Micro-Econometrics and Household Behaviour
Richard Blundell

Report by Adriaan R. Soetevent, University of Groningen[*]

1 Introduction

This paper reviews the lectures given by Richard Blundell at the 27th NAKE Workshop, held in Amsterdam in December 1999.

In this paper, only a concise overview can be given of the topics Blundell discussed and touched on during the week. I will preserve the same order as Blundell and therefore kick off with a discussion about the usefulness of non- and semiparametric methods. These methods are useful for examining the relationship between the expenditure on specific commodities and the total expenditure for an individual and they provide useful benchmarks to test parametric specifications against. In Section 3, the data set considered consists of panel data. In particular, the problem of using the common GMM estimator for estimating a dynamic panel data model when the instruments are weak, is discussed and a solution to this problem is given. Section 4 deals, also on the basis of panel data, with the question how the impact of treatments, like e.g. a labor market program, can be evaluated. This problem can be seen as a missing data problem. The measure proposed is the difference-in-difference estimator. Finally, in Section 5, matters are combined: attempts are made to describe the consumption of an individual over his life cycle by solving the dynamic programming problem. The issue which IV variables to use is elaborated on.

2 Non- and Semiparametric Estimation

The relationship between expenditure on specific commodities and total expenditure at a particular point in time and location (the Engel curve) as stated in the classical Working-Leser form is $\frac{p_j q_j}{x} = \alpha_j + \beta_j \ln x + \varepsilon_j;\ \forall j = 1, \dots, n$. p_j is the price of commodity j, q_j the quantity, x total expenditure and ε_j an error term for which it is assumed that

[*]Faculty of Economics, University of Groningen, P.O. Box 800, 9700 AV Groningen. E-mail: a.r.soetevent@eco.rug.nl

$E(\varepsilon_j|x) = 0 \; \forall j$. A number of studies has indicated that this linear specification of the relationship between the commodity expenditure and log income is inaccurate for an appreciable number of goods. For some commodities, e.g. food, the linear specification is reasonably in accordance with reality but for others, like alcohol, the curvature indicates that further terms in income are required (see Banks, Blundell and Lewbel (1997)).

One way to analyze the actual relationship is with the use of non- and semiparametric methods. These methods are increasingly used in the field of microeconometrics. They are used as a means to relax distributional assumptions. In many cases, non- and semi-parametric analysis is a useful alternative to test a parametric specification against. Two requirements that limit the potential use of nonparametrics are a lot of observations are needed and that the variable x is not measured with errors.

Nonparametric Regression

I will first outline how nonparametric methods are applied to Engel curve analysis. Subsequently, I explain how a particular subset of semi-parametric models, the partial linear models, can be used to include demographic shifts in the regression model. Finally, I will show how outcomes can be compared with the parametric alternative. Details can be found in the article of Blundell and Duncan (1998) which covers a large part of the material.

Suppose that the relation of interest is given by

$$y = g(x) + \varepsilon; \tag{1}$$

where ε is an independent random error satisfying $E(\varepsilon|x) = 0$; $var(\varepsilon|x) = \sigma^2(x)$. In the Engel curve analysis y represents the expenditure share on some good and x the total budget. Nonparametrics allows us to investigate the relationship between y and x without making parametric assumptions on g. The object of interest is the conditional expectation $E(y|x) = g(x)$ which can be estimated by the use of Kernel regression[1]. The conditional expectation can be rewritten as

$$E(y|x) = \int y f(y|x)dy = \frac{\int y f(x; y)dy}{f(x)}: \tag{2}$$

Given observations $\{X_i; Y_i\}$, the quantities on the right hand side to be estimated are $f(x; y)$ and $f(x)$. The general form for the multivariate density estimator is

$$\hat{f}_H(x) = \frac{1}{n}\sum_{i=1} K_H(x_i - x); \tag{3}$$

[1] Note that this is just one out of more non-parametric estimation techniques, which also include for example Nearest Neighbour, Splines and Median Smoothing.

