GAUTHIER LANOT
Umeå University, Sweden

I Z A
World of Labor
Evidence-based policy making

Maximum likelihood and economic modeling

Maximum likelihood is a general and flexible method to estimate the parameters of models in labor economics

Keywords: log-likelihood, economic model, parameter estimates

ELEVATOR PITCH

Most of the data available to economists is observational rather than the outcome of natural or quasi experiments. This complicates analysis because it is common for observationally distinct individuals to exhibit similar responses to a given environment and for observationally identical individuals to respond differently to similar incentives. In such situations, using maximum likelihood methods to fit an economic model can provide a general approach to describing the observed data, whatever its nature. The predictions obtained from a fitted model provide crucial information about the distributional outcomes of economic policies.

Observed and predicted distribution of individual choice of work hours (h) at offered wages (w)
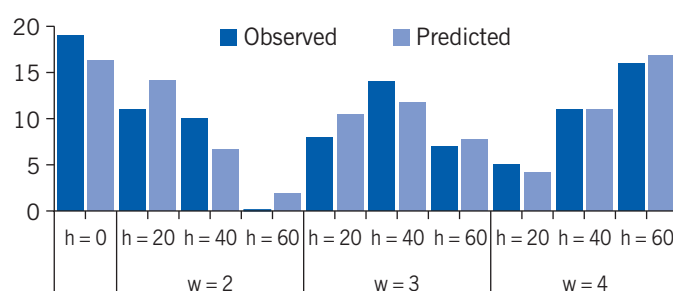
Note: The predicted distribution is obtained at the maximum likelihood estimates of the model parameters.
Source: Author's analysis using fictitious data.

I Z A
World of Labor

KEY FINDINGS

Pros

⊕ Maximum likelihood methods enable the measurement of parameters of complex economic models even in the presence of complex sampling schemes.

⊕ Maximum likelihood can be used to analyze any kind of data, whether observational, experimental, or quasi-experimental.

⊕ By using all the information available in a sample, maximum likelihood methods produce measurements that are as "good as one can get" given the specified model.

⊕ Maximum likelihood methods allow the testing of all testable hypotheses that describe a given model.

Cons

⊖ Maximum likelihood methods require the specification of an economic model that properly describes the probability distribution of the events observed in the sample.

⊖ Using maximum likelihood requires assumptions about the distribution of unobserved components of the model.

⊖ If the modeling assumptions are not satisfied, the estimates will suffer from bias, which affects robustness.

⊖ Maximum likelihood methods rely on numerical methods to evaluate and maximize the likelihood.

AUTHOR'S MAIN MESSAGE

Maximum likelihood is the statistical method of choice for estimating parameters in situations where data are not amenable to the use of statistical techniques that are generally applied to data from experimental studies with a homogeneous population (such as difference-in-differences analysis). A well-specified economic model is a useful tool for understanding the consequences of policy changes that are not necessarily marginal or that affect the distribution of outcomes as well as the mean outcomes. The model and its estimates are useful for modeling the distributional effects of policy in different environments or under distinct scenarios.

MOTIVATION

Much of the recent empirical research in labor economics focuses on the use of field, natural, quasi, or deliberate experiments within a reduced-form framework to identify and measure the causal impacts of economic incentives. This may suggest that progress will come from further exploiting similar methods [1]. Unfortunately for most empirical labor economists, evidence is rarely of this canonical kind: the data may frequently be observational, not experimental, and there may be few, if any, quasi-experimental features that can be exploited. Moreover, it is likely that there are no obvious or convincing instrumental variables available.

Reduced-form framework and structural models

Structural model

Given the assumptions describing individual behavior (for example, to maximize utility given a budget constraint) and the assumptions related to the institutions of the market in which the agents operate, a family of structural economic models describes how endogenous outcomes (for example, the quantity of labor supplied, the amount of effort exerted at work, and the number of job offers rejected) vary in the population. The outcomes are usually determined given the description of exogenous (predetermined) conditions that may or may not be observed. Distinct values of the structural form parameters describe distinct members of a given family of model. The analyst can interpret the parameters of a structural model in terms of the behavioral or the institutional assumptions (such as the response in labor supply to wage changes, the coefficient of risk aversion, and the arrival of job offers).

