37

James J. Heckman (1944–)

Richard Blundell and Flávio Cunha

1 Introduction to the First Six Sections[1,2]

James Heckman's contributions to economics and econometrics are extensive and, at the time when he received the Nobel Prize when these first six sections were written, had already spanned nearly three decades. They continue to proceed apace, stretching the boundary between econometric theory and policy evaluation, between microeconometrics and microeconomics, and between statistics and econometrics. The hallmark of Heckman's contributions is a desire to link economic reasoning with the statistical analysis of observed

---

[1] This chapter is formed of two parts. The first, running from Section 1 to Section 6 inclusive, consists of a reprint with minor changes of Richard Blundell's "James Heckman's Contributions to Economics and Econometrics", *Scandinavian Journal of Economics*, 2001, volume 3, number 2, pp. 191–203. It was written following the award of the Nobel Prize in Economics to Heckman and Daniel McFadden in 2000, and appears here with the kind permission of Wiley. © The editors of the *Scandinavian Journal of Economics*, 2001. The remainder of the chapter, running from Section 7 to Section 11 inclusive, is by Flávio Cunha and reviews Heckman's contributions since 2001.

[2] Thanks are due to Chris Flinn, Bo Honoré and Hide Ichimura for commenting on an earlier version of Blundell's part of the chapter.

R. Blundell
Department of Economics, University College London, London, UK
e-mail: r.blundell@ucl.ac.uk

F. Cunha (✉)
Department of Economics and Texas Policy Lab, Rice University, Houston, TX, USA
e-mail: fc15@rice.edu

behaviour. Never satisfied as long as any stone remains unturned, they show a consistency of style and a strong underlying purpose. Heckman's energy and commitment to his work spill over to those he works with and those he supervises, leaving them mentally drained but intellectually stimulated, while producing a recognisable string of brilliant students and renowned joint papers.

The field of microeconometrics in particular has been the principal beneficiary of Heckman's work. His initial research concerned the analysis of household decision-making, especially labour supply and consumption decisions. This work set the scene in much of the field for the next 25 years and still provides the basis from which new models are developed. In these early papers, we see the origination of the theoretically consistent empirical neoclassical models of labour supply and consumption. Two major contributions stand out: the analysis of female labour supply and the analysis of intertemporal labour supply choices. In both, the insights from Heckman's analysis still form the basis for empirical and policy-based modelling of such decisions.

Hand in hand with the development of these models came the further analysis of their statistical implications. Heckman asked the question: If individuals act as rational agents, what are the implications for statistical analysis of their actions? How should we derive reliable estimates of preference parameters when preferences themselves dictate what we observe? Here, Daniel McFadden was already in the process of his pioneering development of the discrete choice model for economic agents. If actions are discrete and interactive, then there will be further conditions on the econometric model to guarantee it provides a coherent statistical relationship between inputs and response—Heckman coined this condition the "principle assumption". Take the mixture of discrete choices and continuous outcomes that are commonplace in microeconomic data and we get the rigorous development of the selection model. Indeed, the selection model is at the heart of the econometric investigation of individual economic behaviour in non-experimental settings, and Heckman's paper in *Econometrica*, "Sample Selection Bias as a Specification Error" (Heckman 1979), remains essential reading for any scholar of empirical microeconomics.

The increasing use of panel data and longitudinal data in empirical microeconomics put in place a natural step for the next set of developments by Heckman. His work on labour market dynamics and duration analysis in the early 1980s produced a number of substantive theoretical and empirical contributions. Again, one followed from the other. To analyse panel data on labour market states, Heckman sought a robust way of separating state dependence from unobserved heterogeneity. This distinction still remains at the centre of panel data and longitudinal data research. His contributions

highlighted the importance of initial conditions and the difficulties of achieving non-parametric identification.

The evaluation of policy interventions has always been a central motivation for much of Heckman's research. One key principle that comes across from his research is the importance of defining a parameter of interest when investigating individual behaviour in a world with unobserved heterogeneity. This can be seen in his development of the selection model and his duration studies. However, it is revealed most clearly in his extensive analysis of training and other labour market programmes using non-experimental data. If there is no randomised assignment in a policy intervention and if individual responses to the policy intervention are heterogeneous, then deriving a parameter that would be of policy interest requires particular care. Beginning in the 1980s, Heckman began to turn his analysis of panel data and selection models towards this question. In subsequent work with colleagues, he has contributed a number of influential papers that examine precisely what is identified from selection, instrumental variables (IV) and matching procedures in programme evaluation studies.

Much of the policy evaluation work in microeconometrics is a partial analysis that does not consider feedback in the economy. Heckman addressed this issue by placing the microeconometric evaluation model within an economy-wide framework with production and intertemporal decision-making. This "general equilibrium" framework for microeconometric policy evaluation set an ambitious agenda, and one which generated significant new contributions for many years.

Heckman's work covers a wide range and it is probably unfair (and maybe unwise) to confine it to a handful of areas. Nonetheless, in this brief non-technical review, I have done so, focusing on four major areas and a selected sample of publications—labour supply, selection models, labour market dynamics and evaluation methods. I have chosen these areas simply as an organising principle and it is not the only alternative. Equally attractive ways of organising these contributions would be around the concept of causal economic parameters of interest or around the identification and estimation of responses to policy interventions. Instead, I will use these to provide reference points across areas.

2   Labour Supply

Heckman's early work on labour supply gave rise to at least three related and important contributions: first, the integration of consumer theory and the theory of labour supply; second, the development of an empirical life-cycle

setting for labour supply; and third, the statistical analysis of participation, labour supply, and market wages. These studies on labour supply, which originated in the early to mid-1970s, set the scene for the development of his research on selection, labour market dynamics and programme evaluation. They are all empirically oriented but with a keen eye on the identification and estimation of structural economic parameters from micro data.

The initial aim of this work was to estimate the parameters of indifference curves for leisure and consumption. These would make it possible to measure the welfare cost of a tax or welfare intervention and to simulate the impact of new policies. This was already an ambitious agenda. A number of unresolved issues in the literature at that time were highlighted in Heckman's analysis of female labour supply. There was the econometric problem of non-participation and the development of a reservation wage analysis in the presence of child-care costs; there was the variation in hours worked among participants which required a reasonably flexible functional form and there was the issue of missing wages. These are all covered in two of his remarkable publications (Ashenfelter and Heckman 1974; Heckman 1974a). Heckman recognised that simple least squares analyses of hours of work, wages, and participation would not by themselves identify preference parameters. The standard Tobit model alone was also insufficient for dealing with the problem. Here, then, we have the development of an estimation procedure which allows the work decision to be based on interrelated choices over hours of work and the use of formal child-care, each with its own separate source of stochastic variation. This represents the forerunner of the many microeconometric developments in this area and continues to set the standard by which models are judged. Indeed, the development of a likelihood, which captures the sampling information on participation and wages, can be seen as the beginnings of Heckman's investigations into the analysis of endogenously selected samples. But there was much more; the marginal rate of substitution specification for preferences turned out to be a highly innovative way of dealing with non-participation, allowing flexible but heterogeneous preferences. The endogenous choice of formal and informal child-care jointly with hours of work and participation provided a basis for the analysis of multiple regime models.

Perhaps the most important aspect of these studies is the policy motivation. Heckman is clear about why he needs to identify and estimate preferences. It is to allow the separation of preferences from constraints which, in turn, enables policy simulations to be carried out—not just simulations of existing policies that could possibly be analysed by simpler forms of estimation, but simulations of hitherto unseen policies which change the opportunity sets facing individuals in new ways. This has always been the toughest test of

empirical economics and the fundamental reason why structural microeconomic models are required. For example, Heckman (1974b) examines the introduction of a child-care voucher to reduce welfare dependency of low educated mothers. Aware of the empirical requirements for tax design and welfare measurement, he set about recovering the full set of theoretically consistent household labour supply elasticities (Ashenfelter and Heckman 1974).

In all of this "static" labour supply analysis, we find repeated references to the potential importance of a more dynamic setting. In fact, this work was undertaken alongside the development of a life-cycle framework which has its origins in the first essay of his 1971 Princeton University doctoral thesis. This work was partly motivated by the observation that both income and consumption appear to follow a similar hump-shaped path—somewhat out of line with a simple consumption-smoothing model. Heckman (1974a) provides a beautifully simple, yet complete, integration of intertemporal consumption and labour supply theory, showing that with labour supply and uncertainty such choices can easily explain these empirical phenomena.

This life-cycle analysis was taken forward in two different directions. The first was to incorporate human capital choices and human capital accumulation. His article, "A Life Cycle Model of Earnings, Learning and Consumption" (Heckman 1976a), showed that earnings functions which ignored life-cycle labour supply tended to overestimate rates of depreciation. Not surprisingly, this ambition of incorporating human capital choices in a life-cycle framework has been a continuing theme and is to be found in the very recent contributions by Heckman and his co-authors to the debate on human capital investment, reviewed further below.

The second direction was to develop an empirically implementable form of the intertemporal substitution model for labour supply. This path-breaking work, much of it with his former student Tom MaCurdy, has become the basis for an extensive literature. It noticed that with intertemporal optimising agents and standard neoclassical assumptions, the marginal utility of wealth is constant over time but differs across individuals and is clearly correlated with wages. Since labour supply choices could be written in terms of current wages and the marginal utility of wealth, this was a perfect application of a fixed effects estimator for panel data; the idea was applied to the panel data analysis of female labour supply in (Heckman and MaCurdy 1980). Without too much adjustment, this model could also account for uncertainty and so became the prototypical intertemporal model of labour supply. It directly recovered the intertemporal substitution elasticity for labour supply and immediately showed the relationship of this intertemporal elasticity with the

standard Hicksian and Marshallian elasticities, thereby tying together the "static" and life-cycle approaches to labour supply analysis.

The fact that labour supply was censored at zero generated a further econometric problem, as there was little in the way of any existing analysis of panel data estimation of the fixed effects model with censored or discrete data. This was just one of a number of new econometric problems that cropped up during these initial empirical studies and formed the motivation for many of Heckman's contributions to the econometrics of discrete data and selected samples to which we now turn. However, the early 1980s did not mark the end of Heckman's contributions to the labour supply field (see, for example, Heckman and MaCurdy 1981, 1986, Heckman and Killingsworth 1987 and Heckman 1993).

3   The Selection Model and Discrete Choice

As we have seen, Heckman's extensive empirical investigations of individual labour supply behaviour stimulated further analysis of their statistical implications. There are at least two major innovations in econometrics that came out of this work: first, the analysis of selected samples, and second, the estimation of simultaneous multivariate choice models in which the outcomes are a mixture of discrete and continuous decision variables. These two developments most obviously fitted into the study of labour supply. The first concerned learning about the unconditional distribution of market wages when labour market participation itself was a choice that may be based on the distribution of market wages. The second concerned the analysis of household choices where these were a mixture of both discrete and continuous decision variables. Although the motivation from the labour supply area was clear, the general applicability of these two developments was much more general.

