A Brief Discussion on Structural Estimation

1 Definition

The founding members of the Cowles Commission defined **econometrics** as: "*a branch of economics in which economic theory and statistical method are fused in the analysis of numerical and institutional data*". Today, economists refer to models that combine explicit economic theories with statistical models as **structural econometric models**.

In structural econometric models, **economic theory** is used to **develop mathematical statements** about how a set of observable endogenous variables, y, are related to another set of observable explanatory variables, x. Economic theory also may relate the y variables to a set of unobservable variables, ξ. These theoretical relations usually are in the form of equalities: $\mathbf{y} = \mathbf{g}(\mathbf{x}, \xi, \Theta)$, where $g(\cdot)$ is a known function and Θ a set of unknown parameters or functions. Occasionally, economic theory may only deliver inequality relations, such as $y \geqslant g(x, \xi, \Theta)$.

After adding statistical assumptions about the joint distribution of x, ξ and other unobservables in the model, we can use **log-likelihood function**, $l(y, x|\Theta)$, or **conditional moments**, such as $E(y, x|\Theta)$ to estimate the structural model.

There are many kinds of structural models, such as:

1. **Fully Specified Structural Models**

   Fully specified structural models make **explicit assumptions** about the economic actors' objectives and their economic environment and information set, as well as specifying which choices are being made within the model. We call these models fully specified because they allow a complete solution to the individuals optimization problem as a function of the current information set.

2. **Partially Specified Structural Models**
   The model is partially specified, in the sense that there is not enough information to solve for the optimal choice as a function of the information set: for example, the labor supply model resulting from the marginal rate of substitution characterization is silent about expectations for the future, the distribution of shocks, and the functioning of credit markets. However, conditioning on consumption makes the relationship between labor supply and wages valid and dependent upon structural parameters that characterize some aspects of utility.

2 Construct Structural Models

2.1 General Sources of "Structure"

There are two **general sources** of "structure" in structural models – **economics and statistics**. Economics allows the researcher to infer how economic behavior and institutions affect relationships between a set of economic conditions x and outcomes y.

Often these economic models are deterministic, and as such do not speak directly to the distribution of noisy economic data. Thus, structural econometric modelers must add statistical structure in order to rationalize why economic theory does not perfectly explain data.

2.2 Assumptions in the Structural Models

When introducing economic and statistical **assumptions** into the structural models, we should aim to reect economic realities, rationalize what is observed in the data or describe how the data were generated and simplify estimation. For instance, necessary assumptions for a linear regression model to have a causal economic interpretation as a production function should satisify 2 conditions: first, the researcher must specify an economic model of the phenomenon under consideration, including in this case the functional form for the production function; second, he/she must incorporate unobservables into the economic model. The second step should receive significantly more attention, because the assumptions made about the unobservables will impact the consistency of OLS parameter estimates.

2.3 Estimation of the Parameters

Estimation of the parameters in the model can be described in a five-step algorithm:

1. Start with an initial guess at a set of parameter values θ.
2. Numerically solve the model given the initial parameter vector θ.
3. If individuals are *ex ante* identical, simulate the careers of say S individuals using a random number generator for realisations of the stochastic variables, and construct moments from the simulated moments analogous to those constructed from the data. If individuals differ by exogenous observed factors, simulate S careers for each value of the exogenous initial conditions. Similarly if individuals differ by some unobserved characteristic (whose distribution is estimated together with the rest of the model) again simulate S careers for each point of support of the unobservable and then take suitable weighted averages when constructing the moments.
4. Calculate the "criterion function" being minimized. This may be a simple or weighted quadratic distance between the data and the simulated moments.
5. Update the set of parameters θ to minimize the criterion function and return to step 2 and numerically solve the model with the updated parameters.

3 Applications

3.1 Industrial Organization – Estimation of Random-CoefŐfcients Logit Models of Demand (Nevo, 2000)

3.1.1 Introduction of the Method

This method is the random-coeffients (or mixed) logit methodology for estimating demand in differentiated-product markets using market-level data. This method can be used to estimate the demand for a large number of products using market data and allowing for the endogenity of price. While this method retains the benefits of alternative discrete-choice models, it produces more realistic demand elasticities.

3.1.2 the Model

First, the method uses (market-level) price and quantity data for each product, in a series of markets, to estimate the model. The key benefit of this methodology is that we **do not need to observe individual consumer purchase decisions** to estimate the demand parameters.

