political Economy, the Journal of Public Economics, and the Quarterly Journal of Economics.

problem is to cluster the standard errors at the level of treatment assignment (Abadie, Athey, Imbens & Wooldridge, 2023). However, the literature on the correct way of estimating standard errors under different assumptions still leaves many specific details open (A. C. Baker, Callaway, Cunningham, Goodman-Bacon & Sant'Anna, 2026; Roth, Sant'Anna, Bilinski & Poe, 2023).

Secondly, a great deal of work has been carried out with the aim of understanding what a TWFE estimate of a DiD model actually measures when the treatment is staggered (i.e., when multiple events occur at different times and in different places) and when there is heterogeneity in the treatment effect (A. C. Baker et al., 2026; Callaway, 2023; de Chaisemartin & D'Haultfœuille, 2023; Goodman-Bacon, 2021; Roth et al., 2023; Sun & Abraham, 2021). These papers highlighted the fragility of these estimates as TWFE estimators often lack an intuitive average treatment effect interpretation, but rather include, e.g., comparisons between already treated and later treated observations. This strand of literature is important, since many questions lend themselves to the use of a staggered DiD model. For example, labour market regulation often differs between US states and includes policy changes at different points in time. See Dube and Lindner (2024) for a review of applications in the study of minimum wages. For a more detailed discussion, see Section 6.6.3 Panel Data and Event-Study Methods of the Treatment Effects chapter of this book.

An important recent development of DiD applications is the emergence of event study designs. D. L. Miller (2023) provides a practical overview of event study designs. In these settings, the time dimension is measured relative to an event, which can happen in different caledar time for different individuals, such as job loss (Jacobson, LaLonde & Sullivan, 1993) or the birth of a child (Kleven, Landais & Søgaard, 2019). Event study plots have become standard elements of applied papers using DiD methods and these papers are similarly subject to the same issues as other DiD models based on TWFE (Currie et al., 2020). As Roth (2022) point out, it has been a common practice to assess the validity of the parallel trends assumption by plotting pre-trends and reporting simple pre-trend test statistics. However, several papers highlighted that the reliance on this intuitive and illustrative testing for parallel trends raises concerns (e.g., Freyaldenhoven, Hansen & Shapiro, 2019). Kahn-Lang and Lang (2020) show that these tests are neither necessary nor sufficient for the assumption to hold. Roth (2022) highlighted that these tests might be seriously underpowered and reliance on them could further distort estimates. Rambachan and Roth (2023) proposed new inference methods that relaxes the parallel trends assumption to handle these concerns. Roth et al. (2023) provides a summary of recent trends in DiD-style analysis.

17.4.3 Regression Discontinuity Methods

The regression discontinuity design (RDD) originated in psychology when it was used to analyse the impact of merit awards on future academic outcomes (Thistlethwaite & Campbell, 1960). The earliest RDD paper in economics was a theoretical work comprising two unpublished manuscripts by Goldberger (1972a, 1972b) (cited by Cook, 2008). However, the method remained dormant in economics for almost another two decades until the 1990s, when applications of RDD in economics started to increase rapidly together with other methods of the ‘natural experiment’ era. Labour economics was at the forefront of the use of regression discontinuity in economics.

Regression discontinuity methods can be used to estimate treatment effects in non-experimental situations where treatment assignment is determined by whether a unit is below or above the cutoff value of an observable continuous variable (score, or running variable). We compare units just below and just above the cutoff to estimate the treatment effect, based on the assumption that units very close to the cutoff value are very similar in their observable and unobservable characteristics. The continuity framework (Hahn, Todd & van der Klaauw, 2001) assumes that the regression functions of the potential outcomes are smooth functions of the running variable conditional on observed covariates at and around the cutoff and the density of the score near the cutoff is positive. The local randomization framework (Cattaneo, Frandsen & Titiunik, 2015) assumes that within an appropriate window the assignment to treatment is as good as random. Both approaches provide clear intuition for RDD methods. It is also a crucial assumption that individuals cannot manipulate their running variable, nor their treatment status. When there is imperfect compliance, the fuzzy RDD is applicable, which relies on a similar monotonicity condition and exclusion restriction as an IV. Cattaneo and Titiunik (2022) provide a detailed overview of the theoretical background of RDD, updating earlier reviews of the method (Cook, 2008; Imbens & Lemieux, 2008; D. S. Lee & Lemieux, 2010; van der Klaauw, 2008).

RD models require the researcher to fit a regression line separately on both sides of the cutoff value. Key decisions involve the functional form, parametric vs. nonparametric estimation methods and the sample window used for the estimation. Parametric methods and high-order nonparametric polynomial methods have poor performance at the boundaries and a rich literature emerged on inference in RDD. Calonico, Cattaneo and Titiunik (2014a, 2014b) provide a robust data-driven method for bandwidth selection and inference for local polynomial nonparametric estimation that is standard practice in RDD estimation today. Pei, Lee, Card and Weber (2022) suggest a data-driven method for the optimal choice of polynomial order.

Studying the labour supply impact of unemployment insurance (UI) benefits lends itself very easily to RD-style analysis, as receipt of UI benefits is often tied to critical values of an underlying running variable, such as age, earnings or employment history. One of the earliest examples of the use of RD in the UI literature uses a cut-off value in job tenure to identify the impact of a lump-sum severance payment on unemployment duration after job loss (Card, Chetty & Weber, 2007), while Lalive (2008) uses age and geographic discontinuities to estimate the impact of maximum benefit duration on unemployment duration. For more RD applications in the UI literature, see the summary by Le Barbanchon et al. (2024). More examples of the use of RDD in labour research—as well as education—can be found in D. S. Lee and Lemieux (2010).

Card, Lee, Pei and Weber (2015, 2017) introduce the regression kink design[6] (RKD) that is similar to the RDD, but it exploits a change in the slope of a policy rule rather than a change in level. They use fuzzy RKD to estimate the effect of benefits rates on unemployment duration in Austria using kinks in unemployment insurance benefits at minimum and maximum values of past earnings. Once again, there is a running variable and a related policy variable of interest, which is determined by the running variable. There is also a cutoff value in the running variable, where there is a kink in the policy rule. Typical examples include minimum and maximum amounts of tax and benefit schedules at certain past income values, or changes in marginal tax rates.

17.4.4 What Aged Well and Less Well

Labour economics has been a key driver of the 'credibility revolution' and the emergence and widespread adoption of quasi-experimental methods in and since the 1990s. The emphasis on using credible exogenous variations to identify causal effects has aged very well and raised the standards of what constitutes credible empirical strategy, following several economists' warnings about the over-reliance on arbitrary assumptions in empirical economics (e.g., Leamer, 1983, among others). Since then, it has become standard practice in applied economics to report robustness checks and sensitivity analyses, and to rely on multiple methods (Athey & Imbens, 2017). However, estimation frameworks and inference methods have evolved from their original forms, and the econometric literature on IV, DiD and RD methods remains active.

As longer panel data and administrative data became available, new methods, as well as concerns about inference and identification emerged, often leading to new waves of econometric research aimed at providing solutions, just as the original innovations of the quasi-experimental movement itself had flourished due to the increased availability of microdata (Stafford, 1986). Applications of DiD estimation have moved beyond the traditional two-by-two examples and case studies to encompass the estimation of staggered DiD models with TWFE, leading to the adoption of event-study analyses. These developments have highlighted the limitations of existing approaches when it comes to handling serial correlation, interpreting estimates in terms of usual treatment effects, and applying the parallel trends assumption. Regression kink design (RKD) also emerged in response to the increased precision that large administrative data sets allowed when estimating the labour supply impacts of benefit and tax schedules. However, the new methods that emerged in the 2000s and 2010s did not replace more traditional methods such as instrumental variable estimation. Instead, they enriched the toolkit of applied labour economists, who

6 This method was first employed by Nielsen, Sørensen and Taber (2010) in an educational economics setting that closely resembles that of the seminal work of Thistlethwaite and Campbell (1960): it studies the impact of student aid on college enrolment.

increasingly adopted a "collage" approach to using these methods (Currie et al., 2020).

The move towards more sophisticated quasi-experimental methods also stems from a growing recognition that human behaviour is complex, with these methods digging deeper into issues such as heterogeneous treatment effects and the parallel trends assumption. In this sense, the original appeal of the 'simplicity' of these methods (Card, 2022) has aged less well. A recurring concern in applied economics is that, while the "credibility revolution" has greatly improved causal identification (Imbens, 2010), it may sometimes encourage design-based or reduced-form estimates that are weakly connected to explicit economic models or policy-relevant parameters. Critics from the structural side argue that empirical work should more clearly articulate the economic model behind the question, the parameter being identified, and the policy counterfactual to which the estimate relates (Keane, 2010; Nevo & Whinston, 2010). The two traditions have increasingly been combined, whether through sufficient statistics approaches that express welfare effects in terms of estimable reduced-form elasticities (Chetty, 2009; Kleven, 2021), or through structural models estimated and validated with quasi-experimental variation (Todd & Wolpin, 2023).