Knowledge of the structural form is important since it allows a researcher to calculate the distribution of outcomes of the model in a variety of situations, even those that are not observed. Constraints on behavior can be changed or relaxed (for example, income taxes can be added to the model), new constraints on behavior can be introduced (for example, a minimum wage can be instituted or the retirement age can be increased), or different market institutions can be considered. Hence a structural model can describe (given its assumptions) a variety of counterfactual distributions of outcomes.

Reduced form

The reduced form of a structural model describes the relationship between the endogenous variables on the one hand, and the exogenous variables and the unobserved components on the other hand. While the reduced form depends on the same parameters as the structural form, some precise features of the relationship between the endogenous and exogenous variables is not available to the analyst when the model is expressed as a reduced form. In general, it is not possible to interpret the parameters of the reduced form in terms of the parameters of the structural model. Estimation of the parameters of the reduced form is often easier than estimation of the parameters of the structural form.

A model that describes how the demand for and the supply of labor depend on the wage rate has a structural interpretation (the model requires that the analyst knows both the wage elasticity of demand and the wage elasticity of supply). The description of the demand and the supply of labor, together with the assumption that at equilibrium labor supply matches labor demand, determines the quantity and the price of labor at equilibrium. The reduced form of the model (in this case) describes how the equilibrium wage and labor quantity vary as the general economic conditions change. The parameters of the reduced form are only informative about the structural parameter values (the supply and demand wage elasticities) if, for example, some of the variables that describe the economic conditions can be specifically attributed to the demand side or to the supply side.

> ### Instrumental variable
>
> In a regression context, the satisfactory estimation of the model parameters rests on the assumption that the explanatory variables are not correlated with the unobserved components of the model. When legitimate explanatory variables are omitted or when one explanatory variable and the explained variable of the regression model are determined simultaneously, direct estimation methods like ordinary least squares will yield biased measurements of the regression parameters. If the analyst can credibly identify additional variables that correlate well with both the explained and the explanatory variables in the regression model, these variables can be used as instrumental variables (instruments) in order to estimate the model parameters without bias.

Thus, most empirical economists are left to make unpalatable choices: ignore the "wrong kind" of data (observational data) or attempt to recast the data into the acceptable kind, such as by finding a natural experiment to exploit. While attempting to recast the data may occasionally succeed, this approach may not always be possible. Furthermore, in many cases, designing economic policy requires measuring quantities that go beyond the average effect of the treatment on the treated, for example, a wage or an income elasticity. In general, the average treatment effect is likely to be a mixture of parameters that are crucial to the design of a better policy [2], [3].

Economics proposes structural models in which the interactions between observed and unobserved characteristics describe how the data are distributed. Combining structural models with maximum likelihood estimation can provide a useful alternative route to the analysis of both observational and experimental data.

DISCUSSION OF PROS AND CONS

Models and data

An estimable economic model makes a number of assumptions about the relationships between variables that describe the environment (exogenous variables) and variables that describe the outcome of interest (endogenous variables). When there are many endogenous variables, the model describes the relationships among them. Finally, the model describes how the variability from uncontrolled sources determines the distribution of outcomes. The precise distribution of outcomes implied by the model depends on the unknown values of particular parameters. The model assumptions can be used to predict the distribution of the observed outcomes in the sample given the exogenous variables and the parameter values, which in turn allows evaluating the likelihood of the data. The purpose of the maximum likelihood estimation method is to find the parameter values that are most likely to generate the observed data: the parameter values that maximize the likelihood of the observed data.

Labor economists already use maximum likelihood methods to estimate parameter values when the conditional distribution of a given type of outcome is well known—discrete (binary, ordered, multinomial, or count models), censored or truncated (tobit models), or selected (Roy model and its generalizations), for example. Most econometric software provides "ready-to-use" estimation routines to obtain parameter estimates for models that fit each particular instance.

