Aggregation, Structural Change and Cross Section Estimation

by

Thomas M. Stoker

WP# 1485-83 September 1983
Revised October 1984
This paper characterizes the local effects of parametric behavioral model change on relationships between aggregate variables, and presents consistent estimators of such effects using cross section data. Two equivalent interpretations of model change effects are given: an "average-marginal" formulation and a cross section regression formulation. The relation between model change effects and maximum likelihood estimation of the behavioral parameters is explained. Finally, the question of whether $R^2$ (from a cross section OLS regression) is a general measure of the sensitivity of aggregate relationships to model change effects, is addressed.

KEY WORDS: Model Change Effect; Distribution Effect; Maximum Likelihood Estimation; Exponential Family; OLS Regression
Thomas M. Stoker is Associate Professor of Applied Economics, Alfred P. Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02139. This research was funded by National Science Foundation Grants No. SES-8308768 and SES-8410030. This author wishes to thank James Powell for ongoing discussions and the referees for helpful comments.
AGGREGATION, STRUCTURAL CHANGE AND CROSS SECTION ESTIMATION

1. Introduction

In general, movements in an aggregate (averaged or totaled) dependent variable can arise from two different sources. If the microeconomic behavioral function between the dependent variable and independent variables is stable, then movements in the aggregate dependent variable arise only from changes in the distribution of independent variables. Such influences are referred to as distribution effects. Alternatively, an aggregate dependent variable will vary if the microeconomic behavioral function varies. We refer to these influences as structural model change effects.

Distribution effects on aggregate variables have been characterized in detail by Stoker (1982, 1983a, 1983b). The purpose of this paper is to characterize model change effects. Like the above cited papers, we characterize model change effects locally vis-a-vis their relation to (microeconomic) cross section data estimators.

The questions of interest are usefully introduced by a simple example. Suppose that the purchase of a home by a family is described by a standard discrete choice model as in

\[ y = 1 \quad \text{if } u \leq \gamma_0 + X' \gamma_1 \]
\[ = 0 \quad \text{if } u > \gamma_0 + X' \gamma_1 \]

where \( y = 1 \) denotes purchasing, \( X \) denotes a vector of attributes including family income, demographic and interest rate variables which affect purchase decisions, and \( u \) denotes an unobserved random disturbance. For simplicity, if \( u \) is distributed as a univariate normal variable with mean 0 and variance 1, the probability that a family with attributes \( X \) buys a home is given by the familiar probit model.
\[ E(y|X) = \Phi(T_0 + X'T_1) \]

where \( \Phi \) is the cumulative normal distribution function. The aggregate dependent variable is \( E(y) \), the overall proportion of families buying homes, which is given as

\[ E(y) = \int E(y|X) \, p(X|\mu) \, dX \]
\[ = \int \Phi(T_0 + X'T_1) \, p(X|\mu) \, dX \]
\[ = \Phi(T_0, T_1, \mu) \]

where \( p(X|\mu) \) denotes the density of the marginal distribution of \( X \), and \( \mu \) denotes a vector of parameters which indicate how the marginal distribution can vary, which for concreteness we take as the mean of \( X \), \( \mu = E(X) \).

Now suppose that we observe a cross section data set for a particular time period \( t^o \), namely a random sample \( y_k, X_k, k=1,...,K \), where \( \mu = \mu^o \) at time \( t^o \). The distribution effects are the effects of changing the mean \( \mu \) of \( X \) on \( E(y) \); i.e. \( \partial \Phi/\partial \mu^o \). Stoker (1983a) has shown that these effects are consistently estimated by certain cross section instrumental variables coefficients of \( y_k \) regressed on \( X_k \). For example, when \( p(X|\mu) \) can be expressed as a member of a particular exponential family, the ordinary least squares (OLS) coefficients of \( y_k \) regressed on \( X_k \) consistently estimate \( \partial \Phi/\partial \mu^o \). These results are valid regardless of the true form of the behavioral model describing \( y_k \), here taken as a probit model.

In this paper, we analyze the aggregate effects of varying the behavioral parameters \( T_0 \) and \( T_1 \), namely \( \partial \Phi/\partial T_0 \) and \( \partial \Phi/\partial T_1 \), for a given (cross section) distribution of \( X \). Estimates of these effects measure the impact of parametric changes in the behavioral model on the overall proportion \( E(y) \) of families purchasing homes. These estimates are useful for making aggregate judgements based on varying estimates of \( T_0 \) and \( T_1 \). In a similar spirit, if a particular behavioral parameter displays a large model change effect, but its initial
estimate is imprecise, then additional effort is indicated toward sharpening the estimate.

In broader terms, the first motivation for studying model change effects arises from the study of aggregate data in applied economic research. While aggregate (macroeconomic) relationships are of natural interest to the study of economic time series data, ascertaining the relative importance of distribution effects and model change effects with time series data may be difficult because of identification problems in modelling (Stoker 1984) or data based problems such as multicollinearity. The results of this paper provide methods for obtaining outside information on model change effects from cross section data. Moreover, we point out that in cross section data, the information on model change effects is orthogonal to the information on distribution effects, so that the confluence problems associated with time series data may be avoided.

The second overall motivation arises from the study of cross section estimation of behavioral parameters. We show that model change effects arise naturally in maximum likelihood estimation, as intuitive measures of the contribution of the data on the dependent variable to the maximum likelihood estimators of the behavioral parameters.

We begin the exposition in Section 2 with the notation and basic assumptions, together with discussion of the general structure of model change and distribution effects. In Section 3 we present two characterizations of model change effects, with Section 3.1 deriving the "average-marginal" interpretation, and Section 3.2 deriving the cross section regression interpretation. Each interpretation indicates a consistent method of estimating model change effects with cross section data. The cross section regression interpretation shows the relation of model change effects to maximum likelihood estimation.

Finally, in Section 4 we consider the general question of whether a
statistic from cross section data can be regarded as a nonparametric index of the sensitivity of aggregate relationships to structural changes in behavior. Some formulae are presented, which are based on densities of the exponential family, that suggest that $R^2$ from a cross section OLS regression has this character. Upon closer examination, however, we find that $R^2$ does not quite satisfy the model change criterion, with the development highlighting the difference between true random disturbances and structural nonlinearity in behavioral models.

2. The Basic Framework

2.1 The Cross Section Data and The Behavioral Model

The data situation we consider is that of a cross section of observations from a particular time period, say $t = t^0$. Letting $y$ denote a dependent variable of interest, and $X$ denote an $M$ vector of independent variables, the cross section data consists of $K$ observations, $y_k, X_k, k=1,...,K$. These observations are assumed to represent a random sample from a distribution (that is absolutely continuous with respect to $\sigma$-finite measure $\nu$) with density $P_\nu(y,X)$. Moreover, the entire population of $N$ observations is assumed to represent a random sample from the same distribution, with $N \gg K$. We denote the means of $y$ and $X$ at time $t_0$ as $\mu_\nu = E_\nu(y), \mu_\nu = E_\nu(X)$; and the variance-covariance matrix of $y$ and $X$ at time $t_0$ as

$$
\Sigma = \begin{bmatrix}
\sigma_{yy} & \Sigma_{yx} \\
\Sigma_{xy} & \sigma_{xx}
\end{bmatrix}
$$