17.5 Behavioural and Structural Modelling

Many of the central questions in labour economics turn on how people make decisions: on the choices, expectations, and incentives behind observed outcomes. How would a tax-and-transfer reform reshape the labour supply of single mothers? How long would the unemployed search, and at what reservation wage, if benefits became less generous? How much do young people's schooling choices respond to the expected lifetime return rather than to immediate earnings? Questions of this kind ask what would happen under a policy not yet implemented, or in a counterfactual environment no natural experiment isolates. They turn on forward-looking behaviour, dynamic incentives, and equilibrium responses that reduced-form methods, however cleanly identified, are not built to recover. The structural tradition answers by writing down an explicit model of the agent's decision problem, deriving its empirical implications, and estimating the parameters—preferences, offer distributions, transition and constraint rules—that are by construction invariant to the policy under study. With those parameters in hand, the model can simulate behaviour under policies that were never tried and weigh their costs and benefits, neither of which follows from a treatment-effect estimate (Keane, Todd & Wolpin, 2011).

The approach traces to the structural-versus-reduced-form debates of the 1960s and to the Cowles Commission programme of estimating systems of structural equations.[7] What changed from the late 1970s was the feasibility of estimating richer, explicitly dynamic models of individual choice. Two strands developed in parallel. The first was the dynamic discrete choice (DDC) framework, in which a forward-looking

[7] For an overview about the emergence of SEMs see section 6.2 The Structural Equation Paradigm (1940s-1980s) in the Treatment Effects chapter of this book

agent repeatedly chooses among a finite set of alternatives, trading current rewards against the expected discounted value of future states. Rust (1987), studying a bus-depot manager's engine-replacement decisions, gave the canonical formulation: a regenerative optimal-stopping problem, unobserved states entering through additively separable extreme-value shocks, and the nested fixed point algorithm, which solves the agent's dynamic program at each trial parameter vector while maximising the likelihood. The application was not itself a labour problem, but the machinery transferred directly to settings where agents make repeated, forward-looking decisions.

The leading labour applications model schooling, work, occupational choice, and fertility. Wolpin (1984) estimated an early dynamic model of fertility and child mortality, while Moffitt (1984) and Hotz and Miller (1988) estimated early complete life-cycle models of fertility, labour supply, and wages for married women. Eckstein and Wolpin (1989a) estimated a dynamic model of married women's participation with the endogenous accumulation of work experience, and Eckstein and Wolpin (1989b) survey the specification of dynamic stochastic discrete choice models more broadly. Keane and Wolpin (1997) brought the approach to maturity, estimating a life-cycle model in which young men choose each year among schooling, three occupational sectors, the military, and home, accumulating occupation-specific human capital as they go. Imai and Keane (2004) pushed the framework towards labour supply, showing that when human capital accumulates through work experience the observed wage understates the true return to an hour worked, so that a relatively flat life-cycle hours profile is consistent with a substantially larger intertemporal elasticity of substitution than conventional estimates imply. Occupational choice itself was given an early structural treatment by R. A. Miller (1984), who modelled the choice of occupation as a problem of learning about match quality through experience and estimated its parameters from panel data. The human-capital and education margin has since grown into a rich field of its own. Arcidiacono (2005) estimated a dynamic model of college and major choice to study how admission and financial-aid rules shape later earnings; Keane and Wolpin (2001) studied schooling under borrowing constraints and parental transfers; Fu (2014) embedded student application and college admission decisions in an estimated equilibrium model of the college market; and Arcidiacono, Aucejo, Maurel and Ransom (2025) modelled college attrition as a process of gradual learning about one's own ability and productivity from grades and wages. A parallel literature on early skill formation—Cunha and Heckman (2008), Cunha, Heckman and Schennach (2010), and Agostinelli and Wiswall (2025), building on the identification of the skill production technology from noisy measures of latent ability—studies how cognitive and non-cognitive skills are produced by parental and own investment over childhood.

The credibility of such estimates rests as much on the numerical methods used to solve and estimate the model as on the economics it encodes, and a methodological literature grew up to make estimation tractable: the simulation-and-interpolation methods of Keane and Wolpin (1994) that approximate the expected value function rather than computing it exactly; the method of simulated moments embedding simulation in the estimation criterion (McFadden, 1989; Pakes & Pollard, 1989); the conditional-choice-probability estimators of Hotz and Miller (1993) and their extension to models with unobserved heterogeneity by Arcidiacono and Miller (2011), which recover value-function differences from observed choice frequencies and so avoid repeated full solution of the dynamic program (see Aguirregabiria & Mira, 2010); and the discrete-continuous endogenous grid method of Iskhakov, Jørgensen, Rust and Schjerning (2017) for models that combine discrete choices with a continuous saving decision. A distinct strategy is indirect inference (Gourieroux, Monfort & Renault, 1993), which estimates structural parameters by matching the observed data and model-simulated data through the lens of an auxiliary descriptive model; Bruins, Duffy, Keane and Smith (2018) develop a generalised version suited to discrete-choice models.

A second strand carried the same logic into job search. The behavioural foundations of Section 17.3—McCall (1970)'s reservation-wage rule and the equilibrium search frameworks of Mortensen (1970) and Burdett and Mortensen (1998)—were taken into structural estimation, where the objects of interest are the wage-offer distribution, the arrival rate of offers, and the value of non-employment. Wolpin (1987) estimated an early structural search model of the transition from school to work, and Eckstein and Wolpin (1990) estimated a market-equilibrium search model from individual panel data, endogenising the wage-offer distribution as the outcome of firms' and workers' decisions; Eckstein and van den Berg (2007) survey the empirical search literature that followed. Recovering these primitives, rather than reduced-form hazard rates, is what lets one ask how policy reshapes the labour market: Flinn (2006) estimated a search model with Nash bargaining and endogenous contact rates to evaluate the employment and welfare effects of the minimum wage, and Paserman (2008) estimated a job-search model with hyperbolic discounting to evaluate UI policy. Equilibrium search models have also been used to evaluate tax-credit policy directly: Shephard (2017) estimated an equilibrium job-search model with wage posting to study the labour-market effects of the UK Working Families' Tax Credit.

The same dynamic apparatus has been carried across the policy domains that motivate the field—welfare and tax-credit reform, retirement, family formation, and life-cycle labour supply. On welfare and taxation, Keane and Moffitt (1998) estimated an influential static model of participation in multiple programs jointly with labour supply, using simulation methods to handle the combinatorial choice among program bundles; Keane and Wolpin (2010) brought the questions into a fully dynamic model of single mothers' labour supply, marriage, and program participation; Chan (2013) estimated a dynamic model of labour supply and multiple program participation, using variation from actual reforms for identification; Chan (2017) focused on present-biased preferences, recovering heterogeneous time-preference parameters from the time-limit variation in a welfare reform experiment. On the design side, Blundell and Shephard (2012) estimated a structural model of employment and hours for low-income families and used it to characterise the optimal tax-and-transfer schedule, while Blundell, Costa Dias, Meghir and Shaw (2016) estimated a life-cycle model of employment, education, and human capital for women that exploits UK tax and benefit reforms and quantifies the insurance value of alternative programs; Mullins (2026) used a structural model of maternal labour supply and child investment to pool evidence across several welfare-reform experiments in a structural meta-analysis.

Chan and Moffitt (2018) survey this literature. On retirement, French and Jones (2011) estimate a dynamic model in which health insurance and self-insurance shape the timing of retirement, following the structural retirement literature of which Gustman and Steinmeier (1986) and the job-exit model of Berkovec and Stern (1991) are early instances; a parallel semi-structural approach, the option value of continued work introduced by Stock and Wise (1990), trades some of that structure for tractability. Fan, Seshadri and Taber (2024) bring labour supply, human-capital investment, and retirement into a single life-cycle model, letting retirement arise endogenously and the wage path respond to Social Security rules.

A closely related body of work places labour supply inside the family: van der Klaauw (1996) estimated a dynamic model of women's employment jointly with marital-status decisions; Brien, Lillard and Stern (2006) modelled cohabitation, marriage, and divorce as outcomes of a dynamic process of learning about match quality; Adda, Dustmann and Stevens (2017) estimated a life-cycle model of labour supply, fertility, and occupational choice to quantify the career costs of children; Voena (2015) estimated a dynamic limited-commitment model of household labour supply, saving, and divorce, using the staggered adoption of unilateral-divorce and property-division laws to identify how divorce regimes reshape couples' intertemporal choices; Tartari (2015) modelled parents' joint decisions over marital status, conflict, and child investment to identify the effect of divorce on children's cognitive achievement; and Gayle and Shephard (2019) estimated an equilibrium model of marriage, home production, and family labour supply to study optimal joint taxation. The gender earnings gap has been studied with the same tools, for instance by Gayle and Golan (2012), who estimate a dynamic model of experience accumulation and statistical discrimination.