Perhaps the selection model is most renowned. Heckman's analysis of endogenously selected samples laid the groundwork for the analysis of returns to training where training was not randomly assigned, to the study of union wage differentials and to many other microeconometric problems. His approach was innovative but also simple—a winning combination. Starting with an additive regression model—that is, where the unobservables are additive with unconditional mean zero—Heckman noted that, for normal distributions, the conditional mean for a selected sample involved a single additional term which itself was a function of the selection probability. This term or "control function" could therefore be estimated in a first step from the choice

probability model. The Heckman two-step estimator—or Heckit estimator—was born (Heckman 1976b, 1979).

The selection model and the two-step estimator were also much more generally applicable than in the normal distribution case, and semi-parametric extensions were readily forthcoming for the additively separable model. In fact, this estimator not only became widely used in application, it also generated many important semi-parametric estimation and identification results. The easiest general identification results came with identification in the limit, which recognised that if the choice probability could be varied independently of the unconditional mean, then choosing points where the selection probability was close to unity enabled identification of the unconditional mean without parametric assumptions on the selection distribution. This required an exclusion restriction which allowed the selection probability to be varied independently of the conditional mean function. With his former student, Bo Honoré, Heckman derived the general non-parametric identification of the Roy Model—a two-regime generalisation of the additively separable selection model. This again relied on being able to "fix" the selection probability independently of the unconditional mean function; see Heckman and Honoré (1990). Indeed, "matching" observations for which the control function is similar has proved to be a very attractive approach to semi-parametric estimation in selection models (see, for example, Heckman 1990 and discussants' comments therein).

In the areas of labour market dynamics and especially in programme evaluation, Heckman went on to extensively develop and use the ideas in the selection model. Apart from these applications of the selection model, Heckman also showed its usefulness in a number of other important areas. Perhaps some of the most interesting, and most directly related to the original development of the model, were applications with Honoré (1990) and with Guilherme Sedlacek (1985, 1990). This work provided an analysis of aggregate and sectoral wage distributions when individuals self-select into the labour market and into sectors of the economy.

If actions are a mixture of discrete and continuous decision variables—labour supply and consumption, for example—and are in addition simultaneously determined, then there will be a further condition on the econometric model to guarantee that it provides a coherent statistical relationship between inputs and response. This condition, which Heckman labelled as the "principle assumption", was presented in his *Econometrica* paper, "Dummy Endogenous Variables in a Simultaneous Equation System" (Heckman 1978). It is a remarkable article that derives the conditions for a coherent econometric framework. This involved a jump parameter in the basic mean of the latent

variable underlying the discrete choice. This may appear odd, but a moment's thought can relate this directly to the fixed entry cost model.

Many of Heckman's remaining contributions to the econometrics literature fall in the areas of labour market dynamics and/or programme evaluation, and some of these will be described below. It may also be worth pointing out two further contributions. The first relates to the fixed effects censored or discrete choice model. In "Statistical Models for Discrete Panel Data" (Heckman 1981a), Heckman presented a widely cited analysis of the impact of fixed effects in such models and studied the impact of increasing time series sample size. This showed a relatively small bias from static dummy variable Probit estimation (the equivalent of within groups for the linear panel data model) for panels of a reasonable length (say ten time-series observations or more). This remains a standard reference used to argue in favour of using such models in censored and discrete data. The chi-square test as a model validation test, introduced in Heckman (1984), has also remained a popular and attractive method of assessing the reliability in discrete and censored microeconometric models.

4   Labour Market Dynamics

Perhaps one of the most daunting issues in the microeconometrics of panel data, or longitudinal data more generally, is the separation of state dependence and heterogeneity. This was a central theme of Heckman's 1980 World Congress address (Heckman and Singer 1982) and continued to attract Heckman's attention for a good while later (Heckman 1991). This work recognised that 'in the presence of heterogeneity, "exogenous variables" become endogenous'. Unobserved heterogeneity in longitudinal data represents permanent individual attributes of tastes or technology. Since these attributes will be correlated with choices made in the past, variables that represent past choices will be "endogenous". They will be correlated with the unobservables. This includes historic outcomes of the dependent variable and explanatory variables related to those historic outcomes—predetermined or weakly exogenous variables. Indeed, even if there is no direct state dependence in a process, unobserved heterogeneity can induce the appearance of state dependence, so that individual decisions may appear unduly habit-forming and unemployment may appear to "cause" future unemployment even when it does not. Alternatively, true dependence on the past may be masked by offsetting effects of permanent heterogeneity.

This observation and solutions to the resulting estimation and inference problem have become a central theme of much of the subsequent work on panel data with predetermined or weakly exogenous regressors. The Heckman and MaCurdy labour supply model, discussed above, had already stressed the importance of unobserved individual effects and their likely correlation with included regressors. Consequent problems of identification and estimation are exacerbated when the dependent variable under consideration is discrete or censored. The dynamic labour supply model for panel data had the potential for both of these problems: the specification required a fixed effect and the dependent variable—employment or hours of work—was naturally censored or discrete. Heckman's work from this era made at least two valuable contributions which have heavily influenced subsequent research. First, the degree to which the bias was attenuated with increasing length of the panel has already been mentioned in the preceding section; see Heckman (1981a). The use of initial conditions to capture the impact of heterogeneity was a further insight; see (Heckman 1981b). These ideas were already embedded in his World Congress lecture (Heckman and Singer 1982) and have been important for the subsequent development of dynamic discrete choice models and conditional likelihood estimation approaches to panel data models.

This work evolved into the area of event history data and duration analysis. Typically, the data available to economists in analysing labour market durations are not drawn from controlled experiments. Consequently, the econometric analysis of duration data requires accounting for the effect of the sampling plan on the distributions of sample spells. Heckman's 1980s contributions with Burton Singer and also with Chris Flinn are particularly notable. This work developed a structural economic basis for the continuous-time Markov model used in longitudinal studies and considered the identification of the economic parameters of interest in a job arrival model with search. One motivation here was the recoverability of the wage offer distribution, and this work highlighted where parametric assumptions were likely to be important. These studies tied in very closely with the original selection analysis, but now with the added twist of a Markov model of entry and exit. Deriving equilibrium wage distributions in empirical search models continued to be a highly active area in research and much of the statistical development originated in Heckman's early work; see in particular (Flinn and Heckman 1982). This research also showed the importance of multi-spell data and developed the random-effects maximum likelihood estimator for multi-spell models with heterogeneity.

An important contribution of this work on labour market dynamics was in developing a three-state model that could allow the distinction between

unemployment and "out of the labour market" to be tested as different labour market states. This was applied to a sample from the National Longitudinal Surveys of Youth (NLSY) in Flinn and Heckman's 1983 *Journal of Labor Economics* paper. The results firmly rejected the hypothesis that these non-employment states could be regarded as the same. The exit rate from unemployment to employment was found to exceed that from out of the labour market to employment. However, the authors carefully noted that this does not necessarily imply that the rate of arrival of job offers is higher for the unemployed. Again, this is now a heavily cited reference. It has inspired much further empirical and theoretical analysis of the competing risks model for multi-state durations. Indeed, in subsequent work with Honoré, published in *Biometrika* (Heckman and Honoré 1989), the inclusion of regressors was shown to be able to overturn a previous non-identification result for the proportional hazard and accelerated failure time models. The similarities with the identification results for the Roy model of selection mentioned above are clear.

The conventional analysis of single-spell duration data remained based in a random effects approach with unobserved heterogeneity chosen from some parametric family. Although the parametric family chosen was known to influence the resulting parameters, it was not until Heckman and Singer's 1984 *Econometrica* paper (Heckman and Singer 1984a), and subsequent studies (Heckman and Singer 1984b, 1985), that it became apparent that the distribution of unobservables was non-parametrically identified. This work was able to show that, given an assumed functional form for the structural duration distribution and the empirical distribution of durations, it was possible to consistently estimate the distribution function for unobservables in a general class of proportional hazard models. Moreover, identification of the proportional hazard model could be achieved through a restriction on the tail distribution of unobserved heterogeneity. Interestingly, this application showed that when the impact of distributional assumptions on heterogeneity was minimised, the data were consistent with the declining reservation model of search theory.

Heckman's work on labour market dynamics is extensive and concerns more than what has briefly been discussed here. One notable application was to the study of fertility and labour market opportunities. This work with James Walker (Heckman and Walker 1989) and also Joe Hotz (Heckman et al. 1985) used Heckman and Singer's identification and estimation results to analyse the impact of economic variables on the timing and spacing of births.

5 Evaluation of Labour Market Programmes and Human Capital Incentives

Although Heckman's early research had been motivated by the measurement of policy impacts, it was the work published with Richard Robb in the mid-1980s that represented the comprehensive development of these ideas (see Heckman and Robb 1985, 1986). This focused on the analysis of training programmes and saw the beginnings of the enormous body of empirical and theoretical work that Heckman, with his students and colleagues, would produce on the evaluation problem. The papers with Robb are innovative in a number of respects. They extend the control function approach for selection to the policy intervention framework and relate it to the two-regime switching Roy model. These studies also develop the estimation of the treatment on the treated parameter in a random coefficient model for discrete programmes—the heterogeneous response model—and point out the bias from standard instrumental variable procedures. They extend this to the repeated cross-section and panel data framework and develop conditions under which the fixed effect estimator recovers the treatment on the treated parameter.

This was important work and remains so; it emerged as more emphasis was being placed on the use of experimental design of evaluation programmes. It showed the limit of what could be derived from non-experimental settings and set the scene for the ensuing debate on the use of non-experimental econometric methods in the analysis of programme interventions. It is interesting to note the relationship between this line of investigation and the more structural evaluation based on economic theory, such as that in the labour supply analysis described above. In terms of a tax or welfare policy intervention, for example, this line of study is typically interested in the ex-post evaluation of an existing policy. This contrasts with the labour supply analysis aimed at recovering income and substitution effects. In so far as it does not recover preferences or technology separately from constraints, it cannot be used in the evaluation of a new policy, or in the welfare analysis of policy reform. But, in general, the evaluation literature has had much more limited aims.

The advent of experimental data on policy interventions and, in particular, the data developed for the evaluation of participants in the Job Training Partnership Act 1982, provided a perfect empirical setting for assessing alternative approaches to evaluation; the experimental data were used to characterise the bias from relying on non-experimental econometric methods. This

debate had already been taken further in Heckman's work with Joe Hotz (1989), but in his *Econometrica* paper with Hidehiko Ichimura, Jeffrey Smith, and Petra Todd, "Characterizing Selection Bias Using Experimental Data" (1998), Heckman provided a detailed description of the biases in three popular non-experimental methods—matching, selection, and difference-in-differences. Put simply, selection adjusts for unobserved differences that are correlated with participation in the programme and matching adjusts for differences among observables. Difference-in-differences adjusts for unobservables—but only those that are time-invariant—and it also requires repeated observations. Heckman et al. (1997) had already established the relationship between matching and selection estimators. The issue remained in understanding the various biases in comparison to purely experimental studies.

