AVOIDANCE OF SELECTIVITY BIAS THROUGH CLUSTER ANALYSIS

A Thesis submitted in Partial Fulfilment of the Requirements for the Degree of M. Phil in Statistics

by

MD. AZAD HOSSAIN HAZARI

REGISTRATION NO. - 148

SESSION : 1998-1999

DEPARTMENT OF STATISTICS

UNIVERSITY OF DHAKA

DHAKA-1000, BANGLADESH

DECEMBER, 2007

Supervisor's Declaration

This is to certify that the disertation titled "Avoidance of Selectivity Bias Through Cluster Analysis" submitted by Md. Azad Hossain Hazari in partial fulfillment for the degree of M.Phil, in the Department of Statistics, University of Dhaka is an original study carried out by him under my supervision and guidance.

Supervision
Professor Dr. Matiur Rahman
Department of Statistics
University of Dhaka
Dhaka-1000

Prof. Dr. Matiur Rahman
Department of Statistics
University of Dhaka
Dhaka, Bangladesh.

heterogeneity distribution. This means that an increase in current category attractiveness has a positive impact in terms of influencing current purchase. The anticipated stockout time has a negative impact (- 0.09 and - 0.01) on category purchase for both supports, i.e., the longer the time for which the household's current inven-tory will last, lower the household's current probability of buying the category, which makes intuitive sense. Between the two supports, the first is associated with larger magnitudes for the three estimated parameters. This means that the segment of consumers that comprises the more frequent buyers of paper towels (i.e., intercept of 2.77) also consists of users who are more sensitive to both marketing variables (which can also be ascertained by looking at the price elasticities of incidence reported in Table 9, separately for each segment) and stockout pressure in the product category. We present the estimated parameters of the brand choice model of PROPOS in Table 4. The estimates of the price and display parameters have the expected signs (negative and positive respectively) for both supports,

Table 3: Estimated Parameters of the PROPOS Incidence Model and their Standard Errors in Parentheses.

Parameter	Segment1	Segment2
Intercept	2.77 (0.41)	2.00 (0.54)
CA_{ht}	1.42 (0.11)	1.27 (0.23)
$Week_{sht}$	-0.09 (0.01)	-0.05 (0.01)

while the feature parameter is positive (as expected) for the first support but insignificant for the second support. Households in segment 1 are found to prefer major national brands – Bounty, Viva and Scott – compared to the other brands (as indicated by the estimated brand intercepts). Between the two supports of the heterogeneity distribution, the first is associated with larger magnitudes for the marketing mix coefficients, which implies that households in segment 1 are more sensitive to the marketing mix variables compared to households in segment 2 (which can also be ascertained by looking at the price elasticities of brand choice reported in Table 9, separately for each segment).

Table 4: Estimated Parameters of the PROPOS Brand Choice Model and their Standard Errors in Parentheses.

Parameter	Segment1	Segment2
Generic	-1.68 (0.20)	-0.21 (0.14)
Bounty	2.78 (0.16)	0.73 (0.18)
Viva	1.84 (0.19)	-0.53 (0.24)
Sparkle	-0.50 (0.16)	-1.90 (0.23)
Scott	2.59 (0.20)	0.91 (0.23)
Gala	-0.86 (0.16)	-2.17 (0.26)
Mardi	-0.48 (0.17)	-1.75 (0.22)
Price	-6.95 (0.40)	-4.57 (0.44)
Display	1.19 (0.10)	0.99 (0.13)
Feature	0.20 (0.10)	-0.20 (0.14)

The estimated parameters of the quantity model of PROPOS are given in Table 5. All the estimates are statistically significant. To the extent that the estimated intercept is larger for segment 2, we can interpret this segment as the heavy buyer segment in terms of the average quantity bought. Coupled with the results of Table 3, this presents an intriguing finding: even though consumers in segment 2 are less likely to buy paper towels than consumers in segment 1 for a given set of marketing variables (as indicated by the smaller estimated intercept for segment 2 in the incidence model), they are more likely to buy larger quantities of paper towels conditional on buying the product. As expected, the coefficient of inventory has a negative sign (- 0.08 and - 0.02), while the coefficients of consumption rate (1.16 and 0.11) and category utility (0.12 and 0.22) have positive signs on the predicted quantity for both supports of the heterogeneity distribution. The heavy buyer segment (i.e., segment 2) responds less to product inventory (I_{ht}) compared to the other segment. To the extent that the estimated coefficient of V_{hjt} is larger in magnitude for the heavy buyer segment, one may be inclined to believe that the heavy buyer segment responds more than the other segment to brands' marketing variables in its quantity decision. However, as soon as one realizes that the covariates V_{hjt} are constructed using segment-specific brand choice coefficients – which are not similar across the two segments but, in fact, imply higher sensitivity to the marketing mix in brand choices for segment 1 (see Table 4) – it is not clear whether the net effect of marketing variables is to imply higher sensitivity to the marketing mix for segment 2. In fact, once we convert these estimates to price-elasticities of demand for the household's quantity decision (see Table 9 for the price elasticities of demand for quantity outcomes, reported separately for each segment), we find that the heavy buyer segment (i.e., segment 2) is actually less sensitive to the marketing mix compared to segment 1.

Table 5: Estimated Parameters of the PROPOS Quantity Model and their Standard Errors in Parentheses.

Parameter	Segment1	Segment2
Intercept	-0.65 (0.15)	0.61 (0.17)
I_{ht}	-0.08 (0.02)	-0.02 (0.01)
K_h	1.16 (0.13)	0.11 (0.03)
V_{hit}	0.12 (0.04)	0.22 (0.05)

The estimated parameters (and their standard errors) of the incidence model of BENCH are given in Table 6. Category attractiveness has coefficients of 1.31 and 1.10 for the two supports of the heterogeneity distribution, which are both smaller in magnitude compared to their counterparts based on PROPOS (see Table 3). This implies that ignoring correlations in unobservables between the household's incidence and quantity decisions results in an under-estimate of the impact of the marketing mix on the household's incidence decision (which can also be ascertained by looking at the price elasticities of incidence reported in Table 10, separately for each segment).

Table 6: Estimated Parameters of the BENCH Incidence Model and their Standard Errors in Parentheses.

Parameter	Segment1	Segment2
Intercept	3.87 (0.42)	-0.50 (0.45)
CA_{ht}	1.31 (0.10)	1.03 (0.31)
$Week_{sht}$	-0.08 (0.01)	-0.05 (0.01)

We present the estimated parameters of the brand choice model of BENCH in Table 7. For all brands in the category, their estimated intercepts relative to the excluded brand are smaller. This implies that the market's preferences for national brands, relative to the excluded private label, are understated by ignoring correlations among unobservables between brand choice and quantity outcomes.

Table 7: Estimated Parameters of the BENCH Brand Choice Model and their Standard Errors in Parentheses.

Parameter	Segment1	Segment2
Generic	-2.28 (0.20)	0.64 (0.17)
Bounty	2.59 (0.16)	0.43 (0.22)
Viva	1.84 (0.18)	-1.10 (0.28)
Sparkle	-0.78 (0.15)	-2.16 (0.28)
Scott	2.59 (0.20)	-0.37 (0.37)
Gala	-1.15 (0.15)	-2.68 (0.34)
Mardi	-0.66 (0.15)	-2.11 (0.27)
Price	-8.06 (0.42)	-1.99 (0.58)
Display	1.16 (0.09)	1.20 (0.17)
Feature	0.01 (0.10)	0.11 (0.17)

The estimated parameters of the quantity model of BENCH are given in Table 8. The estimate of the coefficient of the brand's deterministic utility, V_{hjt}, is insignificant for each support of the heterogeneity distribution. This implies that ignoring correlations in unobservables between a households's quantity outcomes and its incidence and brand choice outcomes would lead one to falsely conclude that marketing variables do not influence the household's quantity decisions when, in fact, they do (as evidenced by Table 5).

Table 8: Estimated Parameters of the BENCH Quantity Model and their Standard Errors in Parentheses.

Parameter	Segment1	Segment2
Intercept	-0.28 (0.13)	-0.41 (0.10)
I_{ht}	-0.03 (0.01)	-0.02 (0.01)
K_h	0.98 (0.10)	0.28 (0.03)
V_{hjt}	0.02 (0.02)	0.05 (0.05)

In Table 9, we report the price-elasticities of incidence, brand choice and quantity out-comes, based on PROPOS.9 In Table 10, we report the counterparts based on BENCH (note that segment numbers have been assigned based on estimated segment sizes). It is clear that the estimated elasticities may be overstated in terms of their estimated magnitude by as much as 1.50, or understated by as much as 3.36, which indicates that elasticity distortions can be substantial from using the mis-specified benchmark model. Since the estimates of the coef- icients associated with V_{hjt} are not significantly different from zero for either segment under BENCH, quantity elasticities are estimated to be zero under the mis-specified benchmark. This underscores the fact that the substantive consequences of ignoring self-selectivity corrections in the quantity outcomes (the focus of this paper) can be quite severe.

Table 9: Estimated Price Elasticities of Demand based on PROPOS.

	Segment1			Segment2		
Brand	Incidence	Brand Choice	Quantity	Incidence	Brand Choice	Quantity
Private	-1.22	-3.98	-0.96	-0.63	-2.52	-0.04
Generic	-0.68	-2.91	-1.01	-1.15	-0.99	0.07
Bounty	-3.10	-4.83	-0.84	-0.44	-4.18	-0.16
Viva	-1.28	-6.08	-1.03	-0.13	-4.48	-0.46
Sparkle	-0.56	-4.80	-1.16	-0.08	-3.34	-0.52
Scott	-0.41	-8.87	-1.39	-0.17	-5.84	-0.45
Gala	-0.31	-5.26	-1.33	-0.05	-3.55	-0.67
Mardi	-0.48	-5.07	-1.21	-0.08	-3.47	0.51

Table 10: Estimated Price Elasticities of Demand based on BENCH.

	Segment1			Segment2		
Brand	Incidence	Brand Choice	Quantity	Incidence	Brand Choice	Quantity
Private	-1.81	-4.20	0	-017	-1.14	0
Generic	-0.70	-3.39	0	-0.34	-0.46	0
Bounty	-2.67	-6.10	0	-0.20	-1.73	0
Viva	-1.35	-7.06	0	-0.04	-1.94	0
Sparkle	-0.59	-5.57	0	-0.02	-1.46	0
Scott	-0.33	-10.37	0	-0.09	-2.49	0
Gala	-0.31	-6.12	0	-0.01	-1.56	0
Mardi	-0.55	-5.85	0	-0.02	-1.52	0

We develop a joint econometric model of incidence, brand choice and quantity decisions at the household-level, that treats quantity outcomes as discrete as well as corrects for the effects of endogenous self-selectivity in the quantity outcomes. Using scanner panel data on paper towel purchases, we find that the effects of self-selectivity corrections are very impor-tant from the standpoint of predicting purchase quantity outcomes (improving the quality of prediction of these outcomes by 62%), and that the effects of marketing mix on demand are understated if one does not take into account the effects of self-selectivity corrections. There are some possible areas of future research. First, comparing our approach to that of Allenby, Shively, Yang and Garratt (2004) would throw light on the relative advantage of having a statistically flexible model (such as ours) in comparison to a model that is grounded in the economic theory of utility maximization (such as theirs). Second, it will be useful to understand the cross-category generalizability of our empirical findings by estimating the pro-posed model on a wide variety of product categories and noting if cross-category differences emerge (Bell, Chiang and Padmanabhan 1999). Third, it will be of interest to see if one can understand the drivers of the estimated self-selectivity correlations using experimental data.

(B) Selection Bias in College Admissions Test Scores

Data from the two leading college admissions tests—the SAT and the ACT—can provide valuable measures of student achievement for researchers studying the economics of education. In contrast to accountability tests administered by individual states or school districts, for which individual scores are rarely reported (and even less often have consequences for the individual test-taker), college entrance exam scores have important consequences for students, who therefore have incentives to perform as well as they can. In addition, both the SAT and ACT have been administered in essentially the same forms for several decades, whereas states and districts typically substantially revise or discontinue their assessments every few years, making it difficult to assess trends over time. Finally, both the SAT and the ACT are administered nationwide to students from nearly every high school, and therefore permit comparisons both across and within states; alternative tests are often administered in only a single state or only to small samples of students.

However, an important concern about the use of ACT and SAT scores for research is the non-representativeness of test takers. Both exams are taken primarily by college bound students, who most likely perform better than would their non college bound peers. As a result, comparisons of mean SAT or ACT scores between schools or over time, for example, could be biased by the fact that only high achievers' scores are observed. Research has documented evidence of selection bias in state level mean SAT scores and suggested selection- Many states have implemented high-stakes testing policies which reward or sanction schools or teachers based on their students' performance on the state achievement tests. These may be the worst of both worlds: Students do not clearly face incentives to do their best, while teachers may tailor their teaching to the test or simply cheat (Roderick et al., 2002; Jacob and Levitt, 2003).Koretz (2002) for an overview of the research on score inflation under test based accountability policies, correction procedures for analysts working with state level data (Dynarski, 1987; Dynarski and Gleason, 1993). However, this work has focused on the SAT but not the ACT, which has become increasingly prevalent in recent years and which is now the predominant exam in many states. Furthermore, there is no available evidence on the degree of selection bias at the school or district level, and therefore little guidance for researchers conducting increasingly common analyses with student level test score data (for example, Krueger and Whitmore, 2001; Card and Payne, 2002; Rothstein, 2006; Abraham and Clark, 2006; and Hanushek and Taylor, 1990).

We provide new evidence on the degree of selection bias in mean SAT and ACT scores. We begin by reviewing the evidence on state level selection into test participation. Dynarski (1987) documented a strong negative across state correlation between SAT participation rates and mean scores. We show that much of this correlation is driven by the contrast between what we call "SAT states," where most college bound seniors take the SAT rather than the ACT, and "ACT states," where traditionally only a highly selected group of students hoping to attend elite, private, out-of-state colleges took the SAT. Across SAT states, SAT scores and participation rates are essentially uncorrelated. A similar analysis of ACT scores and participation rates among ACT states is also consistent with no selectivity, though point estimates are larger. In contrast, we find stable negative correlations between scores and participation rates on secondary exams (i.e. SAT scores among ACT states and ACT scores among SAT states). The estimated selectivity thus varies in important ways across states: SAT takers are highly positively selected in ACT states but not in SAT states, while ACT takers are somewhat positively selected in SAT states but (perhaps) not in ACT states. Each of these results is robust to controlling for scores on the NAEP exam, given to a random sample of 3 students in each state, as NAEP scores are approximately uncorrelated with test participation rates. We then explore various strategies for estimating the degree of selectivity into SAT- and ACT-taking across and within schools. Our preferred strategy takes advantage of a policy reform in Illinois that substantially increased ACT participation rates but, plausibly, did not have important effects on underlying student achievement.