Second, the estimation allows prices to be correlated with the econometric error term. A product will be defined by a set of characteristics, while producers and consumers are assumed to observe all product characteristics. The researcher, on the other hand, is assumed to observe only some of the product characteristics. Each product will be assumed to have a characteristic that influences demand but that either is not observed by the researcher or cannot be quantified into a variable that can be included in the analysis. The unobserved characteristics will be captured by the econometric error term. Since the producers know these characteristics and take them into account when setting prices, this introduces the econometric problem of **endogenous prices**. The contribution of the estimation method presented below is to transform the model in such a way that

instrumental-variable methods can be used.

1. **Basic Setups**

   Assume we observe $t = 1, \ldots, T$ markets, each with $i = 1, \ldots, I_t$ consumers. For each such market we observe aggregate quantities, average prices, and product characteristics for J products.

   The indirect utility of consumer i from consuming product j in market t, $U(x_{jt}, \xi_{jt}, p_{jt}, \tau_i; \theta)$ is a function of observed and unobserved (by the researcher) product characteristics, x_{jt} and ξ_{jt} respectively; price, p_{jt}; individual characteristics, τ_i; and unknown parameters, θ. The method focuses on a particular specification of this indirect utility

   $$\begin{gathered} u_{ijt} = \alpha_i(y_i - p_{jt}) + x_{jt}\beta_i + \xi_{jt} + \varepsilon_{ijt} \\ i = 1, \ldots, I_t, \quad j = 1, \ldots, J, \quad t = 1, \ldots, T \end{gathered} \tag{1}$$

   where y_i is the income of consumer i, p_{jt} is the price of product j in market t, x_{jt} is a K-dimensional (row) vector of observable characteristics of product j, ξ_{jt} is the unobserved (by the econometrician) product characteristic, and ε_{ijt} is a mean-zero stochastic term. Finally, α_i is consumer is marginal utility from income, and β_i is a K-dimensional (column) vector of individual-specific taste coefficients.

   Observed characteristics vary with the product being considered. Unobserved characteristics, for example, can include the impact of unobserved promotional activity, unquantifiable factors (brand equity), or systematic shocks to demand. Depending on the structure of the data, some components of the unobserved characteristics can be captured by dummy variables.

2. **Model Heterogeneity**

   The individual characteristics consist of two components: demographics, which are referred to as observed, and additional characteristics, which are referred to as unobserved, denoted D_i and v_i respectively, which will be modeled as:

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \Pi D_i + \Sigma v_i, \quad v_i \sim P_v^*(v), \quad D_i \sim \hat{P}_D^*(D) \tag{2}$$

where D_i is a $d \times 1$ vector of demographic variables, m_i captures the additional characteristics discussed in the previous paragraph, $P_v^*(\cdot)$ is a parametric distribution, $\hat{P}_D^*(\cdot)$ is either a nonparametric distribution known from other data sources or a parametric distribution with the parameters estimated elsewhere, Π is a $(K+1)\times d$ matrix of coefficients that measure how the taste characteristics vary with demographics, and Σ is a $(K+1) \times (K+1)$ matrix of parameters.

The specification of the demand system is completed with the introduction of an outside good: the consumers may decide not to purchase any of the brands. Without this allowance, a homogenous price increase (relative to other sectors) of all the products does not change quantities purchased. The indirect utility from this outside option is

$$u_{i0t} = \alpha_i y_i + \xi_{0t} + \pi_0 D_i + \sigma_0 v_{i0} + \varepsilon_{i0t}$$

The mean utility from the outside good, ξ_{0t} is not identified (without either making more assumptions or normalizing one of the inside goods). Also, the coefficients π_0 and σ_0 are not identified separately from coefficients on an individual-specific constant term in equation (1). The standard practice is to set ξ_{0t}, π_0, and σ_0 to zero, and since the term$\alpha_i y_i$ will eventually vanish (because it is common to all products), this is equivalent to normalizing the utility from the outside good to zero. Let $\theta = (\theta_1, \theta_2)$ be a vector containing all the parameters of the model. The vector $\theta_1 = (\alpha, \beta)$ contains the linear parameters, and the vector $\theta_2 = (\Pi, \Sigma)$ the nonlinear parameters. Combining equations (1) and (2), we have