The *Econometrica* study (Heckman et al. 1998) study provides an exceptionally clear breakdown of the major sources of bias in recovering the treatment on the treated parameter from non-experimental data. It also provides a useful guideline for evaluation studies that do not have access to experimental controls. The bias arises from not directly measuring the counterfactual impact of the programme on those who did not receive it. With experimental data, this can be measured directly from the randomised control group. Heckman and his co-authors break the bias down into three components: differences in the support of the regressors, differences in the shape of the distribution of regressors, and pure selection bias. Matching is clearly designed to remove the first two biases and, with non-parametric propensity score matching, this seems to work well. But pure selection bias still remains. This can be minimised by choosing a comparison group that is from the same local labour market, but not eliminated. Labour force history variables and personal characteristics are found to work well as exclusion restrictions for the selection model. However, the normal selection model is not a good approximation and semi-parametric methods which relax the normality assumption appear to work best. Thus, in this work, we see the drawing together of Heckman's early work on the selection model with the subsequent semi-parametric development of estimators for that model.

With heterogeneous responses to programme interventions, it is natural to ask, "What parameters are of direct interest?" In principle, the whole distribution of the response effects would be ideal. Above, we discussed the treatment on the treated parameter—that is, the average response among those who receive the programme, sometimes split down by observable characteristics or the propensity to engage in the programme. But there are other local parameters of interest. In work with Edward Vytlacil (1999, 2000), Heckman took forward his earlier analysis of the implicit assumptions underlying different

models of programme participation and treatment (Heckman 1996, 1997), to consider analysis of treatment effects by relating the treatment effects model directly to the latent variable discrete choice framework. This was extremely enlightening work and showed that the standard parameters of interest—average treatment on the treated, the overall average treatment effect and the local average treatment effect—relate to each other according to the chosen interval for the range of the unobservables determining programme participation. The local estimates relating to the responses of those individuals on the margin of joining the programme, therefore measure the impact of the programme on those most likely to be induced to move by the intervention.

Underlying all this work on evaluation has been an interest in the general equilibrium effects of interventions. All the analyses we have described so far are partial equilibrium. Indeed, experimental analysis is based on the premise of no spillover to the control group. General equilibrium effects imply that an intervention has wider, perhaps economy-wide, impacts. The approach taken by Heckman and his colleagues is to use estimates derived from non-experimental partial equilibrium analysis to parameterise a dynamic general equilibrium model. In important work with Lance Lochner and Chris Taber (1998a, b), he examined the impact of human capital incentives when the market wage is allowed to respond to changes in supply. As with much of Heckman's work, and as can be inferred from the extensive overview with Martin Browning and Lars Hansen, "Micro Data and General Equilibrium Models" (1999), these are wonderfully ambitious studies with many innovative aspects—not least in the solution to determining substitution parameters between workers with different levels of human capital.

6   Summary

James Heckman has made many other important contributions to economic science than those I have discussed in this brief non-technical review. Perhaps the finest of these is simply his willingness to discuss ideas—and that at any time of day or night, by e-mail or whatever means. But there are many other substantive papers too, such as a wonderful *Review of Economic Studies* article with Jose Scheinkman, "The Importance of Bundling in a Gorman-Lancaster Model of Earnings" (Heckman and Scheinkman 1987), which derived conditions on the unit pricing of skills in the competitive labour market. Many researchers write down models with constant skill prices but without thinking through the conditions for their existence. Heckman was clearly not happy

until a precise and complete model was derived—an endearing and enduring aspect of his research.

There is a strong and consistent approach to empirical economics running through Heckman's research. It epitomises the interrelationship between microeconomics and statistics, recognising the importance of individual choice for the statistical analysis of individual data. It is driven by the identification of economically interesting parameters in the study of heterogeneous agents in market economies. It is rigorous and ambitious, yet solidly based on economic data and the study of economic policy—most effectively summarised by his own retrospective review in the *Quarterly Journal of Economics* (Heckman 2000).

The interaction between the economic behaviour of individuals and the micro data so generated was at the heart of Heckman's development of the selection model and has changed the way empirical microeconomics has been conducted since the 1970s. It has been the springboard for the progression of Heckman's own research and for much of the development of microeconometrics. It stands as a truly great contribution and, although there is much besides, it is fitting that his analysis of selected samples has been singled out for citation in the award of his Nobel Memorial Prize.

7   Contributions Since 2001

This review covers Heckman's contributions to economics and econometrics since winning the Nobel Prize in 2001. It is challenging to summarise two decades of work spanning 179 published papers, 16 working papers, and eight books. According to Google Scholar, in the last five years alone, Heckman has accumulated nearly 84,000 citations. As impressive as these numbers are, they pale compared to Heckman's enthusiasm for science—data and theory as he regularly states in every article he has ever written or presented. The firm belief that thorough scientific work can help improve peoples' lives explain why Heckman has advised nearly 100 students throughout his career.

It is impossible not to commit sins of omission when describing the work that has influenced scientists and policy makers alike in a few pages. Still, choices must be made. In this review, I examine three lines of work that Heckman has focused on in the last 20 years. First, I describe his contributions to structural econometrics. Heckman's research in this field has influenced investigations in labour economics, macroeconomics, industrial organisation, international trade, and any other area in economics with a solid empirical tradition.

Second, I turn to the many ways Heckman's research has influenced the study of inequality. I start by describing the impact that his research has had on social scientists' understanding of "human capital". A few years ago, economists used years of completed schooling to measure human capital even though schooling attainment varies across time and place. Heckman's research showed that that approach was misleading because it led to wrong inference about the importance of human capital and limited the set of interventions and policies that society could implement to reduce inequality and increase prosperity.

Third, I describe how Heckman's recasting of human capital beyond education informed his research on understanding human capital formation as a life-cycle skill formation process. I illustrate his theoretical and empirical work in this area, and I show why this work informs interventions that occur in different stages of an individual's life cycle, from early childhood to adulthood. I summarise Heckman's influential research on the long-term impacts of high-quality early childhood programmes. This work has influenced many researchers who have adapted the methodological approaches to studying interventions that foster human capital formation in early childhood, adolescence, or adulthood. Heckman's contributions in this area of research spill-over beyond the ivory tower, and they are a fundamental part of public policy debates around the world. These studies have challenged policy makers to translate research findings into large-scale interventions that promote human development.

8   Structural Econometrics

In structural econometrics, researchers attempt to estimate structural parameters, that is, parameters that are invariant to a counterfactual policy environment. Therefore, whether a parameter is structural or not depends on the counterfactual question a researcher attempts to answer. For example, if the goal of the analysis is to investigate how subsidisation of input will impact the use of that input (e.g. clean energy), then it is arguable that the production function parameters are invariant to the source of energy. However, if the goal is to study how pollution taxes impact pollution, then the same production function parameters are no longer structural because firms may choose to change how they organise production to reduce pollution. In this case, the technology the firm uses is itself an object of choice. The firm's production technology in the state-of-the-art policy scenario is not the same as in the counterfactual scenario in which the government raises pollution taxes

significantly. Heckman's research has pushed forward the frontier on identifying and estimating structural models to evaluate social programmes.

Structural econometrics is crucial when researchers need to produce interpretable estimates of interventions or social programmes. Typically, participation in a social programme (e.g. child-care vouchers) is not mandatory, and many participants choose not to receive benefits from such programmes. However, governments can encourage individuals to participate depending on the costs and (private and social) benefits. For example, the government can encourage participation via many forms, such as cash incentives, reduction of red tape, provision of transportation, and other ways. As participation is voluntary, the methods used in impact evaluation must recognise that participation is subject to selection in levels and gains. Because of the latter, there could be a discrepancy between the various average effects that standard methodologies (e.g. instrumental variable, difference-in-differences, regression discontinuity, matching) recover and the question the evaluator attempts to address. In essence, the standard methodologies usually recover an average effect that may not be relevant for the policy under consideration. The issue arises because different encouragement forms may generate a different set of participants. Given that each participant has a specific gain, different sets of participants could produce a different average impact. Thus, a cash incentive that disproportionately attracts individuals with the smallest gains will produce a small average effect, while a transportation subsidy that focuses on individuals with the most sizeable gains will generate a substantial average impact.

Structural models are also instrumental for evaluating programmes that policy makers have never implemented, which researchers refer to as "ex-ante" evaluation. Ex-ante policy evaluation methods allow for researchers to go beyond the anonymity postulate. The anonymity postulate treats two aggregate distributions as equally good if the overall distribution is the same after income is redistributed among persons. For example, the methods in Carneiro et al. (2003, henceforth CHH) determine, for reforms that are contemplated but have never been implemented, which groups in an initial position benefit or lose, how much they lose, how they would vote in advance of reform, and how they would vote after it is implemented. It goes beyond impact because it identifies potential "losers" and "winners" of proposed policy reforms.

A crucial aspect of any such reform is the identification of the joint distribution of outcomes. Consider a policy that aims to reduce the financial burden of college attendance. Assume that no country, state, or county has ever implemented the policy in question. Therefore, ex-post policy evaluation is not feasible because there is no dataset in which some individuals benefit from

the policy and others do not. The economic approach is to model two aspects of the problem jointly to answer this question. The first aspect is how financial costs affect the decision to graduate college. This model component is the choice equation because it describes how preferences and constraints determine individual choices. In the context of policy evaluation, the alternatives individual can choose are discrete (e.g. to enrol or not to enrol in college). Work by Matzkin (1992) shows that, under some normalisation conditions, it is possible to identify and estimate fairly general models with discrete dependent variables.

The second aspect is to link college graduation to labour earnings. We refer to this object as the wage offer function. In typical applications in empirical economics, the data are longitudinal; that is, we observe earnings (or other outcomes) for each individual for many periods. However, for each individual, we observe only one sequence of earnings. If the individual did not attend college, we would observe the no-college earnings, which we denote by $Y_{i, t, 0}$. If the individual graduated college, we observe the college earnings, or $Y_{i, t, 1}$. The challenge that structural econometricians face is that they need to identify the joint distribution $F\left(\{Y_{i,t,0}\}_{t=1}^{T}, \{Y_{i,t,1}\}_{t=1}^{T}\right)$ to estimate the ex-ante impact of a new policy and uncover the individuals who are "winners" and "losers".

