Journal of Econometrics 46 (1990) 333–346. North-Holland

ESTIMATION OF TIME-DEPENDENT PARAMETERS IN LINEAR MODELS USING CROSS-SECTIONS, PANELS, OR BOTH

Theo NIJMAN and Marno VERBEEK*

Tilburg University, 5000 LE Tilburg, The Netherlands

Received March 1988, final version received June 1989

In this paper we consider the estimation of time-dependent parameters in linear models from panel data, cross-sections, or both. We determine the fraction of individuals that should be reinterviewed each period in order to minimize the variance of the most efficient estimator of linear combinations of the parameters. Moreover we derive simple sufficient conditions for the optimal fraction to be zero or one, respectively.

1. Introduction

In recent years economists often have financial resources at their disposal to have data collected. In this paper we analyse how to spend this money efficiently if the aim is, e.g., to monitor average expenditures on some consumption categories by either interviewing the same individuals in several periods or interviewing different individuals in different periods or a combination of these two approaches. The first approach yields a data set known as a panel, while the second approach gives a series of cross-sections. As is well known [see, e.g., Hsiao (1985)] a panel data set has several advantages compared to a series of cross-sections: if a panel is available additional parameters might be identifiable, omitted variable bias might be reduced, and errors-in-variables models might be estimable without recourse to external instruments [Griliches and Hausman (1986)]. On the other hand, it has recently been shown in the econometric literature that panel data are not indispensible for the identification of parameters in a wide class of models [see, e.g., Deaton (1985), Moffitt (1988), and Heckman and Robb (1985a,b)]. Relatively little attention however seems to have been paid to the analysis of

*We thank Rob Alessie, Arie Kapteyn, Arthur van Soest, Tom Wansbeek, an anonymous referee, and participants of seminars at the INSEE and at the University of Paris–Val de Marne for helpful comments and suggestions. Financial support by the Royal Dutch Academy of Arts and Sciences (K.N.A.W.) and by the Netherlands Organisation for Scientific Research (N.W.O.) is gratefully acknowledged by the first and second author, respectively.

the efficiency of estimators obtained from panels, cross-sections, or a combination of these two data sources, which is the subject of this paper.

In the first part we concentrate on the estimation of linear combinations $\xi'\mu = \sum_{t=1}^{T} \xi_t \mu_t$ of the period means μ_t in the simple analysis of variance model

$$y_{it} = \mu_t + \alpha_i + \varepsilon_{it}, \qquad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{1}$$

where the α_i and ε_{it} are i.i.d. random variables with mean zero and variances σ_α^2 and σ_ε^2, respectively, which are mutually independent of the unknown constants μ_t. Subsequently, we discuss extensions to the analysis of covariance model

$$y_{it} = \mu_t + \beta_t x_{it} + \alpha_i + \varepsilon_{it}, \tag{2}$$

where the x_{it} are observed and independent of α_i and ε_{it} and extensions to (1) or (2) with linear restrictions on the time-dependent parameters. Throughout this paper we assume for simplicity that the parameters σ_α^2 and σ_ε^2 are known *a priori*. If these parameters are unknown and replaced by consistent estimates the same results hold true asymptotically.

Let η denote the relative cost of interviewing T different individuals in T periods compared to interviewing the same individual T times. The value of η of course depends on T and on the problem under consideration, but experts suggest that it will usually be slightly larger than one. The only formal analysis of η in the literature we are aware of is presented by Duncan, Juster, and Morgan (1987) who suggest that the field costs of a cross-sectional survey comparable to an additional wave of the Panel Study of Income Dynamics (PSID) are 30% to 70% higher than the costs of such an additional wave, yielding $0.7 + 0.3T < \eta < 0.3 + 0.7T$. We show in section 2 that a pure panel will yield the most efficient estimate of any linear combination of the period means in (1) if $\eta > 1 + (T-1)\rho$ with $\rho = \sigma_\alpha^2(\sigma_\alpha^2 + \sigma_\varepsilon^2)^{-1}$, while the same holds true for a series of cross-sections if $\eta < 1 - \rho$. If one is estimating changes in means the condition for optimality of panels can be relaxed to $\eta > 1 - \rho$, while in case of an estimate of the average mean cross-sections are already optimal if $\eta < 1 + (T-1)\rho$. Analytical and numerical results are presented for cases in which neither of these conditions is satisfied. In section 3 a numerical illustration is given using Dutch consumer expenditure data. Extensions to the analysis of covariance model (2) are provided in section 4, while section 5 contains some concluding remarks. Note that we do not assume any *a priori* knowledge on the parameters in (1) or (2). If such knowledge would be available, Stein-rule estimators or estimators based on an assumed hyperdistribution would be superior to the estimators to be considered below.