Underlying much of this work is the life-cycle labour supply model, in which consumption, saving, and hours are chosen jointly over the life cycle (see Blundell & MaCurdy, 1999 for a survey). Heckman and MaCurdy (1980) and MaCurdy (1981) estimated its first structural versions—the former for married women's participation in the presence of selection, the latter establishing the canonical approach to recovering the intertemporal substitution elasticity from panel data on hours and wages. This intertemporal substitution (or Frisch) elasticity—the response of hours to an anticipated wage change along the life-cycle path—is a central object of the literature, since it disciplines how labour supply moves with predictable wage growth and, together with wealth effects, informs the response to permanent tax changes. Altonji (1986) showed how consumption data could be used to control for the marginal utility of wealth in estimating it, and Blundell, Meghir and Neves (1993) estimated an intertemporal model of labour supply and consumption that accommodates corner solutions and uncertainty; the micro estimates that emerged were typically small, though Keane and Rogerson (2012) argue that they are nonetheless consistent with much larger aggregate responses once human-capital accumulation and the extensive margin are allowed for. The framework was then extended to incorporate taxation and uninsurable risk: Ziliak and Kniesner (1999) estimate the labour-supply effects of taxes under intertemporal non-separabilities, while Low, Meghir and Pistaferri (2010) and Blundell, Pistaferri and Saporta-Eksten (2016) increasingly treat labour supply as itself a form of insurance—the former separating productivity risk from employment risk and valuing the insurance that social programs provide, the latter showing how the labour supply of two earners smooths permanent wage shocks within the household.

17.5.1 What Aged Well and Less Well

The growth of structural estimation coincided with an expansion of experimental and quasi-experimental evidence in labour economics (discussed in Section 17.4 of this chapter) , and with methodological debates about how the two relate. One question was whether non-experimental estimators could recover what a randomised experiment would. LaLonde (1986) compared selection-correction and other non-experimental estimators against the experimental benchmark of the National Supported Work demonstration and found that they often failed to reproduce the experimental earnings estimate, and that standard specification tests did not reliably detect the failures; the result reinforced Leamer (1983)'s broader warning about the fragility of inferences drawn from non-experimental specification search. Ham and LaLonde (1996) showed that the problem ran deeper for dynamic outcomes: even with experimental data, recovering the effect of training on the duration of employment and unemployment spells requires modelling the sample-selection and initial-conditions problems created by the dynamics, so randomisation alone does not identify these effects. Heckman, Ichimura and Todd (1997), re-examining non-experimental evaluation, attributed much of the earlier discrepancy to non-comparable comparison groups (differences in the surveys, regions, and labour-market histories of treated and untreated units) and showed that matching on a common support of observed characteristics reduced the bias.

A second question concerned the role of explicit behavioural models in policy analysis, and how much structure an empirical analysis should impose. Keane (2010) argued that treatment-effect estimates answer narrowly defined questions, and that structural models are needed for counterfactuals, welfare analysis, and extrapolation beyond the variation observed. Heckman and Urzúa (2010) compared the economic questions that instrumental-variables and structural estimators answer, in the spirit of Marschak's Maxim (Marschak, 1953): the assumptions imposed should be matched to the policy question at hand, since answering a specific question often requires identifying only a combination of structural parameters rather than the complete model. This principle supports the quasi- and semi-structural approaches that occupy a middle ground, identifying a policy-relevant parameter (such as the marginal treatment effect of Heckman & Vytlacil, 2005) without solving a complete dynamic program. How much structure to impose involves a genuine trade-off rather than a single rule. Richer models deliver counterfactuals and welfare measures that more agnostic methods cannot, but they rest on functional-form, distributional, and behavioural assumptions whose validity is harder to assess: a tension often summarised in the observation that all models are wrong, but some are useful (Box, 1976). What any given approach can recover is ultimately a question of identification. French and Taber (2011) survey identification in the standard labour models (the Roy, selection, search, and dynamic discrete choice models), showing how exclusion restrictions, support conditions, functional-form and distributional assumptions, and the longitudinal structure of panel data each contribute, and how some objects, such as the discount factor in a dynamic discrete choice model, are identified only under additional normalisations; Abbring (2010) and Abbring and Heckman (2007) review identification of dynamic discrete choice and dynamic treatment-effect models specifically.

What aged well is the case for structural work when the policy question reaches beyond the specific design that was tried. The negative income tax experiments of the late 1960s and 1970s, which assigned families to guarantees and benefit-reduction rates and measured the labour-supply response, illustrate the limits of experimentation on its own. A given experimental arm identifies the response to that particular guarantee and rate, but not the response to the many untested combinations a national scheme might adopt; and because lowering the benefit-reduction rate raises the income level at which benefits are exhausted, it draws new recipients into the program, so the aggregate effect depends on the resulting composition of participants (Chan & Moffitt, 2018). Nor can a temporary, local experiment reveal the response to a permanent, national programme. A structural model of the underlying income and substitution responses can address these questions: once its parameters are estimated, it can trace out the labour-supply effects across the whole schedule of guarantees and rates, and extrapolate from the experimental setting to the policy of interest (Todd & Wolpin, 2023).

What aged less well is that structural models vary in how heavily they rely on functional-form and distributional assumptions, and that it can be difficult to judge how far a model's conclusions follow from those assumptions. Todd and Wolpin (2023) describe model validation and selection as perhaps the most vexing problem in empirical research. The concern that richly parameterised models can be opaque, and their robustness hard to establish, is a long-standing one in labour econometrics (Moffitt, 1999). A growing body of work responds to it by making the dependence of results on assumptions explicit: Andrews, Gentzkow and Shapiro (2017) measure the sensitivity of parameter estimates to the moments used in estimation, and Christensen and Connault (2023) bound how far counterfactuals can move as the distributional assumptions are relaxed within nonparametric neighbourhoods.

Over the past two decades structural and non-structural approaches have increasingly been combined (Galiani & Pantano, 2022). Heckman (2010) set out a conceptual bridge, relating the treatment-effect parameters that instrumental variables deliver to the policy-relevant parameters a structural model targets, and arguing for putting the policy question ahead of the statistic. One form of combination applies the out-of-sample validation common in statistics: a structural model is estimated on part of the data and its predictions tested against a held-out part, which may be experimental or non-experimental. Todd and Wolpin (2006) estimated a model of schooling and fertility on the control villages of Mexico's PROGRESA experiment and validated it against the experimentally measured impact on the treated villages, and Galiani, Murphy and Pantano (2015) validated a model of neighbourhood choice on a held-out arm of the Moving to Opportunity experiment. A second form uses experimental variation directly in estimation, as an exclusion restriction for identification; Todd and Wolpin (2023) survey the resulting body of work, which now extends across labour, development, public, and urban economics. Viewed as a research programme in the sense of Lakatos (1970), the structural approach has been a developing rather than a static one, its successive refinements and its growing use of experimental and quasi-experimental variation extending the range of questions it answers and the evidence it confronts.

17.6 Earnings Dynamics and Heterogeneity

The structural models of Section 17.5 ask what an individual would choose under a counterfactual policy. A similarly hard question is: how should we describe the earnings an individual receives over a working life? Labour economists have long been interested in earnings mobility, inequality, and persistence, and answering any of these requires a model that separates durable differences between workers from the year-to-year shocks that buffet them. A single cross-section of earnings conflates the two. Two workers with identical lifetime prospects can sit far apart in this year's distribution because one drew a good year and the other a bad one, while two workers in the same place this year may be on sharply diverging paths. Recovering the underlying structure—how much of measured inequality is permanent and how much transitory, how long a shock persists, how much of the dispersion was foreseeable to the worker—is the task that the earnings-dynamics literature set itself.

The panel-data methods that matured since the 1970s made the separation feasible. A worker's log earnings may consist of a persistent component and a transitory component, with the autocovariances determining how fast the transitory part decays. Lillard and Willis (1978) decomposed earnings into an individual fixed effect and a serially correlated transitory error, and Lillard and Weiss (1979) incorporated an individual-specific growth rate (see also Hause, 1980). MaCurdy (1982) introduced a persistent stochastic earnings process, working in first differences, which removed individual fixed effects; he found the persistent component close to a random walk, its shocks accumulating rather than dying out. These two devices for generating persistence—fixed heterogeneous profiles on the one hand, a near-unit-root stochastic component on the other—would later define the central debate of the field. Abowd and Card (1989) modelled the joint covariance structure of earnings and hours in the same error-components tradition. Gottschalk and Moffitt (1994) decomposed the rising cross-sectional variance of US male earnings in the 1970s and 1980s and found that growing transitory variance—earnings instability, not just a widening of durable inequality—accounted for a substantial share of the rise (see also Moffitt & Gottschalk, 2002). M. Baker and Solon (2003), studying Canadian men in longitudinal tax records, allowed heterogeneous individual growth profiles alongside the persistent and transitory shocks. The permanent–transitory distinction had become a tool for reading the anatomy of inequality.

