Generalized Indirect Inference for Discrete Choice Models

Michael Keane
and
Anthony A. Smith, Jr.*

PRELIMINARY AND INCOMPLETE

November 2003
(first version: June 2003)

Abstract

This paper develops and implements a practical simulation-based method for estimating dynamic discrete choice models. The method, which can accommodate lagged dependent variables, serially correlated errors, unobserved variables, and many alternatives, builds on the ideas of indirect inference. In particular, the method uses a descriptive statistical (or auxiliary) model—typically a linear probability model—to summarize the statistical properties of the observed and simulated data. The method then chooses the structural parameters so that the coefficients of the auxiliary model in the simulated data match as closely as possible those in the observed data. The main difficulty in implementing indirect inference in discrete choice models is that the objective surface is a step function, rendering useless gradient-based optimization methods. To overcome this obstacle, this paper shows how to smooth the objective surface. The key idea is to use a function of the latent utilities as the dependent variable in the auxiliary model. As the smoothing parameter goes to zero, this function delivers the discrete choice implied by the latent utilities, thereby guaranteeing consistency. A set of Monte Carlo experiments shows that the method is fast, robust, and nearly as efficient as maximum likelihood when the auxiliary model is sufficiently rich.

*Both authors are in the Department of Economics at Yale University. We thank seminar participants at the Board of Governors of the Federal Reserve Bank, Carnegie Mellon University, University of Rochester, University of Texas at Austin, University of Virginia, Yale University, and the 2003 Annual Meetings of the Society for Economic Dynamics for helpful comments.

1 Introduction

Many economic models have the feature that it is simple to simulate data from the model (given knowledge of the model parameters), but that estimation of the model is extremely difficult. Models with discrete outcomes or mixed discrete/continuous outcomes commonly fall into this category. A leading example is the multinomial probit model, in which an agent chooses from amongst several discrete alternatives the one with highest utility. Simulation of data from this model is trivial: the agents simply construct the utility of each alternative (algebraic operations) and then choose the alternative that gives highest utility. But estimation of the model, via either maximum likelihood (ML) of method of moments (MOM), is exceedingly difficult. This difficulty arises because, from the perspective of the econometrician, the probability that an agent chooses a particular alternative is a high dimensional integral over stochastic factors—unobserved by the econometrician—that affect the utilities that the agent assigns to each alternative. These probability expressions must be evaluated in order to estimate the model by ML or MOM.

Econometricians worked for many years on developing simulation methods to evaluate the choice probabilities in such discrete choice models (see Lerman and Manski 1981) but only relatively recently have sufficiently accurate smooth probability simulators been developed that make ML or MOM based on simulated proofs practical (see McFadden 1989 and Keane 1994).

A very different approach to inference in this type of model is "indirect inference" (II). This approach circumvents the need to construct the choice probabilities generated by the economic model, because the method is not based on either forming the likelihood for the model or forming moments implicated by the model. Rather, the idea of indirect inference is to simulate data from the structural economic model of interest. One then chooses a descriptive statistical model that provides a rich description of the patterns of covariation in the data. Such a descriptive model can be estimated on both the simulated data from the economic model, and on the actual observed data. The indirect inference estimator of the vector of structural parameters of the economic model, β, is that $\hat{\beta}$ which makes the simulated data from the economic model "look like" the observed data—in the sense that the descriptive statistical model estimated on the simulated data "looks like"that same model estimated on the observed data.

Different formal metrics for the similarity of a statistical model when it is estimated on two different data sets lead to different varieties of indirect inference estimators. Of course, to make this approach computationally appealing, one should choose the descriptive model

in such a way that it is easy to estimate, and choose the distance metric in such a way that it is easy to evaluate.

Indirect inference holds out the promise that it will be practical to estimate any economic model from which it is practical to simulate data, even if construction of the likelihood or population moments implied by the model is very difficult or impossible. But this promise has not been fully realized because of limitations in the II procedure itself. It is very difficult to apply II to discrete (or discrete/continuous) choice models for the following reason: Small changes in the structural parameters of such models will, in general, cause the data simulated from the model to change discretely. Such a discrete change causes the parameters of a descriptive model fit to the simulated data to jump discretely. This jump, in turn, causes the metric of distance between the descriptive models estimated on the observed and simulated data to jump discretely too. When the statistical objective function one is seeking to minimize is not a smooth function of the model parameters, one cannot use gradient-based methods of optimization. One must instead rely on much slower random search algorithms (like the simplex). These algorithms typically perform very poorly if there are more than a few parameters.

The non-smooth objective function of II procedures applied to discrete data is a very serious problem, rendering the method of little practical use in such problems. By analogy, simulated ML (SML) and simulated MOM approaches were very little used for discrete choice problems until the advent of smooth simulation algorithms (see McFadden 1989 and Keane 1994) made such applications much more feasible. In this paper we develop a "generalized indirect inference" procedure that addresses this problem.

The idea of the GII procedure is to apply different descriptive statistical models to the simulated and actual observed data. The model applied to the simulated data will not fit the discrete outcomes in that data. Rather, it will fit the underlying continuous latent variables that generate those discrete outcomes. Since these latent variables are smooth functions of the model parameters, the non-smooth objective function problem is resolved. Of course, the same statistical model cannot be estimated on the observed data, because the underlying latent variables generating actual agents' observed choices are not seen by the econometrician. However, as long as the two descriptive statistical models are chosen so that both generate asymptotically the same vector of pseudo true parameter values, the GII estimator based on minimizing the distance between these two models remains consistent and asymptotically normal. Furthermore, the GII estimator is asymptotically efficient provided that each descriptive model satisfies the information matrix equality.

In the remainder of the paper we prove these results, and provide Monte Carlo evidence that the GII procedure performs well on a set of example models. We look at some cases where SML is also feasible, and show that efficiency losses relative to SML are small. We also show how judicious choice of the descriptive (or auxiliary) model is very important for the efficiency of the estimator. This is true not only here, but for II more generally.

2 The Model

In this section, we describe a class of discrete choice models that we will use as test cases for the estimation method that we develop in this paper. As will become clear, however, this method can be used in almost any conceivable model of discrete choice, including models with mixed discrete/continuous outcomes and models in which individuals' choices solve forward-looking dynamic programming problems. In Section 5, we show how to use the estimation method developed in this paper to estimate a selection model with both discrete and continuous outcomes.

In this section, we consider panel data models with N individuals each of whom selects a choice from a set of J discrete alternatives in each of T time periods. Let u_{itj} be the (latent) utility that individual i attaches to alternative j in period t. Without loss of generality, set the utility of alternative J in any period equal to 0.[1] In each period, each individual chooses the alternative with the highest utility. Let y_{itj} be equal to 1 if individual i chooses alternative j in period t and be equal to 0 otherwise. Define $u_{it} \equiv (u_{it1}, \ldots, u_{it,J-1})$ and $y_{it} \equiv (y_{it1}, \ldots, y_{it,J-1})$. The econometrician observes the choices $\{y_{it}\}$ but not the latent utilities $\{u_{it}\}$.

The vector of latent utilities u_{it} is assumed to follow a stochastic process

$$u_{it} = f(x_{it}, y_{i,t-1}, \ldots, y_{i,t-\ell}, \epsilon_{it}; \beta), \quad t = 1, \ldots, T,$$

where x_{it} is a vector of exogenous variables.[2] For each individual i, the vector of disturbances ϵ_{it} follows a Markov process $\epsilon_{it} = r(\epsilon_{i,t-1}, \eta_{it}; \beta)$, where $\{\eta_{it}\}_{t=1}^{T}$ is a sequence of i.i.d. random vectors with known distribution. The functions f and r depend on a set of k structural parameters β. The sequences $\{\eta_{it}\}_{t=1}^{T}$, $i = 1, \ldots, N$, are independent across individuals and independent of x_{it} for all i and t. The initial values ϵ_{i0} and y_{it}, $t = 0, -1, \ldots, 1-\ell$, are fixed exogenously.

[1] In other words, u_{itj} can be interpreted as the difference between the utilities that individual i attaches to alternatives j and J in period t.