Beginning with the high school class of 2002, the ACT exam was required for graduation. As a result, the Illinois ACT participation rate jumped from 71% in 2001 to 99% in 2002. Moreover, the impact of the new requirement varied substantially across schools, with larger increases in participation rates in schools that initially had low participation—so were far from compliance—and smaller increases in schools that already participated at high rates. Initial participation rates varied substantially with the school racial composition, with a strong positive association between the school white share and the participation rate in 2000. By 2004, the association between schools' white shares and their participation rates, while not entirely eliminated, was substantially attenuated. Racial composition is thus a strong predictor of the change in participation rates between 2000 and 2004. Under the plausible assumption that the relationship between a school's racial composition and the latent achievement of its students did not change over this four-year period, the former is a valid instrument for the change in participation rates. Results from this analysis indicate strong positive correlations between test participation and latent scores, both within and across schools: Schools with the highest mean latent achievement have the highest participation rates, and within schools those students who would score highest are most likely to write the exam.2 The one exception is that the specification that includes the NAEP score indicates significant selectivity into ACT participation in ACT states. We consider the implications of our estimates for analyses that do not have access to natural experimental variation in test participation and so do not adjust fully for selection. Across school variation in latent test scores is large relative both to within-school variation and to variation in test participation rates, so the bias in observed means is small relative to the underlying variation. Thus, despite the substantial selectivity into participation, observed mean scores are extremely highly correlated across schools with latent, unselected means. This result suggests that school-level analyses that fail to account for selectivity biases are unlikely to seriously misstate the relationships between admissions test scores and other variables, though these relationships will tend to be attenuated by as much as 25%. We then provide an econometric model of selection into test taking and describe our estimation strategies.

Data

For our school level analysis, we construct school level participation rates and mean scores from data sets containing samples of individual records on SAT and ACT-takers, matched to the high schools that they attended. Our SAT data set includes 100 percent of test takers in California and Texas, 100 percent of black and Hispanic test-takers, and a 25 percent random sample of test-takers of other races. The ACT data set includes 100 percent of ACT takers with missing race data, a 50 percent random sample of nonwhite students with nonmissing race data .5 and a 25 percent random sample of white students. Our school-level analyses use data from the cohorts that graduated from high school in 1996, 2000, and 2004; for reasons that will become clear below, we use 1996 data for state-level analyses.

The SAT reports two subscores for each student, math and verbal, and a composite score. Each of the subscores is reported on a scale with mean 500 and standard deviation around 100 among tested students (with a maximum score of 800 and a minimum of 200). The composite score is the sum of the math and verbal scores, and has mean 1000 and standard deviation around 200. The ACT reports several subscores, including one for a natural science subtest that has no SAT analogue. We focus on those that are analogous to the SAT math (the ACT math) and verbal (the sum of ACT English and reading) subscores, as well as the ACT "composite" score that averages over all subjects. We convert ACT scores from their native scale (ranging from 1 to 36 on each subtest) to the more familiar SAT scale.

Both the SAT and ACT data sets contain high school identifiers that can be linked to records from the Common Core of Data (CCD), an annual census of public schools. From the CCD, we extract the number of 12th graders at the school, the racial composition, and the fraction of students receiving free school lunches (an indicator of poverty). Because sampling rates are so high in both the SAT and ACT data, the simple ratio of the number of (weighted) observations SAT the school to the number of 12th graders provides an accurate estimate of the participation rate at all but the smallest schools. All of our school level analyses exclude schools with participation rates below 0.02. The SAT and ACT are competing exams, and market shares vary substantially across states. In some states—the Northeast, some of the South, and the Pacific coast—most college-bound students take the SAT, while in others—the Midwest and Plains, and portions of the South—the ACT is more common. Colleges that enroll primarily nearby students (public universities, for example) tend to prefer the dominant test in their state, while national, elite universities tend to prefer the SAT but accept either. When a college will accept either test score, students can sometimes arbitrage between the two, taking both tests and reporting only the higher score. As a result, students who take the less popular test in their state—particularly SAT-takers in ACT states—tend to be those who hope to attend the most selective schools.

Our state-level analysis focuses on average test scores and participation rates for the cohort graduating high school in 1996. Test taking rates come from *The Digest of Education Statistics* (National Center for Education Statistics, various years) for the SAT and from the ACT Corporation (2003) for the ACT. Average test scores by state are computed from the student level data, after converting individual ACT scores to the SAT scale as described above. We merge the state-level average scores and participation rates to state-level mean scores from the National Assessment of Educational Progress (NAEP; National Center for Education Statistics, 2006). Representative state-level NAEP scores are available only in selected years in selected subject areas, and only for 4th and 8th graders. We focus on NAEP 8 th grade math scores from 1992, which we merge to SAT and ACT data from 1996—the same cohort of students four years later. The scatterplot of SAT vs. ACT participation rates at the state level in 1996; there are two clear groupings and few states that do not fall into these groupings. We distinguish throughout between "SAT states" and "ACT states," defining each as those states where the participation rate on the relevant exam exceeds 40%. Table 1 presents descriptive statistics for the school level data sets. Not surprisingly, average SAT scores are much higher in ACT states, where only the strongest students write the SAT, than in SAT states. The same is true, to a much lesser degree, for ACT scores in SAT states. However, within SAT states, the selection appears to go the opposite direction: schools with high participation rates tend to have higher scores than those with low participation rates, on both the SAT and the ACT. Similarly, within ACT states, schools with high participation rates tend to have higher scores than those with low participation rates on both exams.

A Model of Selection into Test Taking

We work with a standard Gronau (1974), Heckman (1979) bivariate normal model of selection. As in Gronau (1974), the group (school) level selection probability is observed. We allow for unrestricted school effects on latent test scores and explore different strategies for identifying the parameters of the selection process (and thereby the underlying school effects), relying on progressively weaker assumptions about the relationship between these effects and the participation rate. We suppose that the latent test score of student i from school j in year t is t_{ijt}^*.

The mean latent score in school j at time t is α_{jt}, so

(1) $t_{ijt}^* = \alpha_{jt} + \varepsilon_{ijt}$.

Only a fraction of students actually take the test. Student i writes the exam (and so we observe $t_{ijt} = t_{ijt}^*$) if $\mu_{ijt}^* > 0$ where

(2) $\mu_{ijt}^* = \theta_{jt} + u_{ijt}$.

We assume that ε and u are bivariate normal with mean zero within each j-t cell. We define $\sigma^2 = Var[\varepsilon_{ijt}]$ and $\rho = corr(\varepsilon_{ijt}, u_{ijt})$, both constant across schools and years. We normalize $Var[u_{ijt}] = 1$.

We observe the participation rate in each school year cell, p_{jt}. Note the following relationship:

(3) $p_{jt}=pr[\mu_{ijt}>0 \mid j, t] = pr[\theta_{jt} + u_{ijt} > 0 \mid j, t] = pr[u_{ijt} > -\theta_{jt}] = \Phi(\theta_{jt})$.

Thus, we can write $\theta_{jt}=\Phi^{-1}(p_{jt})$. The mean and variance of ε_{ijt} conditional on t_{ijt} being observed are

(4) $E[\varepsilon_{ijt} \mid \mu_{ijt} >0] = E[\varepsilon_{ijt} \mid u_{ijt}>-\Phi^{-1}(p_{jt})] = \rho\sigma\lambda (p_{jt})$ and

(5) $Var[\varepsilon_{ijt} \mid \mu_{ijt}>0] =\sigma^2 [1 - \rho^2\delta(p_{jt})]$, where $\lambda(p_{jt}) \equiv \varphi(\Phi^{-1}(p_{jt}))/p_{jt}$ is the inverse Mills ratio (hereafter, *IMR*) and $\delta(p_{jt})\equiv\lambda(p_{jt}) [\lambda(p_{jt})+\Phi^{-1}(p_{jt})]$.

We can write observed individual scores as

(6) $t_{ijt}=\alpha_j +\rho\sigma\lambda(p_{jt}) + e_{ijt}$, where $e_{ijt}=\varepsilon_{ijt} - E[\varepsilon_{ijt} \mid \mu_{ijt} >0]$,

with $E[\varepsilon_{ijt} \mid j, t] = 0$ and $Var[\varepsilon_{ijt} \mid j, t]$ given by (5).

Cross-sectional analyses

We hope to estimate the selection parameters, with the ultimate goal of identifying schools' latent mean scores, the α_{jt}s. Depending on the data configuration, relatively weak or strong assumptions are required for this. We begin with the least-well-identified case, where only a single cross section of test scores is observed. A strong assumption, that there is no across school average selectivity—that $corr(\alpha_j, \lambda(p_j))=0$—is then required. If so, we can obtain an unbiased estimate of $\beta=\rho\sigma$ from a school level regression of average scores on the inverse Mill's ratio:

(7) $\bar{t}_j = \alpha + \beta\lambda(p_j) + ((\alpha_j - \alpha) + \bar{\epsilon}_j)$.

This regression is heteroskedastic, both because the number of observations used to compute $\bar{t}_j$ varies across schools and because $var(e_{ij})$ varies with p_j. A heteroskedasticty robust estimator for $V(\hat{\beta})$ is called for. Because p_j is observed directly, we need not account for error in estimation of "first-stage" coefficients as in typical implementations of the Heckman model.

To convert our estimate of β into one of ρ=β/σ, we need an estimate of σ^2. Note that equation (5) implies that

(8) $\sigma^2 = Var[e_{ijt}] + \rho^2\sigma^2\delta(pj)$.

We estimate the first term from the individual deviations from school mean scores:

(9) $\hat{E}(e^2) = (N-k)^{-1}\sum_{i,j}(t_{ij} - \bar{t})^2$, where k is the number of schools in the sample and N the number of student-level test score observations. The second term is estimated from the across-school average, weighted by the number of test-takers:

(10) $\hat{E}(\delta(p)) = \left(\sum_j N_{jt}\right)^{-1}\sum_j N_{jt}\delta(p_{jt})$.

This gives a straightforward estimate of the variance of residual latent scores:

(11) $\hat{\sigma}^2 = \hat{E}(e^2) + \hat{\beta}^2\hat{E}(\delta(p))$.

Our estimator for ρ is then straightforward: $\hat{\rho} = \hat{\beta}/\hat{\sigma}$. Inference on ρ might be by the delta method,

(12) $V(\hat{\rho}) = \hat{\sigma}^{-2}[V(\hat{\beta}) - \hat{\rho}^2 V(\hat{\sigma})] = \hat{\sigma}^{-2}\left[V(\hat{\beta}) - \frac{1}{4}\hat{\rho}^2\hat{\sigma}^{-2}V(\hat{\sigma}^{2})\right]$

though this would frequently indicate confidence intervals that extend above 1 or below -1. We report standard errors only for β; note that ρ equals zero if and only if β does.

First-differenced estimates

The assumption that there is no correlation between latent mean performance and test participation is implausible, as better schools almost certainly have both higher latent scores and higher rates of college going. Indeed, the cross-sectional correlation between IMRs and mean scores is negative, which can only be true if $\rho < 0$ or if $corr(\alpha_j, \lambda(p_{jt})) < 0$.

To weaken our assumption, we can take advantage of multiple observations on the selection test score relationship at the same school, from different years.

Suppose that we have observations from t=1 and t=0. Our identifying assumption is now:

(13) $corr(\alpha_{j1} - \alpha_{j0}, \lambda(p_{j1}) - \lambda(p_{j0})) = 0$.

That is, while there may be a cross-sectional correlation between a school's IMR and its latent average score, there is no correlation between the *change* in a school's IMR and the *change* in its latent mean score.

The difference between the expressions for school mean scores given by (6) in t=0 and t=1 is,

(14) $\bar{t}_{j1} - \bar{t}_{j0} = \beta(\lambda(\rho_{j1}) - \lambda(\rho_{j0}) + \alpha_{j1} - \alpha_{j0} + \bar{\epsilon}_{j1} - \bar{\epsilon}_{j0}$

Thus, with assumption (13) a regression of the change in school mean test scores on the change in IMRs estimates β. We can use the same approach discussed above to estimate σ and ρ.

With consistent estimates of β we can estimate α_{jt} as $\bar{t}_{jt} - \beta\lambda(p_{jt})$. In the first-differenced data, we can use this to compute the point-in-time correlation between α_{jt} and $\lambda(p_{jt})$. As the identifying assumption of the cross-sectional strategy is that this correlation is zero, the first-differenced specification can serve as a test of the cross-sectional specification. We can also estimate $corr(\alpha_{jt}, \theta_{jt})$, which measures across-school selection into participation in the same way that ρ measures within-school selection.

First-differenced estimates, with instruments

Even the assumption that the change in a school's IMR is uncorrelated with the change in its mean latent scores is unattractive, as the two may change together if a school's quality or demographic composition is evolving. With a suitable instrument for the change in the IMR, however, (14) can simply be estimated by instrumental variables.

Below, we form such an instrument by taking advantage of a policy change in Illinois in 1992 that mandated that all high school graduates take the ACT.

Because demographic characteristics of schools were correlated with pre implementation participation rates, the "bite"of these policies varied with school demographics. In particular, schools with higher minority shares had lower pre-implementation participation rates and larger increases in participation (and larger reductions in IMRs) surrounding implementation of the policy. Of course, minority share is correlated with mean latent test scores, but it is plausible that the base-year minority share is uncorrelated with the *change* in mean latent scores surrounding the implementation of the policies. If so, this base-year minority share forms a valid instrument for the change in IMRs surrounding the policy implementation.

State-Level Analysis

We begin by examining the relationship between test-taking rates and average admissions test scores at the state level. However, this inverse relationship is driven almost entirely by the ACT states and the contrast between ACT and SAT states; within the SAT states the participation-score correlation is much smaller. The suggestive of strong positive selection into SAT-taking in the ACT states, as higher participation rates are associated with lower scores, and with weaker selection in SAT states. This is consistent with the idea that in ACT states only high performing students who hope to attend elite out of state colleges take the SAT, while in SAT states all or most college bound students take the exam. The overall negative correlation between test taking rates and average scores shown in the SAT data is not as apparent in the ACT data. However, consistent with the patterns, there does appear to be an inverse relationship between average ACT scores and ACT participation rates in SAT states, although it is not nearly as pronounced as in ACT states in the SAT data.