The models were then developed in two directions. The first concerned the nature of labour income risk and its identification. Does the persistence in earnings come from permanent shocks, or from heterogeneous, individual-specific growth profiles; what aspects of data allow them to be distinguished? Guvenen (2009) discussed the potential empirical difficulty in distinguishing restricted income profiles, in which highly persistent shocks do the work, versus heterogeneous income profiles, in which shock persistence is weaker but growth rates differ across workers. The two also imply very different income risk for a forward-looking agent. Browning, Ejrnæs and Alvarez (2010) argued that earnings processes require far more cross-worker parameter variation than standard specifications allow.

The second concerned distributional implications. Meghir and Pistaferri (2004) let the conditional variance of shocks be stochastic, so that risk, not just earnings, varies across workers and over time. Attention also moved from the variance of shocks to their shape. Guvenen, Ozkan and Song (2014) found that what is countercyclical is not the variance of idiosyncratic shocks but their left-skewness: in recessions, large upward movements become rarer and large drops more common. Drawing on records for millions of workers, Guvenen, Karahan, Ozkan and Song (2021) documented that earnings changes depart sharply from log-normality—strong negative skewness, high kurtosis, and persistence that depends on the sign and size of the shock. These features, which the linear-Gaussian framework assumes away, were important. Arellano, Blundell and Bonhomme (2017) developed a nonlinear quantile-based framework for the earnings process, and Hospido (2012) modelled heterogeneity in wage volatility directly; Halvorsen, Holter, Ozkan and Storesletten (2024) attributed the nonlinear mean reversion in earnings more to the dynamics of hours worked than that of wages. Overall, by quantifying how much of a shock passes through to consumption, the earnings process is closely connected to the partial-insurance question of Section 17.5 (see Blundell, Pistaferri & Preston, 2008); Meghir and Pistaferri (2011) and Blundell (2014) survey this junction of earnings, consumption, and life-cycle choice.

Another strand of the literature located the heterogeneity not in the worker alone but in the worker–firm match. The spread of linked employer–employee data from the late 1990s allowed earnings to be decomposed into additive worker and firm components, with Abowd, Kramarz and Margolis (1999) providing the two-way fixed-effects framework that became a standard tool. Its empirical regularities—a large worker contribution to the variance of earnings, a smaller but non-trivial firm contribution, and a modest worker–firm covariance—reshaped how the profession thought about wage dispersion and informed a literature on sorting and firm pay (Song, Price, Guvenen, Bloom & von Wachter, 2019); Card, Heining and Kline (2013) attributed rising wage inequality in West Germany to a combination of worker heterogeneity, firm pay dispersion, and increasingly assortative worker–firm matching. The variance decomposition is subject to an incidental parameters problem (‘limited mobility bias’): when workers move between firms rarely, the firm and covariance terms are biased, requiring corrections (see Kline, Saggio & Sølvsten, 2020). Informative about how much firms matter, the additive model treats mobility as exogenous and is silent on why workers move. Recent work has attempted to open up the black box of what the firm effect represents, using mobility data (Sorkin, 2018).

A structural strand treats the firm component and job mobility as outcomes of an economic process rather than fixed effects to be estimated. Building on the equilibrium-search foundations of Section 17.3, its point of departure is the sequential-auction model of Postel-Vinay and Robin (2002): workers search on and off the job, and an employer learning of an outside offer competes for the worker through Bertrand bidding rather than, as in wage-posting models, ignoring it. Wages rise either through a poaching move or through a counter-offer that retains a worker who has drawn a rival bid, so the firm enters wages not as an additive premium but through a worker's accumulated history of offers. Cahuc, Postel-Vinay and Robin (2006) generalised the auction to strategic bargaining, in which the worker captures a share of the match surplus, and found between-firm competition to matter more for wages than bargaining power. Put to work on wage dynamics, Bagger, Fontaine, Postel-Vinay and Robin (2014) embedded this mechanism in a tractable model with human-capital accumulation and employer heterogeneity and decomposed wage growth into human-capital and job-search parts, within and between jobs: search drives much of early-career growth—a job-shopping phase—after which workers settle into good matches and use outside offers to lever raises. A related literature models sorting between heterogeneous workers and firms: Lise and Postel-Vinay (2020) let workers differ in cognitive, manual, and interpersonal skills, sort into jobs with matching requirements, and accumulate or lose skills with use, quantifying the cost of mismatch along each dimension. Taber and Vejlin (2020) nested Roy-style comparative advantage, compensating differentials, and search frictions in a single model and asked how much of observed wage dispersion each mechanism explains, finding the Roy and human-capital channels to dominate. Bonhomme, Lamadon and Manresa (2019) focussed on the identification and estimation of firm and worker heterogeneity in non-additive models, classifying firms into discrete latent classes and recovering the distribution of worker types within each. Altonji, Smith and Vidangos (2013) estimated an earnings-dynamics model whose equations approximate the decision rules of structural models of job mobility, search, and labour supply while providing a rich statistical description of the income process.

17.6.1 What Aged Well and Less Well

What aged well is the conceptual core: the separation of earnings into a durable component reflecting persistent heterogeneity and a transitory component, and the use of panel data to measure each. That framework has survived every refinement because the distinction it draws is the one economic questions turn on: how much of inequality is permanent, how much income risk a worker faces. The use of administrative records alongside survey panels also aged well, enabling richer empirical features to be revealed.

What aged less well are the restrictive functional forms of the early specifications. The modelling of income profiles has since leaned towards accommodating more heterogeneous features. The linear-Gaussian framework missed the skewness, kurtosis, and sign-dependent persistence that richer data revealed. The early statistical models also said little about why earnings move: they did not distinguish a job-to-job move from a layoff, or a firm effect from a worker effect, and so could not connect earnings dynamics to the labour-market events behind them. The more recent literatures attempt to close the gap between the statistical and economic features of income processes. Earnings-dynamics modelling remains a core pillar of labour econometrics because it has kept evolving with richer data and more flexible specifications, while the conceptual frame laid down in the 1980s still organises the questions.

17.7 Concluding Themes

Three threads run through the history assembled here. The first is that labour markets have repeatedly generated econometric problems of direct policy relevance. Whether generous benefits prolong unemployment, whether schooling raises earnings, what a tax-and-transfer reform would do to labour supply: each question exposed a gap between what the data showed and what the analyst wished to know, and closing that gap required new method rather than new data alone.

The second is that, as a consequence, labour research has been a productive application area for econometrics. Many of the methods considered here—among them the sample selection and treatment-effect models and the duration and transition models—were developed to answer a labour question and then travelled into other fields.

The third is that the methods coexist rather than succeed one another, because the questions do. What any approach can deliver depends on the question asked of it, and the different approaches have increasingly been combined. That the field draws on the whole range of tools it has accumulated reflects the diversity of the questions labour markets pose and the capacity of these methods to answer them.

References

Abadie, A., Athey, S., Imbens, G. W. & Wooldridge, J. M. (2023). When should you adjust standard errors for clustering? *The Quarterly Journal of Economics*, *138*(1), 1–35.

Abbring, J. H. (2010). Identification of dynamic discrete choice models. *Annual Review of Economics*, *2*, 367–394.

Abbring, J. H. & Heckman, J. J. (2007). Econometric evaluation of social programs, part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation. In J. J. Heckman

& E. E. Leamer (Eds.), *Handbook of Econometrics* (Vol. 6B, chap. 72). Amsterdam: Elsevier.

Abowd, J. M. & Card, D. (1989). On the covariance structure of earnings and hours changes. *Econometrica*, *57*(2), 411–445.

Abowd, J. M., Kramarz, F. & Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, *67*(2), 251–333.

Adda, J., Dustmann, C. & Stevens, K. (2017). The career costs of children. *Journal of Political Economy*, *125*(2), 293–337.

Agostinelli, F. & Wiswall, M. (2025). Estimating the technology of children's skill formation. *Journal of Political Economy*, *133*(3).

Aguirregabiria, V. & Mira, P. (2010). Dynamic discrete choice structural models: A survey. *Journal of Econometrics*, *156*(1), 38–67.

Altonji, J. G. (1986). Intertemporal substitution in labor supply: Evidence from micro data. *Journal of Political Economy*, *94*(3, Part 2), S176–S215.

Altonji, J. G., Smith, A. A. & Vidangos, I. (2013). Modeling earnings dynamics. *Econometrica*, *81*(4), 1395–1454.

Anderson, T. W. & Rubin, H. (1949). Estimation of the parameters of a single equation in a complete system of stochastic equations. *The Annals of Mathematical Statistics*, *20*(1), 46–63.

Andrews, I., Gentzkow, M. & Shapiro, J. M. (2017). Measuring the sensitivity of parameter estimates to estimation moments. *The Quarterly Journal of Economics*, *132*(4), 1553–1592.