[2] The estimation method proposed in this paper can also accommodate models in which the latent utilities in any given period depend on lagged values of the latent utilities.

Although the estimation method proposed in this paper can be applied to any model of this form, we focus on four special cases of the general model. Three of these cases (Models 1, 2, and 4 below) can be feasibly estimated using simulated maximum likelihood, allowing us to compare the performance of the proposed method to that of maximum likelihood.

In Model 1, $I = 2$, $T > 1$, and $u_{it} = \gamma x_{it} + \epsilon_{it}$, where x_{it} is a scalar, $\epsilon_{it} = \rho\epsilon_{i,t-1} + \eta_{it}$, $\eta_{it} \sim iidN(0,1)$, and $\epsilon_{i0} = 0$. This is a two-alternative dynamic probit model with serially correlated errors; it has two unknown parameters γ and ρ.

In Model 2, $I = 2$, $T > 1$, and $u_{it} = \gamma_1 x_{it} + \gamma_2 y_{i,t-1} + \epsilon_{it}$, where x_{it} is a scalar and ϵ_{it} follows the same process as in Model 1. The initial value y_{i0} is set equal to 0. This is a two-alternative dynamic probit model with serially correlated errors and a lagged dependent variable; it has three unknown parameters γ_1, γ_2, and ρ.

Model 3 is identical to Model 2 except that the econometrician does not observe the first $s < T$ of the individual's choices. Thus there is an "initial conditions" problem (see Heckman 1981).

In Model 4, $I = 3$, $T = 1$, and the latent utilities obey:

$$\begin{aligned} u_{i1} &= \gamma_{10} + \gamma_{11} x_{i1} + \gamma_{12} x_{i2} + \eta_{i1} \\ u_{i2} &= \gamma_{20} + \gamma_{21} x_{i1} + \gamma_{22} x_{i3} + c_1 \eta_{i1} + c_2 \eta_{i2}, \end{aligned}$$

where $\{\eta_{i1}\}_{i=1}^N$ and $\{\eta_{i2}\}_{i=1}^N$ are i.i.d. sequences of $N(0,1)$ random variables.[3] This is a static three-alternative probit model; it has eight unknown parameters $\{\gamma_{1i}\}_{i=0}^2$, $\{\gamma_{2i}\}_{i=0}^2$, c_1, and c_2.

3 Generalized Indirect Inference

3.1 The central idea

The method that we propose to estimate the model described in Section 2 is a generalization of indirect inference (see Smith 1990, 1993, Gourieroux, Monfort, and Renault 1993, and Gallant and Tauchen 1996). Indirect inference exploits the ease and speed with which one can simulate data from complicated structural models. The basic idea of indirect inference is to view both the observed data and the simulated data (generated by the structural model given structural parameters β) through the "lens" of an descriptive statistical (or auxiliary) model characterized by a set of p auxiliary parameters θ. The $k \leq p$ structural parameters

[3]Since $T = 1$ in this model, the time subscript has been omitted, so that u_{ij} denotes the latent utility that the ith individual attaches to the jth alternative x_{ik} denotes the kth element of the ith individual's exogenous vector x_i, and η_{ij} denotes the jth element of the random vector η_i.

β are then chosen so as to make the observed data and the simulated data look similar when viewed through this lens.

To formalize these ideas, assume that the observed choices $\{y_{it}\}$, $i = 1, \ldots, N$, $t = 1, \ldots, T$, are generated by the structural model described in Section 2 given a set of structural parameters β_0.

The auxiliary model can be estimated using the observed data to obtain parameter estimates $\hat{\theta}$. Formally, $\hat{\theta}$ solves:

$$\hat{\theta} = \arg\max_{\theta} \mathcal{L}(y; x, \theta),$$

where $\mathcal{L}(y; x, \theta)$ is the likelihood function associated with the auxiliary model, $y \equiv \{y_{it}\}$ is the set of observed choices, and $x \equiv \{x_{it}\}$ is the set of observed exogenous variables.

Given x and structural parameters β, the structural model can be used to generate M statistically independent simulated data sets $\{\tilde{y}_{it}^m(\beta)\}$, $m = 1, \ldots, M$, where $\tilde{y}_{it}^m \equiv (\tilde{y}_{it1}^m(\beta), \ldots, \tilde{y}_{it,J-1}^m(\beta))$. Each of the M simulated data sets is constructed using the same set of observed exogenous variables x. The auxiliary model can then be estimated using each of the simulated data sets to obtain M estimated parameter vectors $\tilde{\theta}_m(\beta)$. Formally, $\tilde{\theta}_m(\beta)$ solves:

$$\tilde{\theta}_m(\beta) = \arg\max_{\theta} \mathcal{L}(y_m(\beta); x, \theta),$$

where the likelihood function associated with the auxiliary model is, in this case, evaluated using the mth simulated data set $\tilde{y}_m(\beta) \equiv \{y_{it}^m(\beta)\}$. Denote the average of the estimated parameter vectors by $\tilde{\theta}(\beta) \equiv M^{-1} \sum_{m=1}^{M} \tilde{\theta}_m(\beta)$. As the observed sample size N grows large (holding M and T fixed), $\tilde{\theta}(\beta)$ converges to a nonstochastic function $h(\beta)$. Gourieroux, Monfort, and Renault (1993) refer to h as the "binding" function.

Loosely speaking, indirect inference generates an estimate $\hat{\beta}$ of the structural parameters by choosing β so as to make $\hat{\theta}$ and $\tilde{\theta}(\beta)$ as close as possible. When generating simulated data sets, the set of random draws $\{\tilde{\eta}_{it}\}$ is held fixed for different values of β. The key idea underlying the consistency of indirect inference is that, as the observed sample size N grows large, $\hat{\theta}$ and $\tilde{\theta}(\beta_0)$ both converge to the same "pseudo" true value $\theta_0 = h(\beta_0)$.

Magnac, Robin, and Visser (1995), An and Liu (2000), and Nagypál (2000) use various implementations of indirect inference to estimate discrete choice models. These papers encounter the difficult computational task of optimizing a step function. This task is very time-consuming and puts severe constraints on the size of the structural models that can be feasibly estimated. Step functions arise when applying indirect inference to discrete choice models because any simulated choice $\tilde{y}_{it}^m(\beta)$ is a step function of β (holding fixed the set

of random draws $\{\tilde{\eta}_{it}\}$ used to generate simulated data from the structural model). Consequently, the estimated set of auxiliary parameters $\tilde{\theta}(\beta)$ is a step function of β.

We propose a generalization of indirect inference that addresses this difficulty. The fundamental idea is that the estimation procedures applied to the observed and simulated data sets need not be identical, provided that they both provide consistent estimates of the same vector of pseudo true parameter values.[4] We exploit this idea to smooth the function $\tilde{\theta}(\beta)$, obviating the need to optimize a step function when using indirect inference to estimate a discrete choice model.

Let $\tilde{u}_{it}^m(\beta)$ denote the set of latent utilities that individual i attaches to the first $J-1$ alternatives in period t of the mth simulated data set, given structural parameters β (recall that the latent utility of the Ith alternative is set to 0). Rather than use the choice $\tilde{y}_{itj}^m(\beta)$ when computing $\tilde{\theta}(\beta)$, we use instead a function $g(\tilde{u}_{it}^m(\beta), j; \lambda)$ of the latent utilities. The function g is chosen so that as the smoothing parameter λ goes to 0, $g(\tilde{u}_{it}^m(\beta), j; \lambda)$ converges to $\tilde{y}_{itj}^m(\beta)$. Letting λ go to 0 at the same time that the observed sample size goes to infinity ensures that $\tilde{\theta}(\beta_0)$ converges to θ_0, thereby delivering consistency of the generalized indirect inference estimator of β_0.