Taken together, the two panels that selection into the secondary test in a state (the ACT in SAT states and the SAT in ACT states) creates bias in estimated mean scores, but that there is little bias in the dominant test. This is again consistent with the idea that students hoping to attend out-of-state schools are more positively selected than are students hoping to attend local colleges.

Of course, it is possible that latent ability is correlated with test taking rates, which would affect the observed correlations between test-taking rates and observed scores. For instance, if among the SAT states those with higher participation rates also tend to have higher mean ability, this could push the participation observed score correlation upward, masking what perhaps would be a clear downward slope. This can be seen as a form of omitted variables bias: The omission of a direct "ability" measure may bias the participation-observed score relationship to the extent that participation is correlated with ability.

To examine this, we turn to state-level NAEP scores, which are taken by random samples of students in each state and are therefore not subject to the selection bias that may be reflected in SAT and ACT averages. The scatterplot of scores on the 1992 8th grade —the same cohort as was in 12th grade in 1996— NAEP math test against SAT- and ACT taking rates, respectively. Within the ACT states, there is no clear relationship between NAEP scores and either the SAT or the ACT test-taking rate. However, within the SAT states, NAEP scores are positively related with SAT-taking rates but negatively related with ACT-taking rates.

Table 1. Summary statistics

	SAT States				ACT States			
	All Schools		Above average Participation	Below average Participation	All Schools		Above average Participation	Below average Participation
	Mean	SD	Mean	Mean	Mean	SD	Mean	Mean
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SAT								
Mean Score	978	100	1037	958	1104	145	1117	1096
School Size	285	201	304	279	239	153	270	218
No. of test-takers	130	109	209	104	35	55	65	15
Participation rate	46%	21%	71%	38%	13%	14%	22%	7%
Fr. white	68%	31%	78%	64%	83%	23%	82%	84%
# of Schools	5167		1291	3876	1882		756	1126
ACT								
Mean Score	970	122	983	962	963	89	999	949
School Size	296	202	254	323	174	144	179	173
No. of test-takers	42	49	56	32	99	92	132	86
Participation rate	16%	15%	26%	9%	57%	16%	75%	50%
Fr. white	68%	30	74%	64%	82%	26%	90%	78%
# of Schools	4143		1654	2489	4050		1147	2903

Notes: All data pertain to the 2000 cohort. Sample includes only public schools with participation rates (on the SAT in the first panel or the ACT in the second panel) above 2%. All statistics are computed from unweighted school-level data. Samples in columns 3,4,7, and 8 are schools with participation rates on the relevant exam above and below the state-level participation rate. School size is the size of a single grade cohort at the school.

The lack of correlation between NAEP scores—interpretable as a measure of latent ability—and test taking rates in the ACT states suggests that the potential endogeneity of the participation rate does not seriously bias the correlations for these states. In contrast, the positive correlation between NAEP scores and SAT participation rates in the SAT states suggests that the relationship between SAT participation rates and SAT mean scores may be biased downward by the omitted "ability" variable.

Table 2 presents regression versions of the cross sectional comparisons in replacing the participation rate measures with the inverse Mills ratios that are indicated by the bivariate normal selection model. The top panel shows results for SAT scores, with three specifications estimated on data from SAT states in the same specifications repeated for ACT states in while the bottom panel repeats the analysis for ACT scores. The first specification in each set is a regression of average scores on the inverse Mills ratio with no additional covariates; the second limits the sample to states for whom NAEP data are available (but does not control for NAEP scores); and the third includes a control for the cohort's average score on the NAEP math exam in 1992.

Table 2. Cross-sectional estimates of state-level selectivity into test-taking, 1996

	SAT States			ACT States		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: SAT						
Inverse Mills Ratio in test-taking rate in state	45.0	-40.7	38.7	110.4	122.0	121.2
NAEP 8th grade math scores, 1992	36.2	26.7	33.8	21.6	26.0	16.4
Exclude states without NAEP data	n	y	y	n	y	y
N (states)	23	19	19	25	21	21
R^2	0.06	0.05	0.39	0.52	0.56	0.78
Underlying parameters						
ρ	0.22	-0.20	0.19	0.52	0.56	0.55
σ	200.4	199.9	199.6	213.8	219.0	218.6
Panel B: ACT						
Inverse Mills Ratio in test-taking rate in state	28.4	30.5	21.2	77.4	77.5	89.9
NAEP 8th grade math scores, 1992	10.9	11.9	18.4	85.2	96.5	25.7
Exclude states without NAEP data	n	y	y	n	y	y
N (states)	23	19	19	25	21	21
R^2	0.11	0.19	0.21	0.04	0.03	0.92
Underlying parameters						
ρ	0.15	0.16	0.11	0.40	0.40	0.46
σ	195	195.3	194.3	193.4	193.4	196.3

Notes: Average ACT scores are computed from ACT micro-data, concorded to the SAT scale using the concordance developed by Dorans et al. 1999.

As expected, the models that do not include controls for NAEP scores are consistent with the results. Selection into participation is strongly positive for the SAT in ACT states, and weaker but still positive for the ACT in SAT states. The selection parameter is indistinguishable from zero for the dominant exam in each group of states. Little changes when the sample is limited to the set of states with NAEP data; in one case the IMR coefficient changes sign but it remains insignificant. Table 2 attempt to absorb the potential correlation between state-level participation rates and mean achievement by controlling for four-year-lagged NAEP scores. These controls have little effect on the estimates. Reflecting the strong correlation between lagged NAEP scores and SAT participation rates in the SAT states, the inverse Mills ratio coefficient for SAT participation in SAT states rises by 80 points, though it remains small and insignificantly different from zero. In the ACT states, estimates of selection bias remain large and significant for the SAT data and become significant for the ACT data.

School Level Analysis – Observational

The above analyses—identification issues aside—are suggestive of the bias that selective participation induces in estimated state-level mean test scores. Selection at the state level combines two distinct margins: Student sorting across schools that vary in their participation rates, and within-school selection. The latter is the most relevant for analyses of school-level test scores. As noted earlier, a school-level analysis has the potential to break apart the two components. We present observational analyses of the relationship between school-level test taking rates and average scores at the school level, first using cross sections of schools and then examining changes over time. Each approach requires implausible assumptions to identify the selectivity parameters, and each in turn yields implausible estimates of the key within school selectivity parameter, corr(ε, u).

Cross-sectional estimates

The scatterplot of school-level participation rates and mean test scores in the 2000 graduating cohort. The four panels show data for the SAT and ACT exams in SAT and ACT states; in each panel, both participation rates and test scores are deviated from state means. In contrast to the across-state relationships in all four school level scatterplots show upward slopes, indicating that schools with higher participation rates also post higher mean scores. The slope is steepest for

SAT scores in SAT states and least steep for SAT scores in ACT states. For ACT scores, we again see a steeper slope in states where the ACT is the dominant exam than in states where it is not. Finally, note that there is much more variability in participation in the exam that is the market leader than in the secondary exam. This is driven partly by low average participation rates in the secondary exam—so that no school's participation rate can be far below the state mean—but there is less variability above the statemean as well.

Table 3 presents regression versions of these cross-sectional comparisons, with similar results. The top panel shows results for SAT scores, with three specifications estimated on data from SAT states in columns 1-3 and the same specifications repeated for ACT states in columns 4-6, while the bottom panel repeats the analysis for ACT scores. The first specification in each trio includes state fixed effects; the second adds controls for the school racial composition and for the fraction of students at the school who are eligible for free school lunches; and the third adds county fixed effects. The inverse Mills ratio coefficients are uniformly negative, reflecting the positive relationship between participation rates and average scores, though all are reduced notably by the addition of the demographic controls.

If school participation rates are independent of mean latent scores, the negative relationship between inverse Mills ratios and average scores indicates negative selection into test participation within schools—that is, that the *weakest* students in each school are the most likely to take the college entrance exams. Table 3 reports the correlation between test-writing propensity and latent scores within schools implied by each regression. This is much more negative for the dominant exam in a state than for the secondary, but estimates of selection into the SAT in SAT states and for selection into the ACT in ACT states are quite similar. The final row in each panel of Table 3 shows the implied across-school correlation between latent mean scores α, and the propensity to write the exam, $\theta = \Phi^{-1}(p)$. The identifying assumption in Table 3 is that $\lambda(p)$ is uncorrelated with α across schools (that is, that school participation rates are independent of mean latent scores). Once this assumption is imposed, corr(α, θ) can only be non-zero due to nonlinearities in the mapping between $\lambda(p)$ and θ. As this mapping is approximately linear (with substantial nonlinearities only as p approaches 0), it is not surprising that the estimated across-school correlations are all approximately zero.

Table 3. Cross-sectional estimates of within-School selectivity into test-taking, 2000

	SAT States			ACT States		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: SAT						
Inverse Mills Ratio in test-taking rate at School	-170.0	-68.5	-77.9	-25.0	-19.5	-28.0
Controls	5.1	4.7	5.4	6.3	5.8	8.8
State FEs	y	y	n/a	y	y	n/a
Racial composition, fr. free lunch	n	y	y	n	y	y
Country FEs	n	n	y	n	n	y
N (Schools)	5167	5167	5167	1882	1882	1882
R^2	0.33	0.65	0.76	0.14	0.34	0.66
Underlying parameters						
ρ	-0.75	-0.36	-0.41	-0.15	-0.11	-0.16
σ	226.6	189.3	191.6	171.2	170.6	171.6
corr(x,θ)	0.013	0.013	0.005	0.001	.001	0.001
Panel B: ACT						
Inverse Mills Ratio in test-taking rate at School	-45.0	-2.7	-1.9	-158.6	-63.8	-80.3
Controls	5.5	4.2	5.9	6.4	5.3	7.5
N (Schools)	4143	4143	4143	4050	4050	4050
R^2	0.12	0.47	0.62	0.44	0.66	0.78
Underlying parameters						
ρ	-0.25	-0.02	-0.01	-0.75	-0.36	-0.44
σ	177.4	172.7	172.7	211.1	179.6	183.5
corr(x,θ)	0.001	0.000	0.002	-0.004	0.009	0.010

Notes: Sample includes public Schools with participation rates on the relevant exam (SAT in panel A: ACT in panel B) above 2% and non-missing demographic information. Data from Schools with missing data for any variable (including those from five states for which free lunch data are unavailable for nearly every school) are excluded.

Of course, the identifying assumption for Table 3, that the test-taking rate at a school tells us nothing about the quality of its students, is quite implausible. Schools differ in their quality, and quality is almost certainly correlated with rates of college attendance (and therefore with admissions test-taking rates). A positive relationship between latent scores and participation rates across schools would bias downward the relationship between observed scores and inverse Mills ratios, likely accounting for the large negative cross-sectional estimates observed here.

Time differenced estimates

One might hope to remove the correlation between test participation rates and school quality by focusing only on changes over time in school-level participation rates and scores. This strategy removes any fixed components of school quality, and identifies the selectivity parameters if changes in test participation are unrelated to changes in a school's latent score.

The changes in schools' participation rates and mean observed scores between 1994 and 2000. There is substantial variation in the changes in both participation rates and mean scores, but in each panel the two are uncorrelated. Table 4 presents regression models corresponding to these scatterplots. The inverse Mills ratio coefficients (and estimated ρ's) are uniformly positive, but generally quite small. This is particularly true in SAT states, where five out of six coefficients are insignificantly different from zero. Coefficients and ρ's are somewhat larger in ACT states, particularly for SAT scores.

Table 4. First-differenced estimates of within-School selectivity into test-taking, 1996 to 2000

	SAT States			ACT States		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: SAT						
Change in Inverse Mills Ratio in test-taking rate at School	4.6	9.3	5.0	18.2	20.0	32.7
Controls	5.1	4.8	5.7	14.0	14.1	20.6
State FEs	y	y	n/a	y	y	n/a
Country FEs	n	n	y	n	n	y
Racial composition, fr., free lunch	n	y	y	n	y	y
N (Schools)	4076	4076	4076	1132	1132	1132
R^2	0.01	0.04	0.31	0.01	0.02	0.45
Underlying parameters						
ρ	0.03	0.05	0.03	0.11	0.12	0.19
σ	181.8	182.0	181.9	172.0	172.1	173.7
corr(x,θ)	0.49	0.50	0.49	-0.04	-0.03	0.02
Panel B: ACT						
Change in Inverse Mills Ratio in test-taking rate at School	-0.4	1.2	7.2	19.4	22.3	18.6
Controls	6.9	6.9	8.9	6.8	6.8	7.8
N (Schools)	3122	3122	3122	3286	3286	3286
R^2	0.02	0.03	0.30	0.01	0.03	0.34
Underlying parameters						
ρ	0.00	0.01	0.04	0.11	0.13	0.11
σ	174.2	174.2	174.3	173.0	173.2	172.9
corr(x,θ)	0.08	0.08	0.10	0.48	0.49	0.48

Notes: Sample includes public Schools with participation rates on the relevant exam (SAT in panel A: ACT in panel B) above 2% and non-missing demographic information in both 1996 and 2000.

The first-differenced specifications do not impose assumptions about the cross-sectional correlation between school mean latent scores (α_j) and propensity to write the exam (θ_{jt}), assuming only that the projection of α_i on the inverse Mills ratio $\lambda(p_{jt})$ is the same in the two years. Thus, the estimates of corr(α_j, θ_{jt}) in the last row of each panel are no longer constrained to zero, though they in practice must be similar in 1994 and 2000. Accordingly, we report in the final row of each panel the average of the estimated corr(α_j, θ_{jt}) in 1994 and 2000. These are substantial for the dominant test in each group of states but small (though still positive) for the secondary test.

Even these differenced results, we believe, cannot be taken very seriously. The larger estimated ρ's (as compared with those in Table 3) suggest that differencing reduces the endogeneity of participation rates, but we have no assurance that it is eliminated. If school quality (or the composition of school enrollment) evolves at all between 1994 and 2000, one would expect that this would produce a positive correlation between the change in θ and the change in α, and that estimates in Table 4 will continue to be biased downward. To eliminate this bias, we need a source of variation in the change in participation rates that is exogenous to changes in school quality.