Angrist, J. D. & Evans, W. N. (1998). Children and their parents' labor supply: Evidence from exogenous variation in family size. *American Economic Review*, *88*(3), 450–477.

Angrist, J. D. & Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, *106*(4), 979–1014.

Angrist, J. D. & Krueger, A. B. (1999). Empirical strategies in labor economics. In *Handbook of Labor Economics* (Vol. 3A, pp. 1277–1366).

Angrist, J. D. & Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic Perspectives*, *15*(4), 69–85.

Angrist, J. D. & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton, NJ: Princeton University Press.

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.

Arcidiacono, P. (2005). Affirmative action in higher education: How do admission and financial aid rules affect future earnings? *Econometrica*, *73*(5), 1477–1524.

Arcidiacono, P., Aucejo, E., Maurel, A. & Ransom, T. (2025). College attrition and the dynamics of information revelation. *Journal of Political Economy*, *133*(1), 53–110.

Arcidiacono, P. & Miller, R. A. (2011). Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica*, *79*(6), 1823–1867.

Arellano, M., Blundell, R. & Bonhomme, S. (2017). Earnings and consumption dynamics: A nonlinear panel data framework. *Econometrica*, *85*(3), 693–734.

Ashenfelter, O. C. (1978). Estimating the effect of training programs on earnings. *The Review of Economics and Statistics*, *60*(1), 47–57.

Ashenfelter, O. C. & Card, D. (1985). Using the longitudinal structure of earnings to estimate the effect of training programs. *The Review of Economics and Statistics*, *67*(4), 648–660.

Athey, S. & Imbens, G. W. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, *31*(2), 3–32.

Atkinson, A. B. & Micklewright, J. (1991). Unemployment compensation and labor market transitions: A critical review. *Journal of Economic Literature*, *29*(4), 1679–1727.

Autor, D. H., Dorn, D. & Hanson, G. H. (2013). The china syndrome: Local labor market effects of import competition in the united states. *American Economic Review*, *103*(6), 2121–2168.

Bagger, J., Fontaine, F., Postel-Vinay, F. & Robin, J.-M. (2014). Tenure, experience, human capital, and wages: A tractable equilibrium search model of wage dynamics. *American Economic Review*, *104*(6), 1551–1596.

Baker, A. C., Callaway, B., Cunningham, S., Goodman-Bacon, A. & Sant'Anna, P. H. C. (2026). Difference-in-Differences designs: A practitioner's guide. *Journal of Economic Literature*, *64*(2), 498–557.

Baker, M. & Solon, G. R. (2003). Earnings dynamics and inequality among Canadian men, 1976–1992: Evidence from longitudinal income tax records. *Journal of Labor Economics*, *21*(2), 267–288.

Bartik, T. J. (1991). *Who benefits from state and local economic development policies?* W.E. Upjohn Institute.

Berkovec, J. & Stern, S. (1991). Job exit behavior of older men. *Econometrica*, *59*(1), 189–210.

Bertrand, M., Duflo, E. & Mullainathan, S. (2004). How much should we trust Differences-in-Differences estimates? *The Quarterly Journal of Economics*, *119*(1), 249–275.

Björklund, A. & Moffitt, R. A. (1987). The estimation of wage gains and welfare gains in self-selection models. *The Review of Economics and Statistics*, *69*(1), 42–49.

Blanchard, O. J. & Katz, L. F. (1992). Regional evolutions. *Brookings Papers on Economic Activity*, *23*(1), 1–76.

Blandhol, C., Bonney, J., Mogstad, M. & Torgovitsky, A. (2026). When is TSLS actually LATE? *The Review of Economic Studies*, 1–35.

Blundell, R. (2014). Income dynamics and life-cycle inequality: Mechanisms and controversies. *The Economic Journal*, *124*(576), 289–318.

Blundell, R., Costa Dias, M., Meghir, C. & Shaw, J. (2016). Female labor supply, human capital, and welfare reform. *Econometrica*, *84*(5), 1705–1753.

Blundell, R., Duncan, A. & Meghir, C. (1998). Estimating labor supply responses using tax reforms. *Econometrica*, *66*(4), 827–861.

Blundell, R. & MaCurdy, T. E. (1999). Labor supply: A review of alternative approaches. In *Handbook of Labor Economics* (Vol. 3A, pp. 1559–1695). Elsevier.

Blundell, R., Meghir, C. & Neves, P. (1993). Labour supply and intertemporal substitution. *Journal of Econometrics*, *59*(1–2), 137–160.

Blundell, R., Pistaferri, L. & Preston, I. (2008). Consumption inequality and partial insurance. *American Economic Review*, *98*(5), 1887–1921.

Blundell, R., Pistaferri, L. & Saporta-Eksten, I. (2016). Consumption inequality and family labor supply. *American Economic Review*, *106*(2), 387–435.

Blundell, R. & Shephard, A. (2012). Employment, hours of work and the optimal taxation of low-income families. *The Review of Economic Studies*, *79*(2), 481–510.

Bonhomme, S., Lamadon, T. & Manresa, E. (2019). A distributional framework for matched employer employee data. *Econometrica*, *87*(3), 699–739.

Bound, J. & Holzer, H. J. (2000). Demand shifts, population adjustments, and labor market outcomes during the 1980s. *Journal of Labor Economics*, *18*(1), 20–54.

Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association*, *71*(356), 791–799.

Brien, M. J., Lillard, L. A. & Stern, S. (2006). Cohabitation, marriage, and divorce in a model of match quality. *International Economic Review*, *47*(2), 451–494.

Browning, M., Ejrnæs, M. & Alvarez, J. (2010). Modelling income processes with lots of heterogeneity. *The Review of Economic Studies*, *77*(4), 1353–1381.

Bruins, M., Duffy, J. A., Keane, M. P. & Smith, A. A. (2018). Generalized indirect inference for discrete choice models. *Journal of Econometrics*, *205*(1), 177–203.

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

Cahuc, P., Postel-Vinay, F. & Robin, J.-M. (2006). Wage bargaining with on-the-job search: Theory and evidence. *Econometrica*, *74*(2), 323–364.

Callaway, B. (2023). Difference-in-Differences for policy evaluation. In K. F. Zimmermann (Ed.), *Handbook of Labor, Human Resources and Population Economics* (pp. 1–61). Cham: Springer.

Calonico, S., Cattaneo, M. D. & Titiunik, R. (2014a). Robust data-driven inference in the regression-discontinuity design. *The Stata Journal*, *14*(4), 909–946.

Calonico, S., Cattaneo, M. D. & Titiunik, R. (2014b). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, *82*(6), 2295–2326.

Card, D. (1990). The impact of the Mariel boatlift on the miami labor market. *Industrial and Labor Relations Review*, *43*(2), 245–257.

Card, D. (1999). The causal effect of education on earnings. In O. C. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics* (Vol. 3A, pp. 1801–1863). Amsterdam: Elsevier.

Card, D. (2022). Design-based research in empirical microeconomics. *American Economic Review*, *112*(6), 1773–1781.

Card, D., Chetty, R. & Weber, A. (2007). Cash-on-Hand and Competing Models of Intertemporal Behavior: New Evidence from the Labor Market. *The Quarterly Journal of Economics*, *122*(4), 1511–1560.

Card, D., Heining, J. & Kline, P. (2013). Workplace heterogeneity and the rise of West German wage inequality. *The Quarterly Journal of Economics*, *128*(3), 967–1015.

Card, D., Kluve, J. & Weber, A. (2018). What works? a meta analysis of recent active labor market program evaluations. *Journal of the European Economic Association*, *16*(3), 894–931.

Card, D. & Krueger, A. B. (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., Lee, D. S., Pei, Z. & Weber, A. (2015). Inference on causal effects in a generalized regression kink design. *Econometrica*, *83*(6), 2453–2483.

Card, D., Lee, D. S., Pei, Z. & Weber, A. (2017). Regression kink design: Theory and practice. In M. D. Cattaneo & J. C. Escanciano (Eds.), *Regression discontinuity designs: Theory and applications* (pp. 341–382). Bingley, UK: Emerald Publishing Limited.

Carneiro, P., Heckman, J. J. & Vytlacil, E. J. (2011). Estimating marginal returns to education. *American Economic Review*, *101*(6), 2754–2781.

Carneiro, P., Lokshin, M. & Umapathi, N. (2017). Average and marginal returns to upper secondary schooling in Indonesia. *Journal of Applied Econometrics*, *32*(1), 16–36.

Castro, R., Lange, F. & Poschke, M. (2025). Labor force transitions. In C. Dustmann & T. Lemieux (Eds.), *Handbook of Labor Economics* (Vol. 6, pp. 819–881). Elsevier.

Cattaneo, M. D., Frandsen, B. R. & Titiunik, R. (2015). Randomization inference in the regression discontinuity design: An application to party advantages in the U.S. senate. *Journal of Causal Inference*, *3*(1), 1–24.

