INDIRECT INFERENCE

(prepared for: *The New Palgrave Dictionary of Economics, Second Edition*)

Abstract

Indirect inference is a simulation-based method for estimating the parameters of economic models. Its hallmark is the use of an auxiliary model to capture aspects of the data upon which to base the estimation. The parameters of the auxiliary model can be estimated using either the observed data or data simulated from the economic model. Indirect inference chooses the parameters of the economic model so that these two estimates of the parameters of the auxiliary model are as close as possible. The auxiliary model need not be correctly specified; when it is, indirect inference is equivalent to maximum likelihood.

Introduction

Indirect inference is a simulation-based method for estimating, or making inferences about, the parameters of economic models. It is most useful in estimating models for which the likelihood function (or any other criterion function that might form the basis of estimation) is analytically intractable or too difficult to evaluate. Such models abound in modern economic analysis and include nonlinear dynamic models, models with latent (or unobserved) variables, and models with missing or incomplete data.

Like other simulation-based methods, indirect inference requires only that it be possible to simulate data from the economic model for different values of its parameters. Unlike other simulation-based methods, indirect inference uses an approximate, or auxiliary, model to form a criterion function. The auxiliary model does not need to be an accurate description of the data generating process. Instead, the auxiliary model serves as a window through which to view both the actual, observed data and the simulated data generated by the economic model: it selects aspects of the data upon which to focus the analysis.

The goal of indirect inference is to choose the parameters of the economic model so that the observed data and the simulated data look the same from the vantage point of the chosen window (or auxiliary model). In practice, the auxiliary model is itself characterized by a set of parameters. These parameters can themselves be estimated using either the observed data or the simulated data. Indirect inference chooses the parameters of the underlying economic model so that these two sets of estimates of the parameters of the auxiliary model are as close as possible.

A formal definition

To put these ideas in concrete form, suppose that the economic model takes the form:

$$y_t = G(y_{t-1}, x_t, u_t; \beta), \quad t = 1, 2, \ldots, T, \tag{1}$$

where $\{x_t\}_{t=1}^T$ is a sequence of observed exogenous variables, $\{y_t\}_{t=1}^T$ is a sequence of observed endogenous variables, and $\{u_t\}_{t=1}^T$ is a sequence of unobserved random errors. Assume that the initial value y_0 is known and that the random errors are independent and identically distributed (i.i.d.) with a known probability distribution F. Equation (1) determines, in effect, a probability density function for y_t conditional on y_{t-1} and x_t. Indirect inference does not require analytical tractability of this density, relying instead on numerical simulation of the economic model. This is not the most general model that indirect inference can accommodate—indirect inference can be used to estimate virtually any model from which it is possible to simulate data—but it is a useful starting point for understanding the principles underlying indirect inference. The econometrician seeks to use the observed data to estimate the k-dimensional parameter vector β.

The auxiliary model, in turn, is defined by a conditional probability density function, $f(y_t|y_{t-1}, x_t, \theta)$, which depends on a p-dimensional parameter vector θ. In a typical application of indirect inference, this density has a convenient analytical expression. The number of parameters in the auxiliary model must be at least as large as the number of parameters in the economic model (i.e., $p \geq k$).

The auxiliary model is, in general, incorrectly specified: that is, the density f need not describe accurately the conditional distribution of y_t determined by equation (1). Nonetheless, the parameters of the auxiliary model can be estimated using the observed data by maximizing the log of the likelihood function defined by f:

$$\hat{\theta} = \arg\max_{\theta} \sum_{t=1}^{T} \log f(y_t|y_{t-1}, x_t, \theta).$$

The estimated parameter vector $\hat{\theta}$ serves as a set of "statistics" that capture, or summarize, certain features of the observed data; indirect inference chooses the parameters of the economic model to reproduce this set of statistics as closely as possible.