Illinois Policy Reform

We turn now to our analysis of data from Illinois, an ACT state, during the years spanning a policy change that made a version of the ACT a graduation requirement. Beginning in the Spring of 2001, all juniors in Illinois public high schools took the "Prairie State Assessment Exam" (PSAE), which included an entire ACT exam as one component. The ACT was administered under the same conditions as are used for ordinary ACT administrations, and scores derived from the ACT component of the Prairie State Assessment are counted as valid for most admissions purposes. The new requirement led to large increases in test participation in Illinois. It would be rash, however, to assume that this means that changes in inverse Mills ratios in Illinois schools surrounding the policy implementation are entirely exogenous to changes in school quality.

First, initial test-taking rates varied substantially across schools. As the new requirement had more "bite" in schools with lower initial participation rates, one might expect that schools with lower quality would have tended to see larger participation increases as they complied with the policy. Second, compliance with the policy was incomplete, with the average participation rate rising only to 89%

as calculated in our data in 2004 (as compared with 57% in 2000). If schools that complied fully differed in their quality from those that did not, this would again be a source of potential endogeneity of the change in test participation rates.

Fortunately for our analysis, however, it is relatively easy to identify clear patterns in the impact of the policy that are plausibly unrelated to changes in school quality. The scatterplot of school-level participation rates against school the fraction white at the school for Illinois high schools in 2000, before the Prairie State Assessment Exam was introduced, and in 2004. We distinguish between the Chicago Public Schools and those in other Illinois districts, as the two groups differed substantially in both dimensions in 2000.

Note that in the left panel, which uses 2000 data, there was a strong relationship between test participation rates and school racial composition, with much higher participation rates in schools with more white students. This pattern is still present, but much weaker, in 2004. Thus, the increase in participation between 2000 and 2004 was most dramatic for the schools with the lowest white shares. A similar pattern would hold if we used inverse Mills ratios in place of participation rates on the vertical axis. We exploit this feature of the data to generate variation in the "bite" of the Prairie State requirement that is plausibly exogenous to changes in school's latent test scores. Specifically, we use measures of the school racial composition in 2000 – the fraction black and the fraction Hispanic – as instruments for the change in inverse Mills ratios between 2000 and 2004. These are valid instruments if the relationship between school racial composition and mean latent score did not change over this time period. Table 5 presents first-stage models of the relationship between school racial composition in 2000 and the change in the ACT participation rate (first panel) or inverse Mills ratio (second panel) between 2000 and 2004.

Table 5. First-stage models for the change in ACT participation rates, 2000 - 2004

	ACT States except IL /CO		IL	
	(1)	(2)	(3)	(4)
Panel A: Change in fr. taking ACT, 2000-2004				
School fr. black, 2000	0.002	0.008	0.157	0.143
	(0.009)	(0.009)	(0.054)	(0.053)
School fr. Hispanic, 2000	0.032	0.045	0.356	0.322
	(0.019)	(0.020)	(0.070)	(0.077)
Chicago			-0.071	-0.052
			(0.053)	(0.054)
Change in skl. fr. black, 2000-2004		-0.206		0.309
		(0.061)		(0.382)
Change in skl. fr. Hispanic,2000-2004		-0.208		0.430
		(0.068)		(0.247)
Constant	-0.013	-0.010	0.265	0.257
	(0.002)	(0.002)	(0.012)	(0.013)
N	4,310	4,310	383	383
R^2	0.04	0.04	0.17	0.18
F statistic, 2000 racial composition	1.5	3.1	13.1	9.1
Change in ACT Inverse Mills Ratio, 2000-2004				
School fr. black, 2000	-0.004	-0.016	-0.284	-0.267
	(0.017)	(0.018)	(0.088)	(0.087)
School fr. Hispanic, 2000	-0.070	-0.096	-0.598	-0.558
	(0.051)	(0.054)	(0.119)	(0.131)
Chicago			0.115	0.091
			(0.086)	(0.087)
Change in skl. fr. black, 2000-2004		0.410		-0.383
		(0.105)		(0.624)
Change in skl. fr. Hispanic, 2000-2004		0.430		-0.525
		(0.159)		(0.403)
Constant	0.024	0.018	-0.412	-0.402
	(0.004)	(0.004)	(0.019)	(0.019)
N	4310	4310	383	383
R^2	0.05	0.06	0.19	0.19
F statistic, 2000 racial composition	1.0	2.1	13.6	10.1

Notes: Sample includes public Schools with ACT participation rates above 2% and non-missing demographic information in each of 1994, 2000, and 2004.

We present two specifications, including in the second controls for the change in minority share between 2000 and 2004. The first two columns show results for all ACT states except Illinois and Colorado (which implemented a policy similar to Illinois' during this period). We would not expect base-year racial composition to be a strong predictor of the change in test taking rates in states that did not change their policies, and this is supported by the regression results. The constants indicate that participation rates at all-white schools fell by about 1.3 percentage points on average between 2000 and 2004. Coefficients on the 2000 racial composition variables are positive but small, and suggest participation rates rose slightly at all-Hispanic schools. The much larger coefficients on the control variables measuring the change in racial composition in Column 2 indicate that a 1% rise in either the black or the Hispanic share at a school was associated with a 0.2% decline in the test participation rate over this period. Not surprisingly, all of these results are reflected in mirror image in the models for IMRs in the second panel.

Columns 3 and 4 repeat this analysis for Illinois high schools, including a dummy for each specification. These estimates are strikingly different from those for states not exposed to the policy: The intercepts indicate that participation rates at all-white schools rose by about 26 percentage points. The growth in participation was more than twice as large in all-Hispanic schools—though only five Illinois high schools had Hispanic shares above 85%—and about 60% larger in all-black schools. The coefficients on the change in racial composition between 2000 and 2004 are positive but extremely imprecisely estimated.

Table 6 presents analyses of the relationships between ACT participation rates and scores in Illinois. The first column reports a cross-sectional model using 2000 data. This is quite similar to the analogous model in Table 3, and like that indicates strong negative selection into test participation.

Table 6. Selection estimates from Illinois data

	Cross-sectional 2000	1st difference, 1996-2000	1st difference, 2000-2004			
	OLS	OLS	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
Inverse Mills Ratio in test-taking rate at School	-194.6					
	(16.8)					
Change in Inverse Mills Ratio in test-taking rate at School		8.6	88.0	149.4	76.4	101.9
		(15.6)	12.5	37.5	11.2	36.9
Chicago	-160.1	12.2	57.7	67.8	51.6	56.2
Controls:	(14.1)	(5.3)	(7.0)	11.8	6.9	10.6
Change in racical composition, 2000-2004	n	n	n	n	y	y
N (Schools)	390	390	390	390	390	390
R^2	0.73	0.01	0.23	0.16	0.30	0.29
ρ	-0.85	0.05	0.48	0.75	0.42	0.55
σ	228.7	174.9	183.4	199.0	181.2	186.3
corr(x,θ),2000	-0.05	0.67	0.75	0.80	0.74	0.76
corr(x,θ),2004			0.48	0.54	0.47	0.49

Notes: Sample includes public Schools with ACT participation rates above 2% and non-missing demographic information in each of 1996, 2000, and 2004. Instruments in columns 4 and 6 are the fraction black and fraction Hispanic at the School in 2000.

The second column shows a first-differenced specification using changes between 1994 and 2000, as in Table 4. Like the earlier first-differenced models, this indicates zero within-school selectivity and strong positive between-school selectivity. The remaining columns of Table 6 report differenced estimates from data spanning the Illinois policy change, taking as the dependent variable the change in school mean test scores between 2000 and 2004 and the key independent variable the change in inverse Mills ratios.

Column 3 presents OLS estimates. These differ dramatically from those in Column 2, indicating a within-school correlation between participation and scores of 0.47. Evidently, the endogenous component of the change in participation rates between 2000 and 2004—that part that derives from changes in school quality—is small relative to the exogenous component generated by the Illinois policy reform, so the downward bias is much smaller than in Column 2.

Nevertheless, there may be a component of the change in participation that is endogenous to the change in school quality. To address this, Column 4 presents estimates that instrument for the change in inverse Mills ratios with the year-2000 racial composition. Assuming that racial.24 composition in 2000 is uncorrelated with the change in latent scores between 2000 and 2004, this provides a consistent estimate of both the within- and across-school correlations. The estimated within-school correlation is 0.75, notably larger than that in Column 3. Within schools, the students who take the exam are evidently strongly positively selected. One threat to the identification strategy here is that racial composition in 2000 may be correlated with the change in racial composition over the next four years, which could be reflected in (or could itself reflect) a school's mean latent score. Columns 5 and 6 repeat the OLS and IV specifications with direct controls for the change in racial composition between 2000 and 2004. Schools that saw increases in their minority shares saw declines in their scores, but the inclusion of these controls serves only to slightly reduce the selectivity parameter of interest.

The last rows of Table 6 present the across-school correlations between participation and latent scores implied by the estimates. The estimates that take advantage of the exogenous variation in participation rates produced by the policy change indicate year-2000 correlations that are notably larger than those seen in the earlier columns. Correlations in 2004 are smaller than in 2000, consistent with the idea that the policy served to reduce the "signal" variation in test participation that might correlate with the quality of the school or its students.

Implications for Analyses of Unadjusted Means

The final component of our analysis is to draw out the implications of our results for analyses that, because they lack a source of exogenous variation in participation like that produced by the Illinois policy change, are unable to correct for the causal effect of test participation on observed scores. We cannot estimate the precise sample selectivity bias without producing the exact specification that a researcher might attempt to estimate, but we can report some relevant correlations. The key parameter for school-level analyses is the correlation between the observed mean score at a school and the underlying latent mean, α. To the extent that this correlation is large, schools with high observed scores will tend also to have high latent scores, and analyses that treat the observed mean score as a measure of the school's quality will not be unduly affected by selection bias. (That is, coefficient signs will likely be correct; as we demonstrate below, the scale of estimated effects will be smaller, as observed means are less variable than are latent means.)

To see this, consider a regression that would be correctly specified if the dependent variable was true mean latent scores α, but suppose that the observed mean score $\bar{t}$ is used instead. Let R be the coefficient from a projection of *jt t* onto á jt. Then if the true model is:

(15) $\alpha_{jt} = X_{jt}*\beta+\varepsilon_{jt}$

we can write (neglecting intercepts)

(16) $\bar{t}_{jt} = \alpha_{jt}R + (\bar{t}_{jt} - \alpha_{jt}R) = X_{jt}*\beta*R + \varepsilon jt + (\bar{t}_{jt} - \alpha_{jt}R)$.

The bias in an estimated regression coefficient (relative to βR) depends on the coefficient of a projection of $(\bar{t}_{jt} - \alpha_{jt}R)$ onto X. This will vary with the choice of variables to include in X, of course, but it will necessarily be small if Var $(\bar{t}_{jt} - \alpha_{jt}R) =$ Var($\bar{t}_{jt}$)*(1-(corr($\bar{t}_{jt}, \alpha_{jt}$))^2) is small; that is, if corr $(\bar{t}_{jt}, \alpha_{jt})$ is large.

Table 7 presents various relevant parameters that are implied by the estimates in Table 6. The first row repeats the within-school correlation, ρ. The next several rows show estimates of the across-school standard deviations of average observed scores, mean latent scores, the school-level participation parameter θ, and the selection term in models for observed average scores, $\sigma*\rho*\lambda(p)$.

Table 7. Implications for studies with imperfect controls for sample selectivity: Estimates from Illinois

	Table 6 Column 4		Table 6 Column 6	
	2000	2004	2000	2004
	(1)	(2)	(3)	(4)
Within-School, individual-level correlation (ρ)	0.74		0.54	
Standard deviation across Schools				
tbr	112.0	113.4	112.0	113.4
α	146.6	128.7	135.0	123.4
θ	0.53	0.60	0.53	0.60
$\rho*\sigma*\lambda(p)$	45.8	28.3	31.2	19.3
Across-School correlations				
Corr(tbr, α), unconditional	0.97	0.98	0.99	0.99
Corr(tbr, α), conditional on racial composition	0.93	0.95	0.96	0.98
Net correlation (unconditional)				
corr($\alpha+\epsilon$, $\theta+u$)	0.75	0.68	0.60	0.52
Controls for racial composition	n	n	y	y

Notes: Estimates in column 1 and 2 derive from the specifications reported in columns 4 and 6 of Table 6, respectively.

All are computed from the same Illinois data used in Table 6. In both 2000 and 2004, the across-school standard deviation of mean observed scores is about 112 SAT points. This is notably smaller than the standard deviation of latent scores, which is around 140 in 2000 and 125 in 2004; comparing this to the within-school standard deviation (σ) of around 185 indicates that about 35% of the variance of latent test scores is between schools. By comparison, the standard deviation of (σ) is around 0.6. As the within-school variation in test participation is normalized to have variance one, this indicates that only about one quarter of the variation in test-taking propensity is between schools. The next rows of Table 7 show the correlation between observed and latent scores, first unconditional and then conditional on the school racial composition. These are extremely high, all well above 0.9 and generally above 0.95. Thus, the omitted selectivity correction in models that use unadjusted observed mean test scores is quite small, with a standard deviation around 20 points. R – the coefficient of a projection of observed means on latent means – ranges from around 0.7 to around 0.9. Coefficients in regressions like (16) will be attenuated relative to the true β by 10-30%, but any additional bias is likely to be quite small.

To illustrate this further, the scatterplot of school mean observed scores against estimated latent means (estimating α_{jt} as $\bar{t}_{jt} - \rho\sigma\lambda(p_{jt})$ in the left-hand panel, and the scatterplot of the residuals of these variables from regressions on school fraction white in the right-hand panel. Both show extremely high correlations, suggesting that qualitative results from regressions that use school mean test scores as the dependent variable are likely to be invariant to corrections for selection bias. On the other hand, reinforces the result that latent scores are notably more variable than are observed means, so the magnitude of estimated coefficients will not be invariant to this substitution. Indeed, while the regression of school mean observed scores on the fraction white—the source of the horizontal axis in the right-hand 26.27 panel has a coefficient of 258, when estimated latent scores are used instead the coefficient is 304. The final row of Table 7 displays the overall selectivity of test participation, combining the across- and within-school components. The estimates range from 0.5 to 0.75, suggesting that test-takers are strongly positively selected on the whole. This is the relevant parameter for correcting state mean scores, supposing that the relative importance of within- and between-school components of selectivity do not vary across states. Of course, all of these estimates are derived from models estimated from data on Illinois ACT-takers in the early 2000s. They are dependent on the validity of the exclusion restrictions underlying our estimates, and even if these are valid, they may not generalize to other states or to SAT scores.