In a few cases, where the data are generated through a quasi or natural experiment, the economic model does not have to be fully formulated in order to measure the response of specific outcomes to a given policy change. Nevertheless, such measurements are meaningful only for characterizing the competing models of the world that the analyst considers when analyzing the experiment that generates the data.

The power of economic analysis lies in its ability to clearly state the assumptions about the factors that determine the behavior of agents, given the constraints they face and the assumptions about the mechanisms that allow individual actions to be consistent which each other. For example, in the simplest and most common economic analysis, economists describe a demand curve and a supply curve for a single market and discuss how their interactions determine the equilibrium quantities and prices that are observed.

Models of this type are useful in practice when the analyst is able to distinguish the demand price elasticity (the ratio of the relative change in quantities demanded in response to a 1% change in price) from the supply price elasticity (the ratio of the relative change in quantities supplied in response to a 1% change in price). This is not always possible, however, since the evidence available (equilibrium prices and quantities) is the result of unobserved interactions between the demand and the supply sides. Two strategies are available to make it possible to use observational data to distinguish the behavior of the supply side from that of the demand side: exclusion restrictions, which posit exogenous sources of variations specific to each side, and restrictions on the shape of the demand or the supply schedules, so that the quantity demanded is a decreasing function of the price while the quantity supplied is an increasing function of the price [4]. While one hypothesis may seem more suitable than the other, determining what can be learned from the data requires developing hypotheses that define the critical parameters to be measured.

Richer data require modeling strategies that represent the main features that can be observed. In particular, heterogeneity is an important characteristic of the data labor economists use. It is quite likely in many studies that unobserved/unexplained variability among otherwise similar observations will account for more than half of the observed variability. Having this much heterogeneity that does not correlate strongly with observed characteristic is a feature of data generated by human behavior: even within a homogeneous group of people, the potential variability is large [5]. The hypothesis is that interactions between the determinants of behavioral responses to economic incentives and individual heterogeneity explain the observed features of the data. The measurement of the model parameters may not be amenable to traditional linear regression but may require methods that rely on the distributional characteristics of the data, like the maximum likelihood method.

An illustrative example

An example can make it easier to understand the process of modeling, estimation, and interpretation. Consider a labor supply model for a single worker [6]. The worker will supply labor if the offered wage is large enough; otherwise, they will not participate in the labor market. If this worker supplies their labor, their hourly wage and hours of work per period may be observed. If they do not supply any labor, there are no observable data on the wage that they might have been offered. Thus, individual behavior generates complex data.

Now consider an imaginary distribution of a random sample of a population of identical individual workers, assuming (to simplify) that three different wage levels are offered and that three positive levels of hours of work are available (see the illustration on p. 1). Non-participants are represented by zero hours, and no wage is recorded for them (because observers would not know the offered wage since the job offer was rejected).

Next, a simple labor supply model is fitted to this observed distribution. The model assumes that each individual receives a wage offer (one of the three potential wages) and then decides on the number of hours of work. Individuals are assumed to choose their preferred option by weighing the enjoyment from consuming their earnings against the disutility of work for each alternative level of hours. While the utility of consumption increases with consumption, the marginal disutility of work increases as the time spent at work increases. In addition, there are unobserved components that introduce randomness to the individual decisions: individuals who are observationally identical (in the example, it could be imagined that all individuals in the population have the same age, education, and marital status) can nonetheless exhibit distinct optimal choice (because of personal characteristics that are not observable, such as an individual's dislike for work). The observed differences in outcomes are thus the result of unobserved differences among individuals. The model depends on three parameters that measure the trade-off between consumption and work as well as the importance of the unobserved component in the distribution of individual choices.

