A Utility Model for Teaching Load Decisions in Academic Departments

WILLIAM F. MASSY* and ROBERT ZEMSKY†

*207 CERAS, School of Education, Stanford University, Stanford, CA 94305, U.S.A.
†Institute for Research in Higher Education, The University of Pennsylvania, 4200 Pine Street, Room 5A, Philadelphia, PA 19104-6216, U.S.A.

Abstract—The authors model the class size and teaching load decisions of academic departments in terms of a departmental utility function. Utility is postulated to be asymmetric around class size and teaching load norms, and variables for curricular structure, disciplinary domain, and institutional type are taken into account. Maximization of the utility function produces decision rules for the number of sections to be offered for each course, and hence the faculty's overall teaching load. A nonlinear estimator is developed for the decision rules' parameters and applied to data from four liberal arts colleges and two research universities. Results are consistent with theories about faculty discretionary time and with expectations about the effects of curricular structure on class size. The paper concludes with a discussion about the effects of enrollment uncertainty on faculty load decisions. [JEL: I21] ©1997 Elsevier Science Ltd

Over the past decade the cost of higher education in the United States has risen dramatically. The twin issues of cost and productivity have emerged as important subjects for research and action. Most discussions, however, are based on speculation rather than fact. There has been remarkably little research on the production function associated with higher education's most basic of services—undergraduate instruction.

Our purpose in this paper is to supply a modeling framework to support real discussion of the undergraduate production function, based on actual behavior of faculty in relation to student enrollments. As we shall see, the path to the production function lies through utility theory. We begin, though, with some simplifying assumptions about faculty attitudes and behaviors. At issue is the behavior of individual faculty members, but for modeling purposes we focus on academic departments.

Faculty members divide their time between classroom teaching and the associated preparation, grading, and administration, and other activities including research and institutional service. In so doing they respond to a complex combination of intrinsic and instrumental values—that is, to personal convictions about what is important and to incentives, mostly promulgated by their institutions and disciplines, that promise rewards or punishment as a function of effort and performance. Unpacking this complex of motivations represents an important research area, but efforts to improve our understanding of departmental behavior need not wait for definitive answers. Our first simplifying assumption, then, is that faculty themselves integrate the motivational complexity to a "bottom line" we call utility, and that certain characteristics of utility can be inferred from departmental behavior.

Our second simplifying assumption is that faculty value discretionary time—at least in part because increases in discretionary time enable increased research output. Research success produces both intrinsic and instrumental benefits for the faculty member, and faculty tend to prefer lower to higher teaching loads because teaching load correlates negatively with discretionary time. Our third simplifying assumption is that faculty are intrinsically interested in educational quality, and that at least some institutionally-based incentives encourage that view. Faculty seek situations that allow them to provide quality instruction, and they believe that quality instruction correlates with class size. (The correlation is especially strong when one controls for pedagogy: e.g. lecture, discussion class, or seminar.) When classrooms are overcrowded, instructional quality declines and faculty discomfort rises. Greater student numbers can also increase faculty work and thus erode discretionary time.

The aforementioned considerations set up a conflict between class size and teaching load. For example, a department can decrease class size—and thus arguably improve educational quality—by increasing the number of course sections fielded in response to

[Manuscript received 29 March 1994; revision accepted for publication 18 August 1996]
exogenously-determined enrollment levels. But if faculty size is fixed, such an increase will increase teaching load—thus reducing the discretionary time for one or more department members.

This paper quantifies the tradeoff between class size and teaching load. Economists have developed language to facilitate discussion and analysis of tradeoff issues. They speak of “utility maximization”, possibly subject to constraints (such as financial bounds) that limit behavior. For practical purposes it is immaterial, or at most a matter of semantic hair-splitting, whether “utility” springs full-blown from strictly personal roots or whether the value associated with the perceived consequences of an action also enters the equation. For example, when economists speak of the “marginal utility of money” they do not imply that money is intrinsically valuable, but rather that the things money can buy are valuable. Likewise, when we model the tradeoffs between class size and teaching load we are merely quantifying the utility tradeoff; why utility is what it is remains a matter for further study.