Heckman attacked this joint identification model throughout his career in highly cited papers (e.g. Heckman and Honoré 1990), but these articles primarily focused on situations where the data were cross-sectional. Adding a time dimension complicates the problem because it increases the dimension of the joint distribution of outcomes. In a cross-sectional framework, we observe as many outcomes as the number of alternatives that the individual can choose. For example, if there are N mutually exclusive alternatives, the researcher must identify the joint distribution of N outcomes. In contrast, when the dataset has a time dimension as well, the researcher needs to identify the joint distribution of $N \times T$ random variables, which is, in practice, intractable if T is large. CHH's innovation was to use factor models to reduce the problem's dimensionality. In their framework, the factors capture both the dependence over time (i.e. the dependence over the time dimension) and the dependence across alternatives. They show that a model with few factors fits well the data both within alternatives and across time.

The factor model approach is not only a convenience for representation and estimation, but it is also crucial for the identification of the joint distribution of outcomes $F\left(\{Y_{i,t,0}\}_{t=1}^{T}, \{Y_{i,t,1}\}_{t=1}^{T}\right)$. To understand why, assume that the choice depends on the sequence of earnings in the two sectors (i.e. college and no-college). Then, the choice observation provides information about the

sequence of earnings in the sector the individual did not choose. To focus on the crucial point, suppose that $T = 1$ and that the individual chooses to go to college if, and only if $Y_{i,1} \geq Y_{i,0}$. This choice model dictates that any individual who went to college has higher earnings in the college sector than the no-college sector, and the opposite is true for any individual that chose not to go to college. If the comparison of earnings streams is the only input in the choice model, then Heckman and Honoré (1990) prove non-parametric identification of the joint distribution F. In the literature, researchers refer to this model as the simple Roy model.

Note that the factor model is unnecessary when the simple Roy model accurately describes the phenomenon that researchers are modelling. However, this case is rare. A more realistic assumption is that econometricians do not observe all random variables influencing an individual's choice. For example, there may be non-pecuniary (or psychic) costs or benefits of attending college. Individuals know how willing they are to attend classes, but such information is rarely—if ever—collected in datasets that researchers collect and analyse. The general Roy model extends the simple Roy model, allowing for such random variables unobserved by the econometrician. Unfortunately, Heckman and Honoré (1990) show that one cannot identify the joint distribution of outcomes without further restrictions. However, when the data are longitudinal and one can use a factor model to capture the dependence between unobserved variables that determine the choice and the residuals in the earnings equations, then the generalised Roy model is identifiable. In this sense, CHH's factor structure is the price researchers must pay to represent preferences and constraints more generally. This result is appealing from a methodological standpoint because CHH shows that the restrictions on the number of factors are relatively weak.

Using a few factors to capture dependence across alternatives and over time has influenced research that separates heterogeneity from uncertainty in life-cycle earnings. Heterogeneity is the variation in labour earnings that we can attribute to differences in human capital across individuals[3] or prices of elements of human capital. Heterogeneity captures inequality in human capital that influences earnings inequality. Therefore, individuals know and act on this heterogeneity that econometricians do not observe when making choices.

On the other hand, uncertainty captures aspects of labour earnings that reflect "luck". For example, variation in earnings across real estate agents may reflect heterogeneity in the capacity to execute sales and temporary variation in supply and demand conditions that may affect the volume or price of real

---

[3] Or differences across markets in the prices of elements of human capital.

estate transactions. From the point of view of economic theory, uncertainty generates welfare costs when markets are incomplete. In terms of policy making, heterogeneity partly reflects differences in the implementation of policies that foster human capital accumulation (e.g. variation in school quality across socio-economic, racial, or ethnic groups). Both cases offer arguments for intervention, but the nature of this intervention depends on the relative contribution of heterogeneity and uncertainty.

Two papers use the CHH framework to decompose earnings inequality into heterogeneity and uncertainty. First, Cunha et al. (2005) applied the CHH framework to analyse the National Longitudinal Survey of the Youth 1979 (NLSY/79) data, a longitudinal study whose participants were a representative sample of individuals born between 1957 and 1964. They show that over 60% of inequality in lifetime earnings is due to heterogeneity and the remaining 40% is due to uncertainty. In contrast, Cunha and Heckman (2016) compare inequality in lifetime earnings evolved for the cohort of participants in the National Longitudinal Survey (NLS/66), whose participants represented the population born between 1942 and 1951, with the NLSY/79 cohort. They show that inequality increased, but most of the increase in no-college earnings was due to uncertainty.

So far, we have summarised work in which the choice takes place at one point in time. However, in most circumstances, individuals have to make choices every period. For example, in higher education, individuals choose whether to enrol or not in the first semester. After they finish the first semester, they can re-enrol for the second semester or drop out of college. Heckman et al. (2018) extend the CHH framework for dynamic discrete choice models, and provide a new interpretation of the classic works by Mincer (1958) and Becker and Chiswick (1966).

Methodologically, the paper bridges the treatment effect approach and the fully structural dynamic discrete choice approach. As in the treatment effect literature, the model does not impose precise rules agents use to make decisions, but there are choice equations that capture individual decision-making processes. Therefore, readers should think of the model as one that generates approximations of decision rules. However, as in the structural literature, there are well-identified margins of choice and a well-defined sequence of choices in which some alternatives are only feasible if individuals have decided to follow some specific paths. For instance, an individual who has not graduated high school cannot choose to jump to a graduate programme without first completing a college degree.

The middle ground opens up the possibility to model heterogeneity in rich ways because it simplifies the identification issues that arise in fully structural

models. This advantage does come with a cost. The approach does not allow endogenous state variables to enter the right-hand side of the outcome equations because these endogenous state variables correlate with the unobserved factors. This restriction, for example, rules out models with learning-by-doing in which an endogenous state variable (e.g. actual experience) is part of the wage offer equation. Such a limitation, however, is not applicable to the types of questions Heckman et al. (2018) pursue in their paper.

Heckman has also used structural models to conduct ex-post policy evaluation. A significant concern in his research is that any policy impacts individuals in a different way. In the literature, researchers refer to this fact as treatment effect heterogeneity. One way to illustrate this issue is to consider the following system of equations:

$$
\begin{aligned}
Y_i &= \alpha + \beta_i D_i + \epsilon_i \\
D_i &= \gamma + \delta Z_i + \eta_i
\end{aligned}
\tag{37.1}
$$

The variable Y_i is the outcome. The binary variable D_i denotes exposure to the policy. So, $D_i = 1$ if individual i was exposed to the policy, and $D_i = 0$, otherwise. The variable Z_i influences exposure to the policy ($\delta \neq 0$), and it does not correlate with ϵ_i, so $\mathrm{E}(\epsilon_i|Z_i) = 0$. In words, Z_i is an instrumental variable.

System (37.1) is distinct from the typical two-stage least squares model because the coefficient β_i varies across individuals. In particular, when D_i represents participation in a social programme (e.g. receiving COVID-19 vaccine), we would expect β_i to correlate with D_i. This phenomenon represents sorting on impacts of the policy or "essential heterogeneity" as defined in Heckman and Vytlacil (2005).

Essential heterogeneity is a significant obstacle for meaningful policy evaluation. Let $\bar{\beta} = \mathrm{E}(\beta_i)$, and rewrite the first equation of system (37.1) in the following way:

$$
Y_i = \alpha + \bar{\beta} D_i + v_i
\tag{37.2}
$$

where $v_i = (\beta_i - \bar{\beta}) D_i + \epsilon_i$. In this situation, the IV approach does not consistently estimate the parameter $\bar{\beta}$ because the instrumental variable Z_i correlates with v_i through D_i.[4] If we impose additional restrictions, the IV approach can estimate the local average treatment effect (LATE, e.g. Imbens

---

[4] If Z_i does not correlate with D_i, then Z_i is not an instrument. In this case, the analysis would not be possible because the first stage breaks down.

and Angrist 1994). LATE estimates the average impact among individuals whose participation is induced by the variation in the instrument Z_i ("compliers"). LATE is interpretable whenever the instrumental variable represents the lever policy makers explore to encourage participation in a programme. For example, suppose that Y_i are hours of hospitalisation, D_i is the COVID-19 vaccination status, and Z_i is the travel time to the nearest vaccination site. In this case, LATE informs policy makers of the impact of vaccination on hospitalisation hours among individuals whose vaccination status is affected by a reduction in travel time. Therefore, if policy makers encourage individuals to get vaccinated by reducing travel time (e.g. improving transportation supply), then LATE is the parameter of interest (or the policy-relevant treatment effect, henceforth PRTE). However, if the policy lever is a cash incentive, the LATE, which uses travel time as the exclusion restriction, does not recover the PRTE because the individuals the transportation improvement policy induces to take vaccines are generally different from those encouraged by the cash incentive. Heckman and Vytlacil (2005) show how to recover the PRTE by modelling outcomes and choices jointly.[5]

9   Human Capital: Beyond Educational Attainment and Cognitive Skills

The General Educational Development (GED) is an eight-hour subject test that establishes the equivalence between high school graduates and high school dropouts. Heckman's work on understanding the GED programme started early in Cameron and Heckman (1993) but continued after 2000. Heckman and Rubinstein (2001) showed that GEDs—individuals who have passed the GED test—have the same cognitive ability, as measured by the Armed Forces Qualification Test (AFQT), as high school graduates who do not go on to college. On average, individuals with a GED credential have higher AFQT scores than regular high school dropouts. The comparison of raw hourly wages yields the same pattern. They also showed a remarkable result: Once one controls for AFQT, the hourly wages of GED recipients are lower than those of high school dropouts.

Adam Smith implicitly refers to human capital as one of the four types of capital used to produce goods and services.[6] In 1928, Pigou referred to it

---

[5] For an application of this methodology, see Carneiro et al. (2011).

[6] '[T]he acquired and useful abilities of all the inhabitants or members of the society' (Smith 1999: 377).

explicitly.[7] Finally, the contributions by Mincer (1958), Schultz (1961), and Becker (1962) crystallised the term "human capital" as a staple across the social sciences. Still, with a few exceptions, human capital was essentially equated to either educational attainment (years of completed schooling) or cognitive ability (e.g. Barro 2001; Griliches 1977). The exceptions to this rule were the works by Bowles and Gintis (1976), Edwards (1976), and Klein et al. (1991), who surveyed employers and showed that job stability and dependability are the skills that they value the most.[8]

Heckman's work on the GED forever changed social scientists' understanding of what constitutes human capital. Heckman's argument was strong because he showed that the GEDs' lower hourly wages reflected their lower stock of non-cognitive skills. Moreover, unlike previous work, his evidence came from data on hourly wages (prices of human capital) which complemented the evidence from the previous research.

