A USER'S INTRODUCTION TO ECONOMETRIC TECHNIQUES FOR ANALYZING PANEL DATA

Johny K. Johansson*

#170

College of Commerce and Business Administration
University of Illinois at Urbana-Champaign
A USER'S INTRODUCTION TO ECONOMETRIC TECHNOLOGIES FOR ANALYZING PANEL DATA

Johny K. Johansson*

#170

*Assistant Professor of Business Administration, University of Illinois, Urbana, Illinois 61801.
Using panel data on brand purchases and prices, various statistical models for the estimation of the impact of price upon purchases are presented. It is shown how each model emerges from the preceding one as more demand for variety across time and individuals is made on the basis of theoretical arguments and empirical findings. The basic steps in applications of the algorithms are indicated.
Introduction

Panel data generally provide brand purchase information for a large number of individuals over several months, and they sometimes also include individuals' ratings of the brand on some attitude scales, or the exposure to the different brands' advertising. Such disaggregated data makes several types of analysis possible -- one can identify over-time patterns, or similarities between individuals, or regularities in brand purchase, for example -- and there have been several recent developments of statistical algorithms designed to utilize the information available as efficiently as possible. This paper surveys these new techniques, focusing especially upon the considerations that help the data analyst to choose one technique over another.

Preliminaries

In what follows we assume that the data available are of the following nature. A panel of 300 families have been keeping track of their weekly purchases for a two year period, indicating in their diaries how much was bought and at what price of the various products and brands.

We assume that the objective of the data analysis is to assess the role of price in individual's purchase decisions for one of the product groups represented in the panel diaries, coffee, say. As we are interested mainly in "our" brand, the modeling is focused upon determining price effects for that brand of coffee.
Next, the dependent and independent variables measurements will have to be settled on. Cutting a long discussion short, it is assumed here that a three week period is found sufficient for coffee re-purchasing and that the price and deal effects are played out within that time span. Thus, it becomes acceptable to define a dependent variable for each individual per every three weeks which is scored a "1" when our brand is purchased, a "0" otherwise. The quantity bought is disregarded. Similarly, the price variable is measured as our brand’s price to the individual relative to the mean market price for all purchasers during that three week period. Deals are translated into dollar "equivalents" and subtracted from the price before relative prices are computed.

An Aggregate Model

With these definitions and measurements of purchase and price a simple regression model is developed. It relates the relative frequencies of purchases of our brand (our "market share") linearly to our relative price and a constant:

\[ p_{0t} = c_0 + c_1 (r_0/r_t) + u_{0t}, \quad t = 1, \ldots, T, \]

where \( p_{0t} \) denotes our market share at time \( t \), \( c_0 \) is a constant unique to our brand, \( c_1 \) is a slope parameter also unique to our brand, \( r_0 \) is our price and \( r_t \) is mean price for our competitors. \( u_{0t} \) is a disturbance term, made up partly of left out variables. The two unknown parameters \( c_0 \) and \( c_1 \) are to be inferred on the

---

1Thus, a simple "market share" measure for our brand is any three-week time period is the number of 1's relative frequency to the total number of purchasers.
basis of the 34 time periods of observations available. Since $p_{0t}$ is a relative frequency measure, $u_{0t}$ cannot be assumed to have a normal distribution -- accordingly, the model is rewritten as

$$3\text{-}\frac{\text{f}_{0t}}{(1-p_{0t})} = c_0 + a_0\left(r_{0t}/r_t\right) + v_{0t},$$

which can be estimated using ordinary least squares (see Johansson, 1973). This model is no longer linear, but exhibits a diminishing return shape for positive relative prices. In management's judgment this representation is better than the linear formulation, since some customers will always buy regardless of price.

**A Disaggregate Model**

We might feel a little uncomfortable about some of the assumptions involved in the initial model, however. In particular, the assumption that individuals are homogenous with respect to their price sensitivity might be unwarranted. In such a case, one would have to carry out the analysis accounting for such differences as a priori are assumed to exist.

Let us assume that we define an individualized measure of the dependent variable purchase which gives the amount of our brand bought during each eight week period relative to the total amount of coffee bought by the individual during these weeks. This measure would be the individual counterpart of the relative frequency measure used above, and the problem with it would be that even with eight weeks it would tend to cluster around zero and one. But let us disregard that problem here and go
on to the more general estimation problem.