2. Analytical and numerical results for the analysis of variance model

Denote the maximum sample size per period, given the available funds, if different individuals are interviewed each period by N and the fraction of the funds used to collect panel data by λ, which implies that $\lambda\eta N$ individuals will be reinterviewed every period, while the remaining $(1-\lambda)N$ individuals will be replaced each period. The analysis of this type of data is advocated, e.g., by Kish (1986) who refers to it as a split panel design (SPD). We will determine the optimal value of λ as a function of the relative cost of the two types of data sets η, the importance of the individual effect ρ and the linear combination $\xi'\mu$ of the μ_t's one is interested in. A similar analysis of the choice between pre-experimental observations and control groups in social experimentation has been presented by Aigner and Balestra (1988).

In order to determine the optimal fraction λ of the budget to be spent on the collection of panel data, we will first derive the efficient estimator and its variance for given values of λ. It is well known [see, e.g., Hsiao (1986, p. 34 ff.)] that the best linear unbiased estimator of $\mu=(\mu_1,\dots,\mu_T)'$ in (1), using only the information on individuals which are reinterviewed every period, is the Aitken estimator $\hat{\mu}_p$ (which in this case coincides with the OLS estimator) and that

$$\operatorname{var}(\hat{\mu}_p)=(\lambda\eta N)^{-1}V_p, \tag{3}$$

where $V_p=\sigma_\varepsilon^2 I_T+\sigma_\alpha^2\iota_T\iota_T'$ and ι_T is a T-dimensional column vector of ones. Analogously, the best linear unbiased estimator based on the cross-section information only is the OLS estimator $\hat{\mu}_{cs}$ for which

$$\operatorname{var}(\hat{\mu}_{cs})=((1-\lambda)N)^{-1}V_{cs}, \tag{4}$$

with $V_{cs}=(\sigma_\varepsilon^2+\sigma_\alpha^2)I_T$. Since $\hat{\mu}_p$ and $\hat{\mu}_{cs}$ are independent, the best linear unbiased estimator which uses all available data is given by

$$\hat{\mu}=\left\{\lambda\eta V_p^{-1}+(1-\lambda)V_{cs}^{-1}\right\}^{-1}\left\{\lambda\eta V_p^{-1}\hat{\mu}_p+(1-\lambda)V_{cs}^{-1}\hat{\mu}_{cs}\right\}. \tag{5}$$

It is easily verified that

$$\operatorname{var}(\xi'\hat{\mu})=N^{-1}\xi'\left(V_{cs}^{-1}+\lambda W\right)^{-1}\xi, \tag{6}$$

where $W=\eta V_p^{-1}-V_{cs}^{-1}$. Since V_{cs}^{-1} is positive definite and W is symmetric there exists a nonsingular matrix Q such that $Q'V_{cs}^{-1}Q=I_T$ and $Q'WQ=D$ with D a diagonal matrix containing the eigenvalues d_t of $V_{cs}W$ and Q containing the (suitably normalized) eigenvectors of $V_{cs}W$ [see, e.g., Gantmacher (1959, p. 313 ff.)]. Therefore the variance of $\xi'\mu$ can be written as

$$\operatorname{var}(\xi'\hat{\mu}) = N^{-1}\delta'[I_T + \lambda D]^{-1}\delta = N^{-1}\sum_{t=1}^{T}\frac{\delta_t^2}{(\lambda d_t + 1)}, \tag{7}$$

with $\delta = Q'\xi$. Straightforward algebra shows that in our case

$$V_{cs}W = (1-\rho)^{-1}\left[(\eta+\rho-1)I_T - \eta\rho\{1+(T-1)\rho\}^{-1}\iota_T\iota_T'\right], \tag{8}$$

with eigenvalues $d_t = \eta(1-\rho)^{-1} - 1 = d$ $(t = 1,\dots,T-1)$ and $d_T = \eta[1+(T-1)\rho]^{-1} - 1$. Using the equality of the first $T-1$ eigenvalues we obtain

$$\begin{aligned}\operatorname{var}(\xi'\hat{\mu}) &= N^{-1}\sum_{t=1}^{T-1}\frac{\delta_t^2}{\lambda d_t + 1} + N^{-1}\frac{\delta_T^2}{\lambda d_T + 1}\\ &= N^{-1}\delta'\delta\left\{\frac{1-\omega}{\lambda d + 1} + \frac{\omega}{\lambda d_T + 1}\right\},\end{aligned} \tag{9}$$

with $\omega = \delta_T^2(\delta'\delta)^{-1}$. Because $V_{cs} = (\sigma_\alpha^2 + \sigma_\varepsilon^2)I_T = QQ'$ and $\sqrt{(\sigma_\alpha^2 + \sigma_\varepsilon^2)}\,\iota_T/\sqrt{T}$ is the eigenvector of $V_{cs}W$ associated with d_T, (9) can finally be rewritten as

$$\operatorname{var}(\xi'\hat{\mu}) = N^{-1}(\sigma_\alpha^2 + \sigma_\varepsilon^2)\xi'\xi\left\{\frac{1-\omega}{\lambda d + 1} + \frac{\omega}{\lambda d_T + 1}\right\}, \tag{10}$$

with $\omega = T^{-1}(\xi'\iota_T)^2/\xi'\xi$.