The parameters of the auxiliary model can also be estimated using simulated data generated by the economic model. First, using a random number generator, draw a sequence of random errors $\{\tilde{u}_t^m\}_{t=1}^T$ from the distribution F. Typically, indirect inference uses M such sequences, so the superscript m indicates the number of the simulation. These sequences are drawn only once and then held fixed throughout the estimation procedure. Second, pick a parameter vector β and then iterate on equation (1), using the observed exogenous variables and the simulated random errors, to generate a simulated sequence of endogenous variables: $\{\tilde{y}_t^m(\beta)\}_{t=1}^T$, where the dependence of this simulated sequence on β is made explicit. Third and finally, maximize the average of the log of the likelihood across the M simulations to obtain:

$$\tilde{\theta}(\beta) = \arg\max_{\theta} \sum_{m=1}^{M} \sum_{t=1}^{T} \log f(\tilde{y}_t^m(\beta) | \tilde{y}_{t-1}^m(\beta), x_t, \theta).$$

The central idea of indirect inference is to choose β so that $\tilde{\theta}(\beta)$ and $\hat{\theta}$ are as close as possible. When the economic model is exactly identified (i.e., when $p = k$), it is, in general, possible to choose β so that the economic model reproduces exactly the estimated parameters of the auxiliary model. Typically, though, the economic model is overidentified (i.e., $p > k$): in this case, it is necessary to choose a metric for measuring the distance between $\hat{\theta}$ and $\tilde{\theta}(\beta)$; indirect inference then picks β to minimize this distance.

As the observed sample size T grows large (holding M fixed), the estimated parameter vector in the simulated data, $\tilde{\theta}(\beta)$, converges to a so-called "pseudo-true value" that depends on β; call it $h(\beta)$. The function h is sometimes called the binding function: it maps the parameters of the economic model into the parameters of the auxiliary model. Similarly, the estimated parameter vector in the observed data, $\hat{\theta}$, converges to a pseudo-true value θ_0. In the limit as T grows large, then, indirect inference chooses β to satisfy the equation $\theta_0 = h(\beta)$. Under the assumption that the observed data is generated by the economic model for a particular value, β_0, of its parameter vector, the value of β that satisfies this equation is precisely β_0. This heuristic argument explains why indirect inference generates consistent estimates of the parameters of the economic model.

Three examples

Example #1: A simple system of simultaneous equations

The first example is drawn from the classical literature on simultaneous equations to which indirect inference is, in many ways, a close cousin. Consider a simple macroeconomic model, adapted from Johnston (1984), with two simultaneous equations: $C_t = \beta Y_t + u_t$ and $Y_t = C_t + X_t$. In this model, consumption expenditure in period t, C_t, and output (or income) in period t, Y_t, are endogenous, whereas nonconsumption expenditure in period t, X_t, is exogenous. Assume that the random error u_t is i.i.d. and normally distributed with mean zero and a known variance; the only unknown parameter, then, is β.

There are many ways to estimate β without using indirect inference, but this example is useful for illustrating how indirect inference works. To wit, suppose that the auxiliary model specifies that C_t is normally distributed with conditional mean θX_t and a fixed variance. In this simple example, the binding function can be computed without using simulation: a little algebra reveals that $\theta = \beta/(1-\beta) \equiv h(\beta)$. To estimate β, first use ordinary least squares (which is equivalent to maximum likelihood in this example) to obtain a consistent estimate, $\hat{\theta}$, of θ. Then evaluate the inverse of h at $\hat{\theta}$ to obtain a consistent estimate of β: $\hat{\beta} = \hat{\theta}/(1+\hat{\theta})$. This is precisely the indirect inference estimator of β. This estimator uses an indirect approach: it first estimates an auxiliary (or, in the language of simultaneous equations, a reduced-form) model whose parameters are complicated functions of the parameters of the underlying economic model and then works backwards to recover estimates of these parameters.

Example #2: A general equilibrium model of the macroeconomy

In this example, the economic model is a dynamic, stochastic, general equilibrium (DSGE) model of the macroeconomy (for a prototype, see Hansen, 1985). Given choices for the parameters describing the economic environment, this class of models determines the evolution of aggregate macroeconomic time series such as output, consumption, and the capital stock. The law of motion for these variables implied by the economic model is, in general, nonlinear. In addition, some of the key variables in this law of motion (e.g., the capital stock) are poorly measured or even unobserved. For these reasons, in these models it is often difficult to obtain a closed-form expression for the likelihood function.

To surmount these obstacles, indirect inference can be used to obtain estimates of the parameters of the economic model. A natural choice for the auxiliary model is a vector autoregression (VAR) for the variables of interest. As an example, let y_t be a vector containing the values of output and consumption in period t (expressed as deviations from steady-state values) and let the VAR for y_t have one lag: $y_{t+1} = Ay_t + \epsilon_{t+1}$, where the ϵ_ts are normally distributed, i.i.d. random variables with mean 0 and covariance matrix Σ.