Redefining the relative price of our brand to correspond to
the eight week period, we could run each individual separately
according to

\[
\frac{p_{0it}}{1-p_{0it}} = C_{0i} + A_{0}\left(\frac{r_{0i}}{r_{i}}\right)_{t} + v_{0it}, \quad t = 1, \ldots, 13
\]

\[
i = 1, \ldots, n'
\]

with the subscript \(i\) denoting the individual.\(^1\) The sample
comprises \(n' \leq 300\) individuals, and the time period is equal to
eight weeks. It should be noted that \(p_{0it}\) has to be strictly
less than 1.0, otherwise the left hand side becomes undefined.

Thus, if we have completely brand loyal individuals in the sample,
model (3) is inappropriate -- a version of model (1) might have
to be used however unsound it is statistically.

Some Analysis of Covariance Models

The disaggregated approach will give us 13 observations
per regression equation, a fairly small number considering the
originally available 104 weeks. Accordingly, it might be useful
to consider some other ways in which the individual differences
could be accounted for. A common assumption is to assume that
the difference is mainly in terms of the intercept, the slope
coefficient (and thus the marginal effect of price) being the
same for each individual. This leads to the model

\[
\frac{p_{0it}}{1-p_{0it}} = C_{0i} + A_{0}\left(\frac{r_{0i}}{r_{i}}\right)_{t} + v_{0it}, \quad t = 1, \ldots, 13
\]

\[
i = 1, \ldots, n'
\]

\(^1\) We assume that every panel member included has bought our brand
at least once during the 8 week period.
When running this model to get the estimates of the $C_0$ and the $A_0$ (a total of $n'+1$ unknown parameters) the usual approach is to introduce separate dummy variables for each of the individuals.\(^1\) This is a considerable saving of observations compared to the $2n'$ parameters to be estimated in model (3). Additionally, if we are not interested in the intercepts but only in the slope $A_0$, using deviations of the relative price from their respective individual mean will give us the identical estimate without the $n'$ constants, thus preserving degrees of freedom even more.

It is easily seen from (3) that the choice of recognizing individual differences can easily be changed into allowing differences over time. If it is assumed that the intercepts are fairly constant across some (homogeneous) individuals, but may be changing as time goes by (e.g. because of the product life cycle progression), the same approach as the one used to derive (3) and (4) but working with deviations from over-time means will allow for such changes to surface in the estimation.

Furthermore, it might be desirable to allow for individual effects as well as time effects through differences in intercepts. This would be the case where the purchases vary across individuals but also in some level over time. That is, the individuals differ in a stable manner throughout the time period under observation, while all their purchases are shifted ("on the mean" as it were) from time period to time period. Then the correct model would be

\[
P_{01t}/(1-P_{01t}) = C_0 + A_0(r_{01}/r_0) t + u_{01} + v_{0t} + e_{01t}.
\]

\(^1\) Notice that this approach is commonly termed the analysis of covariance. The intercepts represent "main" effects, while price is "covaried out", in the language of the analysis of covariance. Here we are interested more in the covarying variables' effect.
with \( i \) and \( t \) indexed as in (4), \( c_0 \), representing the overall intercept, and with \( u, v, \) and \( e \) three error terms representing the "individual effect", the "time effect", and the overall error term, respectively. As error term the \( u \) and the \( v \) are random variables, but one can equivalently interpret them as fixed parameters to be estimated as intercepts. With that view the estimation is rather straightforward generalizing, from the earlier approach. We simply work with variables as deviations from the individual means, the time means, and the overall mean.