Eq. (10) shows how the variance of the best linear unbiased estimator of $\xi'\mu$ depends on the fraction of the budget spent on reinterviews. For the special case where $T = 2$ and $\eta = 1$, it can be easily checked from this expression that $\hat{\mu}_1 + \hat{\mu}_2$ has smallest variance if $\lambda = 0$ (pure cross-sections), that $\mu_2 - \mu_1$ is estimated most efficiently if $\lambda = 1$ (pure panel), while for estimating μ_1 or μ_2 the intermediate value $\lambda = 1 - (1 + \sqrt{1-\rho^2})^{-1}$ is optimal, which are well-known results in the literature [see, e.g., Raj (1968), p. 157) or Cochran (1977, p. 347)].

Eq. (10) however generates more general results. The variance of $\xi'\hat{\mu}$ will be minimized at $\lambda = 1$ if $d_t > 0$ $(t = 1,\dots,T)$, irrespective of ξ. The smallest eigenvalue of $V_{cs}W$ is $d_T = \eta[1+(T-1)\rho]^{-1} - 1$ which implies that a pure panel will yield the most efficient estimate of any linear combination of the period means if $\eta > 1 + (T-1)\rho$. The same holds true for pure cross-sections if $d_t < 0$ $(t = 1,\dots,T)$, or $\eta < 1 - \rho$. If the relative cost of interviewing T different individuals compared to interviewing the same individual T times satisfies $1 - \rho < \eta < 1 + (T-1)\rho$, the optimal sample design will depend on

Table 1

Values of λ for which the variance of the efficient estimator of the period mean is minimized and relative efficiency compared to pure panel or cross sections if $\eta = 1$.

	$T=2$		$T=3$		$T=6$		$T=12$	
ρ	λ^*	Rel. eff.	λ^*	Rel. eff.	λ^*	Rel. eff.	λ^*	Rel. eff.
0.3	0.49	0.98	0.51	0.96	0.57	0.93	0.63	0.89
0.6	0.44	0.90	0.48	0.84	0.56	0.74	0.64	0.65
0.9	0.30	0.72	0.35	0.59	0.43	0.43	0.51	0.32

the parameter of interest. In general it will be difficult to obtain analytical expressions for the optimal value of λ, λ^*. However, λ^* can easily be determined numerically because it will either be a solution to the quadratic first-order condition for a minimum of (10) or a boundary extremum because $\lambda \in [0,1]$. Note also that a split panel design cannot be optimal if the parameter of interest is a change in means, which implies that $\xi'\iota_T = 0$ or $\omega = 0$. In that case a pure panel will be optimal if $\eta > 1 - \rho$, while the series of cross-sections is preferable if $\eta < 1 - \rho$. The counterpart of this result is the case where the parameter of interest is the overall mean, where ξ is proportional to ι_T implying $\omega = 1$ and hence optimality of the series of cross-sections if $\eta < 1 + (T-1)\rho$.

In order to illustrate the fact that the split panel design might be preferable to pure panel or pure cross-section designs we present the optimal percentage of people reinterviewed each period, 100 λ^*, as a function of ρ and T assuming that $\eta = 1$ and that the aim is to estimate the period means as accurately as possible. Moreover we present in table 1 the relative efficiency of the estimator based on this sample compared to an estimator based on a pure panel or on a pure series of cross-sections (which can easily be seen to yield equally efficient estimators in this case).

Cochran (1977, p. 351) showed that if $\eta = 1$ and individuals are included in the sample for not more than two periods, the percentage of reinterviews which minimizes var($\hat{\mu}_t$) tends to 50% if T increases, irrespective of the value of ρ. As evident from table 1, this result no longer holds in the present model. Replacing half of the sample every period was also found to be optimal if T is large, by Raj (1968, p. 162) who however assumed that $\mathrm{E}u_{i,t}u_{i,t-s}$ with $u_{it} = \alpha_i + \varepsilon_{it}$ is a decreasing function of s ($s > 0$).

3. Estimates of Dutch consumer expenditures

In this section we will briefly consider the implications of the results in the previous sections for the estimation of the monthly consumer expenditures of Dutch households. We use the 342 complete monthly observations in 1985 of

Table 2

Minimum (maximum) relative cost of interviewing different individuals every period η for a panel (series of cross-sections) to yield efficient estimates.