In this sense, Heckman et al. (2006) significantly contributed to this research programme. The authors applied the CHH framework to quantify the contributions of cognitive and non-cognitive skills to various socio-economic outcomes. The NLSY/79 data used in their research was uniquely suitable for this study because the NLSY/79 measured cognitive and non-cognitive skills during adolescence, and the outcomes measured when individuals were in their 30s. This aspect of the NLSY/79 reduced concerns about reverse causality (i.e. high non-cognitive skills arise from success in the labour market and not the other way around). However, the authors still needed to address selection, which they did by using the CHH framework. They found that the gaps in non-cognitive skills predicted gaps in educational attainment, labour market outcomes, and other socio-economic outcomes, such as crime participation and teenage pregnancy. In addition, they find that non-cognitive skills are an essential dimension of human capital for many outcomes and cognitive skills explain a small share of the variation in some crucial outcomes, such as crime participation.

Unfortunately, the NLSY/79 measures only a few non-cognitive skills, so the paper's findings probably suggest a lower bound about the importance of dimensions of human capital that still play a minor role in policy making. For example, the National Assessment of Educational Progress still focuses on assessing students' performance on tests that strongly load cognitive skills (e.g. English language, arts and maths). The same is true for the standardised

---

[7] 'There is such a thing as investment in human capital as well as investment in material capital' (Pigou 1928: 29).

[8] In addition, Bowles and Gintis (1976) showed that perseverance, dependability, and consistency are the most important predictors of grades in school.

assessments that states conduct independently from the federal government. Some states, as Texas, issue school-level report cards based on standardised tests of English and maths. Not only do these tests make "teaching to the test" appealing, but they also encourage schools to shift instruction from non-cognitive towards cognitive skills. Therefore, despite the impact of non-cognitive skills on many socio-economic outcomes, education policy makers still overlook these dimensions of human capital when deciding how to allocate resources within schools.

In a series of papers, Heckman dives into the literature on personality psychology and connects it to economics. Then, a paper that summarises this research and ignites the literature in human capital economics is Almlund et al. (2011). The authors discuss the measurement of personality factors, present challenges in operationalising such measurement, and summarise the predictive power and stability of personality traits.

The paper contains a novel contribution as the authors conceptualise personality traits within economic models. First, the authors extend the Roy model, the basis of the CHH framework, to incorporate personality traits as arguments in the returns to and cost of effort functions. In the Roy model, individuals choose different occupations (or tasks) according to their personality traits. Thus, for example, introverted individuals would choose different occupations from extroverted ones. This difference in sorting could arise because of differences in returns (e.g. extroverted individuals may have more connections in their social network) or psychic costs (e.g. introverted individuals have a more challenging time making many friends).

This extension is crucial for our understanding of inequality. For example, a regression of log hourly wages on potential experience, education, and cognitive skills will lead to an R-squared of about 20%. This result indicates that we do not understand 80% of the variance of wages. We know, however, that most of this variance represents heterogeneity (48%; see Cunha et al. 2005), while the remaining part (32%) is uncertainty, which could be due to information frictions in the labour market (e.g. Mortensen and Pissarides 1994). Thus, we do not observe the lion's share of inequality (heterogeneity) even though individuals act on it when deciding how much human capital to accumulate before entering the labour market. Once we develop forms to measure these unobserved dimensions of human capital (and personality traits may be one component of heterogeneity), we will understand how and when (in an individual's life cycle) we will be able to produce these human capital components. Such knowledge will determine policy making towards reducing inequality.

This body of work has implications for the design of policy that promotes long-term sustainable economic growth. In endogenous growth models, human capital plays a vital role in determining technological innovation and, as a result, the balanced path growth rate. Empirical research confirms the significant relevance human capital has in promoting growth (e.g. see the research on the birth of the bioengineering sector by Zucker et al. 1998). However, we have little understanding of how to produce the human capital that makes individuals highly entrepreneurial and innovative. In this sense, Heckman's research on expanding the definition of human capital beyond cognitive skills and educational attainment is necessary for designing human capital formation policies.

In addition to using personality traits in an extended generalised Roy model, Almlund et al. (2011) also discuss their relationship with parameters in structural models. Two critical parameters in understanding intertemporal choice and choice over states of nature are time preference and risk aversion parameters, respectively. Almlund et al. (ibid.) cite the evidence from Daly et al. (2009) to relate time preference parameters to conscientiousness, self-control, and the capacity to imagine the future consequences of present actions. In addition, they link risk aversion to sensation seeking (Zuckerman 1994; Eckel and Grossman 2002), openness (Dohmen et al. 2010), and neuroticism (Borghans et al. 2009; Rustichini et al. 2016). These relationships are significant for policy because heterogeneity in time preference may help explain heterogeneity in the accumulation of human and physical capital, while differences in risk aversion could explain why investors choose different investment portfolios. Moreover, interventions that shift these personality traits could increase wealth at retirement, elevate productivity (through higher human capital stocks) in the labour market, and impact the rate of innovation in the economy.

10   Life-Cycle Skill Formation

In the late 1990s and early 2000s, research on human capital formation focused on investigating the relevance of credit constraints in explaining the socio-economic gaps in college enrolment (e.g. for a review, see Card 1999). Unfortunately, it is difficult to find evidence on credit constraints because we do not observe them directly. Instead, empirical economists adopt an indirect approach, which depends on the predictions from theoretical models that feature credit constraints.

Around the turn of the century, the theoretical models were based on Becker and Tomes (1986, henceforth BT). BT featured a two-period model, in which the first period was the "investment" period (before entering the labour market) and the second period was the "earnings" period (after entering the labour market). In this model, individuals had an endowment that could impact both the costs of and returns to investments in human capital. In the context of the 1990s, this endowment represented a fixed level of cognitive ability. The crucial word here is "fixed" because BT assumed that investments in human capital could not affect an individual's stock of this endowment. This property of the BT model introduced the debate of nature versus nurture in economics. The endowment was nature, fixed at conception and the investment was nurture, subject to the policy environment.

BT also featured inequality in family resources, so some children were born in families with abundant resources that they could allocate to the child's human capital formation in the first period, and other children were born in families whose resources were not enough to invest optimally in their child's human capital.

In such a framework, the existence of credit constraints can generate suboptimal inequality when low-income families have high-endowment children. To understand this point, suppose that there are diminishing marginal returns to investments in children's human capital. Then, consider the situation where there are no credit constraints and families can take on loans to invest in their children's human capital at a risk-free rate equal to r. In this situation, families will invest in human capital up to the point that the marginal benefit of a dollar is equal to the marginal cost of funds, which is r. As a result, there is inequality in investments in human capital because the point at which the marginal benefit equals the marginal cost depends on the child's level of the fixed endowment. However, this inequality is optimal because it reflects efficiency in resource allocation: marginal benefit equals marginal cost.

Next, consider the situation in which there are credit constraints so that families cannot borrow to invest. In this case, low-income families with high-endowment children will invest "too little" in the sense that the marginal benefit of an extra dollar of investment is above the marginal cost r. In this situation, investment inequality is not optimal, and there is room for policy intervention (e.g. government loans).

Empirical researchers relied on this model prediction to investigate the existence of credit constraints for college enrolment. The argument proceeded in two steps. First, researchers estimated the return to a college degree through ordinary least squares (OLS) regressions. The consensus was that such

regressions are upward biased because of omitted variables since researchers did not observe the children's ability endowments (Griliches 1977; Card 1999).

Second, researchers would search for an instrumental variable that impacted the decision to graduate college but did not correlate with the children's endowments. A few prominent examples of instrumental variables were tuition subsidies, distance to the nearest college, and compulsory schooling laws. Finally, if there were no credit constraints, the IV estimator should be lower than the OLS estimator because the latter suffered from the upward bias due to omitted variables (e.g. Kane 2001). However, researchers found that the IV estimator was greater than the OLS estimator. For researchers, this finding meant that the rate of return of investments in human capital was above that of the risk-free rate r.

Carneiro and Heckman (2002) contributed to the literature by critically revisiting the argument that the IV/OLS comparison was sufficient to determine the existence of credit constraints. They offered three arguments. First, the instrumental variables were not valid. Carneiro and Heckman (ibid.) used the NLSY/79 to show that the instruments used in the literature correlated with the "fixed endowment" measures not present in most datasets but available in the NLSY/79.

Second, even if they were valid, the IV/OLS empirical pattern was consistent with a model with selection on comparative advantage even when credit constraints were absent. As we described above, the empirical literature followed BT to interpret their results. However, in the BT model, investment is a continuous variable, and the solution is interior. On the other hand, investment is a discrete variable in the empirical literature (college versus no college), and such situations do not allow for interior solutions. Instead, the empirical approach should model this dependent variable in the context of a discrete choice model and allow for sorting with heterogeneous returns. Carneiro and Heckman (ibid.) showed that when they do so, the model can generate the IV/OLS empirical pattern even when credit constraints are absent. This argument demonstrated that a greater estimate of the effect of education on earnings by the IV methodology (relative to the OLS one) is not sufficient to establish the existence of credit constraints.

Third, Carneiro and Heckman (ibid.) extend the typical discrete choice model to account for college quality. In this extended model, individuals have three alternatives in their choice set: not attend college, attend a low-quality college, or attend a high-quality college. They assume that the higher the college quality, the greater the returns. For example, some individuals may not attend college if there are credit constraints. However, if the government intervenes in this market and removes credit constraints, some individuals

may shift from the "no college" alternative to the "low-quality college" one. Such a movement increases the weight of observations in the "low-quality college" alternative which, in turn, should reduce the estimated returns to college because the empirical literature does not control for quality. Thus, with this model, Carneiro and Heckman (ibid.) show that the IV/OLS empirical pattern is not necessary to establish the existence of credit constraints.

One empirical regularity called Heckman's attention. There was clear sorting on the "skill" endowment in the research on college enrolment (e.g. Cameron and Heckman 2001; CHH; Cunha et al. 2005). When investigating the impact of policies that reduced tuition, Carneiro and Heckman (2002) found that high-ability individuals were more likely to move from the "no college" to the "college" alternative. In his early work on the evaluation of the Job-Training Program, Heckman found a similar empirical regularity: it was individuals with high stocks of endowment that benefited from such interventions (e.g. Heckman et al. 1997). These findings led Heckman to ask, What if policy interventions can impact these individuals' endowments? What if we move from "nature versus nature" towards "nature and nurture"?

Cunha and Heckman (2007, henceforth CH) proposed a model of life-cycle skill formation to answer this question. In their model, human capital is multidimensional. They assume that human capital is a vector with two skills, denoted by cognitive and non-cognitive skills. Each task in the labour market requires different combinations of these types of skills.

Unlike BT, they assume that it takes many periods for families to produce these skills. More importantly, the marginal productivity of investments differs by skills and by stages of the life cycle. For example, the return to investments in one set of skills may be high in early childhood and low in adolescence, while the opposite could be true for another set of skills. These properties of the production function of human capital, which they refer to as the technology of skill formation, introduce the concepts of critical and sensitive periods of development into the economics literature. A critical period is a stage in life when an individual can develop certain specific skills that the same individual cannot develop later. For example, as CH described, the critical period for vision development is a few months after birth, and lack of visual stimulation during this short window of time will compromise full binocular vision development (e.g. Daw 1998). Researchers have also documented the existence of critical periods of development for language acquisition (e.g. Newport 1990).