Thus, if we are not interested in the varying intercepts but only the price coefficient, each observation on the independent price variable (and correspondingly for the dependent variable) is transformed by

\[
(6) \quad \frac{R_{0i}}{R_i} = (\frac{\bar{R}_{0i}}{\bar{R}_i}) + \bar{E}_i - \bar{E}_t + \bar{E}_{it}
\]

where \( \bar{R}_i \) is the mean price at \( i \) over all individuals, \( \bar{E}_i \) is the mean price for individual \( i \) across all time periods, and \( \bar{E}_{it} \) is the overall mean for the whole data set. The reason this overall mean has to be added is that by subtracting the two "sub"-means \( \bar{E}_i \) and \( \bar{E}_t \) the overall mean in fact is subtracted twice also. The estimation simply uses ordinary least squares for the model

\[
(7) \quad \frac{P_{0it}}{(1-P_{0it})} = A_0 (\frac{R_{0i}}{R_i}) + e
\]

For this extension of the analysis of covariance it is necessary that

\[
(8) \quad \sum_{i=1}^{n'} u_{0i} = 0 \quad \text{and} \quad \sum_{t=1}^{T} v_{0t} = 0, \quad \text{which holds since} \quad c_0, \quad \text{is made to represent the overall mean.}
\]
An Error Components Model

In the cases where one would have an interest in the actual magnitudes of the individual and time effects -- perhaps to gain an understanding of how homogeneous the sampled individuals might be, and also to see if great changes in the brand's market share occurs between periods -- it will be necessary to explicitly introduce the dummy variables representing the intercepts in the estimation. Compared to the case depicted in (7) this leads to a considerable drop in the degrees of freedom, with n' + T new parameters to estimate. In addition, one would think that if these variations between time period and individuals are to be taken directly into account, some use of them could also be made in the estimation of the slope coefficient for price.

These two considerations together with others -- do we really know what to do with all those dummy variables? -- have prompted econometricians to develop in recent years other estimation techniques for the "error" or "variance components" models as they are called (see e.g. Nerlove, 1971, and Maddala, 1971). Basically, the thrust of the several alternative methods developed has been to regard the $u_{ci}$ and $v_{ot}$ in (5) as random variables (just as $e_{oit}$, the overall error term is regarded), and then to estimate the parameters necessary to determine the probability distributions of the two error terms. These distributions are generally assumed to be normal, and then a specification of the mean and the variance will suffice to answer the question as to how large the variations between individuals and over time really are.
There are several competing estimation methods for the error components model, each one having some advantage depending upon the exact specifications of the statistical model. The case with which we are dealing here is one of the simpler ones. The regressors are assumed non-stochastic, thereby excluding the case where we introduce a lagged dependent variable on the right hand side for example, and we further assume that

1. $E(u_{0i}) = E(v_{0t}) = E(e_{0it}) = 0$
2. $V(u_{0i}) = \sigma^2_u$
3. $V(v_{0t}) = \sigma^2_v$
4. $V(e_{0it}) = \sigma^2 \mathbf{I}$

where $\mathbf{I}$ is the identity matrix, of order $n'T$. These random variables are furthermore assumed independent and identically distributed, and it is required that $n'>K$ as well as $T>K$, where $K$ stands for the number of independent variables in the model.

Most of these assumptions are direct analogues to the usual multiple regression case. Under these assumptions, the approach suggested by Swamy (1971, Ch. 3) provides the most straightforward solution to the estimation problem. To get at the individual effect, we simply use ordinary least squares to estimate

$$(8) \quad (p_{0it}/(1-p_{0it}))_i = C_{0i} + A_0 (T_0 / \bar{T})_i + u_{0i},$$

where the added i subscript indicates that the variables used are the individual's over time means. Thus, this regression can be estimated using $n'$ observations. As indicated in (8),

---

1 Since we are not interested here in the constants $C_{0i}, i=1, \ldots, n'$; we can furthermore use the deviations from the mean across individuals for the estimation of (8).
the error term does represent the individual effect, and an estimate of its variance $\sigma^2_u$ (its mean is zero by virtue of the ordinary least square fit) can be derived in the usual way from the squared residuals divided by $n^i-K$, summed over the individuals.

In the same manner we can get an estimate of the variance of $v_{0t}$, the time effect, by running a regression using variables measured (again as deviations from the mean) over individuals for each time period. The number of observations here will be $T$, and the residuals can be used to generate an unbiased estimate of the desired parameter. Finally, using overall deviations as indicated in (6) and (7), ordinary least squares will give an estimate of $\sigma^2_u$.