	Change in means	Monthly mean	Quarterly mean	Annual mean	Any linear combination
	$\omega = 0$	$\omega = 1/12$	$\omega = 1/4$	$\omega = 1$	$\omega \in [0,1]$
Food					
Panel if $\eta >$	0.24	7.1	8.9	9.4	9.4
C.S. if $\eta <$	0.24	0.27	0.33	9.4	0.24
Clothing					
Panel if $\eta >$	0.75	1.7	2.6	3.8	3.8
C.S. if $\eta <$	0.75	0.81	0.93	3.8	0.75

the so-called Expenditure Index panel conducted by INTOMART, a private marketing research agency, on two well-defined consumption categories: food expenditures and expenditures on clothing (including shoes, etc.). Precise definitions of these categories are available on request.

The maximum likelihood estimates of $\rho = \sigma_\alpha^2(\sigma_\alpha^2 + \sigma_\varepsilon^2)^{-1}$ in (1) for food and clothing are 0.76 and 0.25 with standard errors 0.005 and 0.002, respectively. These point estimates reflect the fact that food expenditures are relatively stable compared to expenditures on clothing. In model (1) we tested for first-order autocorrelation in the ε_{it}'s with a common autoregressive parameter γ. This model (with $\sigma_\alpha^2 = 0$) was considered in Raj (1968). The Lagrange multiplier test statistic against this alternative can be shown to be equivalent to N times the (noncentered) R^2 of a simple regression (see appendix). The values of this test statistic are 1.44 and 3.52 for food and clothing, respectively, which we do not take as evidence against the null. Unrestricted ML estimation of the covariance matrix of $\alpha_i + \varepsilon_{it}$, assuming only that the observations are independent over individuals, suggests that there is some heteroskedasticity in the data which we have however ignored.

The estimate $\hat{\rho} = 0.76$ for food suggests that the relative cost of interviewing different instead of the same individuals, η, should be smaller than $1 - \hat{\rho} = 0.24$ for a series of twelve monthly cross-sections to yield estimates of any linear combination of the monthly food expenditures that are as accurate as the ones that can be obtained from a panel in which all households are retained for one year. If $\eta > [1 + (T-1)\hat{\rho}] = 9.36$, the panel will be preferable without ambiguity. For clothing these conditions are $\eta < 0.75$ and $\eta > 3.75$, respectively. In section 2 it was shown how these conditions are affected if one restricts attention to linear combinations $\Sigma_{t=1}^{12}\xi_t\mu_t$ with $\Sigma_{t=1}^{12}\xi_t = 0$ (change) or $\xi_1 = \xi_2 = \cdots = \xi_{12}$ (annual average). The numerical results are given in table 2 where we also present the minimum (maximum) value of η for which a panel (series

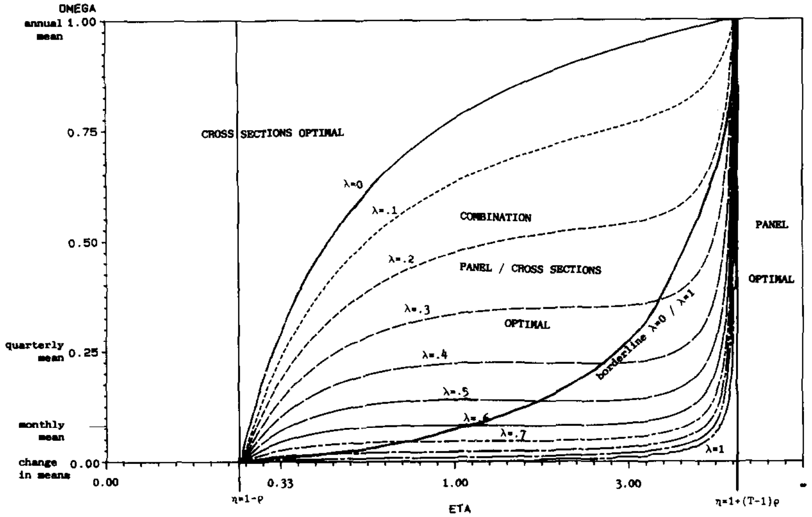

Fig. 1. The optimal panel perunage (λ) for food given relative cost (η) and linear combination of interest (ω).

of cross-sections) will be optimal if the aim is to estimate monthly or quarterly expenditures, respectively. These values can be obtained along the lines described in the previous section.

Because it is evident from (10) that the optimal percentage of households reinterviewed every period, 100 λ^*, depends on T, ρ, η, and $\omega = T^{-1}(\xi'\iota_T)^2/\xi'\xi$ only, an alternative way to present the results in table 2 is to plot the values of λ^* as a function of η and ω if $T = 12$ and $\rho = \hat{\rho}$ as in figs. 1 and 2. The results of table 2 can easily be reconstructed from these figures, and moreover the reader can directly obtain the optimal value of λ for any linear combination of the period means he might be interested in. For comparison we have also indicated for which values of μ and ω a pure panel will be more informative than cross-sections. This is the case if $\eta > 1 - \rho + T\rho\omega$, as can be easily checked from (10). It is unfortunate that the results in figs. 1 and 2 imply that the optimal sample design will strongly depend on the parameter of interest, because one will typically use the same data set for the estimation of many different parameters of interest.