In this example, the binding function maps the parameters of the economic model into the parameters A and Σ of the VAR. To obtain a simulated approximation to the binding function, pick a set of parameters for the economic model, compute the law of motion implied by this set of parameters, simulate data using this law of motion, and then use OLS to fit a VAR to the simulated data. Indirect inference chooses the parameters of the economic model so that the VAR parameters implied by the model are as close as possible to the VAR parameters estimated using observed macroeconomic time series. Smith (1993) illustrates the use of indirect inference to estimate DSGE models.

Example #3: A discrete-choice model

In this example, the economic model describes the behavior of a decision-maker who must choose one of several discrete alternatives. These models typically specify a random utility for each alternative; the decision-maker is assumed to pick the alternative with the highest utility. The random utilities are latent: the econometrician does not observe them, but does observe the decision-maker's choice. Except in special cases, evaluating the likelihood of the observed discrete choices requires the evaluation of high-dimensional integrals which do not have closed-form expressions.

To use indirect inference to estimate discrete-choice models, one possible choice for the auxiliary model is a linear probability model. In this case, the binding function maps the parameters describing the probability distribution of the latent random utilities into the parameters of the linear probability model. Indirect inference chooses the parameters of the economic model so that the estimated parameters of the linear probability model using the observed data are as close as possible to those obtained using the simulated data. Implementing indirect inference in discrete-choice models poses a potentially difficult computational problem because it requires the optimization of a nonsmooth objective function. Keane and Smith (2003), who illustrate the use of indirect inference to estimate discrete-choice models, also suggest a way to smooth the objective surface.

Three metrics

To implement indirect inference when the economic model is overidentified, it is necessary to choose a metric for measuring the distance between the auxiliary model parameters estimated using the observed data and the simulated data, respectively. There are three possibilities corresponding to the three classical hypothesis tests: Wald, likelihood ratio (LR), and Lagrange multiplier (LM).

In the Wald approach, the indirect inference estimator of the parameters of the economic model minimizes a quadratic form in the difference between the two vectors of estimated parameters:

$$\hat{\beta}^{Wald} = \arg\min_{\beta} \, (\hat{\theta} - \tilde{\theta}(\beta))' \, W \, (\hat{\theta} - \tilde{\theta}(\beta)),$$

where W is a positive definite "weighting" matrix.

The LR approach to indirect inference forms a metric using the (approximate) likelihood function defined by the auxiliary model. In particular,

$$\hat{\beta}^{LR} = \arg\min_{\beta} \left(\sum_{t=1}^{T} \log f(y_t | y_{t-1}, x_t, \hat{\theta}) - \sum_{t=1}^{T} \log f(y_t | y_{t-1}, x_t, \tilde{\theta}(\beta)) \right).$$

By the definition of $\hat{\theta}$, the objective function on the right-hand side is nonnegative, and its value approaches zero as $\tilde{\theta}(\beta)$ approaches $\hat{\theta}$. The LR approach to indirect inference chooses β so as to make this value as close to zero as possible. Because the first term on the right-hand side does not depend on β, the LR approach can also be viewed as maximizing the approximate likelihood subject to the restrictions, summarized (for large T) by the binding function h, that the economic model imposes on the parameters of the auxiliary model.

Finally, the LM approach to indirect inference forms a metric using the derivative (or score) of the log of the likelihood function defined by the auxiliary model. In particular,

$$\hat{\beta}^{LM} = \arg\min_{\beta} \; S(\beta)' \, V \, S(\beta),$$

where

$$S(\beta) = \sum_{m=1}^{M} \sum_{t=1}^{T} \frac{\partial}{\partial \theta} \log f(\tilde{y}_t^m(\beta) | \tilde{y}_{t-1}^m(\beta), x_t, \hat{\theta})$$

and V is a positive definite matrix. By definition, $\hat{\theta}$ sets the score in the observed data to zero. The goal of the LM approach, then, is to choose β so that the (average) score in the simulated data, evaluated at $\hat{\theta}$, is as close to zero as possible.

For any number, M, of simulated data sets, all three approaches deliver consistent and asymptotically normal estimates of β as T grows large. The use of simulation inflates asymptotic standard errors by the factor $(1 + M^{-1})^{1/2}$; for $M \geq 10$, this factor is negligible. When the economic model is exactly identified, all three approaches to indirect inference yield numerically identical estimates; in this case, they all choose β to solve $\tilde{\theta}(\beta) = \hat{\theta}$.

When the economic model is overidentified, the minimized values of the three metrics are, in general, greater than zero. These minimized values can be used to test the hypothesis that the economic model is correctly specified: sufficiently large minimized values constitute evidence against the economic model.