To represent behavior, we assume that there exists a behavioral model of the form

\begin{equation}
(2.1) \quad y = f_\nu(X,u)
\end{equation}
where $T$ is a set of structural parameters and $u$ represents unobserved individual differences which affect $y$. The model is completed by a specification of the conditional distribution of $u$ given $X$, denoted $q_{\sigma}(u|X)$, which depends on parameters $\sigma$. For our purposes, the content of these assumptions is captured in the induced conditional distribution of $y$ given $X$, with density denoted $q_{\Delta}(y|X)$, where $\Delta'=(\Upsilon',\sigma')$ denotes all of the parameters. Hereafter, we will refer to $q_{\Delta}(y|X)$ as the "econometric model" and the conditional expectation $E(y|X)=F_{\Delta}(X)$ as the "structural model". The true value of $\Delta$ is denoted as $\Delta^o=(\Upsilon^o',\sigma^o')$, and $q^o(y|X)$ and $F^o(X)$ denote the respective functions evaluated at $\Delta=\Delta^o$. The true density $P^o(y,X)$ factors as $P^o(y,X)=q^o(y|X)p^o(X)$, where $p^o(X)$ is the density of the marginal distribution of $X$ underlying the cross section data.\(^2\)

A traditional econometric analysis of the cross section data involves estimation of the true value $\Delta^o$. For example, the maximum likelihood estimator $\hat{\Delta}_{ML}$ is found by maximizing the log-likelihood function, or equivalently as the solution of the first-order equation\(^3\)

$$\sum \frac{\partial \ln q_{\Delta}(y|X)}{\partial \hat{\Delta}_{ML}} = 0$$

(2.2)

In this paper, we employ the $\Delta$ parameterization for a different purpose, namely to parameterize changes in the econometric model over time. In particular, we derive the aggregate (macroeconomic) relation between the means of $y$ and $X$, and analyze how changes in $\Delta$ affect this relation. While estimation of $\Delta^o$ is not the primary focus of the exposition, we do indicate the connection between model change effects and the maximum likelihood estimator $\hat{\Delta}_{ML}$ of $\Delta=\Delta^o$ obtained from the cross section data.

2.2 Aggregate Functions and Macroeconomic Effects

The aggregate (macroeconomic) function between the means of $y$ and $X$ is
induced by movement in the joint distribution of y and X over time. One source of movement is change in the econometric model, here captured by Δ. The other source is change in the marginal distribution of X.

We assume for simplicity that the density of the marginal distribution of X is parameterized by \( \mu = E(X) \) as \( p(X|\mu) \), where for \( \mu = \mu^0 \), we have \( p(X|\mu^0) = p^0(X) \), the marginal density underlying the cross section data. The joint density of y and X can now be written as

\[
P(y, X|\Delta, \mu) = q(y|X)p(X|\mu)
\]

where for the values \( \Delta^0 \) and \( \mu^0 \), we have \( P(y, X|\Delta^0, \mu^0) = p^0(y, X) \).

With this structure, we derive the aggregate (macroeconomic) function relating \( E(y) \) to \( \Delta \) and \( \mu \) as:

\[
\mu_y = E(y) = \int yP(y, X|\Delta, \mu) \, dy = \phi(\Delta, \mu)
\]

See Stoker (1984) for several examples of aggregate functions.

The framework can be interpreted as follows. For the (cross section) time period \( t^0 \), the population density is characterized by the values \( \Delta^0 \) and \( \mu^0 \), which imply the mean value \( \mu_y^0 = \phi(\Delta^0, \mu^0) \) of y. For time period \( t \neq t^0 \), the density of the y, X distribution is characterized by values \( \Delta^t \) and \( \mu^t \), which imply the mean value \( \mu_y^t = \phi(\Delta^t, \mu^t) \) of y, so that the aggregate function \( \phi \) represents the relation over time between \( \mu_y \) and \( \Delta \) and \( \mu \). A situation where the econometric model is stable over time is represented by setting \( \Delta^t = \Delta^0 \) for all \( t \), whereas a situation where the marginal distribution of X is stable over time is represented by setting \( \mu^t = \mu^0 \) for all \( t \).

The macroeconomic effects of interest in this paper are the derivatives of the aggregate function \( \mu_y = \phi(\Delta, \mu) \) evaluated at time period \( t^0 \), the time period where cross section data is observed. These effects are represented in total derivative form as
where each of the partial derivatives is evaluated at $\Delta = \Delta^o$ and $\mu = \mu^o$. The first term of equation (2.5) represents the influence on $\nu = E(y)$ of parametric model change $d\Delta$, with $\partial \Phi / \partial \Delta^o$ the model change effects. The second term of equation (2.5) represents the influence on $\mu = E(y)$ of (independent variable) distribution shifts $d\mu$, with $\partial \Phi / \partial \mu^o$ the distribution effects.  4

The primary purpose of this paper is to characterize model change effects, in relation to the cross section data. Distribution effects have been characterized in a series of papers by Stoker (1982, 1983a, 1983b). We briefly review some of this work in Section 2.4.

2.3 Formal Assumptions

The framework is purposely general, placing no initial restrictions on the density $P^o(y,X)$ other than the existence of the means of $y$ and $X$. For our development, we require slightly more structure, as summarized in:

**ASSUMPTION A:** For all $(\Delta, \mu)$ in an open convex neighborhood of $(\Delta^o, \mu^o)$;

i) The means of $y$ and $X$ exist, $E(y) = \Phi(\Delta, \mu)$ is differentiable in all components of $\Delta$ and $\mu$.

ii) $P(y,X | \Delta, \mu)$ is second-order differentiable in all components of $\Delta$ and $\mu$.

iii) All variances and covariances of $y$, $X$ and the score vectors $\partial \ln q_{\Delta}(y | X) / \partial \Delta$, $\partial \ln p(X | \mu) / \partial \mu$ exist.

iv) The information matrices

$$I_\Delta = -E \left[ \frac{\partial^2 \ln q_{\Delta}}{\partial \Delta \partial \Delta^t} \right] \quad I_\mu = -E \left[ \frac{\partial^2 \ln p}{\partial \mu \partial \mu^t} \right]$$

exist and are positive definite (where expectation is taken over $y$ and $X$).
These assumptions allow for differentiation of the aggregate function $\phi$, and the use of population covariance values. We also assume two regularity assumptions presented in the Appendix, which, for instance, allow differentiation of $\phi$ under the defining integral.

The information matrices $I_\Delta$ and $I_\mu$ arise naturally in the standard Cramer-Rao theory of CUAN estimation (Rao 1973). Briefly, suppose that our interest was in estimating the values $\Delta^0$ and $\mu^0$ using the cross section data. The Cramer-Rao theory establishes a lower bound on the asymptotic variance of any CUAN estimator $(\hat{\Delta}, \hat{\mu})$ of $(\Delta^0, \mu^0)$, namely $(I_{\Delta^0, \mu^0})^{-1}$, where $I_{\Delta^0, \mu^0}$ is the information matrix formed with respect to $\Delta$ and $\mu$, and evaluated at $\Delta^0$ and $\mu^0$. This information matrix is given as

$$I_{\Delta^0, \mu^0} = -\begin{bmatrix} \mathbb{E}\left[\frac{\partial^2 \ln p}{\partial \Delta \partial \Delta'}\right] & \mathbb{E}\left[\frac{\partial^2 \ln p}{\partial \Delta \partial \mu'}\right] \\ \mathbb{E}\left[\frac{\partial^2 \ln p}{\partial \mu \partial \Delta'}\right] & \mathbb{E}\left[\frac{\partial^2 \ln p}{\partial \mu \partial \mu'}\right] \end{bmatrix} = \begin{bmatrix} I_{\Delta^0} & 0 \\ 0 & I_{\mu^0} \end{bmatrix}$$

The latter equality follows from $P(y, X | \Delta, \mu) = q_\Delta (y | X) p(X | \mu)$, or because $\Delta$ and $\mu$ are $L$-independent (Barndorff-Nielson 1978). This diagonal form implies that $\Delta^0$ and $\mu^0$ can be estimated separately, without loss of information asymptotically. Equivalently, the lower bound on the asymptotic variance of a CUAN estimator $\hat{\Delta}$ of $\Delta^0$ is $(I_{\Delta^0})^{-1}$, and the lower bound on the asymptotic variance of a CUAN estimator $\hat{\mu}$ of $\mu^0$ is $(I_{\mu^0})^{-1}$.