On the other hand, several patterns in the data suggests that the basic conclusion—that selection bias is unlikely to be an important determinant of the qualitative results of analyses that take school- or state-mean test scores as the dependent variable—is likely to be quite robust. Because test-taking rates are less variable than observed scores, and because the two are so highly correlated, observed test scores are unlikely to be seriously misleading about rankings of states or schools even if selection into test participation is much more extreme than our estimates indicate. As a final exercise, we computed latent scores under the assumption that the selection coefficient, ρ, equals one. The correlation between observed and latent school means in Illinois is 0.99 when we use our estimated ρ and falls only to 0.93 when we impose $\rho = 1$.

Table 8 presents several alternative calculations of this correlation, which is extremely robust. Columns 1 and 3 focus on mean test scores, while Columns 2 and 4 convert both observed and latent scores into ranks within the state of Illinois, a common way of comparing schools. We also

explore specifications that narrow the comparision group. While the first row of the Table uses all Illinois schools to compute the bivariate correlation, the second row presents the correlation controlling for school racial composition, and rows 3 and 4 limit the sample to the relatively homogenous subset of schools that serve the Chicago suburbs. All correlations are extremely high: When we use our estimated ρ, they range from 0.94 to 0.99; even when we assume that $\rho = 1$ the range is 0.8 – 0.94. With such high correlations, there is little room for substantive bias even when test scores are not corrected for selection into test participation.

Table 8. Correlations between observed mean ACT scores and estimated latent scores, Illinois, 2000

	Using estimated ρ		Using ρ=1	
	Mean Score	Rank	Mean Score	Rank
	(1)	(2)	(3)	(4)
All Schools				
Scores	0.99	0.97	0.93	0.89
Residual scores	0.97	0.97	0.87	0.87
Chicago suburbs				
Scores	0.97	0.97	0.91	0.90
Residual scores	0.95	0.94	0.83	0.80

Notes: Schools with fewer than 50 students per cohort are excluded. N=348 (94 in Chicago suburbs). "Residual scores" are residuals from (statewide) regressions of mean scores and latent scores on the School fraction white.

Researchers often use scores on college admissions tests as measures of student achievement. But since only a select group of students take these tests, analyses that use selected scores as a proxy for the true latent test score may be biased. Correcting the selection bias has been difficult: important unobserved variables likely impact both the propensity to take the test and the latent test score, and as a result standard selection corrections without an instrument perform poorly. We use a policy change in Illinois requiring all high school juniors to take the ACT as a source of plausibly exogenous variation in changes in school-level participation rates over time. Using this instrument, we find that test takers are strongly positively selected both across and within schools: Participation rates are higher at higher-performing schools, and within schools higher-achieving students are more likely to take the test. The across-school variation in latest test scores is large relative both to within-school variation in latent scores and to across-school variation in participation rates. As a result, observed mean scores are highly correlated with latent means, and selectivity bias serves to attenuate variations in school mean scores but is unlikely to introduce further bias. Thus, despite the strong selectivity of test. participation, simple analyses using observed school-level test scores without controls for selectivity are unlikely to produce misleading results.

Now, we pass on to apply cluster analysis for Australian meat consumption behavior. We have adopted hierarchical method of clustering.

CHAPTER 4

EXPERIMENTAL RESULTS AND ANALYSIS

4.1 INTRODUCTION

In this chapter we adopt tools to empirical data for identifying crucial variables which when accommodated in clustering can make selectivity bias trivial. We also present estimation results and associated analyses.

4.2 CHOICE OF VARIABLES FOR PROBIT ESTIMATION

At the outset we note that we have used data on Australian household meat consumption behavior. Such data were retrieved from the latest available household survey archive.

We have principally included the socio-demographic variables for probit estimation. However, we have also included some variables like duration of unemployment, number of employed and unemployed persons in the households, which have some economic implications. We present the explanatory variables with the rationale used for the probit estimation in the following table:

Table 4.1

VARIABLES CHARACTERISING AUSTRALIAN HOUSEHOLD MEAT CONSUMPTION BEHAVIOR

VARIABLES	LOGIC
CBH=Country of birth of household head	Influence of ethnic background
CBS=Country of birth of household spouse	Influence of ethnic background
DUH=Duration of unemployment	Employed may have different consumption level
OHH=Occupation of household head	Effect of social factor
N_1=Number of persons in the age group up to 19 years	Effect of family composition
N_2= Number of persons in the age group up to 20+	Effect of family composition
NHO=Nature of household occupancy	Owner of a house can afford to spent more on meat
NDIS=Number of dependents including students less than 24 years	Effect of additional expenditures
RSC= State of Residence category	Average consumption differs by region
Y_s=1/(1+No. of years in Australia for spouse(YAS))	Effect of change of tastes
Y_h=1/(1+No. of years in Australia for household head(YAH))	Effect of change of tastes
FRZ= Freezer ownership	Storing tendency
T = Meals out and take-away	Lifestyle effect

Now, we present the probit results for the whole sample followed by an analysis below:

Table 4.2

PROBIT RESULTS FOR THE WHOLE SAMPLE

DEPENDENT VARIABLE	: MEGR
LOG OF LIKELIHOOD FUNCTION	= -712.556
NUMBER OF OBSERVATIONS	= 4492
NUMBER OF POSITIVE OBSERVATIONS	= 4321
PERCENT POSITIVE OBSERVATIONS	= 0.961932
SUM OF SQUARED RESIDUALS	= 164.437
R-SQUARED	= 0.652644
PERCENT CORRECT PREDICTIONS	= 0.961932

Variable	Parameter Estimate	Standard Error	t-statistic
Intercept	1.75138	.374151	4.68095
CBS	-.366227	0.041006	-0.089311
CBH	0.016302	0.040941	0.398184
DUH	-0.062560	0.041328	-1.51373
N_1	0.074839	0.075960	0.985233
N_2	0.019400	0.099750	0.194484
RSC	-0.026490	0.017011	-1.55728
OHH	-0.387871	0.012693	-0.305572
Y_h	0.084636	0.455893	0.185650
Y_s	0.057788	0.236381	0.244471
NDIS	-0.399860	0.088646	-0.451076
NHO	-0.083590	0.043802	-1.97839
FRZ	-0.229768	0.377624	-0.608456
T	**0.430853**	**0.204168**	**2.11029**

4.3 ANALYSIS OF PROBIT RESULTS

On the basis of individual t-values, all the explanatory variables except one, namely, "Take-away food", have resulted to be insignificant. On the whole, we accept the presence of selectivity bias. If we analyze the consumption behavior of households having positive total meat expenditure only, we are not estimating a random sample.

However, in order to make further verification for selectivity bias, we have also performed a joint test in addition to t-test. The test we have applied is the Likelihood Ratio (LR) test. We present the results of the joint test below:

LR=2(Unrestricted log likelihood – Restricted log likelihood)

Restricted log likelihood=$n_1$log(Probability of positive observations) +

(N–n_1) log(Probability of zero observations)

Where, n_1 is the number of positive observations and N=4,492 (Sample size)

$$=4321(-0.038) + 171\ (-3.270) = 726.60$$

Hence, the likelihood ratio test statistic which is a χ^2 as

$$2[-712.556 + 726.6] = 28.08$$

The tabular values of χ^2 at 10%, 5% and 1% levels of significance for 18 degree of freedom are respectively $\chi^2_{.10}(15)=22.30$, $\chi^2_{.05}(15)=24.99$, $\chi^2_{.01}$ (15)=30.57 according to the LR test, there is evidence of selectivity bias at 5% level but not at 1% level of significance.

4.4 GENERAL CONCEPTS FROM TEST FOR SELECTIVITY BIAS FOR THE WHOLE SAMPLE:

There is selectivity bias for the whole sample (N=4492). So, if we use a non random sub sample of households having positive meat expenditure (ME) only, the parameter estimates of the structural model are biased. Then a relevant question arises as to how to estimate some specified model, which necessitates positive values say, budget for meat purchase. So, we address this issue below.

4.5 A PLAUSIBLE ATTEMT TO MINIMISE THE EFFECT OF SELECTIVITY

At this stage, we recall the usefulness of cluster analysis discussed before. We hypothesis that cluster analysis may make selectivity bias a minor problem. This hypothesis arises from the fact that the same features of the data set, which govern selectivity as a problem may also govern cluster formation. As a result, selectivity problem may not remain dominant any more having similar selectivity variable across observation in a particular cluster. Therefore, in line with our previous argument about clustering the consumers, we have performed cluster analysis for Australian households details of which are presented in the next section. So, for further investigation of selectivity bias, we have performed probit analysis clusterwise. We present the clusterwise probit results in the following table followed by an appropriate analysis. We note that we have identified two broad clusters of Australian household meat consumers. These two clusters differ remarkable in some respects such as, in terms of average meat expenditure (both total and meat types), average income and other socio-demographic variables.

4.6 CLUSTERWISE PROBIT RESULTS AND ANALYSIS

Table 4.3

CLUSTERWISE PROBIT RESULTS

CLUSTER 1

DEPENDENT VARIABLE	: MEGR
LOG OF LIKELIHOOD FUNCTION	= -447.5109
NUMBER OF OBSERVATIONS	= 2874
NUMBER OF POSITIVE OBSERVATIONS	= 2766
PERCENT POSITIVE OBSERVATIONS	= 0.962422
SUM OF SQUARED RESIDUALS	= 104.849
R-SQUARED	= 0.955760
PERCENT CORRECT PREDICTIONS	= 0.962122

Variable	Parameter Estimate	Standard Error	t-statistic
C	-1.53389	7026.72	-.218294
CBS	-.197219	.047388	-.041618
CBH	-.017947	.048271	-.371788
DUH	3.41380	7026.72	.485832
N_1	0.035299	.088775	.397621
N_2	-.024297	.116416	-.208703
RSC	-.034819	.020637	1.68723
OHH	-.488864	.015193	-.032178
Y_h	-.040566	.551970	.073494
Y_s	.024929	.300320	.083009
NDIS	.116588	.114485	1.01837
NHO	-.066686	.060844	-1.09601
FRZ	-.389963	.423216	-.921428
T	**.467936**	**.25664**	**1.82325**

CLUSTERWISE PROBIT RESULTS AND ANALYSIS

Table 4.4

CLUSTERWISE PROBIT RESULS

CLUSTER 2

DEPENDENT VARIABLE	: MEGR
LOG OF LIKELIHOOD FUNCTION	= -265.0451
NUMBER OF OBSERVATIONS	= 1618
NUMBER OF POSITIVE OBSERVATIONS	= 1555
PERCENT POSITIVE OBSERVATIONS	= 0.96039
SUM OF SQUARED RESIDUALS	= 98.568
R-SQUARED	= 0.975644
PERCENT CORRECT PREDICTIONS	= 0.96319

Variable	Parameter Estimate	Standard Error	t-statistic
C	-1.021749	0.43574	0.781706
CBS	-0.169008	0.006382	-0.047693
CBH	-0.001645	0.00733	-0.026936
DUH	3.35124	0.43547	0.514168
N_1	0.03954	0.012815	0.587612
N_2	-0.004897	0.01666	-0.014219
RSC	-0.007919	0.004318	-0.12995
OHH	-0.100993	0.0025	-0.273394
Y_h	-0.04407	0.096077	0.1117606
Y_s	0.0322859	0.063939	0.161462
NDIS	0.516448	0.025839	-0.567294
NHO	-0.016904	0.017042	-0.88238
FRZ	-0.160195	0.045592	-0.312972
T	**0.532064**	**0.268824**	**0.28704**

We present an analysis of the above tabular results below:

On the basis of individual t-values, all the explanatory variables have resulted to be insignificant in both the c lusters. It is obvious that clusterwise results of probit estimation differ from those of the whole sample. There are remarkable differences between whole sample test results and clusterwise test results in terms of sign and magnitude of parameter estimates.

However, as in the case of the whole sample, we have also performed LR test clusterwise. We present the results of clusterwise LR test below:

Cluster 1:

The restricted log likelihood is 2766(log 0.962)+108(log 0.038)= – 458.286.

LR=2[– 447.51 +458.28] =21.54

Analysis: The LR test statistic is insignificant even at 10% level of significance. So, we draw the following conclusion.

Conclusion: We can assume no selectivity bias in cluster 1 at 5% level of significance.

Cluster 2:

The restricted log likelihood is 1555(log 0.960)+ 63(log 0.040)= – 264.91.

LR=2[– 261.173 +264.910] =7.48

Analysis: On the basis of both individual t-values and LR test statistic we do not have any evidence of selectivity bias. The LR statistic is still lower than that in cluster 1.

So, we draw the following conclusion.

Conclusion: We can assume no selectivity bias in cluster 2 at 5% level of significance.

Now, we need some interpretation and assessment about the overall test, results for the whole sample as well as for clusters to reach a final strategy. We discuss this issue below.

4.7 COMPARISON BETWEEN THE WHOLE SAMPLE AND CLUSTER SAMPLE TEST RESULTS

If we consider 1% level of significance, we can assume no selectivity bias, neither for the whole sample nor for clusters. At 5% level of significance, we may have selectivity bias when we consider the whole sample but this is not true after having obtained the clusters of households. The test results also suggest that the sample as a whole bears high degree of heterogeneity among households in terms of personal characteristics which lead to unequal probability of sample inclusion of households. The next question arises as to why the whole sample and cluster sample results differ. We discuss this issue below.

4.8 HOW HIERARCHICAL CLUSTERING OF CONSUMERS COULD HAVE MADE SELECTIVITY PROBLEM TRIVIAL?

By clustering the households, we can reduce and/or remove selectivity bias due to internal cohesiveness and external non-cohesiveness of clusters. Clustering brings alike objects (households) together, so it removes (or reduces) the selectivity bias as well (at a particular level of significance). As homogeneous households from a group, within groups selectivity becomes non-influential since probability of sample inclusion becomes identical for all households in a cluster. The selectivity variable $\Phi(X'\beta)/\varphi(X'\beta)$ for positive and $\Phi(X'\beta)/1-\varphi(X'\beta)$ for zero are influenced by X variables. Here Φ and φ indicate the standard normal pdf and cdf respectively evaluated at $(X'\beta)$. Since in a cluster households have similar X values, the selectivity variable which is a function of the probability of sample inclusion is similar for all households. Thus, irrespective of whether a household has positive or zero meat expenditure, the value of the selectivity variable will be alike for households in a cluster. The more homogeneous the households are in a cluster, the more the above argument is true. In our case, the sample selection criterion also governs the formation of clusters. Thus, we did not find any symptom of non trivial selectivity bias after formation of clusters. So, if we omit the household having zero total meat expenditure clusterwise we do not suffer from severe selectivity problem.