Cattaneo, M. D. & Titiunik, R. (2022). Regression discontinuity designs. *Annual Review of Economics*, *14*, 821–851.

Chan, M. K. (2013). A dynamic model of welfare reform. *Econometrica*, *81*(3), 941–1001.

Chan, M. K. (2017). Welfare dependence and self-control: An empirical analysis. *The Review of Economic Studies*, *84*(4), 1379–1423.

Chan, M. K. & Moffitt, R. A. (2018). Welfare reform and the labor market. *Annual Review of Economics*, *10*, 347–381.

Chetty, R. (2009). Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods. *Annual Review of Economics*, *1*(1), 451–488.

Christensen, T. & Connault, B. (2023). Counterfactual sensitivity and robustness. *Econometrica*, *91*(1), 263–298.

Cook, T. D. (2008). "Waiting for Life to Arrive": A history of the regression-discontinuity design in psychology, statistics and economics. *Journal of Econometrics*, *142*(2), 636–654.

Cornelissen, T., Dustmann, C., Raute, A. & Schönberg, U. (2016). From LATE to

MTE: Alternative methods for the evaluation of policy interventions. *Labour Economics*, *41*, 47–60.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, *34*(2), 187–220.

Cunha, F. & Heckman, J. J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, *43*(4), 738–782.

Cunha, F., Heckman, J. J. & Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, *78*(3), 883–931.

Currie, J., Kleven, H. J. & Zwiers, E. (2020). Technology and big data are changing economics: Mining text to track methods. *AEA Papers and Proceedings*, *110*, 42–48.

de Chaisemartin, C. & D'Haultfœuille, X. (2023). Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. *The Econometrics Journal*, *26*(3), C1–C30.

Deming, D. & Silliman, M. (2025). Skills and human capital in the labor market. In *Handbook of Labor Economics* (Vol. 6, pp. 115–157).

Donald, S. G. & Lang, K. (2007). Inference with Difference-in-Differences and other panel data. *The Review of Economics and Statistics*, *89*(2), 221–233.

Doyle, J. J. J. (2007). Child protection and child outcomes: Measuring the effects of foster care. *American Economic Review*, *97*(5), 1583–1610.

Dube, A. & Lindner, A. (2024). Minimum wages in the 21st century. In C. Dustmann & T. Lemieux (Eds.), *Handbook of Labor Economics* (Vol. 5, pp. 261–383). Elsevier.

Duflo, E. (2001). Schooling and labor market consequences of school construction in indonesia: Evidence from an unusual policy experiment. *American Economic Review*, *91*(4), 795–813.

Eckstein, Z. & van den Berg, G. J. (2007). Empirical labor search: A survey. *Journal of Econometrics*, *136*(2), 531–564.

Eckstein, Z. & Wolpin, K. I. (1989a). Dynamic labour force participation of married women and endogenous work experience. *The Review of Economic Studies*, *56*(3), 375–390.

Eckstein, Z. & Wolpin, K. I. (1989b). The specification and estimation of dynamic stochastic discrete choice models: A survey. *Journal of Human Resources*, *24*(4), 562–598.

Eckstein, Z. & Wolpin, K. I. (1990). Estimating a market equilibrium search model from panel data on individuals. *Econometrica*, *58*(4), 783–808.

Fan, X., Seshadri, A. & Taber, C. (2024). Estimation of a life-cycle model with human capital, labor supply, and retirement. *Journal of Political Economy*, *132*(1), 48–95.

Flinn, C. J. (2006). Minimum wage effects on labor market outcomes under search, matching, and endogenous contact rates. *Econometrica*, *74*(4), 1013–1062.

French, E. & Jones, J. B. (2011). The effects of health insurance and self-insurance on retirement behavior. *Econometrica*, *79*(3), 693–732.

French, E. & Song, J. (2014). The effect of disability insurance receipt on labor supply. *American Economic Journal: Economic Policy*, *6*(2), 291–337.

French, E. & Taber, C. (2011). Identification of models of the labor market. In *Handbook of Labor Economics* (Vol. 4A, pp. 537–617). Elsevier.

Freyaldenhoven, S., Hansen, C. B. & Shapiro, J. M. (2019). Pre-event trends in the panel event-study design. *American Economic Review*, *109*(9), 3307–3338.

Fu, C. (2014). Equilibrium tuition, applications, admissions, and enrollment in the college market. *Journal of Political Economy*, *122*(2), 225–281.

Galiani, S., Murphy, A. & Pantano, J. (2015). Estimating neighborhood choice models: Lessons from a housing assistance experiment. *American Economic Review*, *105*(11), 3385–3415.

Galiani, S. & Pantano, J. (2022). Structural models. In K. F. Zimmermann (Ed.), *Handbook of Labor, Human Resources and Population Economics.* Cham: Springer.

Gayle, G.-L. & Golan, L. (2012). Estimating a dynamic adverse-selection model: Labour-force experience and the changing gender earnings gap 1968–1997. *The Review of Economic Studies*, *79*(1), 227–267.

Gayle, G.-L. & Shephard, A. (2019). Optimal taxation, marriage, home production, and family labor supply. *Econometrica*, *87*(1), 291–326.

Goldberger, A. S. (1972a). *Selection bias in evaluating treatment effects: Some formal illustrations.* (Unpublished manuscript, Madison, WI)

Goldberger, A. S. (1972b). *Selection bias in evaluating treatment effects: The case of interaction.* (Unpublished manuscript, Madison, WI)

Goldsmith-Pinkham, P., Sorkin, I. & Swift, H. (2020). Bartik instruments: What, when, why, and how. *American Economic Review*, *110*(8), 2586–2624.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, *225*(2), 254–277.

Gottschalk, P. & Moffitt, R. A. (1994). The growth of earnings instability in the U.S. labor market. *Brookings Papers on Economic Activity*, *1994*(2), 217–272.

Gourieroux, C., Monfort, A. & Renault, E. (1993). Indirect inference. *Journal of Applied Econometrics*, *8*(S1), S85–S118.

Griliches, Z. (1977). Estimating the returns to schooling: Some econometric problems. *Econometrica*, *45*(1), 1–22.

Gronau, R. (1974). Wage comparisons–a selectivity bias. *Journal of Political Economy*, *82*(6), 1119–1143.

Gruber, J. (1994). The incidence of mandated maternity benefits. *The American Economic Review*, *84*(3), 622–641.

Gustman, A. L. & Steinmeier, T. L. (1986). A structural retirement model. *Econometrica*, *54*(3), 555–584.

Guvenen, F. (2009). An empirical investigation of labor income processes. *Review of Economic Dynamics*, *12*(1), 58–79.

Guvenen, F., Karahan, F., Ozkan, S. & Song, J. (2021). What do data on millions of U.S. workers reveal about lifecycle earnings dynamics? *Econometrica*, *89*(5), 2303–2339.

Guvenen, F., Ozkan, S. & Song, J. (2014). The nature of countercyclical income risk. *Journal of Political Economy*, *122*(3), 621–660.

Hahn, J., Todd, P. E. & van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, *69*(1), 201–209.

Halvorsen, E., Holter, H. A., Ozkan, S. & Storesletten, K. (2024). Dissecting idiosyncratic earnings risk. *Journal of the European Economic Association*, *22*(2), 617–668.

Ham, J. C. & LaLonde, R. J. (1996). The effect of sample selection and initial conditions in duration models: Evidence from experimental data on training. *Econometrica*, *64*(1), 175–205.

Hansen, C. B. (2007a). Asymptotic properties of a robust variance matrix estimator for panel data when T is large. *Journal of Econometrics*, *141*(2), 597–620.

Hansen, C. B. (2007b). Generalized least squares inference in panel and multilevel models with serial correlation and fixed effects. *Journal of Econometrics*, *140*(2), 670–694.

Harmon, C. & Walker, I. (1995). Estimates of the economic return to schooling for the united kingdom. *American Economic Review*, *85*(5), 1278–1286.

Hause, J. C. (1980). The fine structure of earnings and the on-the-job training hypothesis. *Econometrica*, *48*(4), 1013–1029.

Heckman, J. J. (1974). Shadow prices, market wages, and labor supply. *Econometrica*, *42*(4), 679–694.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement*, *5*(4), 475–492.

Heckman, J. J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica*, *46*(4), 931–959.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, *47*(1), 153–161.

Heckman, J. J. (2010). Building bridges between structural and program evaluation approaches to evaluating policy. *Journal of Economic Literature*, *48*(2), 356–398.

Heckman, J. J., Ichimura, H. & Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies*, *64*(4), 605–654.

Heckman, J. J. & MaCurdy, T. E. (1980). A life cycle model of female labour supply. *The Review of Economic Studies*, *47*(1), 47–74.