A number of authors have approached the problem of modeling academic department and faculty decision making from the utility-theory point of view. For example, James and Neuberger (1981) argued that academic departments behave as though they are nonprofit labor cooperatives, and Hopkins and Massy (1981) adopted utility maximization as the centerpiece of their microeconomic theory of colleges and universities. Lee et al. (1975) proposed a utility maximization theory based on the gap between desired and existing institutional status; Becker (1975) employed utility theory to explore alternative plans to raise teaching quality; and Huang (1981) incorporated uncertainty into a utility model of the time tradeoffs between research and teaching. Nediger (1980) developed a utility function for individual faculty members within five Canadian institutions, and used these functions to develop indices for the professors’ degree of influence in budgeting and academic policy making. Finally, Hopkins and Massy (1977) constructed an empirical utility model, informed by Stanford University administrators’ responses to a questionnaire, for the tradeoffs among institution-level primary planning variables.

Our research strategy, and the flow of the paper, proceed as follows:

1. We postulate departmental choice behavior as a maximization problem: departments decide on the number of class sections to offer in response to exogenous enrollment by maximizing the “utility” associated with class size and teaching load.
2. We specify the identity relations between class size and teaching load: i.e. that decreases in the former must increase the latter, other things being equal.
3. We derive the necessary conditions for the maximum: the department’s decision rule. The decision rule represents a potentially observable relationship among the variables, which should hold if the model is specified correctly.
4. We postulate the system’s error structure: in our model departments can’t forecast enrollment perfectly and, even if they could, they would not apply the decision rule flawlessly.
5. Having specified the error structure, we proceeded to derive a parameter estimation procedure with desirable statistical properties and apply it to our sample data. The standard statistical methodology, ordinary least squares, is not appropriate for our problem. Hence a special approach had to be devised.

The research reported in this paper is part of our on-going academic department modeling project. Two general publications (Policy Perspectives, 1990; Zemsky and Massy, 1990) describe the core concept on which the model is based: i.e. the so-called “theory of the academic ratchet”. The theory postulates that: (1) professors value discretionary time, which they can use for research and scholarship, (2) group norms determine acceptable class sizes and teaching loads, (3) deviations from the norms are rewarded and punished asymmetrically, and (4) the norms tend to shift over time to allow more discretionary time, thus generating an “output creep” or “ratchet effect” from undergraduate education toward research.

The present model embodies points (1) through (3), which is as far as one can go using the cross-sectional data available to us. This paper details the model’s structure and the parameter estimation procedures needed to deal with the asymmetry called out in point (3). Earlier discussions (Zemsky et al., 1993; Massy and Zemsky, 1994) report first results in relatively nontechnical terms, whereas we present broader findings here in a technical vein. We begin with a specification of the necessary data and variable definitions and then follow with the formal definition of the model itself.

1. DATA AND VARIABLE DEFINITIONS

The database for parameter estimation consists of enrollment records, supplemented by information from faculty teaching rosters and interviews with department chairs, for a sample of four selective liberal arts colleges and two private research universities. We take the numbers of sections as endogenous variables. The exogenous variables include: (1) the department’s list of “course offerings”; (2) the “teaching methodology” used in each course (i.e. whether it is a lecture with breakout sections, a lecture without breakout sections, a discussion class, or a seminar); (3) the course’s “degree of structure” vis à vis the rest of the curriculum; and (4) the number of sections taught by “auxiliary faculty” as opposed to tenure-line faculty. The model allows for errors in the maximization process. These errors originate from: (1)
errors in enrollment forecasting; and (2) errors in the
decision process itself, such as deviations from the
team utility construct and satisfying rather than maxi-
mizing behavior. With respect to the decision error,
the model requires only that departments behave con-
sistently enough to reveal their preferences empiri-
cally, not that they maximize exactly.