where $K_H(x) = \det(H)^{-1}K(H^{-1}x)$ and H the (nonsingular) bandwidth matrix. One form the multidimensional kernel function $K(H^{-1}x)$ may take on is the product kernel $K(H^{-1}x) = K_1((H^{-1}x)_1)\cdot\ldots\cdot K_n((H^{-1}x)_n)$. Plugging the above in to (2) and using this product kernel , after some manipulations, the Nadaraya-Watson estimator

$$\hat{g}_h(x) = \frac{n^{-1}\sum_{i=1}^{n} K_h(x - x_i)y_i}{n^{-1}\sum_{j=1}^{n} K_h(x - x_j)} \tag{4}$$

is obtained. Conditions for the consistency and the asymptotic normality of this estimator are given in Blundell and Duncan (1998) and Härdle (1990). These conditions allow us to derive pointwise confidence bands around the estimated regression curve. Popular choices for $K(\cdot)$ are the Gaussian kernel $K(u) = \frac{1}{\sqrt{2\pi}}\exp^{-u^2/2}$ and the Epanechnikov kernel $K(u) = \frac{3}{4}(1 - u^2 1(|u| \leq 1)$, where $1(\cdot)$ denotes an indicator function. The last one has the useful property that it truncates the distribution.

Particularly important is the choice of the bandwidth h. This can be chosen on the basis of plug-in methods or using the method of Cross Validation (CV). An example of a plug-in method is Silverman's rule of thumb which asymptotically minimizes the expected mean squared error of the estimate $\hat{g}(x)$ of the density $g(x)$ over the range of x. The optimal choice for h is in this case $\hat{h} \approx 1.06\sigma n^{-1/5}$. This bandwidth choice has the disadvantage that it is sensitive to outliers due to the estimator of σ occuring at the right hand side.

In an alternative approach, the method of Cross Validation, the function minimized is

$$CV(h) = \frac{1}{n}\sum_{i=1} w(x_j)(y_j - \hat{g}_{h,j}(x_j))^2;$$

with $w(x_j)$ some trimming function[2]. This criterion function gives the average squared error between the observation y_j and the estimator $\hat{g}_h(x)$, where $\hat{g}_h(x)$ is replaced by the leave-one-out-estimator

$$\hat{g}_{h,j}(x_j) = \sum_{i \neq j} y_i \pi_{ih}^{-j}(x_j);$$

with

$$\pi_{ih}^{-j}(x_j) = \frac{K_h(x_i - x_j)}{\sum_{k \neq j}^{n} K_h(x_k - x_j)};$$

in order to obtain an unbiased estimate of the average squared error. Note that when using $\hat{g}_h(x)$ instead, the criterion function could be made arbitrarily small by letting $h \to 0$.

[2] This trimming function may be used to assign less weight to observations at the tail of the distribution of x.

Semiparametric Regression

Sometimes, it is useful to add to the purely nonparametric part $g(x)$ on the right hand side of (1) a purely parametric part $\gamma' z$ which represents an index in terms of a finite vector of observable exogenous regressors z and unknown parameters γ. Then the partially linear model

$$y = g(x) + \gamma' z + \varepsilon; \tag{5}$$

is obtained, which is a subset in the class of semiparametric models. Assume $E(\varepsilon|z;x) = 0$ and $Var(\varepsilon|z;x) = \sigma^2(z;x)$. One reason to make a partitioning between the explanatory variables x and z may be that the variables x are continuous and the variables z binary or discrete. In the Engel curve analysis this gives a convenient way to account for heterogeneity in the population, in particular differences in family size; Blundell and Duncan (1998) show that the difference in number of children leads to a difference in share equations that is sufficient to consider semiparametric estimation techniques. To obtain an estimate for γ a transformation is made by taking expectations of (5) conditional on x and subtracting from (5), yielding

$$y - E(y|x) = (z - E(z|x))'\gamma + \varepsilon:$$

The conditional expectations in this expression can be estimated using nonparametric kernel regression and subsequently γ can be estimated by applying OLS to the resulting equation. Finally, these estimates and (5) can be used to derive an estimate for $g(x)$.

Parametric versus Nonparametric Models

For the Engel curve estimates, Blundell and Duncan (1998) compared the nonparametric specification with both the Working-Leser specification and the quadratic parametric regression curve. This last specification is extensively discussed in Banks, Blundell and Lewbel (1997). Using the goodness-of-fit statistic derived by Aït-Sahalia, Bickel and Stoker (ABS), Blundell and Duncan, the results given in Table 1 are obtained. Looking at the P-values given in this table, it is not possible to reject linearity for the food share equation, whereas for the alcohol share the quadratic specification is sustained by the empirical evidence.