Although many functional forms could be chosen for g, in this paper we define g as follows:

$$g(\tilde{u}_{it}^m(\beta), j; \lambda) = \frac{\exp(\tilde{u}_{itj}^m(\beta)/\lambda)}{1 + \sum_{k=1}^{J-1} \exp(\tilde{u}_{itk}^m(\beta)/\lambda)}.$$

Because the latent utilities are smooth functions of the structural parameters β, g is a smooth function of β. Moreover, as λ goes to 0, g goes to 1 if alternative j has the highest latent utility and to 0 otherwise.[5]

3.2 Choosing a metric

Implementing indirect inference requires the choice of a formal metric for measuring the "distance" between $\hat{\theta}$ and $\tilde{\theta}(\beta)$. There are three approaches to choosing such a metric, corresponding, roughly, to the three classic approaches to hypothesis testing: Wald, likelihood ratio (LR), and Lagrange multiplier (LM).[6] All three approaches to indirect inference yield

[4] Genton and Ronchetti (2003) use a similar insight to develop robust estimation procedures in the context of indirect inference.

[5] This approach to smoothing in a discrete choice model bears a superficial resemblance to the approach that Horowitz (1992) uses to create a smoothed version of Manski's (1985) maximum score estimator for a binary response model. Whereas Horowitz (1992) focuses on the estimation of a limited set of discrete choice models in a semiparametric setting, in this paper we develop smooth methods to estimate a wide range of fully specified models with mixed discrete/continuous outcomes.

[6] This nomenclature is due to Eric Renault. The Wald and LR approaches were first proposed in Smith (1990, 1993) and later extended by Gourieroux, Monfort, and Renault (1993). The LM approach was first

consistent and asymptotically normal estimates of the true structural parameters β_0. Moreover, in the exactly identified case in which the number of structural parameters k is equal to the number of auxiliary parameters p, all three approaches yield numerically identical estimates of the structural parameters (provided that the same random numbers are used to generate the simulated data in all three approaches).

The Wald approach to indirect inference chooses β to minimize a quadratic form in the vector $\hat{\theta} - \tilde{\theta}(\beta)$:

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

where W is a positive definite "weighting" matrix.

The LR approach to indirect inference forms a metric by using the likelihood function $\mathcal{L}(y; x, \theta)$ associated with the auxiliary model. In particular,

$$\hat{\beta}^{LR} = \arg\max_{\beta} \mathcal{L}(y; x, \tilde{\theta}(\beta)).$$

Finally, the LM approach to indirect inference does not work directly with the estimated auxiliary parameters $\tilde{\theta}(\beta)$ but instead uses the score vector associated with the auxiliary model.[7] Given the estimated auxiliary model parameters $\hat{\theta}$ from the observed data, the score vector is evaluated using each of the M simulated data sets. The LM approach then chooses β to minimize a quadratic form in the average score vector across these data sets:

$$\hat{\beta}^{LM} = \arg\min_{\beta} \left(M^{-1} \sum_{m=1}^{M} \mathcal{L}_{\theta}(\tilde{y}_m(\beta); x, \hat{\theta}) \right)' V \left(M^{-1} \sum_{m=1}^{M} \mathcal{L}_{\theta}(\tilde{y}_m(\beta); x, \hat{\theta}) \right),$$

where $\mathcal{L}_{\theta}$ denotes the vector of partial derivatives of L with respect to θ and V is a positive definite weighting matrix.

Each of the three approaches to indirect inference can be generalized simply by replacing each simulated choice $y^m_{itj}(\beta)$ with its smoothed counterpart $g(u^m_{it}(\beta), j; \lambda)$. In particular, we can state the following proposition (whose proof is to be added):

Proposition 1 Under appropriate regularity conditions, generalized indirect inference (implemented using any of the three metrics) yields consistent and asymptotically normal estimates of the true structural parameter vector β_0, provided that the smoothing parameter λ goes to 0 at an appropriate rate as the observed sample size N goes to infinity (holding fixed the number of simulated data sets M and the number of time periods T).

proposed in Gallant and Tauchen (1996).

[7] When the LM approach is implemented using an auxiliary model that is (nearly) correctly specified in the sense that it provides a (nearly) correct statistical description of the observed data, Gallant and Tauchen refer to this approach as efficient method of moments (EMM).

When the weighting matrices W and V are chosen optimally (in the sense of minimizing asymptotic variance) the Wald and LM approaches to indirect inference (and to generalized indirect inference) have the same asymptotic efficiency. The LR approach, by contrast, produces estimates that are, in general, less efficient asymptotically than under the Wald and LM approaches. If, however, the auxiliary model is correctly specified (in the sense that it provides a correct statistical description of the observed data), then all three approaches are asymptotically equivalent not only to each other but also to maximum likelihood (provided that the number of simulated data sets M is sufficiently large).

For two reasons, we focus in this paper on the LR approach to indirect inference. First, unlike the Wald and LM approaches, the LR approach does not require the estimation of an optimal weighing matrix. In this respect, the LR approach is easier to implement than the other two approaches. Furthermore, because estimates of optimal weighting matrices often do not perform well in finite samples, the LR approach is likely to perform better in small samples than the other two approaches. Second, because the LR approach is asymptotically equivalent to the other two approaches when the auxiliary is correctly specified, the asymptotic efficiency loss of the LR approach relative to the two other approaches is likely to be small when the auxiliary model is chosen judiciously.

3.3 Choosing a value for the smoothing parameter

Implementing generalized indirect inference requires choosing a value for the smoothing parameter λ. This choice is difficult because of two opposing considerations: bias and smoothness. Large values of λ do a good job of smoothing the objective surface but can lead to biased estimates of the structural parameters. Small values of λ reduce this bias but lead to "choppiness" in the objective surface, making it difficult to optimize. For a given value of λ, increasing M, the number of simulated data sets, can mitigate the effects of this choppiness, but only at the cost of increased computation time.

To address these practical problems, we advocate a two-step approach. In the first step, choose a relatively large value for λ so that the objective surface is smooth enough to allow the use of standard gradient-based methods to find its optimum. In this step, the number of simulated data sets M can be small (in fact, as small as $M = 1$) so as to reduce computation time. The first-step estimate $\hat{\beta}_1$ optimizes the objective (i.e., it solves one of the optimization problems in Section 3.2) and is, therefore, a consistent estimate of the structural parameters (where the argument for consistency requires that the smoothing parameter go to 0 as the observed sample size increases). Although $\hat{\beta}_1$ is consistent, in finite samples it tends to have

large bias (because λ is large) and to be contaminated by simulation error (because M is small).

In the second step, choose a relatively small value for λ (to reduce bias) and a relatively large value for M. Choosing a large value for M in the second step serves two purposes. First, it ensures that simulation error does not inflate the standard errors of the estimates. Second, it reduces choppiness in the objective surface, compensating for the relatively small value of λ in the second step. Since M is large in the second step, finding the value of β that optimizes the objective could be very costly in terms of computation time. Rather than optimize the objective, instead take one Newton-Raphson step from the consistent estimate $\hat{\beta}_1$, thereby delivering an estimate $\hat{\beta}_2$ that is asymptotically equivalent to the estimate that optimizes the objective.

The appropriate Newton-Raphson step varies with the approach—Wald, LR, or LM—used to define the generalized indirect inference estimator. The following proposition (whose proof is to be added) shows how to compute the second-step estimate $\hat{\beta}_2$ for the LR approach to indirect inference.

Proposition 2 Let $\hat{\beta}_1$ be a consistent estimate of the true structural parameter vector β_0. Then

$$\hat{\beta}_2 = \hat{\beta}_1 - \left(\hat{J}'\,\mathcal{L}_{\theta\theta}(y; x, \tilde{\theta}(\hat{\beta}_1))\,\hat{J}\right)^{-1} \hat{J}'\mathcal{L}_{\theta}(y; x, \tilde{\theta}(\hat{\beta}_1)),$$

is a consistent and asymptotically normal estimate of β_0, where $\mathcal{L}_{\theta\theta}$ denotes the $p \times p$ Hessian of the likelihood function associated with the auxiliary model and $\hat{J}$ is an estimate of the $p \times k$ Jacobian of the binding function h, evaluated at $\hat{\beta}_1$. Moreover, $\hat{\beta}_2$ is asymptotically equivalent to the generalized indirect inference estimate based on the LR approach.