A sensitive period is a stage in development when we can acquire a particular skill rapidly or, in economics jargon, when investments have very high marginal productivity rates. Unlike skills that have critical periods of development, individuals can develop skills that have sensitive periods in other stages

of the life cycle (earlier or later), but it is more challenging to do so. The sensitive periods of development need not occur in the early childhood stages of the life cycle. For instance, the literature in developmental psychology suggests that secure attachment has a sensitive period in the child's first year of life. However, areas of the brain responsible for social skills have sensitive periods in adolescence. Therefore, naturally, interventions that promote secure attachment (e.g. Nurse-Family Partnership, see Olds 2002) occur in early childhood, while programmes that develop social skills happen in later stages of the life cycle (e.g. Kosse et al. 2020). This discussion shows that sensitive or critical development periods have significant consequences for designing interventions that foster human capital formation. Therefore, if policy makers decide to intervene, they need to pay attention to what and when the intervention will occur.

However, not all components of human capital have sensitive or critical periods of development. Vocabulary growth can continue to occur throughout an individual's lifetime. Furthermore, many forms of knowledge that we accumulate while in school (e.g. history, geography, and maths) do not seem to have sensitive or critical periods of development. Consequently, interventions that aim to foster the accumulation of these skills do not need to worry about the timing (early versus late) because no biological reason prevents individuals from learning these skills at later stages of the life cycle.

CH introduces the concept of the self-productivity of skills, which means that the skills produced in the early stages of the life cycle form the basis of the skills that individuals will develop in later stages. Oral language development helps an individual develop reading skills, and learning how to count numbers helps the child understand arithmetic operations. Therefore, self-productivity is linked with the concept of school readiness, which means that children possess 'the skills, knowledge, and attitudes necessary for success in school' (Head Start Office 2021).

Another essential feature of the CH model is the concept of dynamic complementarity, which means that investments in the early stages of the life cycle increase the returns to investments in later stages (and, crucially, vice versa). Dynamic complementarity offers another view of the empirical regularities that make pre-college skills fundamental for college enrolment and graduation described above. Nevertheless, BT has no room for public policy because these pre-college skills are endowments determined at conception. In Cunha and Heckman (2007), these same skills are the product of a sequence of investments which, combined with genetic makeup, raise or lower the returns to the investment in a college education. From the point of view of a life-cycle skill formation model, the optimal policy that promotes college enrolment

and graduation across socio-economic groups involves tuition subsidies and subsidies to skill formation that occurs in different stages of the life cycle. Skill begets skill.

These forces in the CH model depend on a few parameters of the technology of skill formation. Cunha et al. (2010) discuss the identification and estimation of these parameters, and they zero in on three problems that arise in this literature: the measures of skills do not have a cardinal scale, the measures of skills and investments are error-ridden, and the observed investments likely correlate with unobserved inputs in the production of skills.

The measures of human capital typically do not have a cardinal scale for economic outcomes—the difference between any two values is not itself meaningful. At best, as far as economic outcomes are concerned, these measures are ordinal. This point holds for cognitive (e.g. the Bayley Scale of Infant Development) and non-cognitive measures (e.g. the Child Behavior Checklist). Non-cardinality implies that standard statistical methods may not lead to valid inference—order-preserving transformations of the human capital measures could significantly impact estimates of self-productivity and dynamic complementarity. The lack of robustness is worrisome given the critical role cardinal analyses using these measures have played in guiding both academic research and public policy. To address this problem, Cunha et al. (ibid.) anchor test scores to cardinal outcomes (earnings) to rescale the measures to cardinal units. This research topic is active, and Bond and Lang (2013, 2018) and Nielsen (2015, 2019) build new tools to address the lack of cardinality in test scores.

Measures of human capital and investments have errors. Researchers use such measures as dependent variables, independent variables, or simultaneously as both to estimate the production functions of human capital. The ubiquity of measurement error in early childhood data requires methods that account for such errors, or inference may not be valid. Cunha et al. (2010) explore the fact that for each latent variable of interest (skills or investments), there are multiple measures (proxy variables) that are error-ridden. They show how they can aggregate across measures to address biases that arise from measurement error. The works by Williams (2019, 2020) and Rodríguez Sánchez (2020) further expand on this area of inquiry.

Observed investments correlate with unobserved inputs in the production function of human capital. This problem arises for several reasons. First, current datasets measure a small number of all goods and activities that promote child development. Second, although these datasets measure whether a good is present in the household or whether adults interact with the child in ways that promote development, they do not necessarily measure the frequency in

which the families use these goods or the quality of the interactions, although a few exceptions are starting to appear (e.g. Cunha et al. 2020). Third, children may have unobserved characteristics that simultaneously determine their skills and influence the families' decision of how much to invest in the children. Cunha et al. (ibid.) show how to adapt instrumental variable methods to account for the correlation between investments and omitted inputs. This research topic has benefited from new studies that collect richer investment data within randomised controlled trials, such as Attanasio et al. (2020, 2021).

The authors use the Children of the NLSY/79 dataset (henceforth CNLSY/79) to estimate the technology of skill formation. The CNLSY/79 participants are the children of the NLSY/79's female respondents. In this ongoing study, these children are assessed every two years, starting from birth. Most children were born between the mid-1980s to the mid-1990s, so the children are currently around 30-years-old. They find that the sensitive periods of development for cognitive skills are in early childhood and adolescence for non-cognitive skills. They also find evidence that dynamic complementarity becomes stronger as children age. The research on the identification and estimation of production functions is an active area, and contributions by Attanasio et al. (2020) and Agostinelli and Wiswall (2020) build on these tools.

Heckman's research has focused on estimating the long-term impacts of prototypical early childhood programmes in the last ten years, and this work has influenced policy making from local to state to federal government. This influence is because Heckman was able to transform policy makers' questions into research projects, and answer these questions thoroughly by implementing advanced methodological tools for data analysis and collecting data for new and old programmes. In this sense, Heckman's research goes well beyond the ivory tower, and it shapes the public debate about the tools governments have to enhance human potential, increase economic prosperity, and decrease inequality in socio-economic outcomes.

Heckman has focused primarily on four programmes. First, the Perry Preschool Program (PPP) is a centre-based early childhood project in Ypsilanti, Michigan. At its establishment, PPP taught three- and four-year-old children according to the HighScope Curriculum, the same curriculum the Head Start Program used when it was created. Thus, the PPP's evaluation studies estimate the impact of the Head Start Program if implemented with high fidelity.

Second, Heckman and his team have also studied the Abecedarian Program, a full-day centre-based programme from birth to age five in Chapel Hill, North Carolina. The early focus of the Program, and its intensity, allow us to draw a parallel with the Early Head Start Program. Therefore, evaluation matters not only for the sake of the Abecedarian Program but also because it

allows researchers and policy makers to quantify the impacts that a high-fidelity Early Head Start Program can produce on long-term socio-economic outcomes.

Third, Heckman has also studied the Nurse-Family Partnership, a home visitation programme that targets economically diverse pregnant mothers from the second trimester of pregnancy to a baby's second birthday. The programme's curriculum focuses on promoting a secure attachment between the baby and the mother. Nowadays, the Nurse-Family Partnership programme serves 200,000 families a year in the United States. The evaluation of its prototypical implementation in the twentieth century informs researchers and policy makers about the impacts of a high-fidelity home visitation programme on long-term outcomes.

More recently, Heckman has extended his research to developing countries, specifically China. In this work, he has collaborated with many researchers to culturally adapt and implement a version of the Jamaica Home Visitation Program. This research is fundamental for developing countries because they lack highly trained early childhood teachers. Therefore, researchers need to find a way to overcome a significant hurdle: it takes human capital from members of one generation to help members of another generation produce human capital.

I draw on Heckman et al. (2010) to illustrate Heckman's approach to evaluating the long-term impacts of these prototypical early childhood programmes. In a randomised controlled trial (RCT), the researchers randomly assign the study participants to the control or intervention condition. Thus, there is perfect compliance in textbook RCT cases: individuals in the control group never benefit from the intervention, while individuals in the intervention condition completely adhere to the treatment protocol. However, in reality, compliance is imperfect, either because individuals refuse to participate or because researchers need to modify the assignment rule. For example, a randomisation algorithm may assign one sibling to the control condition and its twin to the intervention one in a centre-based early childhood programme, and researchers may choose to reassign both siblings to the same condition. In addition, when sample sizes are small, unbalance in demographic characteristics due to the random assignment may lead researchers to move participants across groups to achieve the said balance. In any case, compliance is imperfect, and empirical analysis should not overlook any post-assignment swaps the research team executed before the beginning of the intervention.

The textbook RCT case involves assessing the impact of an intervention on a single outcome. In reality, researchers have dozens (if not hundreds) of outcomes of interest. As time goes by, researchers return to fieldwork and reassess

the same outcomes again (and may also include assessing different outcomes). As the number of outcomes increases, the probability of a false positive also increases. Without any adjustment, any evaluation will find impacts on at least some outcomes even if, in reality, there is none. In the statistics literature, this problem is known as multiple hypothesis testing. One powerful solution to it is to sample participants ex-post and relabel them randomly as "control" or "intervention" participants regardless of the authentic original assignment. If all the results are false positives, then this relabeling will not affect the conclusion, only the outcomes for which there is a false positive.

In the programmes evaluated by Heckman and his colleagues, the solution is not straightforward because of imperfect compliance. For example, we cannot simply randomly relabel participants and assume they will comply with the protocol because we know poor compliance existed in the original assignment. In addition, the imperfection is due to researchers executing post-assignment swaps according to maternal working status (to reduce the programme's costs). Heckman et al. (2010) significantly extend the methods by designing data analysis and deriving resampling strategies that simultaneously address the issues of multiple hypotheses testing and imperfect compliance. This research has impacted the analysis of RCTs in economics, and current applied research incorporates these insights when analysing data with imperfect compliance and multiple outcomes.

The application of these procedures shows that the PPP produced mixed results. The Program certainly improved educational and labour market outcomes for females and crime outcomes for males. They report that these findings are consistent with the PPP impacting primarily non-cognitive skills, thus reinforcing the importance of expanding the definition of human capital for research and policy making. They estimate a rate of return of 6% to 10% per year, which is sizeable, but Heckman's more recent research on the PPP shows that these rates of return are downward biased because the Program produced intragenerational and intergenerational impacts (Heckman and Karapakula 2019).

The PPP cohort and the NLSY/79 participants were born around the same years. When Heckman et al. (2010) match participants from the NLSY/79 to those of the Perry study, they conclude that 16% of the US population would be eligible for the PPP. In 2019, there were nearly 20 million children between 0 and 5-years-old in the United States according to the Census Bureau, and the Head Start Program served about 5% of these children. These figures indicate that, although an early childhood programme such as Head Start should not be universal, there is still significant room for growth so that all eligible children can receive such services.