But as indicated, one reason for the use of error components models was the possibility of a more efficient estimate of the slope parameter(s) by better use of the inter-individual, over-time variations. In this instance this efficiency comes about by a pooling of the several estimates of the slope coefficient, $A_0$, that we have generated. As should be clear from the preceding description, the approach will yield one estimate of $A_0$ each time the variance estimation is carried out.

Swamy (1971, p. 70) suggests a simple pooling of the estimates with weights inversely proportional to the variances derived, and shows that this is an Aitken estimator, with the desirable properties of unbiasedness and small (although not necessarily minimum) variance.
A "Seemingly Unrelated" Test Model

So far the discussion has mainly centered on time and individual effects that basically reflect themselves in different intercepts. It is clear, however, that in many cases one will want to allow the price coefficient to vary between individuals and/or over time as well. Reasons for the individual differences lie with the possibility that price sensitivities might vary between people, or between segments. Similarly, time differences might occur because of the product's progress through the life cycle, making price considerations more or less salient.

A general extension of the analysis of covariance use of dummy variables allows the dummy variables to be introduced as interactive with the slope coefficient as well. The procedure amounts to assigning a new variable the value one for "different" individuals, and the value zero for the others. Then a new regressor is created by multiplying this new dummy variable by the price variable, and introducing both this new regressor and the original price variable in the regression model. If more than one different individual or group of individual is wanted, another dummy is created along the same lines, and a third regressor created. A good presentation of the overall approach is given by Gujarati (1970), and for an interesting marketing application of the model the reader is referred to Winter (1973). In the limit, where each individual is seen as different from everyone else, the model simply computes separate regressions for each.
As before, the correspondence to over-time differences follows directly, with the dummy variables now representing different time periods. Also as before, the use of the dummy variables grows increasingly undesirable, as the differences increase, since the degrees of freedom quickly go down. Therefore, a development of a technique assuming random coefficients has recently been undertaken by some econometricians, a parallel development to the error components solution to the analysis of covariance case.

Since random components model are relatively hard to estimate, it is usually desirable to first test whether or not differing slope parameters are in fact desirable. If a priori, theoretical considerations dictate their use, no empirical test is strictly necessary. However, the appropriate empirical test to use is the "seemingly unrelated" model developed by Zellner (1962). Here separate regressions for each individual unit are first run to generate an estimate of the variance-covariance matrix between different individuals, which is used to establish a more efficient estimate of the different slope parameters. Then an approximate F-statistic is generated which tests the deviations of the individual parameter estimates from the overall mean parameter vector (with our single price variable, the vector becomes a scalar), relative to the residual variations "within" the individual over time. A good example of a marketing application of this model is presented by Beckwith (1972).

Most standard econometric and statistical computer packages now offer the "seemingly unrelated" model. It is particularly appropriate, where separate regressions are run for different individuals, but where there is reason to believe that the error term in each equation is related across equations. This would be the case where the factors, other than relative price, that affect market shares would be the same across individuals (e.g. the effect of the introduction of freeze dried coffee).
Whether or not the test indicates that the parameters in fact do differ, one would generally also need to test whether or not they are random. Such a test is described in Swamy (1971, pp. 122-124). It is a likelihood ratio test which requires an iterative estimation procedure which so far is not generally available. Therefore, the tendency among econometricians currently seems to be one of a priori specification of the randomness, and then estimation of the parameters directly.

A Random Coefficient Model

If the test(s) carried out indicates that differences among individual parameters are indeed significant, the use of a random parameter model will be warranted. We will here only describe the simplest such specification, but the reader should be aware that (much as in the case of error component models) several competing alternatives do exist, the appropriateness of which depends upon the particular stochastic nature of the model.

The basic principle underlying random coefficient models is the assumption that the parameters in model (2) are random variables

\[(9) \quad \hat{A}_{0i} = \hat{A}_0 + \hat{d}_{0i}, \quad i = 1, \ldots, n'.\]

with the random increment \(d_{0i}\) added to the overall parameter mean \(\hat{A}_0\). Under the assumptions that the \(d_{0i}\), \(i=1, \ldots, n'\) are identically and independently distributed, with \(E(d_{0i}) = 0\) and \(V(d_{0i}) = \sigma^2\), and the usual assumptions of the multiple regression model, the estimation is relatively straightforward.
First we get separate estimates (via separate runs) of the individual parameters, $\hat{A}_{0i}$, $i=1,...,n'$. An unbiased estimate of $\hat{A}_0$ is derived by taking a weighted estimate of these $n$ estimates, the weights inversely proportional to their variances. Thus, in order to make the procedure operational we have to assess these variances.