A major result of the Cramer-Rao theory is that maximum likelihood estimators are asymptotically efficient, namely that they achieve the Cramer-Rao lower bound. For $\hat{\Delta}_{ML}$ above, under standard conditions one has

$$\sqrt{K}(\Delta_{ML} - \Delta^0) = (I_{\Delta^0})^{-1} \frac{1}{\sqrt{K}} \sum_{k=1}^{K} \frac{\partial \ln q_\Delta (y_k | X_k) / \partial \Delta}{\partial \Delta} + o_p(1)$$

where $\hat{I}_{\Delta^0} = - (1/K) (\Sigma_k (\partial^2 \ln q_\Delta (y_k | X_k) / \partial \Delta \partial \Delta')$. Because $y_k, X_k, k=1, \ldots, K$ is a random sample, we have in general that $\lim \hat{I}_{\Delta^0} = I_{\Delta^0}$ a.s. Asymptotic normality of $\hat{\Delta}_{ML}$ follows from the Central Limit Theorem applied to (2.7) as $K \to \infty$, and
the efficiency result follows from the familiar formula \( \text{Var}(\ln q_{\Delta}(y|X)/\Delta^0) = I_{\Delta^0} \) where this variance is unconditional.

This standard derivation was recalled to motivate a theoretical construct to be used in our characterization of model change effects. Namely, following Efron (1975) and Pitman (1979) we can define the best locally unbiased estimator of \( \Delta \) near \( \Delta^0 \) using the cross section data as

\[
U_{\Delta^0} = (I_{\Delta^0})^{-1} \sum \frac{\ln q_{\Delta}(y_k|X_k)}{\Delta^0} + \Delta^0
\]

It is easy to verify that \( E(U_{\Delta^0}) = \Delta^0 \) and that \( U_{\Delta^0} \) attains the asymptotic Cramer-Rao lower bound. \( U_{\Delta^0} \) is not usually an "estimator" in the familiar sense because it will in general depend nontrivially on the true parameter value \( \Delta^0 \). Notice that the above discussion of maximum likelihood is just a statement of the first order asymptotic equivalence between \( \hat{\Delta}_{ML} \) and \( U_{\Delta^0} \).

Finally, note that all of the above discussion could be equally applied to the estimation of the parameter value \( \mu = \mu^0 \), including the construction of a best locally unbiased estimator.

2.4 Brief Review of Previous Results on Distribution Effects

The distribution effects of equation (2.5) are characterized with respect to cross section data in Stoker (1983a). The major result of interest is that \( \partial \phi / \partial \mu^0 \) is the a.s. limit of slope coefficients from a cross section instrumental variables regression of \( y_k \) regressed on \( X_k \). In particular, we have that \( \delta \equiv \partial \phi / \partial \mu^0 = \lim \hat{d} \) a.s., where \( \hat{d} \) is the estimated slope vector from the equation

\[
y_k = c + \hat{d}'X_k + \hat{u}_k
\]

estimated using \((1, Z_k')'\) as the instrumental variable, where
Following the discussion of Section 2.3, we see that $Z_k$ is a component of the best locally unbiased estimator $U_{\mu_0}$ of $\mu$ near $\mu_0$:

\begin{equation}
U_{\mu_0} = \sum Z_k/K = \bar{Z}
\end{equation}

This characterization of distribution effects depends only on the assumptions listed above, and in particular does not depend on any restrictions testable using the cross section data, such as any functional form assumptions on the true econometric model. A potential problem with implementing the result is that $Z_k$ will in general depend on the true value $\mu_0$, however one can show under general conditions that the above result is valid when $Z_k$ is constructed using any strongly consistent estimator of $\mu_0$ (Stoker 1983a).

A special case of interest occurs when $p(X|\mu)$ can be written in the form of an exponential family with driving variables $X$ (see Section 4 for the definition). In this case $Z_k = X_k$, and $U_{\mu_0} = \sum X_k/K = \bar{X}$, the sample average of $X_k$, $k=1,\ldots,K$. Equation (2.10) represents the OLS regression of $Y_k$ on $X_k$, with $S = (\Sigma_{X})^{-1}\Sigma_{XY} = \beta$. This result underlies the macroeconomic interpretation of cross section OLS coefficients when the true econometric model between $y$ and $X$ is misspecified or unknown (Stoker 1982). For later reference, we display the macroeconomic effects (2.5) explicitly for this case as

\begin{equation}
d\mu_0 = \left(\frac{\partial \phi}{\partial \Delta^0}\right)^d \, d\Delta + \beta^t \, d\nu
\end{equation}

With all of this background, we now turn to the characterization of model change effects.
3. The Macroeconomic Effects of a Changing Behavioral Model

The results reviewed in Section 2.4 characterize distribution effects for an arbitrary, but stable, econometric model. To empirically implement the following characterizations of model change effects, the econometric model \( q_\Delta(\gamma|X) \) must be explicitly specified.

The reader is reminded that the total effects on \( \mu \) are given by \((2.5)\), and that the model change effects studied below are partial effects which hold for \( \mu=\mu^0 \), a stable independent variable distribution. In this regard, one may question whether focusing on model change effects without simultaneously considering distribution effects involves a loss of information. We will show that there is no loss in considering the two types of effects separately -- the technical reason being the diagonal form of the information matrix \((2.6)\). We will return to this point in Section 3.3.

The model change effects \( \partial \phi / \partial \Delta^0 \) are expressible in two equivalent ways which lend themselves to two different interpretations, presented respectively in Section 3.1 and 3.2.

3.1 The Average-Marginal Interpretation

The first interpretation of model change effects is found by taking the expectation \((2.4)\) first conditional on \( X \) and then with respect to \( X \), and differentiating under the integral sign as in:

\[
(3.1) \quad \frac{\partial}{\partial \Delta^0} \left( \phi(\Delta, \mu^0) \right) = \frac{\partial}{\partial \Delta^0} \int F_\Delta(X)p(X|\mu^0)dv
\]

\[
= \int \frac{\partial F_\Delta(X)}{\partial \Delta^0} p(X|\mu^0)dv
\]

\[
= E^0 \left( \frac{\partial F}{\partial \Delta^0} \right)
\]

(3.1) symbolizes a somewhat obvious conclusion, that the effect on the mean \( E(y) \) of a model change \( d\Delta \) is the mean of the marginal changes on the
individual structural model. In other words, \( d \Delta \) induces a change in the value of \( F_{\Delta}(X) \) for each individual, and the effect on \( E(y) \) is just the mean of the individual effects. Moreover, if a change in a particular parameter (say \( \sigma \)) does not affect the structural model, i.e. \( \partial F/\partial \sigma = 0 \) for all \( X \), then it has no (macroeconomic) effect on \( E(y) \).

EXAMPLE 3.1: Standard Linear Model

Suppose that behavior is represented by

\[
y = \phi_\Delta(X, u) = T_0 + X'T_1 + u
\]

where \( T'=(T_0, T_1') \) and the distribution of \( u \) conditional on \( X \) is normal with mean 0 and variance \( \sigma^2 \). \( q_{\Delta}(y|X) \) is therefore a normal distribution with mean \( F_{\Delta}(X) = T_0 + X'T_1 \) and variance \( \sigma^2 \), where \( \Delta=(T', \sigma)' \). The true aggregate function is

\[
E(y) = \phi(\Delta, \mu) = T_0 + \mu'T_1
\]

and so the model change effects are \( \partial \phi/\partial \Delta^o = (1, \mu', 0)' \). Obviously \( \partial \phi/\partial \Delta^o = E(\partial F/\partial \Delta^o) \), since \( \partial F/\partial \Delta^o = (1, X', 0)' \).