García et al. (2020) apply and extend the methods from Heckman et al. (2010) to evaluate the Abecedarian and the Nurse-Family Partnership programme. The evaluation of the Abecedarian Program uses biomedical data that contain measures of many dimensions of health. The authors find that, at age 40, Abecedarian children had better blood pressure, lower blood glucose levels, higher "good" cholesterol levels, lower triglycerides, and smaller waist circumference, all indicating a more positive cardiovascular health.

For the Nurse-Family Partnership programme, Heckman et al. (2017) find positive impacts on cognitive skills for males and females up to age 12. In addition, they show that the children in the Partnership group benefit from growing up in a home environment that offers children more development opportunities. These findings show that skills are malleable and that the home environment significantly fosters human capital formation.

Finally, Heckman et al. (2020) evaluates the causal impacts of an early childhood home visiting programme in China. This work involves an RCT with primary data collection in the field. The authors find that the programme substantially improves child language and cognitive, fine motor, and social-emotional skills development. To the best of my knowledge, the paper is the first to investigate whether an intervention affects the mapping between measurement and latent skills. The authors find that the random assignment to the intervention group affects the relationship between measured and latent skills. This finding indicates that some of the impacts of the programme are not due to improvement in skills, but to changes in the manifestation of skills (e.g. the children in the treatment group may participate in activities that resemble assessment of child development; thus they become better at doing these tasks).

11   Conclusion

I have summarised three areas in which James Heckman's research has been transformational. However, one should not compartmentalise these contributions to specific, separate fields. These contributions maintain a dialogue inside Heckman's head, his co-authors' heads, and his students' heads, forever. For example, Heckman employs the CHH framework to study the importance of non-cognitive skills, thus reshaping our understanding of what human capital is, how we form these skills, and at what point in time we should intervene. The findings from the life-cycle skill formation research, in turn, motivated Heckman to investigate the long-run impacts of early childhood programmes. Armed with this knowledge, Heckman goes back to the

CHH framework and extends it to a dynamic framework. It is chaotic and exhilarating work.

Heckman's research should be understood from a single question that unifies all of his work: How can we use social science to improve lives? The response is a philosophy of work that makes cutting corners unacceptable, not turning every stone inadmissible, and not doing the homework inappropriate. Alternatively, as I have heard many times, 'this is not going to be easy, but we need to bite the bullet'.

References

Main Works by James J. Heckman

Almlund, M., A. Duckworth, J.J. Heckman and T. Kautz (2011). 'Personality Psychology and Economics'. Chapter 1 in E.A. Hanushek, S. Machin and L. Woessmann (eds) *Handbook of the Economics of Education*. Volume 4. Amsterdam: Elsevier: 1–181.

Ashenfelter, O. and J.J. Heckman (1974). 'The Estimation of Income and Substitution Effects in a Model of Family Labor Supply'. *Econometrica*, 42(1): 73–85.

Borghans, L., B.H.H. Golsteyn, J.J. Heckman and H. Meijers (2009). 'Gender Differences in Risk Aversion and Ambiguity Aversion'. *Journal of the European Economic Association*, 7(2/3): 649–658.

Browning, M., L.P. Hansen and J.J. Heckman (1999). 'Micro Data and General Equilibrium Models'. Chapter 8 in J. Taylor and M. Woodford (eds) *Handbook of Macroeconomics*. Volume 1. Amsterdam: Elsevier: 543–633.

Cameron, S.V. and J.J. Heckman (1993). 'The Nonequivalance of High School Equivalents'. *Journal of Labor Economics*, 11(1): 1–47.

Cameron, S.V. and J.J. Heckman (2001). 'The Dynamics of Educational Attainment for Black, Hispanic, and White Males'. *Journal of Political Economy*, 109(3): 455–499.

Carneiro, P., K.T. Hansen and J.J. Heckman (2003). 'Estimating Distributions Of Treatment Effects With An Application To The Returns To Schooling and Measurement Of The Effects Of Uncertainty On College Choice'. *International Economic Review*, 44(2): 361–422.

Carneiro, P. and J.J. Heckman (2002). 'The Evidence on Credit Constraints in Post-Secondary Schooling'. *Economic Journal*, 112(482): 705–734.

Carneiro, P., J.J. Heckman and E.J. Vytlacil (2011). 'Estimating Marginal Returns to Education'. *American Economic Review*, 101(6): 2,754–2,781.

Cunha, F. and J.J. Heckman (2007). 'The Technology of Skill Formation'. *American Economic Review*, 97(2): 31–47.

Cunha, F. and J.J. Heckman (2016). 'Decomposing Trends in Inequality in Earnings into Forecastable and Uncertain Components'. *Journal of Labor Economics*, 34(S2): 31–65.

Cunha, F., J.J. Heckman and S.L. Navarro (2005). 'Separating Uncertainty from Heterogeneity in Life Cycle Earnings'. *Oxford Economic Papers*, 57(2): 191–261.

Cunha, F., J.J. Heckman and S.M. Schennach (2010). 'Estimating the Technology of Cognitive and Noncognitive Skill Formation'. *Econometrica*, 78(3): 883–931.

Flinn, C.J. and J.J. Heckman (1982). 'New Methods for Analyzing Individual Event Histories'. *Sociological Methodology*, 13: 99–140.

Flinn, C.J. and J.J. Heckman (1983). 'Are Unemployment and Out of the Labor Force Behaviorally Distinct Labor Force States?'. *Journal of Labor Economics*, 1(1): 28–42.

García, J.L., J.J. Heckman, D.E. Leaf and M.J. Prados (2020). 'Quantifying the Life-Cycle Benefits of an Influential Early-Childhood Program'. *Journal of Political Economy*, 128(7): 2,502–2,541.

Heckman, J.J. (1974a). 'Shadow Prices, Market Wages, and Labor Supply'. *Econometrica*, 42(4): 679–694.

Heckman, J.J. (1974b). 'Effects of Child-Care Programs on Women's Work Effort'. *Journal of Political Economy*, 82(2): S136–S163.

Heckman, J.J. (1976a). 'A Life Cycle Model of Earnings, Learning and Consumption'. *Journal of Political Economy*, 84(4): S11–S44.

Heckman, J.J. (1976b). 'The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables'. *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. (1981a). 'Statistical Models for Discrete Panel Data'. Chapter 3 in C.F. Manski and D. McFadden (eds) *Structural Analysis of Discrete Data with Econometric Applications*. Cambridge, MA: MIT Press: 114–178.

Heckman, J.J. (1981b). 'The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process'. Chapter 4 in C.F. Manski and D. McFadden (eds) *Structural Analysis of Discrete Data with Econometric Applications*. Cambridge, MA: MIT Press: 179–195.

Heckman, J.J. (1984). 'The χ^2 Goodness of Fit Statistic for Models with Parameters Estimated from Microdata'. *Econometrica*, 52(6): 1,543–1,547.

Heckman, J.J. (1990). 'Varieties of Selection Bias'. *American Economic Review*, Papers and Proceedings, 80(2): 313–318.

Heckman, J.J. (1991). 'Identifying the Hand of Past: Distinguishing State Dependence from Heterogeneity'. *American Economic Review*, Papers and Proceedings, 81(2): 75–79.

Heckman, J.J. (1993). 'What Has Been Learned About Labor Supply in the Past Twenty Years?'. *American Economic Review*, Papers and Proceedings, 83(2): 116–121.

Heckman, J.J. (1996). 'Identification of Causal Effects Using Instrumental Variables: Comment'. *Journal of the American Statistical Association*, 91(434): 459–462.

Heckman, J.J. (1997). 'Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations'. *Journal of Human Resources*, 32(3): 441–462.

Heckman, J.J. (2000). 'Causal Parameters and Policy Analysis in Economics: A 20th Century Retrospective'. *Quarterly Journal of Economics*, 115(1): 45–97.

Heckman, J.J., M.L. Holland, K.K. Makino, R. Pinto and M. Rosales-Rueda (2017). 'An Analysis of the Memphis Nurse-Family Partnership Program'. NBER Working Paper 23610. Cambridge, MA: National Bureau of Economic Research.

Heckman, J.J. and B.E. Honoré (1989). 'The Identifiability of the Competing Risks Model'. *Biometrika*, 76(2): 325–330.

Heckman, J.J. and B.E. Honoré (1990). 'The Empirical Content of the Roy Model'. *Econometrica*, 58(5): 1,121–1,149.

Heckman, J.J. and V.J. Hotz (1989). 'Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training'. *Journal of the American Statistical Association*, 84(408): 862–874.

Heckman, J.J., V.J. Hotz and J.R. Walker (1985). 'New Evidence on the Timing and Spacing of Births'. *American Economic Review*, Papers and Proceedings, 75(2): 179–184.

Heckman, J.J., J.E. Humphries and G. Veramendi (2018). 'Returns to Education: The Causal Effects of Education on Earnings, Health, and Smoking'. *Journal of Political Economy*, 126(S1): 197–246.

Heckman, J.J., H. Ichimura, J. Smith and P. Todd (1998). 'Characterizing Selection Bias Using Experimental Data'. *Econometrica*, 66(5): 1,017–1,098.

Heckman, J.J., H. Ichimura and P. Todd (1997a). 'Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program'. *Review of Economic Studies*, 64(4): 605–654.

Heckman, J.J. and G. Karapakula (2019). 'Intergenerational and Intragenerational Externalities of the Perry Preschool Project'. NBER Working Paper 25889. Cambridge, MA: National Bureau of Economic Research.

Heckman, J.J. and M. Killingsworth (1987). 'Female Labor Supply: A Survey'. Chapter 2 in O. Ashenfelter and R. Layard (eds) *Handbook of Labor Economics*. Volume 1. Amsterdam: North-Holland: 103–204.

Heckman, J.J., B. Liu, M. Lu and J. Zhou (2020). 'Treatment Effects and the Measurement of Skills in a Prototypical Home Visiting Program'. NBER Working Paper 27356. Cambridge, MA: National Bureau of Economic Research.

Heckman, J.J., L. Lochner and C. Taber (1998a). 'Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents'. *Review of Economic Dynamics*, 1(1): 1–58.

Heckman, J.J., L. Lochner and C. Taber (1998b). 'General-Equilibrium Treatment Effects: A Study of Tuition Policy'. *American Economic Review*, 1(1): 1–58.

Heckman, J.J. and T.E. MaCurdy (1980). 'A Life Cycle Model of Female Labor Supply'. *Review of Economic Studies*, 47(1): 47–74.

Heckman, J.J. and T.E. MaCurdy (1981). 'Current Theoretical and Empirical Studies of Labor Supply: Second Generation Studies'. Chapter 4 in R.G. Ehrenberg (ed.) *Research in Labor Economics*. Volume 4. Greenwich, CT: JAI Press: 65–102.