$$\begin{aligned} u_{ijt} &= \alpha_i y_i + \delta_{jt}\left(x_{jt}, p_{jt}, \xi_{jt}; \theta_1\right) + u_{ijt}\left(x_{jt}, p_{jt}, v_i, D_i; \theta_2\right) + \varepsilon_{ijt} \\ \delta_{jt} &= x_{jt}\beta - \alpha p_{jt} + \xi_{jt}, \quad u_{ijt} = \left[-p_{jt}, x_{jt}\right](\Pi D_i + \Sigma v_i) \end{aligned} \tag{3}$$

where $\left[-p_{jt}, x_{jt}\right]$ is a $1 \times (K+1)$ (row) vector. The indirect utility is now expressed as a sum of three (or four) terms. The first term, $\alpha_i y_i$, is given only for consistency with equation (1) and will vanish. The second term, δ_{jt} , which is referred to as the mean utility, is common to all consumers. Finally, the last two terms, $u_{ijt} + \varepsilon_{ijt}$, represent a mean-zero heteroskedastic deviation from the mean utility that captures the effects of the random coefficients.

Consumers are assumed to purchase one unit of the good that gives the highest utility, which means

$$A_{jt}\left(x_{\cdot t}, p_{\cdot t}, \delta_{\cdot t}; \theta_2\right) = \left\{\left(D_i, v_i, \varepsilon_{i0t}, \ldots, \varepsilon_{ijt}\right) | u_{ijt} \geq u_{ilt} \quad \forall l = 0, 1, \ldots, J\right\}$$

The set A_{jt} defines the individuals who choose brand j in market t. Assuming ties occur with zero probability, the market share of the jth product is just an integral over the mass of consumers in the region A_{jt}. Formally, it is given by

$$\begin{aligned} s_{jt}\left(x_t, p_t, \delta_t; \theta_2\right) &= \int_{A_{jt}} dP^*(D, v, \varepsilon) \\ &= \int_{A_{jt}} dP^*(\varepsilon | D, v) dP^*(v | D) dP_D^*(D) \\ &= \int_{A_{jt}} dP_\varepsilon^*(\varepsilon) dP_v^*(v) d\widehat{P}_D^*(D) \end{aligned} \tag{4}$$

Given assumptions on the distribution of the (unobserved) individual attributes, we can compute the integral in equation (4), either analytically or numerically.

3. **Distributional Assumptions**

If we assume that ε_{ijt} are distributed according to a ***Type I extremevalue distribution***, this is the (aggregate) logit model. The market share of brand j in market t, defined by equation (4), is

$$s_{jt} = \frac{\exp\left(x_{jt}\beta - \alpha p_{jt} + \xi_{jt}\right)}{1 + \sum_{k=1}^{J} \exp\left(x_{kt}\beta - \alpha p_{kt} + \xi_{kt}\right)} \tag{5}$$

The price elasticities of the market shares defined by equation (5) are

$$\eta_{jkt} = \frac{\partial s_{jt} p_{kt}}{\partial p_{kt} s_{jt}} = \begin{cases} -\alpha p_{jt}\left(1 - s_{jt}\right) & \text{if } j = k \\ \alpha p_{kt} s_{kt} & \text{otherwise} \end{cases}$$

If in the full model, described by equations (1) and (2), we maintain the i.i.d. extreme-value distribution assumption on ε_{ijt}. Correlation between choices is obtained through the term u_{ijt}. The correlation will be a function of both product and consumer characteristics: the correlation will be between products with similar characteristics, and consumers with similar demographics will have similar rankings of products and therefore similar substitution patterns. Therefore, rather than having to estimate a large number of parameters, corresponding to an unrestricted variance-covariance matrix, we only have to estimate a smaller number.

The price elasticities of the market shares, s_{jt}, defined by equation (4) are

$$\eta_{jkt} = \frac{\partial s_{jt} p_{kt}}{\partial p_{kt} s_{jt}} = \begin{cases} -\frac{p_{it}}{s_{it}} \int \alpha_i s_{ijt}\left(1 - s_{it}\right) d\hat{P}_D^*(D) dP_v^*(v) & \text{if } j = k \\ \frac{p_{tt}}{s_t} \int \alpha_i s_{ijt} s_{ikt} d\hat{P}_D^*(D) dP_v^*(v) & \text{otherwise} \end{cases}$$

Now the own-price elasticity will not necessary be driven by the functional form. The partial derivative of the market shares will no longer be determined by a single parameter, α. Instead, each individual will have a different price sensitivity, which will be averaged to a mean price sensitivity using the individual specific probabilities of purchase as weights. The price sensitivity will be different for different brands.