Finally, assume that the probability that a given wage is offered is a single parameter for each wage level and that the process that generates the wage offer and the process that generates individual differences in taste for work are independent (meaning that the observation of the wage offered does not carry any information about the unobserved factors that determine the individual's choices about work). Therefore, two independent probability parameters characterize the distribution of wages.
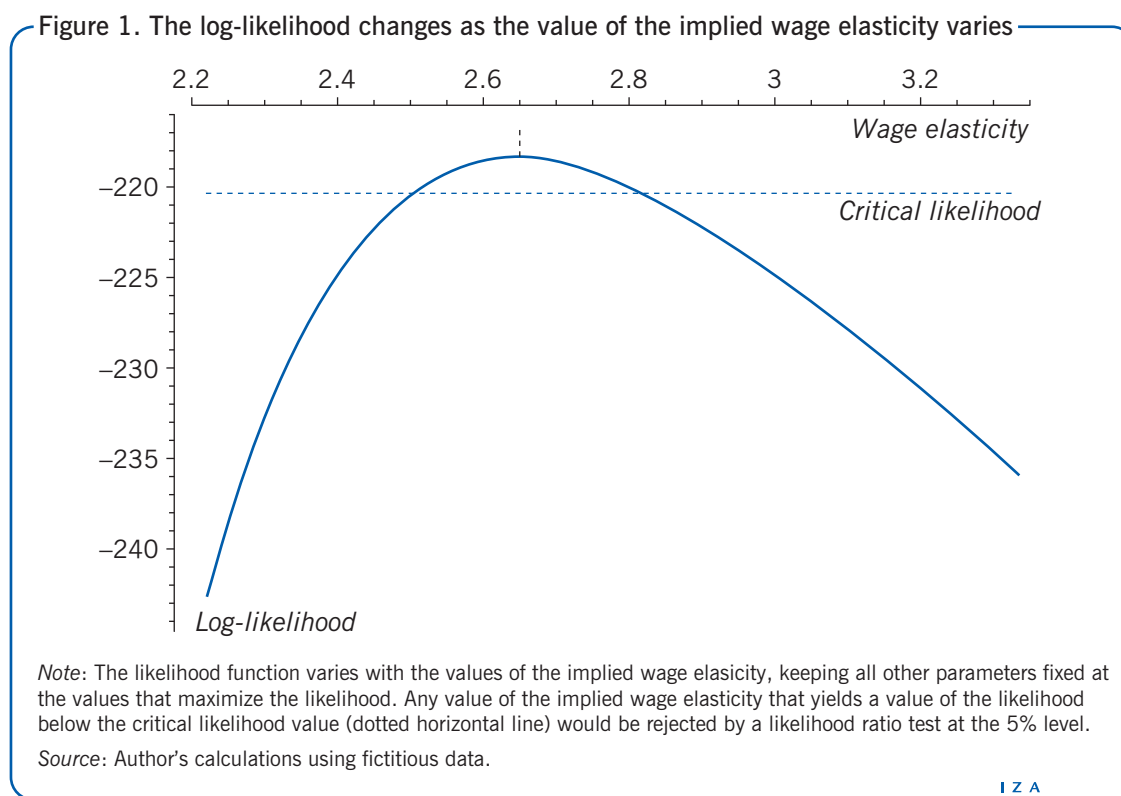
The likelihood of the observed sample is then the product over all observations of the probabilities that the hours of work and the wage are observed. For non-participants, the wage is missing and zero hours are observed. The logarithm of the likelihood simplifies the analysis and the search for the parameter values that yield the largest sample likelihood. In this example, it is the weighted sum of the logarithm of the probability of not working and the weighted sum of the summed logarithms of the probabilities of working positive hours for each possible level of hours and for all possible wages.

The need to evaluate and code such an expression is an obvious disadvantage of the maximum likelihood method: in complex cases, it is difficult and time consuming to establish that the log-likelihood is correctly coded. In the current, simple case, the log likelihood is easily maximized using conventional numerical methods (both commercial statistical software and free, open-source math software like R or SageMath supply an optimizer). The illustration on page 1 displays the observed distribution of outcomes and the predicted distribution using the maximum likelihood method for the five parameters of the model. The predicted distribution is similar, although not identical, to the observed distribution.