Finally, table 3 contains the optimal value of λ if monthly, quarterly, or annual means or changes in means are to be estimated for three values of the relative cost factor η as well as the relative efficiency of the efficient estimator in case of optimal sample design compared to pure cross-sections or pure panels. It is evident from these results that the optimal design can be substantially more informative than the extreme possibilities.

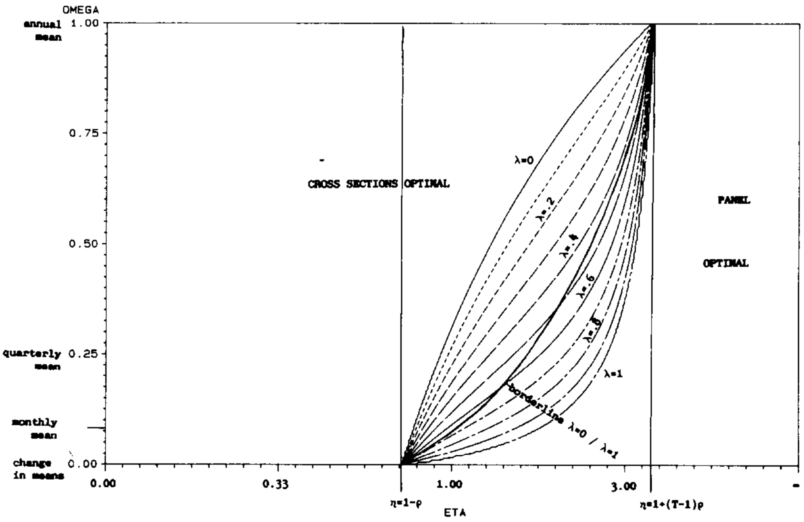

Fig. 2. The optimal panel perunage (λ) for clothing given relative cost (η) and linear combination of interest (ω).

4. Extensions to an analysis of covariance model and to restricted parameters

In this section we first extend the results obtained in section 2 for the analysis of variance model (1) to the analysis of covariance model

$$y_{it} = \mu_t + \beta_t x_{it} + \alpha_i + \varepsilon_{it}, \tag{2}$$

where α_i and ε_{it} are distributed as before and are independent of the observed exogenous variable. Without loss of generality we assume that $\mathrm{E}x_{it} = 0$ ($t = 1,\dots,T$). Subsequently, linear restrictions on the μ_t's and β_t's are incorporated into the analysis and an application to the estimation of marginal budget shares of the consumption categories analysed in section 3 is presented.

As in section 2 we start with the derivation of the variance of the efficient estimator $\hat{\theta}$ of $\theta = (\mu_1,\dots,\mu_T,\beta_1,\dots,\beta_T)'$ given the sample design. Straightforward generalization of (7) implies that, if the eigenvalues of $V_{cs}W = \eta V_{cs}V_p^{-1} - I_{2T}$, with V_{cs} and V_p the variance of $\hat{\theta}$ if only cross-section or only panel observations are used, respectively, are denoted by d_t ($t = 1,\dots,2T$) and the corresponding matrix of eigenvectors is denoted by Q, it holds true that

$$\operatorname{var}(\xi'\hat{\theta}) = N^{-1} \sum_{t=1}^{2T} \delta_t^2(\lambda d_t + 1)^{-1}, \tag{7'}$$

Table 3

Optimal value of λ and relative efficiency in case of optimal sample design versus pure cross-section or panel.

	Change in means			Monthly mean			Quarterly mean			Annual mean		
	$\omega = 0$			$\omega = 1/12$			$\omega = 1/4$			$\omega = 1$		
	$\eta = 0.5$	$\eta = 1$	$\eta = 2$	$\eta = 0.5$	$\eta = 1$	$\eta = 2$	$\eta = 0.5$	$\eta = 1$	$\eta = 2$	$\eta = 0.5$	$\eta = 1$	$\eta = 2$
Food												
Optimal λ	1	1	1	0.57	0.60	0.60	0.30	0.37	0.37	0	0	0
Rel. eff. w.r.t. C.S.	0.48	0.24	0.12	0.75	0.50	0.33	0.92	0.72	0.55	1	1	1
Rel. eff. w.r.t. Panel	1	1	1	0.37	0.50	0.66	0.18	0.29	0.44	0.05	0.11	0.21
Clothing												
Optimal λ	0	1	1	0	0.63	1	0	0.14	0.71	0	0	0
Rel. eff. w.r.t. C.S.	1	0.75	0.38	1	0.91	0.50	1	0.99	0.72	1	1	1
Rel. eff. w.r.t. Panel	0.67	1	1	0.50	0.91	1	0.33	0.66	0.96	0.13	0.27	0.53