The data include a record for each separately-numbered undergraduate course other than independent study or the equivalent. Courses are based in "cat-
alog-departments" (C-departments), which are aca-
demic departments or interdisciplinary teaching pro-
grams listed in the course catalog. Each course is
associated with one and only one C-department,
cross-listings having been eliminated at an earlier
data-processing stage by assigning each such course
to the C-department having the most enrollees. The
database also records each faculty member’s "budget-
department" (B-department), which is defined by the
school’s resource allocation system rather than its
course catalog. (FTE units are used in the case of
joint appointments.) The B-department provides an
organizational home for each faculty member, who
can then teach in a number of different (possibly
inter-disciplinary) C-departments.

We believe that expectations about teaching and
research—and hence normal teaching loads—are for-
med at the B-department level. On the other hand,
sectioning decisions must be made at the C-depart-
ment level, since that represents the teaching-program
locus. Our data pre-processing program generates fac-
ulty mapping matrices showing the number of sections
in each C-department taught by faculty in each
B-department. The model can use these matrices to
take account of the distinction between the two kinds
of departments, but doing so complicates the ex-
position. Therefore, we will describe the model assum-
ing a one-to-one correspondence between the
B- and C-departments. The empirical results reported
later are based on the general case, which is
developed in Appendix A.

Let courses be denoted by \(i = 1, \ldots, n_I\) and depart-
ments by \(j = 1, \ldots, n_J\), where \(ij\) denotes the course
list offered by department \(j\). We define the following
variables for courses and departments.

**Courses:**

- \(s_i\): course sections: departmental decisions
determining the number of regular sections
for course \(i\) (breakout and lab sections are
not included).
- \(E_i\): course enrollment: the total number of
students enrolled in course \(i\).
- \(ACS_i\): average class size: equal to \(E/s_i\).
- \(\eta_{ni}\): normal class size: the department’s norm
for a given teaching methodology (e.g.
seminar or discussion class).
- \(\alpha_{ni}\): normal-class-size shift variables: a vector
of variables that adjust a course’s normal

**Departments:**

- \(S_j\): total sections: the aggregate number of
course sections offered by department
\(j\); \(S_j = \sum_{i \in j} s_i\).
- \(A_j\): auxiliary faculty sections: the number
of sections taught by lecturers and
others not on the tenure line.
- \(F_j\): regular faculty: the number of FTE
tenure-line faculty members in the
department.
- \(\eta_{nj}\): normal teaching load: the norm for the
number of sections taught by regular
faculty; our working definition uses the
median per-FTE teaching load for the
department’s tenure-line faculty.
- \(D TL_j\): departmental teaching load: the average
teaching load for tenure-line faculty; equal to \((S_j - A_j)/F_j\).

There are no shift variables for normal teaching
load.

2. THE MODEL

We postulate that departments maximize a "team
utility function" (James, 1990), representing the value
placed on discretionary time and class size, with
respect to the number of sections offered for each
course in the curriculum. (We are trying to extend
the model to make the number of courses endogenous,
but in this paper we shall deal only with the number of sections per course.) Given a certain enrollment level, the number of sections determines average class size, which has implications for educational quality and also faculty workload—since more students mean more grading, office hours, and the like. Practicality demands the aforementioned team utility function formulation, and also that the model be viewed as approximating utility with an interval-scaled function comparable across departments. Parameters are estimated under the assumption that departments have maximized their utility subject to stochastic disturbances.

The department’s utility function consists of additive terms for each course’s average class size (ACS) and the department’s teaching load (DTL). Organizational sociology’s resource dependence theory (Pfeffer and Salancik, 1978) provides grounding for our specification of ACS- and DTL-utility. The theory focuses on the organization’s so-called “enacted environment”, which represents the group’s views about appropriate behavior given perceived external conditions. In our case, the enacted environment embraces views about normal class size and normal teaching load. The concept of group norms is very important, since we predict qualitatively different behavior when resources constrain the organization’s ability to achieve its norms than when resources are abundant. Academic ratchet theory follows resource dependence theory in predicting asymmetric behavior, and hence asymmetric utility, relative to norms. The utility functions used in most studies are either monotonic (e.g. the logarithmic or exponential function) or symmetric (the quadratic ideal-point function), so they cannot represent asymmetric utility. Hence we had to devise a new utility representation.