EXAMPLE 3.2: Exact Aggregation Models

A simple extension of the above example occurs when behavior takes the form

\[
y = \phi_\Delta(X, u) = b_0(T) + X'b_1(T) + u
\]

where \( b_0 \) and \( b_1 \) are coefficient functions dependent on \( T \). \( T \) here may represent parameters to be estimated, or variables that are constant across all individuals in any given time period (e.g. prices in a study of aggregate demand, see Jorgenson, Lau and Stoker 1982). By adopting the same stochastic structure for \( u \) as in Example 3.1, we have \( F_{\Delta}(X) = b_0(T) + X'b_1(T) \) and

\[
E(y) = \phi(\Delta, \mu) = b_0(T) + \mu'b_1(T)
\]
Consequently, the model change effects are

\[
\frac{\partial \Phi}{\partial \Delta} = \begin{bmatrix} B & 0 \\ T & 0 \end{bmatrix} \begin{bmatrix} 1 \\ \mu^0 \end{bmatrix}
\]

where \( B \) is the differential matrix of \((b_0(T), b_1(T))'\) with respect to \( T \).

In these examples, the heterogeneity parameter \( \sigma \) does not appear in the aggregate function. For an example where \( \sigma \) is important, consider

**EXAMPLE 3.3 Probit Model**

Suppose that \( y \) is a dichotomous random variable, with behavior given as

\[
y = f_T(X, u) = 1 \quad \text{if} \quad u \leq T_0 + X'\bar{T}_1 \\
= 0 \quad \text{if} \quad u > T_0 + X'\bar{T}_1
\]

where \( u \) is distributed (conditional on \( X \)) normally with mean \( 0 \) and variance \( \sigma^2 \). \( q(y|X) \) is a binomial distribution with mean \( \text{Prob}(y=1|X) = F(\frac{T_0}{\sigma} + X'(\bar{T}_1/\sigma)) \), where \( \Phi \) is the cumulative normal distribution function --- a probit model (clearly one must normalize, say \( T_0 = 1 \), to satisfy Assumption A iv)). \( E(y) = \Phi(\Delta, \mu) \) is the true proportion of individuals with \( y=1 \), and the macroeconomic effects of changing \( \sigma \) are given as

\[
\frac{\partial \Phi}{\partial \sigma} = E^0 \left( \frac{\partial F(Y)}{\partial \sigma} \right) = E^0 \left( -\varphi \left( \frac{T_0}{\sigma} + X'\bar{T}_1 \sigma \right) \left( \frac{T_0 + X'\bar{T}_1}{\sigma^2} \right) \right)
\]

where \( \varphi \) is the normal density function. If \( p(X|\mu) \) is a multivariate normal density with mean \( \mu^0 \) and variance \( \Sigma_{xx} \), we have that

\[
\frac{\partial \Phi}{\partial \sigma} = -\varphi \left( \frac{T_0 + \mu^0'\bar{T}_1}{\sigma^2 + \bar{T}_1'\Sigma_{xx}\bar{T}_1} \right) \left( \frac{(T_0 + \mu^0'\bar{T}_1)2\sigma}{(\sigma^2 + \bar{T}_1'\Sigma_{xx}\bar{T}_1)^{3/2}} \right)
\]

which is consistent with the aggregate function for this case (McFadden and Reid 1975):

\[
E(y) = \Phi(\Delta, \mu) = \Phi \left( \frac{T_0 + \mu'\bar{T}_1}{\sqrt{\sigma^2 + \bar{T}_1'\Sigma_{xx}\bar{T}_1}} \right)
\]
We leave as an exercise the effects of changing $\gamma_0$ and $\gamma_1$ for this example.

When the behavioral function is nonlinear, a precise specification of the marginal distribution $p(x|\mu)$ is required in order to obtain a formula for $\partial \Phi / \partial \Delta^\circ$. Often, one can estimate the value of $\partial \Phi / \partial \Delta^\circ$ by using the cross section sample distribution of $X$, instead of specifying $p(x|\mu)$ explicitly, as follows. Suppose that $\hat{\Delta}$ represents a strongly consistent estimate of $\Delta^\circ$, then under our assumptions (see Stoker 1983a, Theorem 7 for details) we have that

\begin{equation}
\frac{\partial \Phi}{\partial \Delta^\circ} = \lim_{K} \frac{1}{K} \sum \frac{\partial F(X_k)}{\partial \hat{\Delta}} \quad \text{a.s.}
\end{equation}

Consequently, if the $\partial F(X)/\partial \Delta$ terms can be evaluated for $\hat{\Delta}$ and each $X_k$, their sample average is a strongly consistent estimator of the effects $\partial \Phi / \partial \Delta^\circ$.

Using this interpretation, the model change effects can be estimated as long as the explicit (analytic) formula for the structural model $F(\Delta)(X)$ is known, and $\Delta^\circ$ can be consistently estimated. The next characterization permits estimation using cross section data when only the econometric model $q_{\Delta}(y|X)$ is known, and may therefore be helpful when $F(\Delta)(X)$ is difficult to solve for.

3.2 The Regression Coefficient Interpretation

The second characterization of model change effects can be found by differentiating under the integral defining (2.4) without first conditioning on $X$. This gives

\begin{equation}
\frac{\partial \Phi}{\partial \Delta^\circ} = \int y \frac{\partial q(\gamma|X)}{\partial \Delta^\circ} p(x|\mu) \, dy
\end{equation}

\begin{align*}
&= \int y \frac{\partial \ln q_{\Delta}(y|X)}{\partial \Delta^\circ} q^\circ(y|X)p(x|\mu) \, dy \\
&= \text{Cov} \left( y, \frac{\partial \ln q_{\Delta}(y|X)}{\partial \Delta^\circ} \right) = \Sigma_{\Delta y}
\end{align*}

so that $\partial \Phi / \partial \Delta^\circ$ is just the covariance between $y$ and the conditional score.
vector \( \Delta \ln q(y|x)/\Delta \alpha \). This leads to a cross section regression interpretation as follows. Consider the best locally unbiased estimator \( U_{\Delta \alpha} \) of \( \Delta = \Delta \alpha \) defined in (2.8). \( U_{\Delta \alpha} \) can be written as a average, \( U_{\Delta \alpha} = \Sigma w_k/K \), where the components are defined as \( w_k = (I_{\Delta \alpha})^{-1} (\Delta \ln q(y_k|x_k)/\Delta \alpha) + \Delta \alpha \). It is easy to see that the model change effects \( \Delta \Phi / \Delta \alpha \) are consistently estimated by the OLS slope coefficients of the cross section regression of \( y_k \) on \( w_k \); if

\[
y_k = a + w_k'b + v_k \quad k=1, \ldots, K
\]

is estimated by OLS, then

\[
\lim \hat{b} = (\Sigma_{w})^{-1} \Sigma_{wv} = ((I_{\Delta \alpha})^{-1})^{-1} (I_{\Delta \alpha})^{-1} \Sigma_{\Delta y} = \Sigma_{\Delta y} = \Delta \Phi / \Delta \alpha
\]

where \( \Sigma_{w}, \Sigma_{wv} \) denote the covariance matrix of \( w \) and the vector of covariances between \( y \) and \( w \) respectively, and we have used \( \Sigma_{w} = (I_{\Delta \alpha})^{-1} \) and \( \Sigma_{wv} = (I_{\Delta \alpha})^{-1} \Sigma_{\Delta y} \), as well as (3.3).