Local Polynomial Regression

Till now, in the kernel regression the Nadaraya-Watson estimator (4) was used, which is in fact a local constant estimator. However, I can do better by using prior information on the shape of the curve. For example, from Table 1 I might conclude that for the food share Engel equation a local linear estimator is appropiate and for the alcohol equation a local

Table 1: Estimates of the ABS test statistic

	Food	Alcohol
H_0: linear	1.679	4.633
	[.195]	[.031]
H_0: quadratic	0.567	0.526
	[.451]	[.468]

Note: The nonparametric estimates are based on a Gaussian kernel with bandwidths chosen by Least Squares CV. The statistics are distributed as χ^2 under H_0 and the values between brackets give the P-values.

quadratic fit. The estimator is now defined as $\hat{g}_h(x) = m\{x; \hat{\alpha}_n(x)\}$ where $m(x_i; \alpha) = \alpha_0 + \alpha_1 x_i + \dots + \alpha_p x_i^p$. The Nadaraya-Watson estimator corresponds to the choice $p = 0$ and the local linear and local quadratic estimator to $p = 1$ and $p = 2$ respectively. When the parametric model is indeed true, using the higher order polynomial estimators will lead to a reduction in the asymptotic bias of the estimator. Another advantage of using local polynomial regression is that it makes the estimation less sensitive to the choice of the bandwidth h.

3 Dynamic Panel Data Estimation

The preceding discussion on non- and semiparametric estimation techniques dealt with cross section data. The subject of the current section is how to act when one has data that combine time series and cross sections. These panel, or longitudinal data sets contain observations of many individuals i, $i = 1, \dots, N$, each observed at several points t in time, with $t = 1, \dots, T$. Typically, T is small compared to N. Panel data models incorporating dynamic effects are called dynamic panel data models.

This section points out why the first differenced GMM estimator may perform poorly in a dynamic panel data context and how this can be solved.

GMM Estimation

The idea of the Method of Moments is, that one estimates k parameters by expressing these parameters in terms of k population moments and one then replaces these population moments with the corresponding sample moments. However, when there are more moment conditions than parameters, the system is overdetermined. One way to reconcile conflicting estimates is to minimize a certain criterion function. Suppose the criterion function has the form $q = \bar{m}(\theta)' A \bar{m}(\theta)$ where the element $\bar{m}_j(\theta)$ of $\bar{m}(\theta)$ denotes the to moment j corresponding sample mean and A a positive definite matrix that is not a function of θ. Then the Generalized Methods of Moments (GMM) estimator is the estimator that minimizes this function (see Greene (1999)).

GMM Estimation and the Dynamic Panel Data Model

Within the context of the dynamic error components model, the standard first differenced Generalized Method of Moments (GMM) estimator performs very poorly in situations characterized by a moderately large autoregressive parameter and a small amount of observations (Blundell and Bond (1998)). Therefore, an alternative linear estimator is considered. Many of the details can be found in Blundell and Bond (1998).

In order to define these estimators, I consider the following AR(1) model

$$y_{it} = \alpha y_{i t-1} + \eta_i + v_{it}; \text{ for } i = 1; \ldots; N \text{ and } t = 1; \ldots; T \tag{6}$$

where $u_{it} \equiv \eta_i + v_{it}$ is the usual 'fixed effects' decomposition of the error term and where it is assumed that η_i and v_{it} are independently distributed across i. The familiar error components structure implies the $\frac{1}{2}(T-1)(T-2)$ orthogonality conditions

$$E(y_{i;t-s} \Delta v_{it}) = 0 \text{ for } t = 3; \ldots; T \text{ and } s \geq 2;$$

where $\Delta v_{it} = v_{it} - v_{i;t-1}$. These conditions rely on the absence of serial correlation in v_{it} and on the initial condition restrictions $E(y_{i1} v_{it}) = 0$ for $i = 1; \ldots; N$ and $t = 2; \ldots; T$. These moment restrictions can be compressed to

$$E(Z_i' \bar{u}_i) = 0;$$

with Z_i a $(T-2) \times m$ instrument matrix given by

$$Z_i = \begin{pmatrix} y_{i1} & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & y_{i1} & y_{i2} & \cdots & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & \cdots & y_{i1} & \cdots & y_{iT-2} \end{pmatrix};$$

and $\bar{u}_i$ the $(T-2)$ vector $(\Delta v_{i3}; \Delta v_{i4}; \ldots; \Delta v_{iT})'$