However, we note that if the clusterwise test results would be different for both the clusters or for either one of them, then we would need a different strategy to adipt. One way could be to go down along the dendogram and find more number of clusters at a lower node which provide more homogeneity within clusters and as a result selectivity bias could be reduced. But, fortunately we could (may be data specific) get rid of selectivity bias (at the usual 5% level of significance) at a higher node which also suggests that the variables and the clustering techniques used for the present purpose are quite appropriate to obtain homogeneous clusters even at a higher node on the dendogram. Thus, in order to arrive at a specific strategy, we address, two important questions viz.,

1. Should we cluster the whole sample first and then consider only households having positive ME clusterwise?
2. Should we include only households having positive ME to cluster and then apply the model clusterwise?

We term the type of clustering for question 1 as unrestricted and that for question 2 as restricted clustering respectively.

Our foregoing results and arguments answer question 1 affirmatively. If we estimate our augmented model (demand system) or only households having positive ME after forming clusters (clusterwise), the parameter estimates do not suffer from selectivity bias. The answer to question 2 is also provided by the test results. Comparison of test results based on the whole sample with those based on clusters prompt to adopt unrestricted clustering. Thus, the homogeneity assumption of the parameters across households is more appropriate in a particular cluster than that for the whole sample.

Now, relevantly we note that if we consider only those hopuseholds, which have positive total meat expenditure (ME) and do clustering there will be different implications on the formation of clusters as well as on structural parameters. It is definitely true that forcible expulsion of households having zero total meat expenditure will affect cluster formation. It may be that in some cases, out of two particular individuals (households) having similar characteristics, one will have zero and the other will have positive expenditure. Clustering with all households will bring these two households together since they are similar in nature. But, removal of households having zero expenditure will bring more dissimilar households than similar households together to form clusters. As a result, degree of homogeneity within cluster will be reduced. This will in turn impact the structural parameter estimates in addition to the effect of selectivity bias.

Thus, we take up the following final strategy out of the selectivity bias test results.

FINAL STRATEGY

We uncover cluster structures first and then apply structural model clusterwise. In doing so, we may not suffer from bias by selecting a non random sub sample, which may be needed for a structural model at hand. However, as a further piece of research we have applied a system model to analyse Australian meat consumption behavior after minimising selectivity bias. We relevantly present such results herein.

We have principally presented the estimation results of the structural model in addition to the results of (i) model selection and (ii) test of the assumption of the homoscedastic error structure. We have done all of the above clusterwise. The major findings were as follows:

Model selection strategy, namely, LR test criterion, enabled us to identify two parsimonious models for two groups of consumers. With respect to meat consumption behavior, the results have some similarities and some dissimilarities The most influential factors determining the household meat consumption behavior in Australia have turned out tobe real total meat expenditure (included by real total income) and expenditure on take-away foods. Among demographic variables, two age groups, namely, household members being 5 –<15 years and 15–64 years, significantly affect consumption behavior of some meat types but not all. Similarly, occupation category of household head and ownership of a freezer or microwave also influence purchase decisions of some meat types although not identically. The least influential factor has turned out to be the ownership of a house. All the meat types appear to be either necessity or luxury goods. In terms of total expenditure elasticity, the two groups of consumers appear to differ remarkably.

The overall results favor cluster-based analysis for meat consumption behavior. Based on the results of test of the assumption of homoscedastic error structure, we can be quite confident about the validity of such an assumption made in our case.

In this thesis we have made an attempt to analyse the determinants of the household meat consumption expenditure pattern in Australia. The data set used has been obtained from the household level survey conducted by the Australian Bureau of Statistics in 1984. We put particular emphasis on household level studies. We have highlighted some statistical facts which provides some motivation for considering Australian meat consumption as an item for research.

We have based our analysis on a modified Engel curve in a system framework derived from the AIDS model of Deaton and Muellbauer (1980). The final model structure is a two-step modeling strategy adaptable to the special features of the data set to be analysed. In the modeling strategy, we have paid special attention to the censoring nature of the dependent variable (meat expenditure).

For model selection, we have adopted the likelihood ratio test criterion. We have performed such tests in four phases, namely, for all the socio-demographic and economic factors except total meat expenditure, only demographic factors, socio economic factors and each individual factor (placewise). The test results support about the incorporation of socio-demographic factors in demand analysis instead of relating share expenditures to total expenditure only. Our final model selections based on placewise LR test results.

The treatment of micro-level survey data in the framework of a system of censored regressions as conducted in the present study can enable us to overcome some problems with demand analysis. One such problem is with deriving demand equations from utility maximisation for which some strict form of utility function is necessary.

Another important aspect when dealing with cross-section data is the treatment of zero expenditure occurrences on one or more consumption items for a particular observation. A modeling and estimation strategy which can not take account of this aspect properly, yields biased results. The modeling approach adopted in the present study, although not the only way, can elegantly deal with such an aspect. In this approach, we provide sufficient statistical properties of the parameter estimates and we overcome the computational problem with a large system of equations as well. Moreover, in our approach we do not have to discard much information except 3.8% of the households which have zero expenditures on all meat types and we are providing an adequate and appropriate analysis.

As Blundell (1988, p16) points out, "Perhaps the most appealing feature of economic research into consumer behavior is the close relationship between theoretical specification and appropriate estimation technique". We believe that the present study satisfies such a requirement quite well. But, we do not claim that the particular modeling strategy chosen in the present work is the only one or the best one. However, it does have a strong theoretical background. Some of the results obtained in the present study are also consistent with those of other studies on Australian household meat consumption.

It is desirable that the Engel curves should be compatible with some recognised demand models. Similarly, it is also desirable that the Engel curve system should satisfy the adding-up constraint. In our derivation of the Engel curve system from a complete demand system (AIDS model), we have such a guarantee. The current model specification takes account of the effects of total meat consumption expenditure as well as socio-demographic factors maintaining the adding up constraint.

For estimating the system model, we have adopted Lee's (1978) instrumental variable estimation technique for a system of censored regressions. We have demonstrated that for our system model where the simultaneity accrues through the error term, Lee's estimators as well as the covariance matrix of the estimators are exactly those of the standard seemingly unrelated regression system. The estimation strategy adopted in the present study provides sufficiently good statistical properties of the parameter estimates as well. The estimation results revealed the existence of distinctive features of group structures of households.

We have tested the assumption of a homoscedastic error structure of the model. In order to perform such a test, we adopted Harvey and Phillips (1981) test statistic. According to the test results, the assumption of a homoscedastic error structure is accepted for all the equations in cluster 1 at the 5% level of significance and for cluster 2 it is accepted for four equations at the 5% level and for all equations at a conservative level i.e., at the 1% level of significance. However, we did not adjust for heteroscedasticity for the two equations rejected at the 5% level and have kept this as a future extension of the present work.

In this thesis we have highlighted some shortcomings. One such shortcoming is the neglect of a group-specific solution. We have shown through the use of cluster analysis that there can not be a unique solution for households having a high degree of heterogeneous characteristics. We have discovered that there are two broad groups of consumers (households) for which two functional forms (parsimonious models) are necessary for an appropriate analysis of meat consumption behavior. We have presented the major differences between these two groups in terms of descriptive statistics, preferences to different meat types as induced by personal characteristics and total (meat) expenditure elasticity measures. We have discovered considerable support about the necessity of clustering the consumers. We have identified that the determining factors of Australian meat consumption behavior differ remarkably between two groups of consumers, namely, "heavy consumers" and "medium consumers" groups. We also demonstrated that the selectivity bias associated with a non-random sub-sample from a bigger random sample can be reduced and/or removed via cluster analysis.

We have argued in this thesis that the aggregation of meat categories in the present work is better than that in previous studies. This is claimed to provide a better analysis of meat consumption behavior. We note that too much aggregation of meat categories on Australian meat consumption and inappropriate aggregation can conceal the facts. So, we have made an attempt to circumvent such drawbacks. We have considered the proportion of households who favour a particular meat type in aggregating the meat types instead of making an adhoc aggregation.

The Australian household level meat consumption were concentrated in a particular region only, whereas ours is a whole countrywide analysis which supposedly can provide more authentic results and interpretation about meat consumption behavior. According to our results, for cluster 1, which include about 64% of the households, the demographic variables play no significant role in determining meat consumption behavior while socio-economic factors do. Such results are quite different from some of the previous studies.

However, for the smaller cluster which includes about 36% of the households, some demographic variables have turned out to be important determinants of meat consumption behavior. We recall that such variables are not many and they are only the age compositions of the household members. Some socio-economic factors play a prominent role for both groups of consumers identified in the present work.

Using the argument that a unique solution is necessary for a homogeneous group of consumers, we have estimated the modified Engel curve system clusterwise. The results obtained have some similarities and some dissimilarities with those on Australian household meat consumption behavior. We highlight a plausible characterisation of household meat purchase behavior in light of the salient findings are below.

The possible changes in the demographic and socio-economic factors identified to be influential in the present study are supposed to affect the future household meat demand in Australia. In this respect the socio-economic factors appear to dominate over demographic factors. The most influential factor turns out to be the rise in real total meat expenditure as induced by a rise in real income. This factor influences all the meat types are either necessary or luxury items and no meat type is an inferior good as far as the present study is concerned. So, we may conclude that at given levels of all other factors, a rise in the real total meat expenditure will cause a rise in the overall household meat demand although not in all meat types identically.

However, we note that the pattern of consumption behavior in general and meat consumption in particular, will depend on income. Total real meat expenditure supposedly will have a relation to real total income. For a given income group, some particular meat types may be more highly favoured than others due to income as well price effects. Thus, the proportion of expenditure on different meat types will vary across different income groups.

A given meat category which is a luxury for a certain income group now, may turn out to be a necessity good when a particular household attains a higher income bracket. But, we must note that even with a rise in income, a particular meat type which is a necessity good currently, may remain the same due to a price effect. It is also true that the demand for different meats will also depend on the proportion of meat purchasers in different income groups at a particular point in time. On the whole, based on our results, we may conclude that meat demand, total as well as by meat types, will supposedly be influenced by income changes and by other factors related to income change. One such income related factor is the life-style factor which is reflected in the consumption of take-away foods. The greater availability of take-aways and fast foods (by types) are likely to affect the overall meat demand more in the future. But, it is true that we can not guarantee that similar types of demand for different meats will persist in the future. It definitely depends on the types, availability, quality and prices of fast foods. The over all demand due to this factor will also depend on the proportion of people in the population who favour different types of fast foods.

The consumption of take-away are related to income, family size and the age composition of the members of the household. Smaller families generally have a tendency to eat food outside home while larger families have a tendency to eat fewer meals outside. Such a tendency arises from the concept of the marginal expenditures and effort needed for cooking. For example, a family with only two members supposedly will incur a greater marginal cost and marginal effort for preparing food at home than does a family with four members. This alertness is related to per member income and per member expenditure. A family with more working members can be expected to have a consumption pattern different from that of a family with more children. The former may prefer more take-away food (to save time and effort) and the latter supposedly may prefer to prepare more meals at home. This is related to income units and expenditure units in the household.

The effect of age composition in the household is also an important factor in the analysis of meat consumption behavior. According to our results, adults appear to disfavor processed meat but, it cannot be concluded that the current young will also disfavor this meat type when they become adults. However, some portion of this young group will become adults and some portion of adults will belong to the next age group (old age group) within the next few years. Observe that both groups are likely to change their preferences when they belong to different cohorts. This change may occur due to health perceptions for the old groups and due to work pressures (e.g. less time, variation in the need for calories) for the adult groups. Thus, at a particular point in time, the size of different cohorts will supposedly have different impacts on the demand for different meats.

The results obtained some similarities with those in that income (reflected in total meat expenditure), size and age composition of the household and the consumption of food outside the home have important influences on meat purchase behavior.

However, the present trend in family planning towards smaller family supposedly will have some impact on meat demand since at different points in time, cohort compositions will vary. Current small families with children supposedly will be small families of adults sooner or latter. Due to change in different factors of life, such as work force participation and awareness of health hazards, demand for different meats will supposedly change. Such types of families will have a greater tendency towards the consumption of take-aways than will larger families. As a result, some types of meat will be more affected than others. But, total meat demand is influenced by the proportion of such families in the population of meat purchasers.

A factor related to income and preferences is the occupation of household head and that of other members of the household. Depending on the nature of work done by the household head and other members of the household, preference for take-aways (by types) and the purchase of different types of meat supposedly will vary. For example, working people have less time and energy to spend in the kitchen and have a preference for meat types which are easy to cook. Such choices are also related to needs of calories and perceptions about nutrient values of different meats. Thus, depending on the pattern of work force participation, the consumption of different meat types vary and not all meats very identically. This again is related to the proportion of workforce participants in different occupations. For example, according to the present results, a manual worker appears to favor a particular meat type (beef and veal). Thus, the influence of this occupation on this meat type depend on how dominant this occupation group is in the whole workforce at a particular point in time.

On the basis of the results, taking the effects of all other factors away, occupations of household heads are supposed to have some impact on meat demand. This arises from types of work (mental or manual). For example, a manual worker is observed to prefer beef and veal to processed meat. It may be that a manual worker considers beef and veal to have more money value and is richer in nutrient value than processed meat. This again raises the issue of the composition of a particular meat type. If the contents of the processed meat changes, consumption patterns of this meattype may also change. Such consideration has some relation to prices as well. Similarly, we cannot rule out the role of other members of the household in this respect since it has relation to the financial position of the household as well as to variations in tastes and needs.

An important socio-economic factor related to meat purchase decision is the use of modern electronic appliances such as food storing devices, easy and fast cooking devices. Freezer ownership, microwave-oven ownership or other easy cooking methods are likely to influence meat demand although not all meat types equally. Depending on prices, availability and tastes, consumers store some particular meat types. Likewise, easy cooking methods may determine the demands of some particular meat types. Likewise, easy cooking methods may determine the demands of some particular meat types. In the present world where life is very fast and competitive, it is quite natural that people will have a tendency to favor easy and comfortable eating. In the present study, we have found some evidence of this fact which has some similarities with that of Bartley et al. (1988).