The connection between model change effects \( \Delta \Phi / \Delta \alpha \) and the maximum likelihood estimator \( \hat{\Delta \alpha} \) can be seen via the following development. We begin by interpreting the analysis of variance decomposition implied by (3.4) by methods analogous to Stoker (1983a). Formally, the large sample variance decomposition implied by (3.4) is:

\[
\sigma_{yv} = \left( \frac{\partial \Phi}{\partial \Delta \alpha} \right)' (I_{\Delta \alpha})^{-1} \left( \frac{\partial \Phi}{\partial \Delta \alpha} \right) + \sigma_{yv}
\]

Suppose, for the moment, that \( \Delta \) consists of a single parameter. The it is easy to check that the first term above, the average "fitted sum of squares", is the asymptotic covariance between the sample average of \( y \), \( \bar{y} = \Sigma y_k/K \), and the best locally unbiased estimator \( U_{\Delta \alpha} \) defined in (2.8); namely

\[
K \text{ Cov}(\bar{y}, U_{\Delta \alpha}) = \left( \frac{\partial \Phi}{\partial \Delta \alpha} \right)^2 (I_{\Delta \alpha})^{-1}
\]
This implies that the squared asymptotic correlation between \( \bar{y} \) and \( U_{\Delta^o} \) is the large sample \( R^2 \) from the regression (3.4), \( R^2 = 1 - (\sigma_{\bar{y}U}/\sigma_{\bar{y}y}) \).

When \( \Delta \) consists of several parameters, the above interpretation extends as follows. Since (3.4) represents an OLS regression, \( (\partial \phi/\partial \Delta^o)'U_{\Delta^o} \) is the linear combination of components of \( U_{\Delta^o} \) that is most highly correlated with the sample average \( \bar{y} \). Consequently, \( R^2 \) is the asymptotic squared (first) canonical correlation between \( \bar{y} \) and \( U_{\Delta^o} \). The model change effects \( \partial \phi/\partial \Delta^o \) represent the proper asymptotic weights on the components of \( U_{\Delta^o} \). In this sense, the effects \( \partial \phi/\partial \Delta^o \) measure the contribution of the \( y \) data to the locally unbiased estimator \( U_{\Delta^o} \).

The connection to maximum likelihood is evident by recalling that the maximum likelihood estimator \( \hat{\Delta}_{ML} \) is asymptotically first order equivalent to the best locally unbiased estimator \( U_{\Delta^o} \). Consequently we can replace \( U_{\Delta^o} \) by \( \hat{\Delta}_{ML} \) in the above discussion, with \( (\partial \phi/\partial \Delta^o)'\hat{\Delta}_{ML} \) the linear combination of components of \( \hat{\Delta}_{ML} \) which is most highly correlated with \( \bar{y} \), asymptotically.

Thus we conclude that the (large sample) \( R^2 \) from the regression (3.4) is the asymptotic (first) canonical correlation between \( \bar{y} \) and \( \hat{\Delta}_{ML} \), and the model change effects \( \partial \phi/\partial \Delta^o \) provide the proper weights on the components of \( \hat{\Delta}_{ML} \). The model change effects \( \partial \phi/\partial \Delta^o \) therefore measure the contribution of the \( y \) data to the maximum likelihood estimator of the behavioral parameters \( \hat{\Delta}_{ML} \).

We now present two examples which show that the covariance representation (3.3) coincides with the "average-marginal" representation (3.1) of model change effects.

**EXAMPLE 3.4: Normal Additive Disturbance**

Suppose that behavior is represented as

\[
y = f_T(X,u) = g_T(X) + u
\]

where \( u \) is distributed normally (conditional on \( X \)) with mean 0 and variance
\( \sigma^2 \). Clearly \( F_\Delta(X) = g_T(X) \). We have that

\[
\ln q_\Delta(y|X) = -\ln 2\pi - \ln \sigma - \frac{1}{2\sigma^2}(y - g_T(X))^2
\]

so that

\[
\frac{\partial \ln q_\Delta(y|X)}{\partial \Delta} = \begin{bmatrix}
\frac{\partial \ln q(y|X)}{\partial \Delta}
\frac{\partial \ln q(y|X)}{\partial \sigma}
\end{bmatrix} = \begin{bmatrix}
\frac{1}{\sigma^2}(y - g_T(X))
\frac{1}{\sigma^2}(y - g_T(X))^2
\end{bmatrix}
\]

By (3.3), the model change effects are given as \( \Sigma_\Delta = \text{Cov}(y, \frac{\partial \ln q_\Delta(y|X)}{\partial \Delta}) \), which are calculated as

\[
\Sigma_\Delta = \begin{bmatrix}
E_\Delta \left[ \frac{1}{\sigma^2}(y(y - g_T(X))) \right]
E_\Delta \left[ \frac{y(y - g_T(X))^2}{\sigma^2} \right]
\end{bmatrix} = E_\Delta \begin{bmatrix}
\frac{\sigma^2}{\sigma^2} \frac{8g_T(X)}{\partial T}
0
\end{bmatrix}
\]

\[
= E_\Delta \begin{bmatrix}
\frac{\partial F_\Delta(X)}{\partial T}
\frac{\partial F_\Delta(X)}{\partial \sigma}
\end{bmatrix} = E_\Delta \begin{bmatrix}
\frac{\partial F_\Delta(X)}{\partial \Delta}
\end{bmatrix}
\]

where the latter equalities follow by taking the expectation first conditional on \( X \) and then with respect to \( X \).

**EXAMPLE 3.5: Discrete Choice Model**

Suppose that \( y \) takes on the values 0 and 1, and that for given \( X \), the probability that \( y = 1 \) is \( E(y|X) = F_\Delta(X) \). We have that

\[
\ln q_\Delta(y|X) = y \ln F_\Delta(X) + (1 - y) \ln (1 - F_\Delta(X))
\]

The score vector is
The model change effects are given by (3.3) as

\[ \frac{\partial \ln q_\Delta(y \mid x)}{\partial \Delta} = \left[ \frac{y}{F_\Delta(x)} - \frac{(1 - y)}{1 - F_\Delta(x)} \right] \frac{\partial F_\Delta(x)}{\partial \Delta} \]

as promised.

As with the interpretation given in Section 3.1, the covariance representation (3.3) motivates a consistent estimator of the model change effects \( \frac{\partial \Phi}{\partial \Delta} \) using the sample distribution of \( X \) observed in the cross section data. If particular, if \( \hat{\Delta} \) is any strongly consistent estimator of \( \Delta = \Delta^0 \) (such as \( \hat{\Delta}_{ML} \)), we have that

\[ \lim_{K} \frac{1}{K} \sum_{k=1}^{K} \frac{\partial \ln q_\Delta(y_k \mid x_k)}{\partial \Delta} = \frac{\partial \Phi}{\partial \Delta^0} \text{ a.s.} \]

This sample covariance is particularly easy to compute once maximum likelihood estimation of \( \Delta^0 \) has been performed, as the (estimated) score vectors are usually utilized to compute the asymptotic covariance matrix of the maximum likelihood estimates. Also, it is easy to verify that this estimator is asymptotically equivalent to the slope coefficients of (3.4), obtained by regressing \( y_k \) on \( \tilde{w}_k \), where \( \tilde{w}_k \) is constructed using the strongly consistent estimator \( \hat{\Delta} \).

The sample covariance estimator in (3.9) has some practical advantages over the "average-marginal" estimator in (3.2). To implement (3.2), the formula for the conditional expectation \( E(y \mid x) = F_\Delta(x) \) must be known. In complicated nonlinear models, the score vector formula \( \frac{\partial \ln q_\Delta(y \mid x)}{\partial \Delta} \) may be easier to obtain.