Heckman, J. J. & Robb, R. (1985). Alternative methods for evaluating the impact of interventions. In J. J. Heckman & B. Singer (Eds.), *Longitudinal analysis of labor market data* (pp. 156–245). Cambridge University Press.

Heckman, J. J. & Urzúa, S. (2010). Comparing IV with structural models: What simple IV can and cannot identify. *Journal of Econometrics*, *156*(1), 27–37.

Heckman, J. J. & Vytlacil, E. J. (1999). Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences*, *96*(8), 4730–4734.

Heckman, J. J. & Vytlacil, E. J. (2001a). Local instrumental variables. In C. Hsiao, K. Morimune & J. L. Powell (Eds.), *Nonlinear statistical modeling: Proceedings of the thirteenth international symposium in economic theory and econometrics: Essays in honor of takeshi amemiya* (pp. 1–46). Cambridge: Cambridge University Press.

Heckman, J. J. & Vytlacil, E. J. (2001b). Policy-relevant treatment effects. *American Economic Review*, *91*(2), 107–111.

Heckman, J. J. & Vytlacil, E. J. (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica*, *73*(3), 669–738.

Heckman, J. J. & Vytlacil, E. J. (2007). Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. In J. J. Heckman & E. E. Leamer (Eds.), *Handbook of Econometrics* (Vol. 6B, pp. 4875–5143). Amsterdam: Elsevier.

Honoré, B. E. & Hu, L. (2020). Selection without exclusion. *Econometrica*, *88*(3), 1007–1029.

Hospido, L. (2012). Modelling heterogeneity and dynamics in the volatility of individual wages. *Journal of Applied Econometrics*, *27*(3), 386–414.

Hotz, V. J. & Miller, R. A. (1988). An empirical analysis of life cycle fertility and female labor supply. *Econometrica*, *56*(1), 91–118.

Hotz, V. J. & Miller, R. A. (1993). Conditional choice probabilities and the estimation of dynamic models. *The Review of Economic Studies*, *60*(3), 497–529.

Imai, S. & Keane, M. P. (2004). Intertemporal labor supply and human capital accumulation. *International Economic Review*, *45*(2), 601–641.

Imbens, G. W. (2010). Better LATE than nothing: Some comments on Deaton (2009) and Heckman and Urzua (2009). *Journal of Economic Literature*, *48*(2), 399–423.

Imbens, G. W. & Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, *62*(2), 467–475.

Imbens, G. W. & Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, *142*(2), 615–635.

Imbens, G. W. & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, *47*(1), 5–86.

Iskhakov, F., Jørgensen, T. H., Rust, J. & Schjerning, B. (2017). The endogenous grid method for discrete-continuous dynamic choice models with (or without) taste shocks. *Quantitative Economics*, *8*(2), 317–365.

Jacobson, L. S., LaLonde, R. J. & Sullivan, D. G. (1993). Earnings losses of displaced workers. *American Economic Review*, *83*(4), 685–709.

Jovanovic, B. (1979). Job matching and the theory of turnover. *Journal of Political Economy*, *87*(5, Part 1), 972–990.

Kahn-Lang, A. & Lang, K. (2020). The promise and pitfalls of differences-in-differences: Reflections on 16 and pregnant and other applications. *Journal of Business & Economic Statistics*, *38*(3), 613–620.

Kaplan, E. L. & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, *53*(282), 457–

481.

Katz, L. F. & Meyer, B. D. (1990). Unemployment insurance, recall expectations, and unemployment outcomes. *The Quarterly Journal of Economics*, *105*(4), 973–1002.

Keane, M. P. (2010). Structural vs. atheoretic approaches to econometrics. *Journal of Econometrics*, *156*(1), 3–20.

Keane, M. P. & Moffitt, R. A. (1998). A structural model of multiple welfare program participation and labor supply. *International Economic Review*, *39*(3), 553–589.

Keane, M. P. & Neal, T. (2023). Instrument strength in IV estimation and inference: A guide to theory and practice. *Journal of Econometrics*, *235*(2), 1625–1653.

Keane, M. P. & Rogerson, R. (2012). Micro and macro labor supply elasticities: A reassessment of conventional wisdom. *Journal of Economic Literature*, *50*(2), 464–476.

Keane, M. P., Todd, P. E. & Wolpin, K. I. (2011). The structural estimation of behavioral models: Discrete choice dynamic programming methods and applications. In *Handbook of Labor Economics* (Vol. 4A, pp. 331–461). Elsevier.

Keane, M. P. & Wolpin, K. I. (1994). The solution and estimation of discrete choice dynamic programming models by simulation and interpolation: Monte carlo evidence. *The Review of Economics and Statistics*, *76*(4), 648–672.

Keane, M. P. & Wolpin, K. I. (1997). The career decisions of young men. *Journal of Political Economy*, *105*(3), 473–522.

Keane, M. P. & Wolpin, K. I. (2001). The effect of parental transfers and borrowing constraints on educational attainment. *International Economic Review*, *42*(4), 1051–1103.

Keane, M. P. & Wolpin, K. I. (2010). The role of labor and marriage markets, preference heterogeneity, and the welfare system in the life cycle decisions of black, hispanic, and white women. *International Economic Review*, *51*(3), 851–892.

Kézdi, G. (2004). Robust standard error estimation in Fixed-Effects panel models. *Hungarian Statistical Review*, *82*(Special English Volume 9), 95–116.

Khanna, G. (2023). Large-scale education reform in general equilibrium: Regression discontinuity evidence from india. *Journal of Political Economy*, *131*(2), 549–591.

Killingsworth, M. R. & Heckman, J. J. (1986). Female labor supply: A survey. *Handbook of Labor Economics*, *1*(C), 103-204.

Kleven, H. J. (2021). Sufficient statistics revisited. *Annual Review of Economics*, *13*(1), 515–538.

Kleven, H. J., Landais, C. & Søgaard, J. E. (2019). Children and Gender Inequality: Evidence from Denmark. *American Economic Journal: Applied Economics*, *11*(4), 181-209.

Kline, P., Saggio, R. & Sølvsten, M. (2020). Leave-out estimation of variance components. *Econometrica*, *88*(5), 1859–1898.

Kling, J. R. (2006). Incarceration length, employment, and earnings. *American Economic Review*, *96*(3), 863–876.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91–195). Cambridge: Cambridge University Press.

Lalive, R. (2008). How do extended benefits affect unemployment duration? a regression discontinuity approach. *Journal of Econometrics*, *142*(2), 785–806.

LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review*, *76*(4), 604–620.

Lancaster, T. (1979). Econometric methods for the duration of unemployment. *Econometrica*, *47*(4), 939–956.

Leamer, E. E. (1983). Let's take the con out of econometrics. *American Economic Review*, *73*(1), 31–43.

Le Barbanchon, T., Schmieder, J. F. & Weber, A. (2024). Job search, unemployment insurance, and active labor market policies. In C. Dustmann & T. Lemieux (Eds.), *Handbook of Labor Economics* (Vol. 5, pp. 435–580). Elsevier.

Lechner, M. (2011). The estimation of causal effects by Difference-in-Difference methods. *Foundations and Trends in Econometrics*, *4*(3), 165–224.

Lee, D. S. & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, *48*(2), 281–355.

Lee, L.-F. (1978). Unionism and wage rates: A simultaneous equations model with qualitative and limited dependent variables. *International Economic Review*, *19*(2), 415–433.

Lee, L.-F. (1979). Identification and estimation in binary choice models with limited (censored) dependent variables. *Econometrica*, *47*(4), 977–996.

Lester, R. A. (1946). Shortcomings of marginal analysis for wage-employment problems. *American Economic Review*, *36*(1), 63–82.

Lillard, L. A. & Weiss, Y. (1979). Components of variation in panel earnings data: American scientists 1960–1970. *Econometrica*, *47*(2), 437–454.

Lillard, L. A. & Willis, R. J. (1978). Dynamic aspects of earnings mobility. *Econometrica*, *46*(5), 985–1012.

Lise, J. & Postel-Vinay, F. (2020). Multidimensional skills, sorting, and human capital accumulation. *American Economic Review*, *110*(8), 2328–2376.

Low, H., Meghir, C. & Pistaferri, L. (2010). Wage risk and employment risk over the life cycle. *American Economic Review*, *100*(4), 1432–1467.

MaCurdy, T. E. (1981). An empirical model of labor supply in a life-cycle setting. *Journal of Political Economy*, *89*(6), 1059–1085.

MaCurdy, T. E. (1982). The use of time series processes to model the error structure of earnings in a longitudinal data analysis. *Journal of Econometrics*, *18*(1), 83–114.

Maestas, N., Mullen, K. J. & Strand, A. (2013). Does disability insurance receipt discourage work? using examiner assignment to estimate causal effects of ssdi receipt. *American Economic Review*, *103*(5), 1797–1829.

Marschak, J. (1953). Economic measurements for policy and prediction. In W. C. Hood & T. C. Koopmans (Eds.), *Studies in econometric method* (pp.