with $\delta = Q'\xi$. For model (2) it can be easily checked that

$$V_{cs}W = \left(\frac{\eta}{1-\rho} - 1\right) I_{2T} - \frac{\eta}{1-\rho}\,\frac{\rho}{1+(T-1)\rho}\begin{pmatrix} \iota_T \iota_T' & 0 \\ 0 & \Omega \end{pmatrix}, \tag{11}$$

with $\Omega_{ts} = \mathrm{E}x_{it}x_{is}/\mathrm{E}x_{it}^2$ $(t, s = 1,\ldots,T)$.

The eigenvalues of $V_{cs}W$ are $d_t = \eta(1-\rho)^{-1}\{1 - \rho[1+(T-1)\rho]^{-1}d_{xt}\} - 1$, with $d_{xt} = 0$ $(t = 1,\ldots,T-1)$, $d_{xT} = T$, and d_{xt} $(t = T+1,\ldots,2T)$ the eigenvalues of Ω. Optimality of a series of cross-sections for any linear combination of μ_t's and β_t's requires $d_t < 0$ $(t = 1,\ldots,2T)$ or $\eta < 1 - \rho$ as in section 2, since $0 < d_{xt} < T$. Similarly, a pure panel is preferable without ambiguity if the relative cost factor η satisfies $\eta > 1 + (T-1)\rho$, in which case $d_t > 0$ $(t = 1,\ldots,2T)$. If the attention is restricted to linear combinations of the β_t's only, less stringent conditions can be obtained because $\delta_t = 0$ for $t = 1,\ldots,T$. Therefore, for any linear combination of the β_t's a series of cross-sections will be the optimal design if $d_t < 0$ $(t = T+1,\ldots,2T)$, or

$$\eta < (1-\rho)\left(1 - \rho\frac{d_x^{\min}}{1+(T-1)\rho}\right)^{-1}, \tag{12}$$

while a panel is optimal if

$$\eta > (1-\rho)\left(1 - \rho\frac{d_x^{\max}}{1+(T-1)\rho}\right)^{-1}, \tag{13}$$

where $d_x^{\min}$ and $d_x^{\max}$ denote the smallest and the largest eigenvalue of Ω, respectively. In applications $d_x^{\min}$ and $d_x^{\max}$ can simply be estimated consistently if panel observations on x_{it} are available.

In order to obtain some feeling for these results we consider two special cases. A first special case is the one where x_{it} can be assumed to be generated by the analysis of variance model that was discussed in the previous sections,

$$x_{it} = \mu_{xt} + \alpha_{xi} + \varepsilon_{xit}, \tag{14}$$

where the α_{xi} and ε_{xit} are i.i.d. random variables with mean zero and variances $\sigma_{x\alpha}^2$ and $\sigma_{x\varepsilon}^2$, respectively, which are mutually independent and independent of the unknown constants μ_{xt}, which yields $d_x^{\min} = 1 - \rho_x$ and $d_x^{\max} = 1 + (T-1)\rho_x$, with $\rho_x = \sigma_{x\alpha}^2(\sigma_{x\alpha}^2 + \sigma_{x\varepsilon}^2)^{-1}$. For this special case conditions (12) and (13) reduce to the earlier ones if $\rho_x = 1$. If, on the other hand, there is no individual effect in the exogenous variable ($\rho_x = 0$), either a series of cross-sections or a pure panel will be optimal because the right-hand sides of (12) and (13) coincide. If neither (12) nor (13) holds, the optimal value of λ can be obtained along the lines sketched in section 2.

In the second special case that we consider we only assume

$$\mathrm{E}x_{it}x_{is} \geq 0, \qquad s, t = 1, \ldots, T. \tag{15}$$

This condition appears to be satisfied for many economic variables. If (15) holds, the eigenvalues of Ω, f_t $(t = 1, \ldots, T)$, satisfy

$$0 \leq f_t \leq 1 + (T-1) \max_{t \neq s} \Omega_{ts}, \tag{16}$$

because every element of Ω is nonnegative and the right-hand side is the largest eigenvalue of the matrix with diagonal elements equal to 1 and off-diagonal elements equal to $\max_{t \neq s} \Omega_{ts}$, which bounds every element of Ω. Using (16), it is straightforward to check that if the attention is restricted to linear combinations of the β_t's sufficient conditions for optimality of pure panels and series of cross-sections are

$$\eta > (1-\rho)\left(1 - \rho \frac{1 + (T-1) \max_{t \neq s} (\Omega_{ts})}{1 + (T-1)\rho}\right)^{-1} \tag{17}$$

and

$$\eta < 1 - \rho, \tag{18}$$

respectively.