According to our results, ownership of a freezer and microwave have some impact on meat purchase decisions although not on all meat types identically.

However, we recognise that a greater segregation of electric devices, such as freezer ownership and microwave ownership treated separately, could give a picture of the pattern of demand for different meats (separate storing effect and easy cooking effect).

As Australia has a multicultural society, the impact of ethnic background on consumption patterns in general and on meat consumption in particular, can not be ruled out. However, the results do not show any significant impact of this factor. This result has some similarity with that of Bartley et al. (1988). The result has some similarity with the findings of a consumer survey conducted by the Department of Primary Industry in 1978 as well which concluded that country of birth had no noticeable effect on the amount of fish and sea-foods consumed.

As regards to the influence of ethnic background on meat demand, we refer to Pender and Erwood (1970, p. 34). They point out that the influence of trend factors like immigrants from high pork consuming countries on pork demand in Australia may not remain as strong as it was 16 years before. We have found evidence of such facts for all meat types. We may argue that due to needs and other factors, the effect of ethnic backgrounds on meat demand became homogenised during 1980s. But, we recognise that some differences between Australian born and non-Australian born families exist, at least in terms of some particular meat types, in a more disaggregated forms than considered. These differences may be perceived to exist due to non familiarity with some particular meat type. For example, Asian-born families are not usually accustomed to tinned or conserved foods. Similarly, such families may also differ from Australian born families in respect of meat types like offal, game, dried fish and some particular types of sea food. However, such consideration of finer disaggregation of meat types is not considered and may be taken up in the future.

On the whole, the results provide more support to socio-economic factors than to demographic factors being influential in determining Australian household meat demand behavior, by total as well as by meat types.

We now highlight some generalisations which can be derived from the present study and which may be true for any study consumption behavior.

A general implication is the necessity of cluster analysis when dealing with this kind of massive data set. It may be that functional heterogenity for markedly differing groups of consumers needs to be considered for analysing consumption behavior as we have identified. Recall that we identified two functional forms, one for the "heavy consumers" group and another for the "medium consumers" group. A similar situation may prevail for other consumption behavior as well. Thus, in empirical analysis, it is necessary to investigate the existence of group structuring if there is any. We have identified support in favor of cluster analysis for dealing with a massive data set. Such support accrues from different points of view such as different functional relationships for different groups, perceptions to different meat types (luxury, necessity, inferior), economic factors (income, total expenditures, duration of unemployment, number of employed, occupations etc), social factors (ownership of modern electric devices, take-aways etc.) and demographics. Also recall that through cluster analysis we have identified and dropped some outliers in our data set.

However, it is equally important to identify the strategic factors which contribute to group structuring (observations or consumption items). For example, in the socio-demographic factors turn out to be important for cluster analysis. In some other cases, such factors may be different. These need to be identified.

Another generalisation is to emphasis that when dealing with individual level survey data, it is important to pay proper attention to zero expenditure occurrences. Of course, the approaches used to treat such zero expenditures should be commodity specific depending on durable and non-durable. For non-durable, the censored dependent variable approach is quite meaningful and easily manageable from a computational point of view. We emphasis that computational feasibility of a structural model is an important consideration for conducting empirical analyses. In our view, it is of no use constructing a highly theory-specific model, which is computationally intractable. A compromise can be to sacrifice some aspects of economic theory while providing sufficient statistical properties which is the case in this study.

The next generalisation is to emphasise the importance of an analysis based on a system framework. Without this, the interdependency of marginal utilities of consumption items are ignored which leads to a deficient analysis. We have a vivid proof of this point. The same factor influences different items(say, meat) differently, in sign and magnitude. If we did not use a systems approach, the results would be different due to the ignorance of the interdependent structure. We recognise that the appropriate modeling strategy for analysing consumption behavior of, say meat, is in the systems framework involving at least other food items and other consumption commodities. However, such a task is too broad and it is beyond the scope.

Another generalisation about consumption behavior based on argue that when income rises, food consumption rises but only to a certain extent. Higher income groups spend more on easy going food preparation and comfortable eating. We have found that "heavy consumers" have lower income than "medium consumers" and that "medium consumers" having higher income spend more on outside foods (take-aways) and spend less on meat purchases. This shows that lower income groups are more conscious about marginality of expenditure than higher income groups. That is quite natural. Probably the lower income groups consider that meat purchases and the preparation of food at home gives greater money value and food value.

It is also apparent from the present results that demographic factors are minor compared to socio-economic factors with respect to purchasing capacity as well as coping with a fast going lifestyle. For example, if someone has to feed himself/herself properly which involves cost and one has to cope with a busy life, ethnic background or persistence to ethnic eating habits becomes a minor matter. One puts more importance on needs and less on customs, beliefs or tradition. So, traditional eating habits and customs become trivial considerations.

Purchasing capacity to feed oneself well and ready made foods become the major consideration.

However, according to the present results, it is true that cohort effects can be hypothesised to play role in determining consumption behavior because of differences between needs and lifestyle. At a particular point in time, a particular cohort group may remain a dominant purchasing group of consumption item. So, we can perceive that, in addition to the factors considered new food products and services, new cooking methods and devices, new and cheaper preservation methods will supposedly have impact on consumption patterns.

In the present case of meat consumption, depending on the varieties of meat "cut types" people supposedly prefer one cut to another. These preferences may have some relation to easier and faster cooking adjustable with fast going and busy life style of the present time. According to the results, we can also postulate that such preferences will have a link with relative prices of different meat cuts as well. This is because of the fact that we identified real total meat expenditure to be an important determining factor of meat demand.

As an additional comment, we can postulate that due to weight and diet consciousness of the present days, "meat cuts" with reduced fats may also attract customers and induce propensity to consume them. As people become more aware of different cuts, their preservation and easy (fast) cooking, greater consumption may be expected. Thus, we can conclude that different meat industries will face demand patterns differently in this respect depending on how successful an industry is to reach different cohort groups having different attitudes and needs.

CHAPTER 5

CONCLUSION

In the thesis we have principally concentrated on selectivity issues. We have made attempt to clarify basic notion behind selectivity bias and its impacts on model estimation results. We have also put forth various strategies meant for testing the presence of selectivity bias. We have demonstrated how cluster analysis could be an useful strategy for making selectivity bias a trivial matter when one wishes to analyse consumption behaviour. In doing so we have clarified various conceptual frameworks of cluster analysis and its usefulness. We have firmly shown that selectivity bias disappears or becomes a minor matter when we analyse a situation after finding cluster structures instead of dealing with whole sample. However, we have also provided analysis of application of structural model after minimising selectivity bias by finding cluster structures. Such results are believed to provide more authentic and realistic consumption behavior. Similar arguments also hold good in many practical situations. We do hope that present strategy may help practitioners in resolving many applied problems. Such attempts make analysis of consumption behavior more meaningful and realistic.

We uncover cluster structures first and then apply structural model clusterwise. In doing so, we may not suffer from bias by selecting a non random sub sample, which may be needed for a structural model at hand. However, as a further piece of research we have applied a system model to analyse Australian meat consumption behavior after minimising selectivity bias. We relevantly present such results herein.

We have principally presented the estimation results of the structural model in addition to the results of (i) model selection and (ii) test of the assumption of the homoscedastic error structure. We have done all of the above clusterwise. The major findings were as follows:

Model selection strategy, namely, LR test criterion, enabled us to identify two parsimonious models for two groups of consumers. With respect to meat consumption behavior, the results have some similarities and some dissimilarities The most influential factors determining the household meat consumption behavior in Australia have turned out tobe real total meat expenditure (included by real total income) and expenditure on take-away foods. Among demographic variables, two age groups, namely, household members being 5 –<15 years and 15–64 years, significantly affect consumption behavior of some meat types but not all. Similarly, occupation category of household head and ownership of a freezer or microwave also influence purchase decisions of some meat types although not identically. The least influential factor has turned out to be the ownership of a house. All the meat types appear to be either necessity or luxury goods. In terms of total expenditure elasticity, the two groups of consumers appear to differ remarkably.

The overall results favor cluster-based analysis for meat consumption behavior. Based on the results of test of the assumption of homoscedastic error structure, we can be quite confident about the validity of such an assumption made in our case.

FURTHER HORIZONS OF RESEARCH

Some plausible horizons for extending the present work are the followings:

Modeling strategy:

An attempt to deal with the situation like ours where a substantial number of households have zero expenditure on or more meat types, can be based on the Multivariate Tobit Model without a budget constraint. In addition a comparison of competing models such as AIDS versus MAIDS (Cooper and McLaren, 1991) versus the approach of Wales and Woodland, 1983, (Kuhn Tucker Model) is another possibility for future research.

Estimation strategy:

Joint estimation of the two steps of the current modeling strategy can be done by FIML. Also estimation of the current structural model with more age break ups, more break ups of occupation of household head and that of other members, may be an interesting empirical investigation of household meat demand patterns.

Model selection strategy:

Other methods of model selection in addition to one used in the current work are Theil's (1967) information inaccuracy measures, Akaike's (1973) information criterion (AIC), the information criterion of Sawa (1978) and cross validation (Shibata, 1981).

Specification tests:

An interesting piece of research would be to test the independence of regressors and the error term of the model by applying Hausman's (1978) specification test. Similarly, for testing the goodness of fit of the system model, the correlation coefficient of Carter and Nagar (1977) can be a good strategy.

Prediction interval test:

A prediction interval test in line with Heien and Wassels (1988) to determine how well the current modeling strategy with cross section data can predict the change over time is another potentially useful topic.

Welfare assessment :

Another interesting piece of work would be to use demand analysis to compare welfare measures between households. Such an analysis may provide a systematic basis for policy implementation for family allowances or income subsidies.

Tests for clustering methods:

It would also be of interest to test the clustering method adopted in the thesis in line with McClain and Rao (1975); Arnold (1979) and Milligan(1981).

Now, we highlight some plausible lines for further research which can be performed based on the availability of further information. This basically refers to improving the design of the questionnaire used in the survey.

The first point concerns the zero expenditures on one or more meat types. More detailed information about zero expenditure occurrences can enable one to perform a better analysis. For example, zero expenditure on a meat type for a particular household does not necessarily mean a non-consumer. Zeros may be fortuitous or feigned zeros. A genuine consumer of some types of meat of which we have 19, might have not bought those items during that short survey period. One example in our case is the sea food which has 100% zero expenditures. This raises some doubts that these might have been feigned zeros. Some of the zeros definitely can be genuine non-consumption as well. Thus, explicit identification of such information through proper questionnaire design is necessary for better insights into consumption behavior.

It would be useful to collect information on health perceptions and attitudes towards different meats. This information can be obtained for the past and present. For example, someone may not consume a particular meat type at present due to health reasons, such as the awareness of the effect of cholesterol, but did consume before. Such information could help provide a better analysis of meat consumption behavior.

Similarly, the information on the types of occupation of household members can also help provide a better analysis because, depending on types of work, tastes and preferences as well as the needs of the household as a whole may vary and may impact on the demand for meat types. We also note that the education level of household heads and that of other members of the households may have some relation to their awareness of food values and nutrition. Thus, information on education level of household heads and that of other members of the households may be useful in helping explain meat consumption behavior.

Similarly, conducting surveys during different seasons of the year may provide information which differs across seasons. These differences may arise from the supply side, from interactions of consumers' characteristics with market forces at different points of time of the year. Variation in prices of different meat types during different seasons may cause differences in purchase patterns. An analysis based on the pooling of such seasonal survey information can provide a more authentic and realistic conclusion about meat consumption behavior.

Another interesting future phase of this thesis would be to look for differences between rural and urban households. These differences may arise from supply side and from differences in tastes and preferences. Some particular meat types may be more available in some areas than others and this may be more true during some particular seasons.

Thus, collecting information in line with the above and then making an analysis may provide more comprehensive insights about meat consumption patterns.

We also note that a comparative analysis of data sets collected in surveys in different years can be another interesting phase of research to uncover how the impacts of different characteristics of households vary from one survey point of time to another. A further line of research may be to consider adult equivalent scales to represent household characteristics which we could not cover in the present work.

Last but not least, we emphasis that the effects of price changes (relative prices of different meats), domestic supplies and export demand may also play an important role in future household meat demand in Australia.

References

1. Allen, R.G.D. (1959), *Mathematical Economics,* 2nd Ed. London: Macmillan and Co., Ltd.
2. Allen, R.G.D., and A.L. Bowley (1935), *Family Expenditure,* London: P.S. King.
3. Allen, R.G.D., and A.L. Bowley (1935), *Family Expenditure,* London: P.S. King.
4. Almon, S. (1965), "The distributed Lag between Capital Appropriations and Expenditures." *Econometrica,* 33: 178-96.
5. Anderson, T.W. (1971), *The Statistical Analysis of Time Series.* New York: John Wiley & Sons, Inc.
6. Ball, R.J. (1968), "Econometric Model Building", in Mathematical Model Building in Economics and Industry, London: Charles Griffin & Co., Ltd.
7. Barnett, V.D. (1973), Comparative Statistical Inference. New York: John Wiley & Sons, Inc.
8. Baumol, W.J. (1967), *Business Behavior, Value and Growth,* Rev. Ed. New York: Harcourt Brace Jovanovich.
9. Beach, E.F.(1957), Economic Models: An Exposition, New York: John Wiley & Sons, Inc.
10. Bergstrom, A.R. (1967), Selected Economic Models and Their Analysis. New York: American Elsevier Publishing Company, Inc.
11. Box, G.E.P., and G.M. Jenkins (1970), *Time Series Analysis; Forecasting and Control.* San Francisco: Holden-Day & Co.
12. Box, G.E.P., and G.M. Jenkins (1970), Time Series Analysis; Forecasting and Control. San Francisco: Holden-Day.
13. Brown, J.A.C., and A.S. Deaton (1972), "Surveys in Applied Economics: Models of Consumer Behavior." *Economic Journal,* 82: 1143-1236.
14. Brown, J.A.C., and A.S. Deaton (1972), "Surveys in Applied Economics: Models of Consumer Behavior." *Economic Journal,* 82: 1143-1236.
15. Brown, M. and D. Heien (1972), "The S-Branch Utility Tree: A Generalization of the Linear Expenditure System." *Econometrica,* 40: 737-47.
16. Brown, R.G. (1963), *Smoothing, Forecasting and Prediction of Discrete Time Series,* Englewood Cliffs, N.J.: Prentic-Hall, Inc.
17. Cain, G.G., and H.W. Watts, Eds. (1973), *Income Maintenance and Labor Supply.* New York: Academic Press, Inc.