Moreover, suppose that the aggregate variable of interest is not \( E(y) \),
but rather the mean of some function of \( y \), say \( E(g(y)) \). For example, a study of labor market behavior might set \( y \) equal to log wages, but the (arithmetic) mean wage was the policy variable of interest, then \( g(y) = \exp(y) \).

To implement (3.2), one must solve for the conditional expectation \( E(g(y) \mid X) \) as a function of \( \Delta \). However, to implement (3.9), one just takes the sample covariance between \( g(y_w) \) and the score vector estimates \( \ln q_{\Delta}(y_w \mid X_w) / \Theta \). In the same spirit, the sample covariance estimator in (3.9) can be used to compute model change effects on several aggregate variables using the same score vector estimates. Namely, if one was interested in the impact of model change on the first few moments of \( y \), say \( E(y \mid X), E(y^2 \mid X), E(y^3 \mid X) \), etc., then the sample covariance between the estimated score vectors and \( y_w, y_w^2, y_w^3 \), etc., could be computed.

3.3 Cross Section Orthogonality of Model Change and Distribution Effects

As mentioned above, there is no loss in generality from considering the effects of changing the behavioral parameters \( \Delta \) separately from the effects of changing the distribution parameters \( \nu \). This can be seen directly from the cross section regression interpretations of model change and distribution effects, by recognizing the diagonal structure of the joint information matrix (2.6).

Formally, if we construct the best locally unbiased estimator of \( (\Delta, \nu) \) at \( (\Delta^0, \nu^0) \), we have

\[
U_{\Delta^0, \nu^0} = (I_{\Delta^0, \nu^0})^{-1} \left[ \begin{array}{c} \sum \ln q_{\Delta} \\ \sum \ln p_{\nu} \end{array} \right] + \left[ \begin{array}{c} \Delta^0 \\ \nu^0 \end{array} \right] = \left[ \begin{array}{c} U_{\Delta^0} \\ U_{\nu^0} \end{array} \right] = \left[ \begin{array}{c} \bar{w} \\ \bar{z} \end{array} \right]
\]

where \( \bar{w} = \Sigma w / K \) and \( \bar{z} = \Sigma z / K \) are the averages that appear in the separate interpretations of model change and distribution effects. Since \( I_{\Delta^0, \nu^0} \) represents the joint covariance matrix of \( (w, z) \), \( w \) and \( z \) are uncorrelated, so that the regression of \( y_w \) on \( w \) and \( z \) results in the same coefficients.
asymptotically as separate regressions of $y_k$ on $w_k$ and $y_k$ on $Z_k$. In this sense, separate treatment of the effects of changing $\Delta$ and $\mu$ implies no loss in generality.

4. A Curiosity Concerning $R^2$ and Densities of the Exponential Family

Section 3 is devoted to the estimation of model change effects consistent with the particular $\Delta$ parameterization of the microeconomic behavioral model. This section is concerned with the broader question of whether there exists a statistic from cross section data that provides a general index of sensitivity of aggregate variables to changes in the underlying behavioral model. Some formulae are developed that suggest that the multiple squared correlation coefficient $R^2$ from the cross section OLS regression of $y$ on $X$ has this character. However, on closer inspection, we find that the apparent connection between $R^2$ and model change effects is not correct, with the correspondence turning on the source of the cross section residuals that underly $R^2$.

The characterization of $R^2$ is based on distribution movement modeled via densities of the exponential family form, with driving variables $y$ and $X$. This structure is introduced via:

**ASSUMPTION B:** The model describing the changes in the joint density of $y$ and $X$ can be written in the form of an exponential family with driving variables $y$ and $X$ as

$$P(y,X|\pi_y,\pi) = C(\pi_y,\pi) P^0(y,X) \exp(\pi_y y + \pi X)$$

where $C(\pi_y,\pi)^{-1} = \int P^0(y,X) \exp(\pi_y y + \pi X) dv$ exists and is positive in a neighborhood of $(\pi_y,\pi)=0$.

The term "driving variables" refers to the variables that appear in the exponent of (4.1), here $y$ and $X$, which control the exact form of density movement implied by (4.1). Alternately, the exponential family form with
driving variables $X$ is represented by (4.1) constrained by $\pi_y = 0$.

The expression (4.1) replaces the previous parameters $\Delta$ and $\mu$ by the
"natural parameters" $\pi_y$ and $\pi$. We can equivalently parameterize (4.1) using
the means $(E(y), E(X)) = (\mu_y, \mu)$, by solving for $\mu_y$ and $\mu$ in terms of $\pi_y$ and $\pi$
as

$$
\begin{pmatrix}
    \mu_y \\
    \mu
\end{pmatrix} = \int \begin{pmatrix}
    y \\
    x
\end{pmatrix} \mathcal{C}(\pi_y, \pi) \mathcal{P}(y, X) \exp(\pi_y y + \pi x) \ dv \equiv H(\pi_y, \pi)
$$

$H$ is invertible and can be used to parameterize the joint density by inserting
$(\pi_y, \pi')' = H^{-1}(\mu_y, \mu)$ into (4.1). The local behavior of the mapping $H$ at the
cross section parameter value $(\pi_y, \pi')' = 0$ is given via (Stoker 1982)

$$
\begin{align}
(4.3a) & \quad d\mu_y = \Sigma_{yy} d\pi_y + \Sigma_{xy} d\pi \\
(4.3b) & \quad d\mu = \Sigma_{xy} d\pi_y + \Sigma_{xx} d\pi
\end{align}
$$

which relate the local changes in $\mu_y$ and $\mu$ to changes in $\pi_y$ and $\pi$.

As noted in Section 2, when the marginal density of $X$ varies in the form
of an exponential family with driving variables $X$, or when $\pi_y$ is held constant
at $\pi_y = 0$, OLS slope coefficients from the cross section regression of $y$ on $X$
consistently estimate the distribution effects of changes in $\mu = E(X)$ on
$\mu_y = E(y)$. More formally, the constraint $\pi_y = 0$ induces the aggregate relation
$\mu_y = \phi(\mu)$ between $\mu_y$ and $\mu$, and the distribution effects $\partial \phi/\partial \mu$ are consistently
estimated by the OLS slope coefficients $b$ of the cross section regression:

$$
(4.4) \quad y = \hat{a} + X \hat{b} + \hat{e}_n
$$

To verify this from the above formulae, set $d\pi_y = 0$ in (4.3a,b), solve (4.3b) for
d$\pi$ as

$$
(4.5) \quad d\pi = (\Sigma_{xx})^{-1} d\mu
$$

and insert into (4.3a) as

21
(4.6) \[ d\nu = \Sigma_{xy}(\Sigma_{xx})^{-1} d\mu = \beta' d\mu \]

where \( \beta = (\Sigma_{xx})^{-1} \Sigma_{xy} \), which is the a.s. limit of \( \hat{\beta} \). \( \beta \) therefore represents the distribution effects when \( d\pi = 0 \).