Using these conditions and the assumed absence of serial correlation in the v_{it}, the two-step GMM estimator $\hat{\alpha}$ can be calculated. In case $T = 3$, this estimator is just identified which means that no weighting improvements are available to enhance the asymptotic efficiency.

The Problem of Weak Instruments

The instruments in the standard first-differenced GMM estimator become less informative when α in (6) is close to 1 or when the relative variance of the fixed effects η_i gets large. In case $T = 3$, the GMM estimator $\hat{\alpha}$ reduces to a simple Instrumental Variable (IV) estimator with y_{i1} as instrument for Δy_{i2} with the corresponding reduced form equation

$$\Delta y_{i2} = \pi y_{i1} + r_i \text{ for } i = 1, \ldots, N. \tag{7}$$

Notice that (6) implies

$$\Delta y_{i2} = (\alpha - 1) y_{i1} + \eta_i + v_{i2} \text{ for } i = 1, \ldots, N. \tag{8}$$

Since η_i and y_{i1} are correlated, $\alpha - 1$ will be biased upward, causing the reduced form coefficient to be biased towards zero, such that the instrument y_{i1} is only weakly correlated with Δy_{i2}. One can show that under the assumption of stationarity and denoting $var(\eta_i) = \sigma_\eta^2$ and $var(v_{it}) = \sigma_v^2$, the plim of $\hat{\pi} \to 0$ as $\alpha \to 1$ or as $(\sigma_\eta^2/\sigma_v^2) \to \infty$ (Blundell and Bond (1998)).

Imposing Restrictions on the Initial Conditions

One way of gaining precision of the GMM estimator in finite samples and asymptotically, is to use additional nonlinear moment conditions. A number of these conditions can be found in Blundell and Bond (1998). Another way of improving precision, is to impose some mild conditions on the initial conditions and using the $T - 3$ linear moment conditions

$$E(u_{it} \Delta y_{i,t-1}) = 0 \text{ for } t = 4, 5, \ldots, T \tag{9}$$

This amounts to using the lagged differences of y_{it} as instruments for the equations in levels. Since I observe Δy_{i2}, I can use the additional restriction

$$E(u_{i3} \Delta y_{i2}) = 0. \tag{10}$$

This equation can be written as $E[(\eta_i + v_{i3})(y_{i2} - y_{i1})] = E[(\eta_i + v_{i3})(v_{i2} + (\alpha - 1)u_{i1})] = 0$ which requires restrictions on the initial condition y_{i1}. When y_{i1} is expressed as $y_{i1} = \frac{\eta_i}{1-\alpha} + u_{i1}$, (10) is equivalent to $E[(\eta_i + v_{i3})(v_{i2} + (\alpha - 1)u_{i1})] = 0$, and necessary conditions for (10) to hold are then $E(u_{i1}\eta_i) = E(u_{i1}v_{i3}) = 0$ for $i = 1, \ldots, N$, so that the key requirement is that the deviations of the initial conditions from $\eta_i/(1-\alpha)$ are uncorrelated with the level of $\eta_i/(1-\alpha)$ itself (Blundell and Bond (1998)). When this requirement is satisfied, Δy_{i2} stays informative as an IV estimator for y_{i2} as α increases towards 1.

Comparisons of the asymptotic variance of the GMM estimator using (9) and (10) with that of the standard first differenced GMM estimator and the nonlinear GMM estimator, shows that large efficiency gains are obtained when using the former and T is small.