3.4 Choosing an auxiliary model

The main consideration when choosing an auxiliary model with which to conduct indirect inference is efficiency. As discussed in Section 3.2, indirect inference (generalized or not) has the same asymptotic efficiency as maximum likelihood when the auxiliary model is correctly specified in the sense that it provides a correct statistical description of the observed data. From the perspective of efficiency, then, it is important to choose an auxiliary model (or a class of auxiliary models) that is flexible enough to provide a good description of the data.

Another important consideration when choosing an auxiliary model is computation time. For the Wald and LR approaches to indirect inference, the auxiliary parameters must be estimated repeatedly using different simulated data sets. For this reason, it is critical to use

an auxiliary model that can be estimated quickly and efficiently. This consideration is less important for the LM approach to indirect inference, since the LM approach does not work directly with the estimated auxiliary parameters but instead uses the first-order conditions (the score vector) that define these estimates.

To meet the twin criteria of statistical and computational efficiency, we let the auxiliary model be a linear probability model (or, more accurately, a set of linear probability models). This class of models is flexible in the sense that an individual's current choice can be allowed to depend on polynomial functions of lagged choices and of current and lagged exogenous variables. This class of models can also be very quickly and easily estimated using ordinary least squares. Section 4 describes in detail how we specify the linear probability models for each of the structural models described in Section 2.

4 Monte Carlo Results

This section conducts a set of Monte Carlo experiments to assess the performance, in terms of bias, efficiency, and computation time, of the estimation method (referred to hereafter as GII, for generalized indirect inference) described in Section 3. The parameters of each of the four models described in Section 2 are estimated a large number of times using "observed" data generated by the respective models. For Models 1, 2, and 4, the parameters are estimated in each Monte Carlo replication using both GII and simulated maximum likelihood (SML) in conjunction with the Geweke, Keane, and Hajivassiliou (GHK) smooth probability simulator.[8] Model 3, which cannot easily be estimated via SML, is estimated using only GII. For each model, the Monte Carlo experiments are conducted for several sets of parameter configurations.

To optimize objective functions, we use a version of the Davidon-Fletcher-Powell algorithm (as implemented in Chapter 10 of Press et al 1996). The initial parameter vector in the hillclimbing algorithm is the true parameter vector. Most of the computation time in generalized indirect inference lies in computing ordinary least squares (OLS) estimates. The main cost in computing OLS estimates lies, in turn, in computing the $X'X$ part of $(X'X)^{-1}X'Y$. We use blocking and loop unrolling techniques to speed up the computation of $X'X$ by a factor of 2 to 3 relative to a "naive" algorithm.[9]

[8] Lee (1997) uses SML with the GHK simulator to estimate versions of Models 1, 2, and 3.

[9] To avoid redundant calculations, we also precompute and store for later use those elements of $X'X$ that depend only on the exogenous variables. We are grateful to James MacKinnon for providing code that implements the blocking and loop unrolling techniques.

4.1 Results for Model 1

As described in greater detail in Section 2, Model 1 is a dynamic two-alternative probit model with serially correlated errors and a single exogenous regressor. It has two unknown parameters: γ, the coefficient on the exogenous regressor, and ρ, the serial correlation parameter. We set $\gamma = 1$ and we consider three values for ρ: 0, 0.4, and 0.85. In the Monte Carlo experiments for this model, the number of individuals $N = 1000$ and the number of time periods $T = 5$. When implementing generalized indirect inference, we use the two-step approach described in Section 3.3 and summarized in Proposition 2. When computing the first-step estimates of the structural parameters, the smoothing parameter $\lambda = 0.03$ and the number of simulated data sets $M = 10$. When computing the second-step estimates, $\lambda = 0.003$ and $M = 300$. The exogenous variables (the x_{it}'s) are i.i.d. draws from a $N(0, 1)$ distribution; each Monte Carlo replication has a new set of draws.

The auxiliary model consists of a set of T linear probability models, one for each time period. The right-hand side variables in the linear probability models differ across time periods. This structure serves two purposes: one, it allows the parameters of the auxiliary model to depend on time; two, it allows the auxiliary model to incorporate more lagged information in later time periods.

Let z_{it} be the vector of regressors in the auxiliary model for individual i in time period t. The vector z_{it} consists of polynomial functions of lagged choices and of current and lagged exogenous variables. Equation t in the auxiliary model is $y_{it} = z_{it}'\alpha_t + \nu_{it}$, where $\nu_{it} \sim iidN(0, \sigma_t^2)$ and α_t and σ_t^2 are parameters to be estimated. The auxiliary model, then, is characterized by a set of parameters $\theta = (\alpha_t, \sigma_t^2)$, $t = 1, \ldots, T$. Since the errors in the linear probability models are assumed to be independent across both time and individuals, the likelihood function associated with the auxiliary model is $\mathcal{L}(y; x, \theta) = \Pi_{i=1}^{N}\Pi_{t=1}^{T} L(y_{it}; z_{it}, \alpha_t, \sigma_t^2)$, where L is a normal probability density function with mean $z_{it}'\alpha_t$ and variance σ_t^2.

In order to examine how increasing the "richness" of the auxiliary model affects the efficiency of the structural parameter estimates, we conduct Monte Carlo experiments using four nested auxiliary models. In all four auxiliary models, we impose the restrictions $\alpha_t = \alpha_q$ and $\sigma_t^2 = \sigma_q^2$, $t = q + 1, \ldots, T$, for some $q < T$. We impose these restrictions because the time variation in the estimated coefficients of the linear probability models is small after the first few time periods. Most of the time variation in the initial time periods comes from the non-stationarity of the errors in the structural model (we assume that the initial error is not drawn from the stationary distribution implied by the law of motion for the errors).

In auxiliary model #1, $q = 1$ and the regressors in the linear probability model are given

by: $z_{it} = (1, x_{it}, y_{i,t-1})$, $t = 1, \ldots, T$, where the unobserved y_{i0} is set equal to 0. We use this very simple auxiliary model, which has only 4 parameters, to illustrate how generalized indirect inference can produce very inefficient estimates if one uses a poor auxiliary model (i.e., one that does a poor job of capturing the statistical properties of the observed data). In auxiliary model #2, $q = 4$ and the regressors in the linear probability models are defined as follows:

$$\begin{aligned} z_{i1} &= (1, x_{i1}) \\ z_{it} &= (1, x_{it}, y_{i,t-1}, x_{i,t-1}), \quad t = 2, \ldots, T \end{aligned}$$

This auxiliary model has a total of 18 parameters. Auxiliary model #3 adds additional terms to the vectors of regressors in auxiliary model #2:

$$\begin{aligned} z_{i1} &= (1, x_{i1}, x_{i1}^3) \\ z_{i2} &= (1, x_{i2}, y_{i1}, x_{i1}) \\ z_{i3} &= (1, x_{i3}, y_{i2}, x_{i2}, y_{i1}, x_{i1}) \\ z_{it} &= (1, x_{it}, y_{i,t-1}, x_{i,t-1}, y_{i,t-2}, x_{i,t-2}, y_{i,t-3}), \quad t = 4, \ldots, T \end{aligned}$$

This auxiliary model has a total of 24 parameters. Finally, in auxiliary model #4, $q = 5$ and there are 35 parameters:

$$\begin{aligned} z_{i1} &= (1, x_{i1}, x_{i1}^3) \\ z_{i2} &= (1, x_{i2}, y_{i1}, x_{i1}) \\ z_{i3} &= (1, x_{i3}, y_{i2}, x_{i2}, y_{i1}, x_{i1}) \\ z_{i4} &= (1, x_{i4}, y_{i3}, x_{i3}, y_{i2}, x_{i2}, y_{i1}, x_{i1}) \\ z_{it} &= (1, x_{it}, y_{i,t-1}, x_{i,t-1}, y_{i,t-2}, x_{i,t-2}, y_{i,t-3}, x_{i,t-3}, y_{i,t-4}), \quad t = 5, \ldots, T \end{aligned}$$

Table 1 presents the results of six sets of Monte Carlo experiments, each with 2000 replications. The first two sets of experiments report the results for simulated maximum likelihood, using 25 draws (SML #1) and 50 draws (SML #2) in the GHK probability simulator. The remaining four sets of experiments report the results for generalized indirect inference, where GII #i refers to generalized indirect inference using auxiliary model #i. In each case, we report the average and the standard deviation of the parameter estimates. We also report the efficiency loss of GII #i relative to SML #2 in the columns labelled $\sigma_{GII}/\sigma_{SML}$, where we divide the standard deviations of the GII estimates by the standard deviations of the estimates for SML #2. Finally, we report the average time (in seconds)

required to compute estimates (we use the Intel Fortran Compiler Version 7.1 on a 2.2GHz Intel Xeon processor running Red Hat Linux).