In applied work often *a priori* restrictions on the parameters in (2) such as $\beta_1 = \cdots = \beta_T = \beta$ will be imposed. If the restrictions are linear such that $\varphi = R'\theta$ is the new set of parameters, the eigenvalues of $\eta R' V_{cs} R (R' V_p R)^{-1} - I$ can be used instead of the eigenvalues of $\eta V_{cs} V_p^{-1} - I$ to obtain sufficient conditions for a pure panel or a series of cross-sections to be optimal. Because the minimal eigenvalue of the first matrix is not smaller than the minimal eigenvalue of the latter and analogously the maximal eigenvalue of the first matrix is bounded by the maximal eigenvalue of the latter, the sufficient conditions for optimality of pure panels or cross-sections obtained above will still be sufficient in case of linear restrictions on the parameters.

In order to illustrate the results above, we consider the estimation of the marginal budget shares of the consumption categories food and clothing assuming that (2) is valid where y_{it} denotes the expenditures on one of the two consumption categories and x_{it} denotes total monthly expenditures on nondurables. The model can be motivated by a two-stage budgeting argument where the total expenditures on nondurables in every month are determined prior to the decision on how to split them over the various categories. The maximum likelihood estimates of ρ in (2) for food and clothing are 0.74 and

0.16 with standard errors 0.005 and 0.001, respectively. The LM test statistics against the hypothesis of first-order autocorrelation in the ε_{it}'s introduced in section 3 equal 0.41 and 3.73 respectively. If (14) is imposed on the expenditures on non-durables the ML estimate of ρ_x is 0.41 with standard error 0.003. The LM test statistic against first order autocorrelation in ε_{xit} in (14) takes the insignificant value of 0.10.

Using (12) and (13), the estimates of ρ and ρ_x suggest that a pure panel will be optimal for any linear combination of the marginal budget shares if $\eta > 0.47$ for food and $\eta > 1.23$ for clothing. Cross-sections are optimal if $\eta < 0.27$ and $\eta < 0.87$, respectively. Note that these results imply that, if the aim is to analyse marginal budget shares, the optimal design does not depend on the linear combination of the parameters one is interested in for realistic values of η, contrary to the results on period means in section 3. The largest unrestricted ML estimate of Ω_{ts} $(t \neq s)$ is 0.93. If this value is used the conditions for optimality of a pure panel will change into $\eta > 2.86$ and $\eta > 2.41$ for food and clothing, respectively. Note however that (16) yields only a rough bound of the largest eigenvalue of Ω. If this eigenvalue is estimated directly, the lower bounds for optimality will reduce to 0.51 and 1.29, respectively, which again imply optimality of a panel design. Using the minimal eigenvalue to obtain upper bounds of η for a series of cross-sections to be optimal yields $\eta < 0.26$ and $\eta < 0.89$.

5. Summary

In this paper we derived a number of simple conditions which can be used to assess whether a panel or a series of cross-sections or a combination of both will yield most efficient estimates of some linear combination of time-dependent parameters in a linear model. These results can be generalized in a straightforward manner to other models. Similar results for the optimal design of rotating panels have been obtained by Nijman, Verbeek, and van Soest (1988).

In the empirical analysis it was shown that if one is estimating period means, the type of data which is preferable will often strongly depend on the linear combination of the time means to be estimated. Only if the relative costs of a panel data set compared to a series of cross-sections are fairly high or fairly low, one of these data sets will be unambiguously preferable. If an exogenous variable with a relatively small individual effect, such as total expenditures on nondurables, is included in the model and attention is restricted to the effects of this variable, the optimal design will be somewhat simpler to obtain. In many cases a panel will be preferable for likely values of the relative cost parameter, irrespective of the linear combination of regression parameters one is interested in.

Appendix: Derivation of the LM test statistic

The Lagrange multiplier test against first-order autocorrelation in the ε_{it} in (1) is a test against the alternative

$$y_{it} = \mu_t + \alpha_i + u_{it}, \qquad u_{it} = \gamma u_{i,t-1} + \varepsilon_{it},$$

where

$$V\{\alpha_i \iota_T + u_i\} = \Omega = \sigma_\varepsilon^2 \begin{pmatrix} 1 & \gamma & \gamma^2 & \cdots & \gamma^{T-1} \\ \gamma & 1 & \gamma & \cdots & \vdots \\ \gamma^{T-1} & & & \cdots \gamma & 1 \end{pmatrix} + \sigma_\alpha^2 \iota_T \iota_T'.$$