The model for the preferences would yield a labor supply function that takes a simple (log) linear form, such that the logarithm of the hours of work is linearly related to the logarithm of the wage rate. If the observations with zero hours are ignored, the wage elasticity can be estimated by a simple regression over the positive hours (and positive

wages) only. The maximum likelihood estimates give an implied elasticity of 2.63, while the regression approach gives an estimate of about 0.66. The difference between the two estimates suggests that the parameter estimates are sensitive to the precise specification of the model and to the sample composition. Since the wage offer distribution and the distribution of the unobserved taste components are independent, the difference between the two estimates arises because of sample selection (the exclusion from the regression estimation sample of the observations for non-participating individuals) and because the observed hours are not determined according to the labor supply function.

The shape of the log-likelihood describes how precisely the parameters are estimated. Figure 1 shows how the log-likelihood changes as the value of the implied wage elasticity varies, keeping all other parameters at their estimated values. Values of the implied wage elasticity smaller than 2.5 and larger than 2.85 are not supported by the evidence in the data. Furthermore, the 0.66 value obtained by a simple (ordinary least squares) estimation applied to the data with positive hours is far from the maximum likelihood estimate based on the model that accounts for all observations.

Figure 1. The log-likelihood changes as the value of the implied wage elasticity varies

Note: The likelihood function varies with the values of the implied wage elasicity, keeping all other parameters fixed at the values that maximize the likelihood. Any value of the implied wage elasticity that yields a value of the likelihood below the critical likelihood value (dotted horizontal line) would be rejected by a likelihood ratio test at the 5% level.

Source: Author's calculations using fictitious data.

I Z A
World of Labor

Model and policy evaluation

Economic models and likelihood methods can be used together in evaluation studies. Suppose a policymaker wants to understand the labor supply response to the introduction of a minimum wage. The minimum wage does not operate in the first region in either period 1 or period 2, and it operates in the second region only in period 2. Assume, to simplify, that four independent random samples are observed, one for each region and time period, as shown in the upper part of Figure 2. The distribution of the random sample in the second region and in the second period across labor supply and wages

Figure 2. Observed and predicted distribution of labor supply and wages before and after introduction of a minimum wage (w_{min} = 3)

Hours of work	Region 1: Wage offer without minimum wage			Region 2: Wage offer with minimum wage in period 2 only		
	w = 2	w = 3	w = 4	w = 2	w = 3	w = 4
Observed distribution (n = 404)						
Period 1: Before minimum wage introduction						
h = 0	19			20		
h = 20	11	8	5	10	9	5
h = 40	10	14	11	9	15	10
h = 60	0	7	16	0	8	15
Period 2: After minimum wage introduction (in region 2 only)						
h = 0	18			—	8[a]	
h = 20	12	7	4	—	19	5
h = 40	11	13	12	—	21	14
h = 60	0	6	17	—	14	20
Predicted distribution (n = 404)						
Period 1: Before minimum wage introduction						
h = 0	17.0			17.0		
h = 20	14.2	10.6	4.4	14.2	10.6	4.4
h = 40	6.7	11.5	11.0	6.7	11.5	11.0
h = 60	1.9	7.4	16.3	1.9	7.4	16.3
Period 2: After minimum wage introduction (in region 2 only)						
h = 0	17.0			—	8.2[a]	
h = 20	14.2	10.6	4.4	—	19.3	5.5
h = 40	6.7	11.5	11.0	—	20.9	13.5
h = 60	1.9	7.4	16.3	—	13.6	20.0

Note: The upper part of the table shows the distribution of observations across (hours × wages) cells. Non-participants are represented by zero hours and by a missing wage. a. The introduction of the minimum wage restricts the wage offer distribution in that w = 2 will not be offered. Non-participants do not report their potential wage offer, hence it could be w = 3 or w = 4.

Source: Author's analysis using fictional data.

I Z A
World of Labor

after the introduction of the minimum wage is shown in the bottom right part of the table (region 2 in period 2). Each cell in the table reports the number of observations for a given number of hours and wage rate. Non-participants in the labor market (h = 0) do not report their wage, and therefore the number of participants corresponds to h = 0 only and not to a particular wage rate.