Table 1 contains several key findings. First, both SML and GII generate estimates with very little bias (provided that sufficiently many draws are used in SML when the serial correlation in the errors is large).

Second, GII is less efficient than SML, but the efficiency losses are small provided that the auxiliary model is sufficiently rich. For example, auxiliary model #1 leads to large efficiency losses, particularly for the case of high serial correlation in the errors ($\rho = 0.85$). For models with little serial correlation ($\rho = 0$), however, auxiliary model #2 is sufficiently rich to to make GII almost as efficient as SML. When there is more serial correlation in the errors, auxiliary model #2 leads to reasonably large efficiency losses (as high as 30% when $\rho = 0.85$), but auxiliary model #3, which contains more lagged information in the linear probability models than does auxiliary model #2, reduces the worst efficiency loss to 13%. Auxiliary model #4 provides almost no efficiency gains relative to auxiliary model #3.

Third, GII is faster than SML: computing a set of estimates using GII with auxiliary model #3 takes about 30% less time than computing a set of estimates using SML with 50 draws. Note too that we are using conservative choices for the number of simulated data sets in the first and second steps of the estimation (10 and 300, respectively). Preliminary results suggest that by reducing the number of simulated data sets (say, to 5 and 150, respectively), it will be possible to reduce the computation time of GII without adversely affecting its finite sample properties.

For generalized indirect inference, we also compute (but do not report in Table 1) estimated asymptotic standard errors. In all cases, the averages of the estimated standard errors across the Monte Carlo replications are very close to (within a few percent of) the actual standard deviations of the estimates, suggesting that the asymptotic results provide a good approximation to the behavior of the estimates in finite samples of the size that we use.

4.2 Results for Model 2

As described in greater detail in Section 2, Model 1 is a dynamic two-alternative probit model with serially correlated errors, a single exogenous regressor, and a lagged dependent variable. It has three unknown parameters: γ_1, the coefficient on the exogenous regressor, γ_2, the coefficient on the lagged dependent variable, and ρ, the serial correlation parameter. We set $\gamma_1 = 1$, $\gamma_2 = 0.2$, and we consider three values for ρ: 0, 0.4, and 0.85. In the Monte

Carlo experiments for this model, the number of individuals $N = 1000$ and the number of time periods $T = 10$. When implementing generalized indirect inference, we use the two-step approach described in Section 3.3 and summarized in Proposition 2. When computing the first-step estimates of the structural parameters, the smoothing parameter $\lambda = 0.05$ and the number of simulated data sets $M = 10$. When computing the second-step estimates, $\lambda = 0.003$ and $M = 300$. When implementing GII for Model 2, we use the same set of four auxiliary models that we use for Model 1. As in Model 1, the exogenous variables (the x_{it}'s) are i.i.d. draws from a $N(0, 1)$ distribution and each Monte Carlo replication has a new set of draws.

Table 2 presents the results of six sets of Monte Carlo experiments, each with 1000 replications. The first two sets of experiments report the results for simulated maximum likelihood, using 25 draws (SML #1) and 50 draws (SML #2) in the GHK probability simulator. The remaining four sets of experiments report the results for generalized indirect inference, where GII #i refers to generalized indirect inference using auxiliary model #i.

The results for Model 2 are similar to those for Model 1. Both SML and GII generate estimates with very little bias (provided at least 50 draws are used in SML, especially when the errors are persistent). SML is more efficient than GII, but the efficiency loss is small (15% at most) when the auxiliary model is sufficiently rich. Auxiliary model #1 can lead to large efficiency losses, but auxiliary model #4 provides few efficiency gains relative to auxiliary model #3. When the errors are not persistent ($\rho = 0$), auxiliary model #3 leads to no efficiency gains relative to auxiliary model #2. Finally, GII using auxiliary model #3 is about 25% faster than SML using 50 draws.[10]

4.3 Results for Model 3

As described in greater detail in Section 2, Model 3 is a dynamic two-alternative probit model with serially correlated errors, a single exogenous regressor, and an "initial conditions" problem: the econometrician does not observe individuals' choices in the first s time periods. Model 3 has the same three unknown parameters as Model 2: γ_1, the coefficient on the exogenous regressor, γ_2, the coefficient on the lagged dependent variable, and ρ, the serial correlation parameter. As for Model 2, we set $\gamma_1 = 1$, $\gamma_2 = 0.2$, and we consider three values for ρ: 0, 0.4, and 0.85. In the Monte Carlo experiments for this model, the number of individuals $N = 1000$, the total number of time periods $T = 15$, and the number of

[10]For generalized indirect inference, we also compute estimated asymptotic standard errors. As for Model 1, in all cases the averages of these standard errors are close to the actual standard deviations of the estimates.

unobserved time periods $s = 5$ (so that the econometrician observes each individual's choices in periods $s+1 = 6$ through $T = 15$). When implementing generalized indirect inference, we use the two-step approach described in Section 3.3 and summarized in Proposition 2. When computing the first-step estimates of the structural parameters, the smoothing parameter $\lambda = 0.05$ and the number of simulated data sets $M = 10$. When computing the second-step estimates, $\lambda = 0.003$ and $M = 300$. As in Models 1 and 2, the exogenous variables (the x_{it}'s) are i.i.d. draws from a $N(0,1)$ distribution and each Monte Carlo replication has a new set of draws.

Auxiliary model #1 is the same one that we use in the Monte Carlo experiments for Models 1 and 2: $q = 1$ and the regressors in the linear probability model are given by: $z_{it} = (1, x_{it}, y_{i,t-1})$, $t = s+1, \ldots, T$, where the unobserved y_{i0} is set equal to 0. In auxiliary model #2, $q = 4$ and the regressors in the linear probability models are defined as follows:

$$\begin{aligned} z_{i,s+1} &= (1, x_{i,s+1}, x_{is}) \\ z_{it} &= (1, x_{it}, y_{i,t-1}, x_{i,t-1}), \quad t = s+2, \ldots, T \end{aligned}$$

This auxiliary model has a total of 19 parameters (note that the exogenous variables are observed in the first s time periods even though the choices are not). In auxiliary model #3, $q = 4$ and there are 27 parameters:

$$\begin{aligned} z_{i,s+1} &= (1, x_{i,s+1}, x_{i,s+1}^3, x_{is}, x_{i,s-1}) \\ z_{i,s+2} &= (1, x_{i,s+2}, y_{i,s+1}, x_{i,s+1}, x_{is}) \\ z_{i,s+3} &= (1, x_{i,s+3}, y_{i,s+2}, x_{i,s+2}, y_{i,s+1}, x_{i,s+1}) \\ z_{it} &= (1, x_{it}, y_{i,t-1}, x_{i,t-1}, y_{i,t-2}, x_{i,t-2}, y_{i,t-3}), \quad t = s+4, \ldots, T \end{aligned}$$