The full model also allows for flexible substitution patterns, which are not constrained by a priori segmentation of the market (yet at the same time can take advantage of this segmentation by including a segment dummy variable as a product characteristic).

3.2 Labor economics – The Career Decisions of Young Men (Keane, M. P. and K. I. Wolpin., 1997)

3.2.1 Introduction of the Paper

The paper provides structural estimates of a dynamic model of schooling, work, and occupational choice decisions based on 11 years of observations on a sample of young men from the 1979 youth cohort of the National Longitudinal Surveys of Labor Market Experience (NLSY) using suitably extended human capital investment model.

First, the authors ***estimate parameters*** that may be of interest in their own right, such as those of technology. In the current context, the framework isolates the quantitative importance of school attainment and occupation-specific work experience in the production of occupation-specific skills. Second, because they explicitly solve an optimization problem and thus determine decision rules, they can ***quantify the effect on decisions of altering specific parameters of the environment***. Moreover, because schooling, work, and occupational choice are interrelated, they can ***estimate the impact of an intervention***, such as a college tuition subsidy, on subsequent occupational choice decisions. Previous research has generally treated school and work decisions in isolation and therefore has been limited in its ability to address such questions. Finally, structural estimation permits welfare analysis, allowing to calculate the distributional consequences of interventions on lifetime wealth and utility.

3.2.2 the Model

First, they use the basic human capital model and present the estimates of this model and evaluate its ability to fit the data. Since the basic model fails to capture quantitatively some important features of the data, the authors present an extended version of the model, which is proved to be fitted substantially better to the data.

3.2.3 Estimation Results of the Basic Model

In assessing the basic model, the authors consider three criteria: (1) the reasonableness of the parameter values, (2) within-sample fit, and (3) out-of-sample fit.

First, they get that the basic model generates parameter values that appear to be within reasonable ranges. Second, based on a simulation of 5,000 individuals, they use five figures graphically depict the fit of the basic model to the choice data. The largest discrepancies occur with respect to the schooling, military, and home alternatives, although the qualitative age patterns are replicated, such as **Figure** 1 below, they also depict percentage blue-collar by age, percentage in the military by age, percentage in school by age and percentage at home by age.

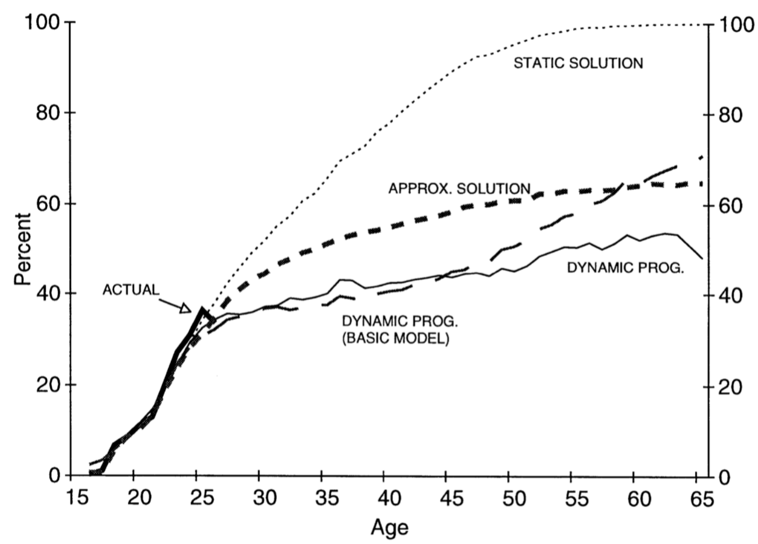

Figure 1: Percentage white-collar by age

Moreover, they found that the impact on accepted wages of the steep wage offerexperience gradient is more apparent outside of the sample ages.

3.2.4 Extended Model

Since the basic human capital model does not provide a good fit to the quantitative features of the NLSY data. To attempt to better fit the data, teh authors extended the basic human capital model in directions motivated by its specific failures. The extensions serve to improve the overall choice distribution and the pattern of persistence.