18. Christ, C. (1966), *Econometric Models and Methods.* New York: John Wiley & Sons, Inc.
19. Christ, C.F. (1966), Econometric Models and Methods. New York: John Wiley & Sons, Inc.
20. Cootner, P.H., Ed. (1964), *The random Character of Stock Market Prices,* Rev. Ed. Cambridge: MIT Press.
21. Cramer, J.S. (1969), *Empirical Econometrics.* Amsterdam: North-Holland Publishing Co.
22. Cramer, J.S. (1969), *Empirical Econometrics.* Amsterdam: North-Holland Publishing Co.
23. Data Resources, Incorporated (1976), *The Data Resources National Economic Information System.* Amsterdam: North-Holland Publishing Co.
24. David, P.A., and T. van de Klundert (1965), "Biased Efficiency Growth and Capital-Labor Substitution in the U.S., 1899-1960." *American Economic Review,* 55: 357-94.
25. De Groot, M. (1970), Optimal Statistical Decisions. New York: McGraw-Hill Book Company.
26. Dhrymes, P. (1970), *Econometrics.* New York: Harper & Row, Publishers.
27. Doman, E.D. (1957), Essays in the theory of Economic Growth. New York: Oxford University Press.
28. Draper, N.R. and H. Smith (1966), Applied Regression Analysis, New York: John Wiley & Sons, Inc.
29. Durbin, J. (1954), "Errors in Variables." *Review of the International Statistical Institute,* 22:23-32.
30. Durbin, J. (1970), "Testing for Serial Correlation in Least-Square Regression when some of the regressions are Lagged Dependent Variables." *Econometrica,* 38:410-21.
31. Enke, S. (1951), "Equilibrium among Spatially Separated Markets: Solution by Electric Analogue. "Econometrica, 19: 40-47.
32. Fienberg, S.E. and A. Zellner, Eds. (1975), Studies in Bayesian Econometrics and Statistics, Amsterdam: North-Holland Publishing Co.
33. Fisher W.D. (1961), "A Note on Curve Fitting with Minimum Deviations by Linear Programming," Journal of the American Statistical Association, 56: 359-62.
34. Fishman, G.S. (1969), *Spectral Methods in Econometrics.* Cambridge: Harvard University Press.

35. Forrester, Jay (1961), Industrial Dynamics. Cambridge: MIT Press; New York: John Wiley & Sons, Inc.

36. Forrester, Jay (1969), Urban Dynamics. Cambridge: MIT Press.

37. Forrester, Jay (1971), World Dynamics. Cambridge: Wright-Allen Press.

38. Freund, R.J. (1963), "A Warning of Round-off Errors in Regression," American Statistician 17: 13-15.

39. Friedman, M. (1957), *A Theory of the Consumption Function.* Princeton: Princeton University Press.

40. Goldberger, A. S., and O.D. Duncan, Eds. (1973), *Structural Equation Models in the Social Sciences,* New York: Seminar Press.

41. Goldberger, A., and T. Gamaletsos (1970), "A Cross Country Comparison of Consumer Expenditure Patterns." *European Economic Review,* 1: 357-400.

42. Granger, C. W. J., and M. Hatanaka (1964), *Spectral Analyses of Economic Time Series.* Princeton: Princeton University Press.

43. Graybill, F.A. (1961), An introduction to Linear Statistical Models. New York: McGraw-Hill Book Company.

44. Green, H.A.J. (1964), *Aggregation in Economic Anaqlysis: An introductory Survey.* Princeton: Princeton University Press.

45. Griliches, Z., Ed. (1971), *Price Indexes and Quality Change.* Cambridge: Harvard University Press.

46. Hamilton, H. R., S. E. Goldstone, J.W. Milliman, A.L. Pugh, E.R. Ronerts, and A. Zellner (1969), *Systems Simulation for Regional Analysis: An Application to River Basin Planning.* Cambridge: MIT Press.

47. Hannan, E.J. (1960), *Time Series Analysis.* New York: John Wiley & Sons, Inc.

48. Hoel, P.G. (1971), Introduction to Mathematical Statistics, 4th Ed. New York: John Wiley & Sons, Inc.

49. Horowitz, I. (1963), "An Econometric Analysis of Supply and Demand in the Synthetic Rubber Industry." *International Economic Review,* 4: 325-45.

50. Houthakker, H.S. (1957), "An International Comparison of Household Expenditure Patterns Commemorating the Centenary of Engel's Laws." *Econometrica,* 25: 532-51.

51. Houthakker, H.S. (1957), "An International Comparison of Household Expenditure Patterns Commemorating the Centenary of Engel's Laws." *Econometrica,* 25: 532-51.

52. Houthakker, H.S. and L.D. Taylor (1970), "Consumer Demand in the United States, 1929-1970, 2nd Enlarged Ed. Cambridge: Harvard University Press.

53. Intriligator, M.D. (1971), Mathematical Optimization and Economic Theory. Englewood Cliffs, N.J. : Prentice-Hall, Inc.

54. Intriligator, M.D. (1971a), *Mathematical Optimization and Economic Theory.* Englewood Cliffs, N.J. : Prentice Hall Inc.

55. Intriligator, M.D. (1971b), *"Econometrics and Economic Forecasting."* in J.M. Engligsh, Ed., *The Economics of Engineering and Social Systems.* New york: John Wiley & Sons, Inc.

56. Intriligator, Michael D. (1971), Mathematical Optimization and Economic Theory, Englewood Cliffs, N.J.: Prentice-Hall, Inc.

57. Johnston, J. (1972), *Econometric Methods,* 2nd Ed. New York: McGraw-Hill Book Company.

58. Kaldor, N. (1940), "A Model of the Trade Cycle." *Economic Journal,* 50: 78-92.

59. Kendall, M.G. (1968), "Introduction to Model Building and its Problems", in *Mathematical Model Building in Economics and Industry.* London: Charles Griffin & Co., Ltd.

60. Kershaw, D. and J. Fair (1976), *The New Jersey Income-Maintenance Experiment.* New York: Academic Press, Inc.

61. Klein, L.R., and A.S. Goldberger (1955), *An Econometric Model of the United States,* 1929-1952. Amsterdam: North-Holland Publishing Co.

62. Kogiku, K. C. (1967), "A Model of the Raw Materials Market. "International Economic Review, 8: 116-20.

63. Konijn, H.S. (1962), "Identification and Estimation in a Simultaneous Equations Model with Errors in Variables." *Econometrica,* 30: 79-87.

64. Kuznets, S. (1966), *Modern Economic Growth.* New Haven: Yale University Press.

65. Leamer, E.E. (1977), Specification Searches. New York: John Wiley & Sons, Inc.

66. Leontief, W.W. (1971), "Theoretical Assumptions and Non-observed Facts," *American Economic Review,* 61:1-7.

67. Lindley, D.V. (1965), Introduction to Probability and Statistics from a Bayesian Viewpoint. New York: Cambridge University Press.

68. Lindley, D.V. (1971), Bayesian Statistics: A Review. Philadelphia: Society for Industrial and Applied Mathematics.

69. Malinvaud, E. (1970), Statistical Methods of Econometrics, 2nd Rev. Ed. Amsterdam: North-Holland Publishing Co.

70. Malinvaud, E. (1970), *Statistical Methods of Econometrics,* 2nd Rev. Ed. Amsterdam: North-Holland Publishing Co.

71. Malinvaud, E. (1970), Statistical Methods of Econometrics. 2nd Rev. Ed. Amsterdam: North-Holland Publishing Co.

72. McFadden, D. (1974), "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka, Ed. *Frontiers in Econometrics.* New York: Academic Press.

73. Meyer, J. and E. Kuh (1957), "How Extraneous Are Extraneous Estimates?" *Review of Econometrics and Statistics,* 39: 380-93.

74. Mood, A.M., and F.A. Graybill (1963), Introduction to the theory of Statistics, 2nd Ed. New York: McGraw-Hill Book Company.

75. Morehouse, N.F., R.H. Strotz, and S.J. Horwitz (1950), "An Electro-Analog Method for Investigating Problems in Economic Dynamics: Inventory Oscillations." *Econometrica,* 18: 313-28.

76. Morgenstern, O. (1963), *On the Accuracy of Economic Observations,* 2nd Ed. Princeton: Princeton University Press.

77. Nerlove, M. (1964), "Spectral Analysis of Seasonal Adjustment Procedures," *Econometrica*, 32: 241-86.

78. Orcutt, G.H. (1962), "Microanalytic Models of the United States Economy: Need and Development," *American Economic Review,* 52:229-40.

79. Orcutt, G.H., M. Greenberger, J. Korbel, and A.M. Rivlin (1961), *Microanalysis of Socioeconomic Systems: A Simulation Study.* New York: Harper & Row, Publishers.

80. Parnes, H.S. (1975), "The National Longitudinal Surveys: New Vistas for Labor Market Research," *American Economic Review,* 65: 244-49.

81. Phelps, E.S., Ed., (1970). *Microeconomic Foundations of Employment and Inflation Theory.* New York: W.WE. Norton & Company, Inc.

82. Phillips, A.W. (1958), "The relation between Unemployment and the Rate of Change of Money Wages in the United Kingdom, 1861-1957." Economica, 25: 283-99.

83. Phlips, L. (1972), "A Dynamic Version of the Linear Expenditure Model." *Review of Economics and Statistics,* 64: 450-58.

84. Phlips, L. (1974), "Applied Consumption Analysis, Amsterdam: North-Holland Publishing Co.

85. Phlips, L. (1974), *Applied Consumption Analysis.* Amsterdam: North-Holland Publishing Co.

86. Plackett, R.L. (1960), Regression Analysis, Oxford: Clarendon Press.

87. Pollak, R.A. and T.J. Wales (1969), "Estimation of the Linear Expenditure System." *Econometrica,* 37: 611-28.

88. Pollak, R.A. and T.J. Wales (1969), "Estimation of the Linear Expenditure System." *Econometrica,* 37: 611-28.

89. Price, Derek J. de Solla (1961), Science Since Babyion, New Haven: Yale University Press.

90. Price, Derek J. de Solla (1963), Little Science, Big Science, New York: Columbia University Press.

91. Rao, P., and R.L. Miller (1971), Applied Econometrics. Belmont, Calif.: Wadsworth Publishing Co., Inc.

92. Rao, P., and R.L. Miller (1971), *Applied Econometrics.* Belmont, Calif.: Wadsworth Publishing Co., Inc.

93. Samuelson, Paul. (1947), Foundations of Economic Analysis. Cambridge: Harvard University Press.

94. Scheffe, H. (1959), The Analysis of Variance. New York: John Wiley & Sons, Inc.

95. Schmidt, P. (1976), Econometrics. New York: Marcel Dekker, Inc.

96. Sewell, W.P. (1969), "Least Squares, Conditional Predictions, and Estimator properties," Econometrica, 37: 39-43.

97. Smith, Adam (1776), The Wealth of Nations. Edited by Edwin Cannan (1937). New York: The Modern Library.

98. Smith, V.L. (1961), *Investment and Production.* Cambridge: Harvard University Press.

99. Smithies, A. (1957), "Economic Fluctuations and Growth." Econometrica, 25: 1-52.

100. Stigler, G.J., and J.K. Kindahl (1970), The Behavior of Industrial Prices. National Bureau of Economic Research. New York: Columbia University Press.

101. Stone, R. (1954), "The Measurement of Consumer's Expenditure and behaviour in The United Kingdom, 1920-1938, Cambridge: Cambridge University Press

102. Stone, R. (1954a), *The measurement of Consumers' Expenditure and Behavior in the United Kingdom,* 1920-1938. New York: Cambridge University Press.

103. Stone, R. (1954a), *The Measurement of Consumers' Expenditure and Behavior in the United Kingdom,* 1920-1938. New York: Cambridge University Press.

104. Stone, R. (1954b),"Linear Expenditure Systems and Demand Analysis; An Application to the Pattern of British Demand." *Econometric Journal,* 64: 511-27.

105. Stone, R. (1965), "The Analysis of Economic Systems." Scripta Varia, 28: 1-88.

106. Stone, R.E. (1954), *The Measurement of Consumers' Expenditure and Behavior in the United Kingdom,* 1920-1938. New York: Cambridge University Press.

107. Strotz, R.H., J.C. McAnulty, and J.B. Naines, Jr. (1953), "Goodwin's Nonlinear Theory of the Business Cycle: An Electro—Analog Solution." Econometrica, 21:390-411.

108. Suits, D., (1963), The Theory and Application of Econometric Models. Athens: Center of Economic Research.

109. Takayama, A. (1974), *Mathematical Economics.* Hinsdale, Ill.: The Dryden Press.

110. Theil, H. (1971), Principles of Econometrics. New York: John Wiley & Sons, Inc.

111. Theil, H. (1971), *Principles of Econometrics.* New York: John Wiley & Sons, Inc.

112. Theil, H. (1972), *Statistical Decomposition Analysis.* Amsterdam: North Holland Publishing Co.

113. Theil, H., and J.C.G. Boot (1962), "The Final Form of Econometric Equation Systems." *Review of the International Statistical Institute,* 30: 136-52, reprinted in Zellner, Ed. (1968).

114. Tinbergen, J., and H.C. Bos (1962), *Mathematical Models of Economic Growth.* New York: McGraw-Hill Book Company.

115. Tustin, A. (1953), *The Mechanism of Economic Systems.* Cambridge: Harvard University Press.

116. Whittle, P. (1963), Prediction and Regulation by Linear Least Square Methods. New York: Van Nostrand-Reinhold.

117. Williams, E. J. (1959), Regression Analysis. New York: John Wiley & Sons, Inc.

118. Wilson, T.A., and O. Eckstein (1964), "Short Run Productivity Behavior in U.S. Manufacturing." *Review of Economics and Statistics,* 46: 41-56.

119. Wonnacott, R.J., and T.H. Wonnacott (1970), Econometrics. New York: John Wiley & Sons, Inc.

120. Zellner, A., Ed. (1968), *Readings in Economic Statistics and Econometrics.* Boston: Little, Brown and Company.