Heckman, J.J. and T.E. MaCurdy (1986). 'Labor Econometrics'. Chapter 32 in Z. Grilliches and M.D. Intriligator (eds) *Handbook of Econometrics*. Volume 3. Amsterdam: North-Holland: 1,917–1,977.

Heckman, J.J., S.H. Moon, R. Pinto, P. Savelyev and A. Yavitz (2010). 'Analyzing Social Experiments as Implemented: A Reexamination of the Evidence from the HighScope Perry Preschool Program'. *Quantitative Economics*, 1(1): 1–46.

Heckman, J.J. and R. Robb, Jr. (1985). 'Alternative Methods for Evaluating the Impact of Interventions'. Chapter 4 in J.J. Heckman and B. Singer (eds) *Longitudinal Analysis of Labor Market Data*. Cambridge: Cambridge University Press: 156–246.

Heckman, J.J. and R. Robb, Jr. (1986). 'Alternative Identifying Assumptions in Econometric Models of Selection Bias'. In G.F. Rhodes, Jr. and D.J. Slottje (eds) *Innovations in Quantitative Economics: Essays in Honor of Robert L. Basmann. Advances in Econometrics*. Volume 5. Greenwich, CT: JAI Press: 243–287.

Heckman, J.J. and Y. Rubinstein (2001). 'The Importance of Noncognitive Skills: Lessons from the GED Testing Program'. *American Economic Review*, Papers and Proceedings, 91(2): 145–149.

Heckman, J.J. and J. Scheinkman (1987). 'The Importance of Bundling in a Gorman-Lancaster Model of Earnings'. *Review of Economic Studies*, 54(2): 243–255.

Heckman, J.J. and G. Sedlacek (1985). 'Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market'. *Journal of Political Economy*, 93(6): 1,077–1,125.

Heckman, J.J. and G. Sedlacek (1990). 'Self-Selection and the Distribution of Hourly Wages'. *Journal of Labor Economics*, 8(1): S329–S363.

Heckman, J.J. and B. Singer (1982). 'The Identification Problem in Econometric Models for Duration Data'. Chapter 2 in W. Hildenbrand (ed.) *Advances in Econometrics*. Proceedings of Fourth World Congress of Econometric Society. Cambridge: Cambridge University Press: 39–78.

Heckman, J.J. and B. Singer (1984a). 'A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data'. *Econometrica*, 52(2): 271–320.

Heckman, J.J. and B. Singer (1984b). 'The Identifiability of the Proportional Hazard Model'. *Review of Economic Studies*, 51(2): 231–241.

Heckman, J.J. and B. Singer (1985). *Longitudinal Analysis of Labor Market Data*. Cambridge: Cambridge University Press.

Heckman, J.J., J. Smith and N. Clements (1997b). 'Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts'. *Review of Economic Studies*, 64(4): 487–535.

Heckman, J.J., J. Stixrud and S. Urzua (2006). 'The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior'. *Journal of Labor Economics*, 24(3): 411–482.

Heckman, J.J. and E. Vytlacil (1999). 'Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects'. *Proceedings of the National Academy of Sciences of the United States of America*, 96(8): 4,730–4,734.

Heckman, J.J. and E. Vytlacil (2000). 'The Relationship between Treatment Parameters Within a Latent Variable Framework'. *Economics Letters*, 66(1): 33–39.

Heckman, J.J. and E. Vytlacil (2005). 'Structural Equations, Treatment Effects, and Econometric Policy Evaluation'. *Econometrica*, 73(3): 669–738.

Heckman, J.J. and J.R. Walker (1989). 'Forecasting Aggregate Period-Specific Birth Rates: The Time Series Properties of a Microdynamic Neoclassical Model of Fertility'. *Journal of the American Statistical Association*, 84(408): 958–965.

Other Works Referred To

Agostinelli, F. and M.J. Wiswall (2020). 'Estimating the Technology of Children's Skill Formation'. NBER Working Paper 22442. Cambridge, MA: National Bureau of Economic Research.

Attanasio, O., S. Cattan, E. Fitzsimons, C. Meghir and M. Rubio-Codina (2020). 'Estimating the Production Function for Human Capital: Results from a Randomized Control Trial in Colombia'. *American Economic Review*, 110(1): 48–85.

Attanasio, O., F. Cunha and P. Jervis (2021). 'Subjective Parental Beliefs: Their Measurement and Role'. NBER Working Paper 26516. Cambridge, MA: National Bureau of Economic Research.

Barro, R.J. (2001). 'Human Capital and Growth'. *American Economic Review*, Papers and Proceedings, 91(2): 12–17.

Becker, G.S. (1962). 'Investment in Human Capital: A Theoretical Analysis'. *Journal of Political Economy*, 70(5): 9–49.

Becker, G.S. and B.R. Chiswick (1966). 'Education and the Distribution of Earnings'. *American Economic Review*, 56(1/2): 358–369.

Becker, G.S. and N. Tomes (1986). 'Human Capital and the Rise and Fall of Families'. *Journal of Labor Economics*, 4(3): S1–S39.

Bond, T.N. and K. Lang (2013). 'The Evolution of the Black-White Test Score Gap in Grades K-3: The Fragility of Results'. *Review of Economics and Statistics,* 95(5): 1,468–1,479.

Bond, T.N. and K. Lang (2018). 'The Black-White Education Scaled Test-Score Gap in Grades K-7'. *Journal of Human Resources*, 53(4): 891–917.

Bowles, S. and H. Gintis (1976). *Schooling in Capitalist America*. New York, NY: Basic Books.

Card, D. (1999). 'The Causal Effect of Education on Earnings'. Chapter 30 in O. Ashenfelter and D. Card (eds) *Handbook of Labor Economics*. Volume 3. Amsterdam: Elsevier: 1,801–1,863.

Cunha, F., M. Gerdes and S. Nihtianova (2020). 'Language Environment and Maternal Expectations: An Evaluation of the LENA Start Program'. Available at: https://www.cesifo.org/sites/default/files/events/2020/ee20_Cunha.pdf.

Daly, M., C.P. Harmon and L. Delaney (2009). 'Psychological and Biological Foundations of Time Preference'. *Journal of the European Economic Association*, 7(2–3): 659–669.

Daw, N. (1998). 'Critical Periods and Amblyopia'. *Archives of Ophthalmology*, 116(4): 502–505.

Dohmen, T., A. Falk, D. Huffman and U. Sunde (2010). 'Are Risk Aversion and Impatience Related to Cognitive Ability?'. *American Economic Review*, 100(3), 1,238–1,260.

Eckel, C.C. and P.J. Grossman (2002). 'Sex Differences and Statistical Stereotyping in Attitudes Toward Financial Risk'. *Evolution and Human Behavior*, 23(4): 281–295.

Edwards, R.C. (1976). 'Individual Traits and Organizational Incentives: What Makes A "Good" Worker?'. *Journal of Human Resources*, 11(1): 51–68.

Griliches, Z. (1977). 'Estimating the Returns to Schooling: Some Econometric Problems'. *Econometrica*, 45(1): 1–22.

Head Start Office (2021). 'Head Start Approach to School Readiness – Overview'. Available at: https://eclkc.ohs.acf.hhs.gov/school-readiness/article/head-start-approach-school-readiness-overview.

Imbens, G.W. and J.D. Angrist (1994). 'Identification and Estimation of Local Average Treatment Effects'. *Econometrica*, 62(2): 467–475.

Kane, T. (2001). 'College-Going and Inequality: A Literature Review'. Working Paper, Russel Sage Foundation. Available at: https://www.russellsage.org/research/reports/college-going.

Klein, R., R. Spady and A. Weiss (1991). 'Factors Affecting the Output and Quit Propensities of Production Workers'. *Review of Economic Studies*, 58(5): 929–953.

Kosse, F., T. Deckers, P. Pinger, H. Schildberg-Hörisch and A. Falk (2020). 'The Formation of Prosociality: Causal Evidence on the Role of Social Environment'. *Journal of Political Economy*, 128(2): 434–467.

Matzkin, R.L. (1992). 'Nonparametric and Distribution-Free Estimation of the Binary Threshold Crossing and the Binary Choice Models'. *Econometrica*, 60(2): 239–270.

Mincer, J. (1958). 'Investment in Human Capital and Personal Income Distribution'. *Journal of Political Economy*, 66(4): 281–302.

Mortensen, D.T. and C.A. Pissarides (1994). 'Job Creation and Job Destruction in the Theory of Unemployment'. *Review of Economic Studies*, 61(3): 397–415.

Newport, E. (1990). 'Maturational Constraints on Language Learning'. *Cognitive Science*, 14(1): 11–28.

Nielsen, E.R. (2015). 'Achievement Gap Estimates and Deviations from Cardinal Comparability'. Finance and Economics Discussion Series 2015-040. Board of Governors of the Federal Reserve System. Available at: https://www.federalreserve.gov/econresdata/feds/2015/files/2015040pap.pdf.

Nielsen, E. (2019). 'Test Questions, Economic Outcomes, and Inequality'. Finance and Economics Discussion Series 2019-013. Board of Governors of the Federal Reserve System. Available at: https://www.federalreserve.gov/econres/feds/files/2019013pap.pdf.

Olds, D.L. (2002). 'Prenatal and Infancy Home Visiting by Nurses: From Randomized Trials to Community Replication'. *Prevention Science*, 3(3): 153–172.

Pigou, A.C. (1928). *A Study in Public Finance*. London: Macmillan.

Rodríguez Sánchez, A. (2020). 'Is It Just Noise? Measuring Unobservable Cognitive Abilities in Early Childhood'. *Personality and Individual Differences*, 110162. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0191886920303512.

Rustichini, A., C.G. DeYoung, J.E. Anderson and S.V. Burks (2016). 'Toward the Integration of Personality Theory and Decision Theory in Explaining Economic Behavior: An Experimental Investigation'. *Journal of Behavioral and Experimental Economics*, 64(October): 122–137.

Schultz, T.W. (1961). 'Investment in Human Capital'. *American Economic Review*, 51(1): 1–17.

Smith, A. (1999). *The Wealth of Nations*. Books I–III. London: Penguin.

Williams, B. (2019). 'Controlling For Ability Using Test Scores'. *Journal of Applied Econometrics*, 34(4): 547–565.

Williams, B. (2020). 'Identification of the Linear Factor Model'. *Econometric Reviews*, 39(1): 92–109.

Zucker, L.G., M.R. Darby and M.B. Brewer (1998). 'Intellectual Human Capital and the Birth of US Biotechnology Enterprises'. *American Economic Review*, 88(1): 290–306.

Zuckerman, M. (1994). *Behavioral Expressions and Biosocial Bases of Sensation Seeking*. New York, NY: Cambridge University Press.