The null hypothesis is H_0: $\gamma = 0$ and the loglikelihood is given by

$$L = \sum_{i=1}^N L_i = \text{constant} - \tfrac{1}{2} \sum_{i=1}^N \log|\Omega| - \tfrac{1}{2} \sum_{i=1}^N (y_i - \mu)'\Omega^{-1}(y_i - \mu).$$

Let

$$\psi' = (\psi_1, \psi_2, \psi_3) = (\sigma_\varepsilon^2, \sigma_\alpha^2, \gamma),$$

then

$$\frac{\partial L_i}{\partial \psi_k} = -\tfrac{1}{2} \sum_{t=1}^T \sum_{s=1}^T \frac{\partial \omega_{ts}}{\partial \psi_k} \frac{\partial \log|\Omega|}{\partial \omega_{ts}} - \tfrac{1}{2} \sum_{t=1}^T \sum_{s=1}^T \frac{\partial \omega^{ts}}{\partial \psi_k} (y_{it} - \mu_t)(y_{is} - \mu_s),$$

where ω_{ts} and ω^{ts} are the (t, s)-elements of Ω and Ω^{-1}, respectively. Using

$$\frac{\partial \log|\Omega|}{\partial \omega_{ts}} = \omega^{ts} \quad \text{and} \quad \frac{\partial \Omega^{-1}}{\partial \psi_k} = -\Omega^{-1} \frac{\partial \Omega}{\partial \psi_k} \Omega^{-1},$$

we can write

$$\frac{\partial L_i}{\partial \psi_k} = -\tfrac{1}{2} \operatorname{trace}\left(\frac{\partial \Omega}{\partial \psi_k} \Omega^{-1} \right) + \tfrac{1}{2}(y_i - \mu)'\Omega^{-1} \frac{\partial \Omega}{\partial \psi_k} \Omega^{-1}(y_i - \mu),$$

which is straightforward to compute under the null. Since the Fisher information matrix is block-diagonal with respect to ψ and μ, the LM test statistic for $\gamma = 0$ can be written as [see, e.g., Engle (1984)]

$$\xi_{LM} = \left(\Sigma \frac{\partial L_i}{\partial \psi'}\right)\left(\Sigma \frac{\partial L_i}{\partial \psi}\frac{\partial L_i}{\partial \psi'}\right)^{-1}\left(\Sigma \frac{\partial L_i}{\partial \psi}\right),$$

to be evaluated under H_0. Consequently ξ_{LM} can be calculated as N times the noncentered R^2 of a regression of ι_N on $\partial L_i / \partial \psi_k$ $(k = 1,2,3)$. As is well known, under the null hypothesis ξ_{LM} converges in distribution to a central χ^2 distribution with one degree of freedom. The test against autocorrelation in (2) can be derived along similar lines.

References

Aigner, D.J. and P. Balestra, 1988, Optimal experimental design for error components models, Econometrica 56, 955–972.

Cochran, W., 1977, Sampling techniques, 3rd ed. (Wiley, New York, NY).

Deaton, A., 1985, Panel data from time series of cross sections, Journal of Econometrics 30, 109–126.

Duncan, G.J., F.T. Juster, and J.N. Morgan, 1987, The role of panel studies in research on economic behavior, Transportation Research 21 A, 249–263.

Engle, R.F., 1984, Wald, likelihood ratio and Lagrange multiplier tests in econometrics, in: Z. Griliches and M.D. Intriligator, eds., Handbook of econometrics, Vol. 2 (North-Holland, Amsterdam) 775–826.

Gantmacher, F., 1959, The theory of matrices, Vol. 1 (Chelsea Publishing Comany, New York, NY).

Griliches, Z. and J.A. Hausman, 1986, Errors in variables in panel data, Journal of Econometrics 31, 93–118.

Heckman, J. and R. Robb, 1985a, Alternative methods for evaluating the impact of interventions, in: J. Heckman and B. Singer, eds., Longitudinal analysis of labor market data (Cambridge University Press, Cambridge) 156–245.

Heckman, J. and R. Robb, 1985b, Alternative methods for evaluating the impact of interventions: An overview, Journal of Econometrics 30, 239–267.

Hsiao, C., 1985, Benefits and limitations of panel data, Econometric Reviews 4, 121–174.

Hsiao, C., 1986, Analysis of panel data (Cambridge University Press, Cambridge).

Kish, L., 1986, Timing of surveys for public policy, Australian Journal of Statistics 28, 1–12.

Moffitt, R., 1988, Estimating dynamic models with a time series of repeated cross sections, Mimeo. (Brown University, Providence, RI).

Nijman, Th.E., M. Verbeek, and A. van Soest, 1988, The optimal design of rotating panels in a simple analysis of variance model, Research memorandum FEW 318 (Tilburg University, Tilburg).

Raj, D., 1968, Sampling theory (McGraw-Hill, New York, NY).