4 Evaluation Methods

In the previous section, a problem related to the estimation of dynamic panel data was discussed. In this section, the econometric approaches to evaluation methods will be explored, also in the context of panel data. The evaluation problem central in this discussion is the measurement of the impact of a program, e.g. a labor market program, on each type of individual. The difficult point is constructing the right counterfactual for assessing the impact of a particular treatment and therefore, the evaluation problem can be seen as a missing data problem. Blundell and Costa Dias (1998) and Blundell and MaCurdy (1998) give a profound discussion on these subjects.

Social Experiments and Difference-in-Differences

The evaluation problem arises because one is unable to observe the outcome variable for participants in a particular program had they not participated and the same goes the other way round for members of the control group. In appropriately defined social experiments, the measurement problem can be overcome by randomly assigning individuals out of a particular group to the treatment. However, often, experimental data are not available and even when they are, side effects occur like people dropping out in a nonrandom way or a change in the behavior of the participants due to external factors or caused by the experiment itself. Other disadvantages of experiments are that they are difficult to extrapolate; they might be expensive to administer and the ethical approval might be doubtful — can you deny someone a promising new treatment which will likely cure him from a life threatening disease?

On the other hand, social experiments have some clear advantages, like a minimization of statistical assumptions and that the answers can be easily understood by non-economists. The counterpart of social experiments are natural experiments, where the experimental and control group are put together in a natural way. One example is the length of the time someone enjoys schooling, which is dependent on the year and month of birth. Another is an enquiry into the influence of minimum wage laws; the different states in the US introduced different minimum wage laws, whereas they are subject to the same macro economical influences.

One way to measure the impact of a treatment in the setting of a natural experiment, is using the difference in difference (DID) estimator. To apply this estimator, longitudinal or repeated cross section data are needed, with at least one wave before and one wave after the program change. Assume that treatment takes place in period k, then the outcome equation can be written as

$$\begin{aligned} Y_{it} &= X_{it}\beta + d_i\alpha + U_{it} \quad \text{if} \quad t > k \\ Y_{it} &= X_{it}\beta + U_{it} \quad \text{if} \quad t \leq k: \end{aligned} \tag{11}$$

In this equation, d_i is a dummy which takes on the value 1 if individual i participates in the program and 0 if otherwise. α measures the impact of treatment; the parameters β define the relationship between the exogeneous variables X and the dependent variable Y. U_{it} is an error term of mean zero which is assumed to be uncorrelated with X (Blundell and Costa Dias(1998)). When t^0 denotes the pre-program period and t^1 the after program period, the DID estimator can be written as

$$\hat{\alpha}_{DID} = (\bar{Y}^T_{t^1} - \bar{Y}^T_{t^0}) - (\bar{Y}^C_{t^1} - \bar{Y}^C_{t^0});$$

where $\bar{Y}^T$ and $\bar{Y}^C$ are the mean outcomes for the treatment and comparison groups, respectively.

However, for the DID estimator to yield consistent estimates of the treatment effect, two assumptions have to be made. The first is that the composition of groups needs to be time invariant and the second is, that macro effects must affect the two groups in the same way. If the first assumption is violated, differencing does not eliminate averages of the individual effects in both the experimental as the control group. Violation of the second assumption — for example by differences in demographic composition of the experimental and control group — may contaminate the estimate of α.

The solution put forward by Blundell and Costa Dias (1998) to solve the latter problem, is the use of an additional time interval t^* to t^{**} over which a similiar macro trend has occured. The resulting differentially adjusted DID estimator can be written as

$$\hat{\alpha}_{DADID} = [(\bar{Y}^T_{t^1} - \bar{Y}^T_{t^0}) - (\bar{Y}^C_{t^1} - \bar{Y}^C_{t^0})] - [(\bar{Y}^T_{t^{**}} - \bar{Y}^T_{t^*}) - (\bar{Y}^C_{t^{**}} - \bar{Y}^C_{t^*})];$$

5 Consumption Growth

The life-cycle model plays an important role in understanding consumer behaviour. Within the life-cycle framework, it is assumed that consumers maximize their expected discounted sum of period-specific utilities conditional on the information set at time t. The life-cycle hypothesis implies that households will allocate consumption expenditures in such a way that the marginal utility of wealth λ_t stays constant over time. This variable is unobservable in practice so that the consequences of imposing the life- cycle hypothesis can only be inferred from observing expenditures on individual goods. Usually, this is done by specifying a parametric utility function and then to derive the Euler equation that governs individual expenditures (Blundell, Browning and Meghir (1994)).