To get an estimate of the variance we first need to account for the equation disturbance terms. An estimate of the individual equation variance is as usual given by the residuals:

$$s_{ii} = \frac{(Y_{di} - Y_{pi})^2}{T - K}, \quad i=1,...,n',$$

where $s_{ii}$ is the variance estimate of the equation for the $i$'th individual, $i=1,...,n$, $Y_{di}$ is the actual sample value of the dependent variable (here $p_{0it}/(1-p_{0it})$), and $Y_{pi}$ is the predicted value.

Next we treat the individual parameter estimates $\hat{A}_{0i}$ as a random sample of size $n$ and compute its variance $S_A$. If we assume the parameter variances to be the same for all individuals, an estimate of the population parameter variance would be

$$\hat{V}(d_{0i}) = \hat{V}(\hat{A}_0) = S_A/(n-1).$$

Accounting for the equation disturbances allows us to decrease this variance, however. As Swamy (1971, p. 107) shows, the unbiased estimate of the variance is

$$\hat{V}(\hat{A}_0) = \hat{\sigma}^2 = S_A/(n-1) - \frac{1}{n} \sum_{i=1}^{n} s_{ii} (X_i'X_i)^{-1},$$

where $X_i$ refers to the data on the independent variables (here only relative price) for the $i$'th individual.
The variances of the estimates $\hat{A}_{0i}$, $i=1,\ldots,n'$, are then the sum of the variation in the parameters themselves and the sampling variation:

$$V(\hat{A}_{0i}) = \delta^2 + s_{ii} \langle X_i'X_i \rangle^{-1}, \quad i = 1,\ldots,n',$$

where the sampling variation is computed similarly to the usual regression case.

Finally, we derive the weights of the estimators $\hat{A}_{0i}$, $i=1,\ldots,n'$, from the estimator variance:

$$W(\hat{A}_{0i}) = \frac{\sum_{j=1}^{n} [\hat{\sigma}_j^2 + s_{jj} \langle X_j'X_j \rangle^{-1}]^{-1}} {\sum_{i=1}^{n} \hat{\sigma}_i^2 + s_{ii} \langle X_i'X_i \rangle^{-1} - 1}.$$ 

Accordingly, the estimate of $\hat{A}_0$ becomes

$$\hat{A}_0 = \frac{\sum_{i=1}^{n} \hat{A}_{0i}} {W(\hat{A}_{0i})}, \quad i=1,\ldots,n',$$

which is a minimum variance, unbiased estimator. Its variance is given by

$$V(\hat{A}_0) = \frac{\sum_{i=1}^{n} [\hat{\sigma}_i^2 + s_{ii} \langle X_i'X_i \rangle^{-1}]} {\sum_{i=1}^{n} \hat{\sigma}_i^2 + s_{ii} \langle X_i'X_i \rangle^{-1}},$$

that is, a simple sum of the component variances (since the individual estimates $\hat{A}_{0i}$, $i=1,\ldots,n'$, are considered independent, their covariances are zero).

As before, the approach applies directly to the estimation of parameters where the random fluctuations occur over time rather than across individuals.

**Variable Parameter Regression Models**

It should come as no surprise to the reader that the next developments in panel data analysis involves slope parameters which vary over time as well as across individuals. One development, by Hsiao (1973), expresses the parameters as sums of time effects and individual effects, much in the manner of the covariance
models developed earlier. In another development by Rosenberg (1973), a much more general framework is presented. Here the individual parameter vectors can be either identical across the population (or a subset of it) or differing between individuals, and over time the parameters are seen as either fixed or following a stochastic process. The differences across individuals at any point in time are treated analogous to the Swamy approach just presented. The stochastic process that represents the over time changes allows for many different specifications provided only the processes are stationary. The number of time periods in the sample would generally set an upper limit on the order of the process, however; one would like to have several repeated realizations of the complete process in the data. The actual form of the process can either be imposed a priori from theoretical considerations, or can be estimated from the data directly. In most cases the process would need to exhibit at least two particular features: a random shock term providing the mechanism for inter-individual differences, and a convergence factor, allowing for individual similarities to assert themselves over time. A simple stochastic mechanism that will do this would be

\[
A_{o(t+1)} = A_{ot} + c (A_{uit} - \bar{A}_u) + w_{uit} + i = 1, \ldots, n_i, t = 1, \ldots, T.
\]