This result is reflected in the more general setting by considering the local effects on \( \mu \) of changing the "mixed" parameter \((\pi, \mu)\) (see Barndorff-Neilsen (1978) for this terminology). To calculate these effects, solve the local equation (4.3b) for \( d\pi \) as

(4.7) \[ d\pi = (\Sigma_{xx})^{-1} d\mu - (\Sigma_{xx})^{-1} \Sigma_{xy} d\pi \]

and insert into (4.3a) as

(4.8) \[ d\nu = \sigma_{yy} d\pi + (\Sigma_{xy})'((\Sigma_{xx})^{-1} d\mu - (\Sigma_{xx})^{-1} \Sigma_{xy} d\pi) \]

\[ = (\sigma_{yy} + (\Sigma_{xy})'((\Sigma_{xx})^{-1} \Sigma_{xy})) d\pi + \beta' d\mu \]

\[ = \sigma_{\pi\pi} + \beta' d\mu \]

where \( \sigma_{\pi\pi} \) is the large sample residual variance from the OLS regression (4.4); namely

(4.9) \[ \sigma_{\pi\pi} = \text{plim} \frac{\hat{\Sigma}_{\pi x}^2}{k} = \sigma_{yy} + \Sigma_{xy}'(\Sigma_{xx})^{-1} \Sigma_{xy} \]

Formula (4.8) says that for \( \mu \) given, the sensitivity of \( \mu \) to changes in \( \pi \) is measured by the large sample residual variance \( \sigma_{\pi\pi} \) from the cross section regression (4.4). \( \pi \) reflects the presence of \( y \) as a determinant of changes in the joint density of \( y \) and \( x \) (namely as a driving variable), and \( \sigma_{\pi\pi} \) measures the maximum distributional impact of influences on \( y \) which are uncorrelated with \( x \). Because \( \sigma_{\pi\pi} \) is not scale invariant, we rewrite (4.8) as

(4.10) \[ d\nu = \sigma_{yy}(1 - R^2) d\pi + \beta' d\mu \]

where \( R^2 = 1 - \sigma_{\pi\pi}/\sigma_{yy} \) is the large sample squared multiple correlation coefficient.
In other words, this development says that a poorly fitting (low $R^2$) cross section regression signals a situation where the mean of $y$ is very sensitive to density changes controlled by (driving) variables that are uncorrelated with $X$ but correlated with $y$.\(^6\) This is intuitively appealing, because a low $R^2$ statistic indicates that much of the variation in $y$ is uncorrelated with $X$, so that misspecification of the driving variables can result in substantial distribution effects on $\mu_y$ not accounted for by the cross section regression of $y$ on $X$.\(^7\) It should be noted that these remarks do not depend on the form of the true behavioral model $q^o(y|X)$ or the marginal density of $X$, $p^o(X)$, which underly in the cross section data.

The curiosity regarding (4.8) arises because of its similarity to (2.12), the general decomposition of macroeconomic effects into model change and distribution components. It is quite tempting to interpret $\beta' \delta u$ as a general exponential family distribution effect, and $\sigma_{\nu y} (1 - R^2) d\nu$ as a general model change effect, implying that $R^2$ is an index of sensitivity of $\nu_y$ to behavioral model change. While very tempting, this interpretation is not in correspondence with the development of Section 3. We now verify this, because it is informative as to the aggregate implications of structural nonlinearity in the behavioral model.

It is true that $\pi_y$ represents a model change parameter as previously defined. Formally, the conditional density of $y$ given $X$ implied by (4.1) is

\begin{equation}
q_{\pi_y}(y|X) = \tilde{C}(\pi_y,X) q^o(y|X) \exp(\pi_y y)
\end{equation}

where $\tilde{C}(\pi_y,X)^{-1} = \int q^o(y|X) \exp(\pi_y y) d\nu$ is assumed to exist. This conditional density depends only on $\pi_y$, and satisfies $q_{\pi_y}(y|X) = q^o(y|X)$ for $\pi_y = 0$, so that $\pi_y$ represents a model change parameter as $\Delta$ did before. More specifically, changes in $\pi_y$ alter both the conditional density (4.11) and the conditional expectation $F_{\pi_y}(X) = E(y|X,\pi_y)$, with the latter effects represented by
where $V_{\nu}(X) = E[(y-F_{\pi_{\nu}}(X))^2|X,\pi_{\nu})$ is the variance of $y$ conditional on the value of $X$. Formula (4.12) is obtained by noticing that for $X$ given, (4.11) is the exponential family density generated by $\phi(y|x)$ with driving variable $y$, and then differentiating under the integral defining $F_{\pi_{\nu}}(X)$. Therefore, for each $X$, $F_{\pi_{\nu}}(X)$ is altered in proportion to $V_{\nu}(X)$, which can represent a wide range of (one parameter) model changes, depending on the precise structure of the variance. For illustration, consider

**EXAMPLE 4.1:** Suppose that the true (cross section) behavioral model is given as $y = F_0(X) + u$, where the distribution of $u$ conditional on $X$ has mean 0 and variance $V_{\nu}(X)$. Consider the following three cases for $V_{\nu}(X)$:

a) $V_{\nu}(X) = \sigma^2 > 0$, constant with respect to $X$.

b) $V_{\nu}(X) = \sigma^2 F_0(X)$, proportional to $F_0(X)$, where $F_0(X) > 0$ for all $X$.

c) $V_{\nu}(X) = \Phi(X)$, a general integrable function of $X$, with $\Phi(X) > 0$ for all $X$.

Under case a), (4.12) can be written as

$$dF_{\pi_{\nu}}(X) = \sigma^2 d\pi_{\nu}$$

Consequently, changes in $\pi_{\nu}$ are equivalent to changes in a constant term for all $X$; in our previous notation, $F_{\pi_{\nu}}(X) = \mathcal{T} F_0(X)$, with $d\mathcal{T} = \sigma^2 d\pi_{\nu}$. Under case b), (4.12) can be written as

$$dF_{\pi_{\nu}}(X)/F_{\pi_{\nu}}(X) = d \ln F_{\pi_{\nu}}(X) = \sigma^2 d\pi_{\nu}$$

Here changes in $\pi_{\nu}$ are equivalent to changes in a scaling factor; in our previous notation, $F_{\pi_{\nu}}(X) = (\mathcal{T} + 1) F_0(X)$, where $d\mathcal{T} = \sigma^2 d\pi_{\nu}$. Case c) represents virtually any one parameter model change desired, with the condition that $\Phi(X) > 0$ insuring the positive monotonicity of $F_{\pi_{\nu}}(X)$ in $\pi_{\nu}$.

While $\pi_{\nu}$ is indeed a model change parameter, the problem in equating the interpretations of (2.12) and (4.10) arises from the assumed change in the
marginal distribution of $X$. When $\pi_\nu = 0$, the density of the marginal
distribution of $X$ is represented as the exponential family with $X$ as driving
variables. However, when $\pi_\nu \neq 0$, the density of the marginal distribution of $X$
implied by (4.1) is

$$p(X | \pi_\nu, \nu) = \frac{C(\pi_\nu, \pi(\pi_\nu, \nu))}{C(\pi_\nu, X)} p^0(X) \exp(\pi(\pi_\nu, \nu) \cdot X)$$

where $\pi = \pi(\pi_\nu, \nu)$ reflects the mixed parameterization of (4.1) in terms of $\pi_\nu$
and $\nu$. In general, $p$ depends on $\pi_\nu$, so we cannot assert that $\theta' d\nu$ represents
the unambiguous impact on $\mu_\nu$ of changes in the marginal distribution of $X$.