The new representation extends the familiar quadratic utility function shown in Equation (1). The functions have the characteristics needed to meaningfully represent departmental preferences: they will be convex given expected parameter values, they are asymmetric, and they are mathematically tractable. Whether the representation is in fact meaningful is a matter for empirical investigation, which is the subject of this paper.

\[
u_{ACS}(S_i) = \lambda_{ACS,0} \max_i \left[ \frac{E_i}{\eta_{s,i}} - \alpha z_i, 0 \right] + \lambda_{ACS} \left[ \frac{E_i}{\eta_{s,i}} - \alpha z_i \right]
\]

\[
u_{DTL}(S_j) = \lambda_{DTL,0} \max_i \left[ \frac{S_j - A_j}{\eta_{s,j} F_j} - 0 \right] + \lambda_{DTL} \left[ \frac{S_j - A_j}{\eta_{s,j} F_j} \right]
\]

The quantity \( E_i/\eta_{s,i} \) is the ratio of the actual average class size \( (E_i/s_j) \) to normal class size. Likewise, \((S_j - A_j)/\eta_{s,j} F_j\) is the ratio of the department’s actual teaching load to its normal teaching load. \( S_j \) is the sum of the \( s_i \) for all \( i \neq j \), as noted earlier. The use of logarithms permits scale normalization without loss of mathematical tractability. Because utility is interval-scaled, not ratio-scaled, one of the \( \lambda \)'s can be set to a fixed value without loss of generality: we will set \( \lambda_{ACS,0} = -1 \) when estimating parameters.

The max functions produce the requisite asymmetry. When \( \log[E_i/\eta_{s,i}] - \alpha z_i > 0 \), for example, the quadratic term will dominate the i-summation as \( s_i \) gets large. A strongly negative value of \( \lambda_{ACS,0} \) relative to \( \lambda_{ACS} \) means the department will feel accelerating discomfort as enrollment rises. Only the ACS-linear terms enters the equation when \( \log[E_i/\eta_{s,i}] - \alpha z_i < 0 \), however. The effect is logarithmic but it is shown as linear in Figure 1.) The slope will be negative if professors prefer classes smaller than the norm, zero if they are indifferent to class size variations below the norm, and positive if they feel accountable for attracting the normal number of students to their courses.

We have postulated that the normal class size applicable to a particular course depends on \( structure, domain, and school \), which shift the crossover point to the right or left depending on \( \alpha \). For example, if the structure effect is negative, as argued in Massy and Zemsky (1994), the crossover point will be shifted right as the degree of structure increases.

Similar logic holds for the terms involving departmental teaching load. We postulate that loads above the norm produce accelerating discomfort, whereas the effect of load changes below the norm may be positive, negative, or zero. A weak department or one in a school under financial duress might actually prefer teaching loads close to the norm—neither much larger nor much smaller—because it fears loss of faculty lines if loads are too light. We expect \( \lambda_{DTL,0} \) to be negative but again make no prediction about the sign of \( \lambda_{DTL} \).

The department’s overall utility function can be written as:

\[
\max_{s} \sum_{i \neq j} u_{ACS}(S_i) + u_{DTL}(S_j) \quad \text{subject to} \quad s > 0
\]

\[
\lambda_{ACS,0} < 0, \quad \lambda_{ACS} = 0, \quad \lambda_{ACS,0} > 0
\]

Figure 1. The effect of average class size on utility.
where \( s = \{s_1, s_2, \ldots\} \). The necessary conditions for the maximum when \( s \) is continuous are:

\[
\frac{\partial U_j}{\partial s_i} = \frac{\partial uACS_i}{\partial s_i} + \frac{\partial s_j}{\partial s_i} \frac{\partial uDTL_j}{\partial s_j} = 0, \text{ for all } i \in j.
\]

Because fractional course is impossible, the necessary conditions should be viewed as following the long-standing modelers’ tradition of approximating a discreet process with a continuous one.