Finally, in auxiliary model #4, $q = 5$ and there are 41 parameters:

$$\begin{aligned} z_{i,s+1} &= (1, x_{i,s+1}, x_{i,s+1}^3, x_{is}, x_{i,s-1}, x_{i,s-2}) \\ z_{i,s+2} &= (1, x_{i,s+2}, y_{i,s+1}, x_{i,s+1}, x_{is}, x_{i,s-1}) \\ z_{i,s+3} &= (1, x_{i,s+3}, y_{i,s+2}, x_{i,s+2}, y_{i,s+1}, x_{i,s+1}, x_{is}) \\ z_{i,s+4} &= (1, x_{i,s+4}, y_{i,s+3}, x_{i,s+3}, y_{i,s+2}, x_{i,s+2}, y_{i,s+1}, x_{i,s+1}) \\ z_{it} &= (1, x_{it}, y_{i,t-1}, x_{i,t-1}, y_{i,t-2}, x_{i,t-2}, y_{i,t-3}, x_{i,t-3}, y_{i,t-4}), \quad t = s+5, \ldots, T \end{aligned}$$

Table 3 presents the results of four sets of Monte Carlo experiments, each with 1000 replications. Because SML is not feasible in this model (due to the initial conditions problem), we estimate Model 3 using only generalized indirect inference, where GII #i refers to generalized indirect inference using auxiliary model #i. Table 3 contains two key findings. First,

as for Models 1 and 2, GII generates estimates with very little bias. Second, increasing the "richness" of the auxiliary model leads to large efficiency gains relative to auxiliary model #1, particularly when the errors are persistent. As for Models 1 and 2, however, auxiliary model #4 provides few efficiency gains relative to auxiliary model #3.[11]

4.4 Results for Model 4

As described in greater detail in Section 2, Model 4 is a (static) three-alternative probit model with eight unknown parameters: three coefficients in each of the two equations for the latent utilities ($\{\gamma_{1i}\}_{i=0}^{2}$ and $\{\gamma_{2i}\}_{i=0}^{2}$) and two parameters governing the covariance matrix of the errors in these equations (c_1 and c_2). We set $\gamma_{10} = \gamma_{20} = 0$, $\gamma_{11} = \gamma_{12} = \gamma_{21} = \gamma_{22} = 1$, $c_2 = 1$, and we consider two values for c_1: 0 (implying that the errors in the equations for the latent utilities are independent) and 4/3 (implying that the errors have a correlation of 0.8). In the Monte Carlo experiments for this model, we set the number of individuals $N = 2000$. When implementing generalized indirect inference, we use the two-step approach described in Section 3.3 and summarized in Proposition 2. When computing the first-step estimates of the structural parameters, the smoothing parameter $\lambda = 0.03$ and the number of simulated data sets $M = 10$. When computing the second-step estimates, $\lambda = 0.003$ and $M = 300$. The exogenous variables (the x_{ij}'s) are i.i.d. draws from a $N(0, 1)$ distribution; each Monte Carlo replication has a new set of draws.

The auxiliary model is a pair of linear probability models, one for each of the first two alternatives:

$$\begin{aligned} y_{i1} &= z_i'\alpha_1 + \nu_{i1} \\ y_{i2} &= z_i'\alpha_2 + \nu_{i2}, \end{aligned}$$

where the vector z_i consists of polynomial functions of the exogenous variables $\{x_{ij}\}_{j=1}^{3}$ and $\{(\nu_{i1}, \nu_{i2})\}_{i=1}^{N}$ is an i.i.d. sequence of normally distributed random vectors with mean 0 and variance Σ. The auxiliary model, then, is characterized by a set of parameters $\theta = (\alpha_1, \alpha_2, \Sigma)$. The likelihood function associated with the auxiliary model is: $\mathcal{L}(y; x, \theta) = \Pi_{i=1}^{N} L(y_{i1}, y_{i2}; z_i, \alpha_1, \alpha_2, \Sigma)$, where L is a bivariate normal density with mean $(z_i'\alpha_1, z_i'\alpha_2)$ and variance Σ. Since the auxiliary model forms a system of seemingly unrelated regression equations with the same right-hand side variables in both equations, its parameters θ can be estimated via ordinary least squares.

[11] For generalized indirect inference, we also compute estimated asymptotic standard errors. As for Models 1 and 2, in all cases the averages of these standard errors are close to the actual standard deviations of the estimates.

We conduct Monte Carlo experiments using four nested versions of the auxiliary model. In auxiliary model #1, the vector $z_i = (1, x_{i1}, x_{i2}, x_{i3})$, i.e., the two linear probability models depend linearly on the three exogenous variables. Auxiliary model #1 has a total of 11 parameters. Auxiliary model #2 adds all second-order terms as well as one third-order term to z_i, i.e.,

$$z_i = (1, x_{i1}, x_{i2}, x_{i3}, x_{i1}^2, x_{i2}^2, x_{i3}^2, x_{i1}x_{i2}, x_{i1}x_{i3}, x_{i2}x_{i3}, x_{i1}x_{i2}x_{i3}),$$

for a total of 25 parameters. In auxiliary model #3, z_i contains all terms up to third order (for a total of 43 parameters) and in auxiliary model #4, z_i contains all terms up to fourth order (for a total of 67 parameters).

Tables 4 and 5 presents the results of six sets of Monte Carlo experiments, each with 1000 replications. The first two sets of experiments report the results for simulated maximum likelihood, using 25 draws (SML #1) and 50 draws (SML #2) in the GHK probability simulator. The remaining four sets of experiments report the results for generalized indirect inference, where GII #i refers to generalized indirect inference using auxiliary model #i. Table 4 shows the results for the case where the correlation between the errors in the structural model is 0 and Table 5 shows the results for the case where this correlation is 0.8.

Tables 4 and 5 contain several key findings. These findings are qualitatively similar to those for Models 1, 2, and 3 described previously. First, both SML and GII generate estimates with very little bias (provided that at least 50 draws are used with SML, particularly when the correlation between the errors in the structural model is large). Second, auxiliary model #1, which contains only linear terms, leads to large efficiency losses relative to SML (as large as 50%). Auxiliary model #2, which contains terms up to second order, reduces the efficiency losses substantially (to no more than 15% when the errors are uncorrelated and to no more than 26% when the errors have a correlation of 0.8). Auxiliary model #3, which contains terms up to third order, provides additional small efficiency gains (the largest efficiency loss is reduced to 20%), while auxiliary model #4, which contains fourth-order terms, provides few, if any, efficiency gains relative to auxiliary model #3. Finally, computing estimates using GII with auxiliary model #3 takes about 30% less time than computing estimates using SML with 50 draws.[12]

[12] As for Models 1, 2, and 3, we also compute estimated asymptotic standard errors for the GII estimates. Once again, in all cases the averages of these standard errors are close to the actual standard deviations of the estimates.

5 A Selection Model with Both Discrete and Continuous Outcomes

In this section, we discuss how to apply generalized indirect inference to a model with both discrete and continuous outcomes. As an example, consider a selection model with two equations, the first of which determines an individual's wage and the second of which determines the individual's latent utility from working at the given wage:

$$\begin{aligned} w_i &= \gamma_{10} + \gamma_{11} x_{1i} + \eta_{1i} \\ u_i &= \gamma_{20} + \gamma_{21} x_{2i} + \gamma_{22} w_i + \eta_{2i}, \end{aligned}$$

where x_{1i} and x_{2i} are exogenous regressors and $\{(\eta_{i1}, \eta_{i2})\}_{i=1}^N$ is an i.i.d. sequence of normally distributed random vectors with mean 0 and covariance matrix Ω. The econometrician does not observe the latent utility u_i, but observes instead whether the individual works, i.e., the econometrician observes $y_i \equiv I(u_i \geq 0)$. In addition, the econometrician observes the individual's wage w_i if and only if he works (i.e., if $y_i = 1$).