If the weighting matrices W and V are chosen appropriately, then the Wald and LM approaches are asymptotically equivalent in the sense that they have the same asymptotic covariance matrix; by contrast, the LR approach, in general, has a larger asymptotic covariance matrix. If, however, the auxiliary model is correctly specified, then all three approaches are asymptotically equivalent not only to each other but also to maximum likelihood (for large M). Because maximum likelihood is asymptotically efficient (i.e., its asymptotic covariance matrix is as small as possible), the LM approach is sometimes called the "efficient method of moments" when the auxiliary model is close to being correctly specified; in such a case, this name could also be applied to the Wald approach.

When estimating the parameters of the auxiliary model is difficult or time-consuming, the LM approach has an important computational advantage over the other two approaches. In particular, it does not require that the auxiliary model be estimated repeatedly for different values of the parameters of the economic model. To estimate continuous-time models of asset prices, for example, Gallant and Tauchen (2002) advocate using a seminonparametric (SNP) model as the auxiliary model. As the number of its parameters increases, an SNP model provides an arbitrarily accurate approximation to the data generating process, thereby permitting indirect inference to approach the asymptotic efficiency of maximum likelihood. For this class of auxiliary models, which are nonlinear and often have a large number of parameters, the LM approach is a computationally attractive way to implement indirect inference.

Concluding remarks

Indirect inference is a simulation-based method for estimating the parameters of economic models. Like other simulation-based methods, such as simulated moments estimation (see, for example, Duffie and Singleton, 1993), it requires little analytical tractability, relying instead on numerical simulation of the economic model. Unlike other methods, the "moments" that guide the estimation of the parameters of the economic model are themselves the parameters of an auxiliary model. If the auxiliary model comes close to providing a correct statistical description of the economic model, then indirect inference comes close to matching the asymptotic efficiency of maximum likelihood. In many applications, however, the auxiliary model is chosen not to provide a good statistical description of the economic model, but instead to select important features of the data upon which to focus the analysis.

There is a large literature on indirect inference, much of which is beyond the scope of this article. Gouriéroux and Monfort (1996) provide a useful survey of indirect inference. Indirect inference was first introduced by Smith (1990, 1993) and later extended in important ways by Gouriéroux, Monfort, and Renault (1993) and Gallant and Tauchen (1996). Although indirect inference is a classical estimation method, Gallant and McCulloch (2004) show how ideas from indirect inference can be used to conduct Bayesian inference in models with intractable likelihood functions. There have been many interesting applications of indirect inference to the estimation of economic models, mainly in finance, macroeconomics, and labor economics. Because of its flexibility, indirect inference can be a useful way to estimate models in all areas of economics.

Anthony A. Smith, Jr.

References

Duffie, D. and K.J. Singleton (1993), "Simulated Moments Estimation of Markov Models of Asset Prices", *Econometrica* 61, 929–952.

Gallant, A.R. and R. McCulloch (2004), "On the Determination of General Scientific Models", manuscript (Duke University).

Gallant, A.R. and G. Tauchen (1996), "Which Moments to Match?", *Econometric Theory* 12, 657–681.

Gallant, A.R. and G. Tauchen (2002), "Simulated Score Methods and Indirect Inference for Continuous-time Models", forthcoming in: Ait-Sahalia, Y. and L. Hansen, eds., *Handbook of Financial Econometrics* (North-Holland).

Gouriéroux, C. and A. Monfort (1996), *Simulation-Based Econometric Methods* (Oxford University Press).

Gouriéroux, C., A. Monfort, and E. Renault (1993), "Indirect Inference", *Journal of Applied Econometrics* 8, S85–S118.

Hansen, G.D. (1985), "Indivisible Labor and the Business Cycle", *Journal of Monetary Economics* 16, 402–417.

Johnston, J. (1984), *Econometric Methods, Third Edition* (McGraw-Hill Book Company).

Keane, M. and A.A. Smith, Jr. (2003), "Generalized Indirect Inference for Discrete Choice Models", manuscript (Yale University).

Smith, Jr., A.A. (1990), *Three Essays on the Solution and Estimation of Dynamic Macroeconomic Models*, Ph.D. dissertation (Duke University).

Smith, Jr., A.A. (1993), "Estimating Nonlinear Time–series Models Using Simulated Vector Autoregressions", *Journal of Applied Econometrics* 8, S63–S84.