We can decompose changes in $\mu_\nu$ under (4.1) into model change effects and
distribution effects, in an informative way. To derive the correct
decomposition, explicitly write the mean $\mu_\nu$ of $y$ as

$$\mu_\nu = \int y p(y, x | \pi_\nu, \pi(\pi_\nu, \nu)) d\nu = \int f_{\pi_\nu}(X) \tilde{p}(X | \pi_\nu, \nu) d\nu$$

The differential change in $\mu_\nu$ at $\pi_\nu = 0$, $\mu = \mu^0$ is given as

$$d\mu_\nu = E^0 \left[ \frac{\partial F(\pi_\nu)}{\partial \pi_\nu} \right] d\pi_\nu + E^0 \left[ F(\pi_\nu) \frac{\partial \ln \tilde{p}}{\partial \pi_\nu} \right] d\pi_\nu + E^0 \left[ F(\pi_\nu) \frac{\partial \ln \tilde{p}}{\partial \nu} \right] d\nu$$

Formula (4.15) gives the proper decomposition, with the first term the model
change effect and the latter two terms the effect of changing the marginal
distribution of $X$. From (4.12), the first term can be written as

$$E^0 \left[ \frac{\partial F(\pi_\nu)}{\partial \pi_\nu} \right] d\pi_\nu = E^0 (V(\pi_\nu)) d\pi_\nu$$

so that the coefficient of $d\pi_\nu$ is just the mean conditional variance of $y$ (or
the "within" variance of $y$). The second term is shown through some very
tedious calculations (left as an exercise) to be

$$E^0 \left[ F(\pi_\nu) \frac{\partial \ln \tilde{p}}{\partial \pi_\nu} \right] d\pi_\nu = E^0 [(F(\pi_\nu) - \mu_\nu)^2 - \theta' \Sigma_{xx} \theta] d\pi_\nu$$

The third term is the easiest, being just the $d\nu$ term of (2.12), i.e.

$$E^0 [F(\pi_\nu) (\partial \ln \tilde{p} / \partial \nu)] d\nu = \theta' d\nu.$$
This long series of calculations can be summarized as follows. The overall impact of changing $\pi_{\nu}$ on the mean $E(y) = \mu_{\nu}$ is given by $\sigma_{\epsilon \epsilon}$, which leads to $R^2$ as a natural measure of sensitivity of $\mu_{\nu}$ to $\pi_{\nu}$. The model change effect associated with $\pi_{\nu}$ is given by (4.17), namely the part of $\sigma_{\epsilon \epsilon}$ due purely to the "true" residuals - departures of $y$ from the true structural model $E(y|x) = F^0(x)$. The effect of $\pi_{\nu}$ on $\mu_{\nu}$ which is induced from changes in the marginal distribution of $x$ is given by (4.17), which is the non-negative part of $\sigma_{\epsilon \epsilon}$ due to nonlinearity in the structural model $F^0(x)$. This term vanishes only when the true structural model is exactly linear. Consequently, while $\sigma_{\epsilon \epsilon}$ (and $R^2$) indicates the precise impact of $\pi_{\nu}$ on $\mu_{\nu}$ under the exponential family density (4.1), the separation of this impact into a model change effect and a marginal distribution effect depends on the source of the departures of $y$ from the regression line $\alpha + X'\beta$ (i.e. the OLS residuals), namely as true disturbance or structural nonlinearity. $R^2$ exactly measures the model change effect only when the true structural model is linear, namely $E(y|x) = \alpha + X'\beta$. The interpretation of $R^2$ as a measure of sensitivity of $\mu_{\nu}$ to model change is therefore only warranted when there is little or no evidence of structural nonlinearity in the behavioral model.

5. Summary and Conclusions.

In this paper we have provided characterizations of the effects of behavioral model changes on relationships among aggregate variables, and shown how they can be consistently estimated with cross section data. Two formulations were presented; the "average-marginal" characterization of Section 3.1, and the cross section regression interpretation of Section 3.2. In addition to providing consistent estimators of model change effects using cross section data, the generic connection between model change effects and maximum likelihood estimators was discussed. Finally, we explored the connection between the OLS goodness of fit statistic $R^2$ and the sensitivity of...
aggregate variables to changes in the behavioral model.

These results, together with Stoker (1983a), provide a complete characterization of the local parametric influences on aggregate relationships -- distribution effects and model change effects -- in terms of cross section data statistics. The major restrictions of the framework are that the microeconomic behavioral model is correctly specified, and that the cross section data represents a random sample from the distribution applicable to the entire population of individual agents. This work, together with Stoker (1984) on aggregate data modeling, is intended to provide a unified theoretical framework for incorporating all relevant evidence into the estimation of models of aggregate data relationships, including data on individual agents as well as aggregate economic time series.
Appendix: Technical Regularity Conditions

To justify the differentiation of expectations under the integral sign, define the difference quotients as

\[ D_i(y, X, h) = \frac{y \ln q_{\Delta^0 + h\varepsilon_i}(y | X) - \ln q^0(y | X)p^0(X)}{h} \quad i = 1, \ldots, L \]

where \( \Delta \) is assumed to be an \( L \)-vector of parameters and \( \varepsilon_i \) is the unit vector with \( i \)-th component 1. We assume

**ASSUMPTION AP.1:** There exists \( \nu \)-integrable functions \( g_i(y, X), i = 1, \ldots, L \), such that for all \( h \) where \( 0 \leq |h| < h_0 \),

\[ |D_i(y, X, h)| < g_i(y, X) \]

for \( i = 1, \ldots, L \).

To insure the consistency of the estimators defined in (3.2) and (3.9), which use the strongly consistent estimate \( \hat{\Delta} \) of \( \Delta^0 \), we denote the \( i \)-th component of \( \partial \ln q_{\Delta}(y | X)/\partial \Delta \) as \( \partial \ln q_{\Delta}(y | X)/\partial \Delta_i \) and assume

**ASSUMPTION AP.2:** There exists measurable functions \( G_i(y, X), i = 1, \ldots, L \), such that

\[ |y \partial \ln q_{\Delta}(y | X)/\partial \Delta_i| \leq G_i(y, X) \]

for all \( \Delta \) in an open neighborhood of \( \Delta^0 \), where \( E|G_i|^{1+\tau} \) is bounded for some \( \tau > 0 \), \( i = 1, \ldots, L \).
Footnotes

1. The results also characterize the aggregate effects of independent variables which are common to all individual agents in each time period (e.g. prices, general economic conditions, common time trends, etc.). This interpretation of model change parameters is not used in the exposition (excepting Example 3.2), however it is relevant for certain applications. Not relevant to this interpretation are the results involving parameter estimation, since common independent variables would be observed.

2. For simplicity, we restrict attention to a single dependent variable $y$, however all of results of Section 3 extend to the case where $y$ represents a vector of dependent variables. In this case $q^o(y|x)$ is the joint density of $y$ conditional on $x$, which can represent a general (linear or nonlinear) simultaneous equations model. The characterizing equations (3.1) and (3.3) apply directly to each component of $y$.

3. This statement requires some general regularity conditions - see Huber(1965), Rao(1973), Barnett(1976) and White(1980), among others.

4. Equation (2.5) also points out why we focus on parameterized model change effects. Without fairly complete parametric modeling, there is a lack of identification between different types of model change effects and distribution effects in aggregate data. For example, with aggregate data one could not identify the difference between a linear microeconomic model that shifts from period to period, and a stable, nonlinear microeconomic model. However, with sufficient modeling, one could reject certain specific nonlinear forms of the microeconomic model.
5. This result applies for all parameter changes in an open neighborhood of \( \eta = 0 \), as well to parameter changes in connected subsets of this open neighborhood. Two examples of such subsets are those represented by constraining some components of \( \eta \) to equal 0 (only certain components of \( X \) as driving variables - see Stoker(1983b)), or those represented by a continuous nonlinear constraint \( g(\eta) = 0 \) (curved exponential families - see Efron(1975)). Restriction to any of these subsets will induce constraints between the means \( E(X) = \mu \).

6. Formally, \( R^2 \) refers to the large sample value defined in (4.10). Consistent estimators of this value include the standard \( R^2 \) statistic from OLS regression, as well as \( \bar{R}^2 \), which includes the degree of freedom adjustment. The latter statistic may be preferable because it accommodates for the number of independent variables, however it should be noted that the standard justification for a degree of freedom adjustment is based on the linear model with normally distributed errors, which is not assumed here.

7. The precise definition of "high" and "low" depend on the relative sizes of \( \sigma_{yy} \) and \( \beta' d \mu \) of equation (4.10).
References


McFadden, D., and F. Reid (1975), "Aggregate Travel Demand Forecasting From Disaggregated Behavioral Models," Transportation Research Record, No. 534.