3.2.5 Estimation Results of the Extended Model

The extended model not only provides a much better fit of the pattern of choices relative to the basic model, but also more closely captures the features of the data

Figure 1 also shows the forecasts of the three models through age 65, well beyond the actual data. These out-of-sample forecasts diverge considerably. The static model predicts unrealistically rapid changes in the choice constellation with age, culminating in almost everyone opting for white-collar employment by age 50. Neither the approximation model nor the dynamic programming model (the basic and extended models) forecasts such extreme outcomes. However, the approximation model tends to extrapolate trends more closely in within-sample age profiles.

3.2.6 More Discussions

Skill endowment heterogeneity is potentially an important determinant of inequality in lifetime welfare. The authors admit that although they cannot determine each individuals actual type, they can use ***Bayess rule*** to compute the probability distribution of the endowment types conditional on choices, wages, and initial schooling. Then, having calculated these endowment type probabilities for each individual, they can determine the extent to which observed family background characteristics are related to type.

Moreover, they use the estimates to **predict** the impact of a $2,000 college **tuition subsidy** on schooling decisions and other life cycle outcomes. They found that such a subsidy would have a negligible impact on the expected present value of lifetime utility. Those who would benefit most are the types with high endowments of white-collar and schoolrelated skills, that is, those who, for the most part, would have gone to college even without the subsidy. Those who are induced to attend college by the subsidy are primarily those with a comparative advantage in blue-collar occupations and poor endowments of school-related skills. Because most of the subsidy is needed simply to bring such people to the margin of indifference between college attendance and other options (in the model individuals are not financially constrained with respect to college tuition costs), it will tend to have little effect on their lifetime wealth. Tuition subsidies of this magnitude do little to compensate for utility differences arising from endowments.

4 Strengths

First, a structural model can be used to **estimate unobserved economic or behavioral parameters** that could not otherwise be inferred from nonexperimental data. Examples of structural parameters include: marginal cost; returns to scale; the price elasticity of demand; and the impact of a change in an exogenous variable on the amount demanded or on the amount supplied.

Second, structural models can be used to **perform counterfactuals or policy simulations**. In counterfactuals, the researcher uses the estimated structural model to predict what would happen if elements of the economic environment change. For example, suppose that one has estimated the demand for a product and the monopolists cost function. He could then, with added assumptions, use these estimates to calculate how market prices and quantities would change if an identical second firm entered the monopoly market.

Finally, structural models can be used to **compare the predictive performance of two competing theories**. For example, we could compare the performance of quantity-setting versus price-setting models of competition.

5 Limitations

The structural estimation **depends heavily on the economic theories**. Sometimes we may use the wrong model, also there are many interesting applications where there is little or no useful economic theory to guide empirical work.

Structural parameter estimates may be **sensitive to the specific model assumptions**, because we cannot test economic models independent of functional form assumptions for a finite number of observations. For example, if we were trying to estimate consumer surplus, we should be aware that it might make a tremendous difference that we assumed demand was linear, as opposed to constant elasticity.

The "**structure**" in structural models also can affect statistical inferences about economic primitives. For example, if a researcher wants to fail to reject a null hypothesis, then she should specify an extremely rich functional form with plenty of variables that are not part of the economic theory. Such a strategy will likely *decrease the power* of the statistical test.

References

[1] Reiss, P. C. and F. A. Wolak. 2007. "Structural Econometric Modeling: Rationales and Examples from Industrial Organization," *Handbook of Econometrics*, Vol. 6A

[2] Low, H. and C. Meghir. 2017. "The Use of Structural Models in Econometrics," *The Journal of Economic Perspectives*, Vol. 31(2)

[3] Hood, W.C., Koopmans, T.C. (1953). Studies in Econometric Method, Cowles Commission Monograph no. 14. John Wiley, New York.

[4] Aviv Nevo. 2000. "A Practitioners Guide to Estimation of Random Coefficients Logit Models of Demand", *Journal of Economics & Management Strategy*, 9(4), 513-548.

[5] Keane, M. P. and K. I. Wolpin. 1997. "The Career Decisions of Young Men," *Journal of Political Economy*, Vol. 105(3)

[6] Jiaming MAO. "Foundations of Causal Inference"
$https://github.com/jiamingmao/data-analysis/blob/master/Lectures$