The model assumes that the preferences between work hours and consumption are unaffected by the introduction of the minimum wage; therefore, the preferences are common to all regions and periods. Furthermore, the wage offer distribution with a minimum wage (in region 2) is parametrized independently from the wage offer distribution without a minimum wage (in region 1). This explains why the predicted values for region 1 are the same before and after the introduction of a minimum wage in region 2. Relative to the estimates based on the data described in the illustration on page 1, the behavioral parameters are almost unchanged, while the implied labor supply elasticity is slightly larger, at 2.7.

The evidence for the observed distributions reported in Figure 2 suggests that the introduction of the minimum wage had a sizable effect on the distribution of the hours supplied, in particular, by reducing non-participation in region 2: only eight people are not working ($h = 0$) in period 2 compared with 20 people in period 1. These distributional changes may not substantially affect the averages and may thus not be captured by a difference-in-differences (DiD) approach, which compares the average change in outcome over time in the case with and the case without the minimum wage, a method that attempts to mimic experimental conditions [7]. But the distributional changes are captured by the fitted economic model, as shown in the lower part of Figure 2: the estimated model predicts a reduction in the number of people who are not working from 17 (without the minimum wage) to 8.2 (when it is introduced).

The predicted distribution implied by the model can be used to calculate the effect of the minimum wage, as measured by the DiD approach. The ability to reproduce the results of the less demanding DiD analysis provides a test of the robustness of the model and of the parameter estimates obtained using the maximum likelihood method. The difference in the mean hours of work (or mean positive hours or wage) with or without the minimum wage can be predicted from the specified model. These predicted differences are complex mixtures of the parameters describing the behavioral responses and of the parameters describing the wage offers with and without a minimum wage.

The DiD analysis applied to measure the effect of the introduction of the minimum wage on the hours supplied or wages paid fails to detect any statistically significant (additive) effect on hours, log hours, or wage (when the focus is on positive hours), but it does find a significant (proportional) effect on the average log wage (an increase of about 12%; Figure 3). The predictions reported in Figure 3 from the fitted model and the point estimates from the DiD analysis are in close agreement, which supports the specification of the economic model.

Figure 3. Comparison of model predicted differences and estimated difference-in-differences (DiD)

Variable	Predicted differences	Estimated (OLS) DiD	Robust standard error	Number of observations	R^2 of estimated equations
h, all	5.93	6.14	4.05	404	0.019
h, $h > 0$	2.74	1.91	3.34	339	0.004
$\ln(h)$, $h > 0$	0.08	0.05	0.09	339	0.003
w, $h > 0$	0.31	0.31	0.16	339	0.032
$\ln(w)$, $h > 0$	0.12	0.12	0.05	339	0.043

Note: The table compares the predicted differences based on the estimated model between a region where the minimum wage applies and a region where it does not with the estimated (ordinary least squares—OLS) differences, estimated using a simple DiD approach on the data described in Figure 2.

Source: Figure 2.

I Z A
World of Labor

LIMITATIONS AND GAPS

The model in the example discussed above fits the data well because the data structure is relatively simple (in the simplest case, the model needs to predict the probability across only ten cells and does not account for the possibility that the introduction of the minimum wage may reduce the number of job offers received) and because the model depends on a large enough number of parameters (five in the initial case and six when accounting for the effect of the minimum wage). In more realistic and complex situations, it will be harder to understand the features that the model should contain in order to provide a faithful representation of the data and of the mechanisms that generate them. The specification search will include a search for a more flexible representation of the behavioral responses and a search for a parsimonious description of the unobserved variability among individual agents that the model seeks to represent. This search is usually difficult to carry out successfully.

More complex models bring further difficulties when a likelihood-based approach is used. Determining or evaluating the likelihood can be difficult, and numerical issues may arise in seeking to locate the maximum of the likelihood. Such numerical issues can require substantial effort to solve. Often the choice is between a simpler model in which the likelihood estimates are easily obtained and a more complex model in which the estimates are more difficult to characterize or time consuming to obtain.