1–26). New York: Wiley.
McCall, J. J. (1970). Economics of information and job search. *The Quarterly Journal of Economics*, *84*(1), 113–126.
McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, *57*(5), 995–1026.
Meghir, C. & Pistaferri, L. (2004). Income variance dynamics and heterogeneity. *Econometrica*, *72*(1), 1–32.
Meghir, C. & Pistaferri, L. (2011). Earnings, consumption and life cycle choices. In O. C. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics* (Vol. 4, pp. 773–854). Elsevier.
Meyer, B. D. (1990). Unemployment insurance and unemployment spells. *Econometrica*, *58*(4), 757–782.
Meyer, B. D. (1995). Natural and quasi-experiments in economics. *Journal of Business & Economic Statistics*, *13*(2), 151–161.
Miller, D. L. (2023). An introductory guide to event study models. *Journal of Economic Perspectives*, *37*(2), 203–230.
Miller, R. A. (1984). Job matching and occupational choice. *Journal of Political Economy*, *92*(6), 1086–1120.
Mincer, J. (1974). *Schooling, experience, and earnings*. New York: National Bureau of Economic Research.
Moffitt, R. A. (1984). Profiles of fertility, labour supply and wages of married women: A complete life-cycle model. *The Review of Economic Studies*, *51*(2), 263–278.
Moffitt, R. A. (1985). Unemployment insurance and the distribution of unemployment spells. *Journal of Econometrics*, *28*(1), 85–101.
Moffitt, R. A. (1999). New developments in econometric methods for labor market analysis. In O. C. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics* (Vol. 3A, pp. 1367–1397). Amsterdam: Elsevier.
Moffitt, R. A. & Gottschalk, P. (2002). Trends in the transitory variance of earnings in the united states. *The Economic Journal*, *112*(478), C68–C73.
Mogstad, M., Santos, A. & Torgovitsky, A. (2018). Using instrumental variables for inference about policy relevant treatment parameters. *Econometrica*, *86*(5), 1589–1619.
Mogstad, M. & Torgovitsky, A. (2024). Instrumental variables with unobserved heterogeneity in treatment effects. In *Handbook of Labor Economics* (Vol. 5, pp. 1–114).
Mortensen, D. T. (1970). Job search, the duration of unemployment, and the Phillips curve. *The American Economic Review*, *60*(5), 847–862.
Mullins, J. (2026). A structural meta-analysis of welfare reform experiments and their impacts on children. *Journal of Political Economy*, *134*(1).
Nevo, A. & Whinston, M. D. (2010). Taking the dogma out of econometrics: Structural modeling and credible inference. *Journal of Economic Perspectives*, *24*(2), 69–82.

Newey, W. K., Powell, J. L. & Walker, J. R. (1990). Semiparametric estimation of selection models: Some empirical results. *American Economic Review*, *80*(2), 324–328.

Nielsen, H. S., Sørensen, T. & Taber, C. (2010). Estimating the effect of student aid on college enrollment: Evidence from a government grant policy reform. *American Economic Journal: Economic Policy*, *2*(2), 185–215.

Obenauer, M. L. & von der Nienburg, B. M. (1915). *Effect of minimum-wage determinations in Oregon* (Bulletin of the United States Bureau of Labor Statistics No. 176). Washington, DC: United States Bureau of Labor Statistics.

Oreopoulos, P. (2006). Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review*, *96*(1), 152–175.

Pakes, A. & Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica*, *57*(5), 1027–1057.

Paserman, M. D. (2008). Job search and hyperbolic discounting: Structural estimation and policy evaluation. *The Economic Journal*, *118*(531), 1418–1452.

Pei, Z., Lee, D. S., Card, D. & Weber, A. (2022). Local polynomial order in regression discontinuity designs. *Journal of Business & Economic Statistics*, *40*(3), 1259–1267.

Postel-Vinay, F. & Robin, J.-M. (2002). Equilibrium wage dispersion with worker and employer heterogeneity. *Econometrica*, *70*(6), 2295–2350.

Rambachan, A. & Roth, J. (2023). A more credible approach to parallel trends. *Review of Economic Studies*, *90*(5), 2555–2591.

Rosenzweig, M. R. & Wolpin, K. I. (1980). Testing the quantity-quality fertility model: The use of twins as a natural experiment. *Econometrica*, *48*(1), 227–240.

Rosenzweig, M. R. & Wolpin, K. I. (2000). Natural "natural experiments" in economics. *Journal of Economic Literature*, *38*(4), 827–874.

Roth, J. (2022). Pretest with caution: Event-study estimates after testing for parallel trends. *American Economic Review: Insights*, *4*(3), 305–322.

Roth, J., Sant'Anna, P. H. C., Bilinski, A. & Poe, J. (2023). What's trending in Difference-in-Differences? a synthesis of the recent econometrics literature. *Journal of Econometrics*, *235*(2), 2218–2244.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, *66*(5), 688–701.

Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of harold zurcher. *Econometrica*, *55*(5), 999–1033.

Shephard, A. (2017). Equilibrium search and tax credit reform. *International Economic Review*, *58*(4), 1047–1088.

Solon, G. R. (1985). Work incentive effects of taxing unemployment benefits. *Econometrica*, *53*(2), 295–306.

Song, J., Price, D. J., Guvenen, F., Bloom, N. & von Wachter, T. (2019). Firming up inequality. *The Quarterly Journal of Economics*, *134*(1), 1–50.

Sorkin, I. (2018). Ranking firms using revealed preference. *The Quarterly Journal of Economics*, *133*(3), 1331–1393.

Stafford, F. P. (1986). Forestalling the demise of empirical economics: The role of microdata in labor economics research. In *Handbook of Labor Economics* (Vol. 1, pp. 387–423). Amsterdam: Elsevier.

Staiger, D. & Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, *65*(3), 557–586.

Stock, J. H. & Trebbi, F. (2003). Retrospectives: Who invented instrumental variable regression? *Journal of Economic Perspectives*, *17*(3), 177–194.

Stock, J. H. & Wise, D. A. (1990). Pensions, the option value of work, and retirement. *Econometrica*, *58*(5), 1151–1180.

Stock, J. H. & Yogo, M. (2005). Testing for weak instruments in linear IV regression. In D. W. K. Andrews & J. H. Stock (Eds.), *Identification and inference for econometric models: Essays in honor of thomas rothenberg* (pp. 80–108). New York: Cambridge University Press.

Sun, L. & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, *225*(2), 175–199.

Taber, C. & Vejlin, R. (2020). Estimation of a Roy/Search/Compensating differential model of the labor market. *Econometrica*, *88*(3), 1031–1069.

Tartari, M. (2015). Divorce and the cognitive achievement of children. *International Economic Review*, *56*(2), 597–645.

Thistlethwaite, D. L. & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, *51*(6), 309–317.

Todd, P. E. & Wolpin, K. I. (2006). Assessing the impact of a school subsidy program in mexico: Using a social experiment to validate a dynamic behavioral model of child schooling and fertility. *American Economic Review*, *96*(5), 1384–1417.

Todd, P. E. & Wolpin, K. I. (2023). The best of both worlds: Combining randomized controlled trials with structural modeling. *Journal of Economic Literature*, *61*(1), 41–85.

van den Berg, G. J. (2001). Duration models: Specification, identification and multiple durations. In J. J. Heckman & E. E. Leamer (Eds.), *Handbook of Econometrics* (Vol. 5, pp. 3381–3460). Elsevier.

van der Klaauw, W. (1996). Female labour supply and marital status decisions: A life-cycle model. *The Review of Economic Studies*, *63*(2), 199–235.

van der Klaauw, W. (2008). Regression-discontinuity analysis: A survey of recent developments in economics. *Labour*, *22*(2), 219–245.

Vella, F. (1998). Estimating models with sample selection bias: A survey. *Journal of Human Resources*, *33*(1), 127–169.

Voena, A. (2015). Yours, mine, and ours: Do divorce laws affect the intertemporal behavior of married couples? *American Economic Review*, *105*(8), 2295–2332.

Vytlacil, E. J. (2002). Independence, monotonicity, and latent index models: An equivalence result. *Econometrica*, *70*(1), 331–341.

Wolpin, K. I. (1984). An estimable dynamic stochastic model of fertility and child mortality. *Journal of Political Economy*, *92*(5), 852–874.

Wolpin, K. I. (1987). Estimating a structural search model: The transition from school to work. *Econometrica*, *55*(4), 801–817.

Wooldridge, J. M. (2003). Cluster-sample methods in applied econometrics. *American Economic Review*, *93*(2), 133–138. (Papers and Proceedings)

Wright, P. G. (1928). *The tariff on animal and vegetable oils*. New York: MacMillan.

Ziliak, J. P. & Kniesner, T. J. (1999). Estimating life-cycle labor supply tax effects. *Journal of Political Economy*, *107*(2), 326–359.

Index