To illustrate how to estimate this structural model using generalized indirect inference, consider the following auxiliary model:

$$\left[\begin{array}{c} w_i \\ y_i \end{array} \right] = z_i' \alpha + \nu_i, \quad \nu_i \sim iidN(0, \Sigma),$$

where z_i is a polynomial function of the exogenous regressors and $\theta \equiv (\alpha, \Sigma)$ is the vector of auxiliary model parameters. To estimate the parameters of this model using the observed data, use the observed wage for those individuals who work and set the wage to an arbitrary constant, say, zero, for those individuals who do not work.[13]

Given a set of parameters $\beta \equiv (\gamma_{10}, \gamma_{11}, \gamma_{20}, \gamma_{21}, \gamma_{22}, \Omega)$ for the structural parameters, the simulated data consists of both simulated latent utilities $\tilde{u}_i^m(\beta)$, each of which yields a simulated choice $\tilde{y}_i^m(\beta)$, and simulated wages $\tilde{w}_i^m(\beta)$, where m denotes the mth simulated data set given the observed exogenous regressors. When estimating the parameters of the auxiliary model using the simulated data, use $g(\tilde{u}_i^m(\beta); \lambda)$ in place of $\tilde{y}_i^m(\beta)$ (as in earlier sections of the paper) and use $g(\tilde{u}_i^m(\beta); \lambda)\tilde{w}_i^m(\beta)$ in place of $\tilde{w}_i^m(\beta)$. Because the simulated latent utilities and the simulated wages are smooth functions of β, the estimated parameters of the auxiliary model using the simulated data are also smooth functions of β. Moreover, as the smoothing parameter λ goes to 0, $g(\tilde{u}_i^m(\beta); \lambda)\tilde{w}_i^m(\beta)$ converges either to $\tilde{w}_i^m(\beta)$ (if

[13] One could also, for example, set the wage for those individuals who do not work to the average observed wage.

the simulated latent utility $\tilde{u}_i^m(\beta) \geq 0$) or to 0 (if $\tilde{u}_i^m(\beta) < 0$). Thus, as λ goes to 0 at the same time that the sample size grows large, the estimates of the auxiliary model parameters using either the observed data or the simulated data converge to the same vector of pseudo true values (given that the simulated data is generated using the true vector of structural parameters). As before, this result guarantees the consistency of generalized indirect inference for the true structural parameter vector.

We are currently conducting Monte Carlo experiments that implement this approach to estimating models with mixed discrete/continuous outcomes. We will report on the results of these Monte Carlo experiments in subsequent versions of the paper. Since this model can also be estimated using maximum likelihood, we intend to focus, as in earlier sections of the paper, on a comparison between generalized indirect inference and maximum likelihood. Finally, it is important to point out that, although maximum likelihood may be hard to implement in richer versions of this model (say, ones in which the individual has more than two alternatives from which to choose), it is straightforward to use generalized indirect inference in such environments.

6 Conclusion

Discrete choice models play an important role in many fields of economics, from labor economics to industrial organization to macroeconomics. In this paper we develop and implement a new simulation-based method for estimating models with discrete or mixed discrete/continuous outcomes. The method is a generalization of indirect inference. The key innovation of this generalization is that, unlike in previous applications of indirect inference to discrete choice models, the objective surface to be optimized is a smooth function of the structural parameters. This smoothness renders indirect inference practical as a method for estimating discrete choice models. We use a set of Monte Carlo experiments to illustrate the practical usefulness of generalized indirect inference (GII). In addition to being robust and fast, GII yields estimates with good properties in small samples. In particular, the estimates display very little bias and are nearly as efficient as maximum likelihood estimates (in those cases where simulated versions of maximum likelihood can be used) provided that the auxiliary model is chosen judiciously.

Although in the Monte Carlo experiments in this paper we do not estimate discrete choice models derived from dynamic programming problems solved by forward-looking agents, GII is sufficiently flexible to accommodate almost any conceivable model of discrete choice. We hope that applied economists from a variety of fields find GII a useful and easy-to-implement

method for estimating discrete choice models.

References

An, M.Y. and M. Liu (2000), "Using Indirect Inference to Solve the Initial-Conditions Problem", *Review of Economics and Statistics* 82, 656–667.

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

Genton, M.G. and E. Ronchetti (2003), "Robust Indirect Inference", *Journal of the American Statistical Association* 98, 67–76.

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

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

Horowitz, J.L. (1992), "A Smoothed Maximum Score Estimator for the Binary Response Model", *Econometrica* 60, 505–531.

Keane, M.P. (1994), "A Computationally Practical Simulation Estimator for Panel Data", *Econometrica* 62, 95–116.

Lee, L.-F. (1997), "Simulated Maximum Likelihood Estimation of Dynamic Discrete Choice Statistical Models: Some Monte Carlo Results", *Journal of Econometrics* 82, 1–35.

Lerman, S. and C.F. Manski (1981), "On the Use of Simulated Frequencies to Approximate Choice Probabilities", in: Manski, C.F. and D. McFadden (editors), *Structural Analysis of Discrete Data with Econometric Applications*, MIT Press (Cambridge, MA), 305–319.

Manski, C. (1985), "Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Score Estimator", *Journal of Econometrics* 27, 313–334.

McFadden (1989), "A Method of Simulated Moments for Estimation of Discrete Choice Models Without Numerical Integration", *Econometrica* 57, 995–1026.

Magnac, T. J.-M. Robin, and M. Visser (1995), "Analysing Incomplete Individual Employment Histories Using Indirect Inference", *Journal of Applied Econometrics* 10, S153–S169.

Nagypál, E. (2000), "Learning-by-Doing Versus Learning about Match Quality: Can We Tell

Them Apart?", manuscript (Northwestern University).

Press, W.H., B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling (1993), *Numerical Recipes*. Cambridge University Press, Cambridge.

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

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

Table 1
Monte Carlo Results for Model 1

	Mean		Std. dev.		$\sigma_{GII}/\sigma_{SML}$		Time
	γ	ρ	γ	ρ	γ	ρ	(sec.)
			$\gamma = 1$, $\rho = 0$				
SML #1	1.000	−0.002	0.0387	0.0454	—	—	0.76
SML #2	1.001	−0.000	0.0373	0.0468	—	—	1.53
GII #1	0.998	0.002	0.0390	0.0645	1.05	1.37	0.67
GII #2	0.993	0.001	0.0386	0.0490	1.03	1.05	0.72
GII #3	0.992	0.001	0.0393	0.0490	1.05	1.05	0.91
GII #4	0.988	0.001	0.0390	0.0485	1.05	1.04	0.99
			$\gamma = 1$, $\rho = 0.4$				
SML #1	0.995	0.385	0.0400	0.0413	—	—	0.78
SML #2	0.999	0.392	0.0390	0.0410	—	—	1.54
GII #1	0.998	0.399	0.0454	0.0616	1.16	1.50	0.70
GII #2	0.993	0.396	0.0410	0.0456	1.05	1.11	0.72
GII #3	0.991	0.395	0.0417	0.0432	1.07	1.05	0.91
GII #4	0.987	0.392	0.0416	0.0432	1.07	1.05	0.97
			$\gamma = 1$, $\rho = 0.85$				
SML #1	0.984	0.833	0.0452	0.0333	—	—	0.74
SML #2	0.993	0.842	0.0432	0.0316	—	—	1.47
GII #1	0.994	0.846	0.0791	0.0672	1.83	2.13	0.71
GII #2	0.991	0.845	0.0511	0.0412	1.18	1.30	0.74
GII #3	0.992	0.846	0.0492	0.0357	1.14	1.13	0.93
GII #4	0.988	0.841	0.0490	0.0357	1.13	1.13	1.00