Choosing a utility function with constant relative risk aversion leads to a constant intertemporal substitution elasticity (ISE)[3] .

[3] The ISE is defined as $\Phi_t = \partial \ln C_t / \partial \ln(1 + i_t)$, that is, the percentage change in consumption in period t when the real price level increases with one per cent in period t (Blundell, Browning and Meghir (1994)).

Estimation of the ISE is done by applying GMM. I use the orthogonality conditions $E(\varepsilon_{it+1}|\Omega_{it})$, where the ε_{it}'s are residuals from regressing the first difference of the logarithm of consumption for individual i at time t on the real interest rate and income for individual i at time t (with $t = 1, \ldots, T$ and $i = 1, \ldots, N$). Ω_{it} is the information set available to individual i in period t. For estimation purposes, I use the observable instruments $z_{it} = \{r_{it}, r_{it-1}, \ldots, x_{it}, x_{it-1}, \ldots\}$, with $z_{it} \subset \Omega_{it}$ and where r and x denote the nominal interest rates and income, respectively. With the help of these instruments, the GMM estimator of the ISE can be obtained.

Naturally, you may worry about measurement erors in the logarithm of consumption, which may urge you to lag instruments a further period, thereby running the risk that the problem of weak instruments pops up.

Estimation with Cross-Section, Panel and Pseudo Panel Data

However, in most cases, income shocks will be correlated with current consumption growth such that $E(v_{it}|\Omega_{it}) \neq 0$, where the v_{it}'s denote the residuals after regressing the first difference of the logarithm of consumption on the instrumental variables z_{it}. This is for example the case if the income process is modelled as $x_{it+1} = \alpha_i m_{t+1} + \xi_{it+1}$, with m_{it+1} a macro economic shock and ξ_{it+1} idiosyncratic risk. In this case $cov_i(v_{it+1}, z_{it}) \neq 0$ and one needs to average over a long time series of macro economic shocks.

This long time series can be obtained in two different ways. Firstly, when available, one can use long panels. That means, a series of observations on the same observational units for a substantial amount of years. Drawbacks of these data are that the measurement error is typically large and that in most cases, only observations on food are tracked for a long time.

An alternative is to use time series of repeated cross-sections. From these data pseudo-panel can be constructed by grouping individuals into cohorts, e.g. on basis of their date of birth. This has the convenience of having a long time series and maintaining the 'life-cycle' evolution of cohorts and information. On the other hand, disadvantages are that by aggregating over cohorts, some micro variation is smoothed out and that the assumption is made that there is no systematic entry or exit from the cohort.

Blundell, Browning and Meghir (1994) come with the help of such a series of repeated cross-sections to the conclusion that household characteristics, like the number of children, are of great importance in explaining the growth of consumption over a households' life-cycle and that controlling for these characteristics is sufficient to eliminate excess sensibility of consumption growth to predicted income growth.

References

Banks, J. W., R. W. Blundell and A. Lewbel (1997), “Quadratic Engel curves, indirect tax reform and welfare”, Review of Economics and Statistics, 79, 527–539.

Blundell, R.W., M. Browning and C. Meghir (1994), “Consumer demand and the life-cycle allocation of household expenditures”, Review of Economic Studies, 61, 57–80.

Blundell, R. W. and A. Duncan (1998), “Kernel regression in empirical microeconomics”, Journal of Human Resources, 33, 63–87.

Blundell, R.W. and S. Bond (1998), “Initial conditions and moments restrictions in dynamic panel data models”, Journal of Econometrics, 87, 115–143.

Blundell, R.W. and M. Costa Dias (1998), “Evaluation methods for non-experi-mental data”, unpublished manuscript.

Blundell, R.W. and T. MaCurdy (1998), “Labor supply: A review of alternative approaches”, in O. Ashenfelter and D. Card, editors, Handbook of Labor Economics, North Holland, 1559–1695.

Greene, W. H. (1999), Econometric Analysis, 4th edition, Prentice Hall, New Jersey.

Härdle, W. (1990), Applied Nonparametric Regression, Econometric Society Monographs No. 19, Cambridge University Press.