Here \(c\) is a coefficient (0 < \(c\) < 1) representing the rate of convergence of the individuals to the population average or "norm" \(\bar{A}_u\), whereas \(w\) depicts the random disturbance (with mean zero) that makes for individual diversity. In the example \(c\) is seen as identical for at least some individuals in the population, a necessary assumption if \(c\) is to be estimated from the data,
rather than specified a priori. In our example, the diversity among individuals might simply arise because of differing price sensitivity among the individuals. The convergence towards some kind of "norm" or average, on the other hand, might be attributable to our brand having established a "norm" market share and sales level, where the role of price is a constant parameter across the buyers.

As one would suspect, estimation in Rosenberg's "convergent parameter" model becomes quite a bit more complicated than in the Swamy approach.1 Basically, the specification of the stochastic parameter process (16) has to take place first. Then, estimates of the individual parameters at any time period are possible to compute, conditional upon the process specified. The method used is a recursive maximum likelihood procedure which starts at time \( t=1 \) and then "updates" the initial estimates through succeeding time periods up through \( T \). Since only one time period is available at \( t=1 \), a "starting problem" exists. The solution is discussed by Rosenberg (1973) at length (if a Bayesian approach is used, these initial values would, of course, be given by an informative or diffused prior) and we will not go into it here. Then these parameter values at \( t=1 \) are used in the stochastic mechanism (16) to predict next period's parameter values. These new parameter values are then used in conjunction with the observations on the independent variables at \( t=2 \) to generate a forecast of the dependent variable for that period. Finally, the difference between this forecasted value and the actual value of the dependent variable

---

1The Rosenberg model will give the Swamy solution as a special case (when the parameter mean stays constant over time, and \( c=0 \) in (16)).
at \( t=2 \) is used to revise the parameter estimate for that period (if this difference is zero, no revision will naturally occur).

For more information on the algorithm, the reader is referred to the Rosenberg (1973) paper. Presently, the computer program is not generally available, but it will be available in the near future. It should be noted that in the Rosenberg approach the explicit values of the different individuals' parameters at different points in time will be available, and it becomes possible to trace the development of one particular individual throughout the sample history. It should perhaps also be noted that these parameters will generally be different for different lengths of the sample history. That is, the revision of the initial estimate on the basis of the forecast error mentioned, can be carried out not only for the present period, but also for earlier periods, back to \( t=1 \).

Extensions

There are many other developments within these main areas that deserve coverage but cannot be encompassed here. The applied researcher when faced with statistical specifications which deviate from the ones treated here should be able to use the references cited to go further and formulate the appropriate method. These cases include the problem of serial correlation in random coefficient models which is well treated by Swamy (1971, pp. 127-131), and error components model with lagged dependent variables introduced on the right hand side, which are discussed by Nerlove (1971).

In our discussion we have consistently focused upon our brand alone. It is clear that an identical approach can serve for the
analysis of any one of the brands in the market. One might, however, use the data across brands even better if the brands were introduced as a dimension equivalent to time and individuals. Then one could derive a "brand" effect in exactly the same manner as we earlier analyzed "time" and "individual" effects. It should be clear that where the number of brands is small, some algorithms will not work. For the random coefficient estimation, for example, it is essential that the number of brands be greater than the number of independent variables in the model. The introduction of across brands variations in the variable parameter model would create computational difficulties (since each time period's iteration would have to account for both inter-individual and brand variation in the parameters) but could probably be done.

Overall, there is every reason to hypothesize that in the next few years this area will see a rapid growth, especially with respect to the availability of efficient computing algorithms. Furthermore, the marketers ought to have much to contribute since in many respects their data bases are the most sophisticated and complete. After all, if we spend all that money collecting the data, we ought also to use the detailed information it contains to the fullest.
References