Table 2
Monte Carlo Results for Model 2

	Mean			Std. dev.			$\sigma_{GII}/\sigma_{SML}$			Time
	γ_1	ρ	γ_2	γ_1	ρ	γ_2	γ_1	ρ	γ_2	(sec.)
$\gamma_1 = 1$, $\rho = 0$, $\gamma_2 = 0.2$										
SML #1	1.000	0.001	0.200	0.0274	0.0357	0.0355	—	—	—	2.47
SML #2	1.002	0.002	0.199	0.0273	0.0362	0.0365	—	—	—	4.89
GII #1	0.999	0.001	0.199	0.0267	0.0571	0.0437	0.98	1.58	1.20	2.72
GII #2	0.996	0.000	0.199	0.0267	0.0379	0.0379	0.98	1.05	1.04	2.80
GII #3	0.995	0.001	0.199	0.0269	0.0377	0.0376	0.99	1.04	1.03	3.66
GII #4	0.993	0.000	0.198	0.0270	0.0377	0.0375	0.99	1.04	1.03	4.06
$\gamma_1 = 1$, $\rho = 0.4$, $\gamma_2 = 0.2$										
SML #1	0.994	0.379	0.214	0.0278	0.0314	0.0397	—	—	—	2.42
SML #2	0.999	0.389	0.206	0.0287	0.0316	0.0397	—	—	—	4.82
GII #1	0.997	0.397	0.198	0.0339	0.0587	0.0544	1.18	1.86	1.37	2.73
GII #2	0.994	0.396	0.198	0.0293	0.0386	0.0462	1.02	1.22	1.16	2.82
GII #3	0.993	0.396	0.197	0.0289	0.0343	0.0431	1.01	1.09	1.09	3.64
GII #4	0.991	0.395	0.196	0.0289	0.0348	0.0434	1.01	1.10	1.09	4.02
$\gamma_1 = 1$, $\rho = 0.85$, $\gamma_2 = 0.2$										
SML #1	0.974	0.831	0.220	0.0321	0.0174	0.0505	—	—	—	2.78
SML #2	0.987	0.840	0.208	0.0327	0.0159	0.0507	—	—	—	5.47
GII #1	1.000	0.854	0.183	0.0952	0.0633	0.1185	2.91	3.98	2.34	3.01
GII #2	0.992	0.852	0.190	0.0417	0.0266	0.0721	1.28	1.67	1.42	2.92
GII #3	0.992	0.851	0.191	0.0383	0.0179	0.0547	1.17	1.13	1.08	3.68
GII #4	0.990	0.850	0.188	0.0379	0.0175	0.0548	1.15	1.10	1.09	4.06

Table 3
Monte Carlo Results for Model 3

	Mean			Std. dev.			Time
	γ_1	ρ	γ_2	γ_1	ρ	γ_2	(sec.)
$\gamma_1 = 1$, $\rho = 0$, $\gamma_2 = 0.2$							
GII #1	0.997	−0.000	0.200	0.0272	0.0532	0.0387	3.91
GII #2	0.994	−0.001	0.200	0.0271	0.0387	0.0347	4.01
GII #3	0.993	−0.001	0.199	0.0272	0.0385	0.0345	4.81
GII #4	0.991	−0.001	0.199	0.0275	0.0389	0.0347	5.38
$\gamma_1 = 1$, $\rho = 0.4$, $\gamma_2 = 0.2$							
GII #1	0.994	0.397	0.198	0.0361	0.0518	0.0493	3.99
GII #2	0.991	0.397	0.197	0.0309	0.0363	0.0430	4.00
GII #3	0.990	0.396	0.196	0.0306	0.0317	0.0399	4.80
GII #4	0.987	0.395	0.196	0.0302	0.0318	0.0400	5.35
$\gamma_1 = 1$, $\rho = 0.85$, $\gamma_2 = 0.2$							
GII #1	0.993	0.851	0.184	0.0936	0.0403	0.1289	4.41
GII #2	0.986	0.851	0.191	0.0546	0.0249	0.0905	4.37
GII #3	0.987	0.850	0.189	0.0430	0.0140	0.0598	4.93
GII #4	0.984	0.849	0.185	0.0411	0.0136	0.0597	5.56

Table 4

Monte Carlo Results for Model 4

($\gamma_{10} = 0$, $\gamma_{11} = 1$, $\gamma_{12} = 1$, $\gamma_{20} = 0$, $\gamma_{21} = 1$, $\gamma_{22} = 1$, $c_1 = 0$, $c_2 = 1$)

	SML		GII				$\sigma_{GII}/\sigma_{SML}$			
	#1	#2	#1	#2	#3	#4	#1	#2	#3	#4
					Mean					
γ_{10}	0.007	0.005	0.003	0.002	0.002	0.002	—	—	—	—
γ_{11}	1.000	1.001	0.995	0.994	0.992	0.990	—	—	—	—
γ_{12}	1.000	1.003	0.998	0.997	0.995	0.992	—	—	—	—
γ_{20}	−0.001	−0.003	−0.006	−0.004	−0.004	0.004	—	—	—	—
γ_{21}	1.006	1.007	1.001	0.999	0.997	0.996	—	—	—	—
γ_{22}	1.005	1.007	1.004	1.000	0.998	0.996	—	—	—	—
c_1	0.020	0.010	0.007	0.005	0.005	0.006	—	—	—	—
c_2	1.004	1.003	1.006	1.001	1.001	1.002	—	—	—	—
					Std. dev.					
γ_{10}	0.0630	0.0628	0.0720	0.0666	0.0656	0.0665	1.15	1.06	1.04	1.06
γ_{11}	0.0686	0.0686	0.0872	0.0764	0.0741	0.0743	1.27	1.11	1.08	1.08
γ_{12}	0.0572	0.0574	0.0719	0.0667	0.0632	0.0646	1.25	1.16	1.10	1.13
γ_{20}	0.0663	0.0657	0.0745	0.0686	0.0677	0.0676	1.13	1.04	1.04	1.03
γ_{21}	0.1065	0.1050	0.1395	0.1128	0.1095	0.1099	1.33	1.07	1.04	1.05
γ_{22}	0.1190	0.1174	0.1593	0.1285	0.1249	0.1244	1.36	1.09	1.06	1.06
c_1	0.1091	0.1107	0.1303	0.1276	0.1224	0.1265	1.18	1.15	1.11	1.14
c_2	0.1352	0.1325	0.1991	0.1509	0.1439	0.1421	1.50	1.14	1.09	1.07
Time	11.5	23.1	7.1	10.4	16.4	34.1	—	—	—	—

Table 5

Monte Carlo Results for Model 4

($\gamma_{10} = 0$, $\gamma_{11} = 1$, $\gamma_{12} = 1$, $\gamma_{20} = 0$, $\gamma_{21} = 1$, $\gamma_{22} = 1$, $c_1 = 1.33$, $c_2 = 1$)

	SML		GII				$\sigma_{GII}/\sigma_{SML}$			
	#1	#2	#1	#2	#3	#4	#1	#2	#3	#4
					Mean					
γ_{10}	−0.031	−0.017	0.000	−0.001	−0.000	−0.001	—	—	—	—
γ_{11}	0.998	1.000	0.993	0.993	0.991	0.989	—	—	—	—
γ_{12}	1.016	1.011	0.998	0.998	0.996	0.994	—	—	—	—
γ_{20}	−0.011	−0.010	−0.011	−0.007	−0.007	−0.006	—	—	—	—
γ_{21}	0.992	0.999	1.000	0.997	0.995	0.991	—	—	—	—
γ_{22}	1.004	1.008	1.006	1.001	0.999	0.995	—	—	—	—
c_1	1.269	1.306	1.347	1.338	1.335	1.330	—	—	—	—
c_2	1.025	1.011	0.993	0.993	0.995	0.997	—	—	—	—
					Std. dev.					
γ_{10}	0.0693	0.0698	0.0789	0.0776	0.0758	0.0757	1.13	1.11	1.09	1.08
γ_{11}	0.0587	0.0588	0.0696	0.0658	0.0632	0.0636	1.18	1.12	1.07	1.08
γ_{12}	0.0745	0.0737	0.0883	0.0801	0.0781	0.0782	1.20	1.09	1.06	1.06
γ_{20}	0.0766	0.0764	0.0900	0.0801	0.0786	0.0780	1.18	1.05	1.03	1.02
γ_{21}	0.0884	0.0886	0.1140	0.0969	0.0952	0.0943	1.29	1.09	1.07	1.06
γ_{22}	0.1106	0.1103	0.1471	0.1204	0.1176	0.1153	1.34	1.09	1.07	1.05
c_1	0.1641	0.1707	0.2454	0.2152	0.2049	0.2041	1.44	1.26	1.20	1.20
c_2	0.1229	0.1206	0.1599	0.1387	0.1338	0.1311	1.33	1.15	1.11	1.09
Time	12.7	25.6	7.4	10.8	17.1	34.4	—	—	—	—