The specification of a model and the use of a likelihood-based estimation method are not enough to solve the common identification questions that trouble most empirical applications in labor economics. Researchers still need to provide careful arguments for or against the presence of a given explanatory variable in their model.

SUMMARY AND POLICY ADVICE

The use of an economic model and the estimation of its parameters using maximum likelihood can be valuable tools for labor economists. That approach allows for the analysis of data from most sources and can provide insight into the determinants of the usual summary (a parameter estimate in a regression) that a less demanding method, like DiD, will obtain.

Drawing on the example presented in this article, the model could, for instance, be used to calculate a predicted distribution of the population, assuming that the wage distribution is very different, or to assess the effect of the introduction of a minimum wage or an income tax system, assuming that the estimated preferences remain constant. This approach is useful for broader prospective policy analysis. Recent contributions to the literature provide more examples and discussions of the possibilities and difficulties [8], [9], [10], [11], [12].

Finally, the model and the estimates are useful for modeling and understanding the distributional effects of policy in different environments or under distinct scenarios. The illustration of the analysis of the log-likelihood in Figure 1 suggests that such simulation exercises should be carried out over several nearly as likely alternative values for the model parameters. Modern information technology and computer power make the analysis possible for models significantly more complex and valuable than the one presented here.

In summary, a methodological approach that uses a model plus maximum likelihood allows for the analysis of a more general class of data (observational as well as experimental) and for a more in-depth analysis of policy design given the parameter estimates.

Competing interests

The IZA World of Labor project is committed to the *IZA Guiding Principles of Research Integrity*. The author declares to have observed these principles.

REFERENCES

Further reading

Christensen, B. J., and N. M. Kiefer. *Economic Modeling and Inference*. Princeton, NJ: Princeton University Press, 2009.

Pudney, S. *Modelling Individual Choice: The Econometrics of Corners, Kinks and Holes*. Oxford: Basil Blackwell, 1990.

Key references

[1]   Angrist, J. D., and J.-S. Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press, 2009.

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

[3]   Deaton, A. "Instruments, randomization, and learning about development." *Journal of Economic Literature* 48:2 (2010): 424–455.

[4]   Manski, C. F. *Identification Problems in the Social Sciences*. Cambridge, MA: Harvard University Press, 1999.

[5]   Browning, M., M. Ejrnaes, and J. Alvarez. "Modelling income processes with lots of heterogeneity." *Review of Economic Studies* 77:4 (2010): 1353–1381.

[6]   Blundell, R., T. MaCurdy, and C. Meghir. "Labor supply models: Unobserved heterogeneity, nonparticipation and dynamics." In: Heckman, J., and E. Leamer (eds). *Handbook of EconometricsVolume 6A*. Amsterdam: Elsevier, 2007.

[7]   Blundell, R., and M. Costa Dias. "Alternative approaches to evaluation in empirical microeconomics." *Journal of Human Resources* 44:3 (2009): 565–640.

[8]   van den Berg, G. J., and J. Vikström. "Monitoring job offer decisions, punishments, exit to work, and job quality." *The Scandinavian Journal of Economics* 116:2 (2014): 284–334.

[9]   Geyer, J., P. Haan, and K. Wrohlich. "The effects of family policy on maternal labor supply: Combining evidence from a structural model and a quasi-experimental approach." *Labour Economics* 36:C (2015): 84–98.

[10] Nevo, A., and M. D. Whinston. "Taking the dogma out of econometrics: Structural modeling and credible inference." *Journal of Economic Perspectives* 24:2 (2010): 69–82.

[11] Iskhakov, F. "Structural dynamic model of retirement with latent health indicator." *The Econometrics Journal* 13:3 (2010): S126–S161.

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

Online extras

The **full reference list** for this article is available from:
http://wol.iza.org/articles/maximum-likelihood-and-economic-modeling

View the **evidence map** for this article:
http://wol.iza.org/articles/maximum-likelihood-and-economic-modeling/map