We postulate that faculty feel angst when the necessary conditions fail to hold, that the effect is monotonic in the discrepancy, and that it is minimized subject to the constraint that \( s \) is a positive integer. Substituting the utility specifications given in Equation (1) and collecting terms yields the following departmental decision rule:

\[
0 = -2\lambda_{ACS,Q} \max \left( \frac{E_i}{\eta_i s_i} - \alpha' z_i, 0 \right) - \lambda_{ACS,L} + \lambda_{DTL,Q} \max \left( \frac{S_j - A_j}{s_j - A_j} - \alpha' z_j, 0 \right) + \lambda_{DTL,L} \frac{s_j}{s_j - A_j}, \text{ for all } i \in j.
\]  

(3)

Departments where regular faculty members do not teach at least one course have been removed from the database, so \( S_j - A_j > 0 \) always.

Equation (3) represents our base model, but we also consider two alternative models for departments’ views about teaching load. The simplest approach is to assume that course sections are considered to be free goods: i.e. that they do not enter the utility function. Then the sum over \( uACS \) can be maximized unconditionally with respect to the \( s_i \). This DTL-free model and its necessary conditions are:

\[
U_j = \sum_{i \in j} uACS_i
\]

\[
\frac{\partial U_j}{\partial s_i} = \frac{\partial uACS_i}{\partial s_i} = 0, \text{ for all } i \in j.
\]

(4)

The decision rule consists of only the first line in Equation (3).

More realistically, we might assume that departments consider the total number of sections to be fixed: i.e. they believe themselves to possess a fixed teaching capacity which must be allocated over courses. Assuming capacity equals the observed \( S_j \), we have the following DTL-fixed model:

\[
U_j = \sum_{i \in j} uACS_i \text{ subject to } S_j = \sum_{i \in j} s_i
\]

\[
\frac{\partial U_j}{\partial s_i} = \frac{\partial uACS_i}{\partial s_i} + \lambda_0 = 0, \text{ for all } i \in j
\]

(5)

where \( S_j \) now is exogenous and \( \lambda_0 \) is a Lagrangian multiplier. The decision rule consists of the first line of Equation (3) with \(+\lambda_0\) appended.

All the models involve the marginal utility of \( ACS \) with respect to the number of sections allocated to each course. The DTL-free model’s unconstrained optimization carries the derivatives to zero. In the DTL-fixed model, the derivatives all equal a constant \( \lambda_0 \), the constant being the marginal utility of \( DTL \) at its fixed value. The base model’s ACS-derivatives vary depending on the DTL-utility function and the value of \( DTL \). This model is the most realistic, so we shall focus our attention mainly on it.

3. PARAMETER ESTIMATION

The combination of utility asymmetry and the model’s error structure complicate parameter estimation to the point where custom estimation procedures are required. In this section we describe the error structure and derive a constrained two-stage least-squares estimator with the requisite properties.

3.1. Error Structure

We consider two independent sources of stochastic variation: errors in the departments’ enrollment forecasts, and errors in the optimizing behavior itself. The error terms are defined as follows.

\( u \): enrollment forecast error; the difference between the course enrollment used by the department in its optimization and the actual enrollment that eventually transpires. We assume that: (1) \( u \) is approximately normal with mean zero and variance \( \sigma_u^2 \), and (2) \( \sigma_u \) is some percentage \( \theta \) of the mean enrollment over the courses that use a given teaching method. The second assumption implies that \( \sigma_u = \theta \bar{E}_h \), where \( \bar{E}_h \) is the sample mean enrollment for teaching method \( h \) and \( \theta \) is a parameter to be estimated from the data.

\( \nu \): decision-rule error; noise due to imperfections in the department’s optimization process. We assume that \( \nu \) is approximately normal, with mean zero and variance \( \text{var}[\nu] \), which will be estimated from the data.

Differences between the integer values of \( s \) and their continuous expectations show up in \( \nu \), the decision-rule error. We make no assumptions about the normality, or even the symmetry, of either error term; hence the fact that \( s \) is integer does not bias the estimation.

The estimation procedure will be derived only for the base model, but extension to the other models is self-evident. Inserting the error terms into Equation (3) and fixing \( \lambda_{ACS,Q} = -1 \), we have:

\[
2 \max \left( \frac{E_i}{\eta_i s_i} - \alpha' z_i + u_i, 0 \right) - \lambda_{ACS,L}
\]

(3)
for all \( i \) and \( j \). Estimation involves minimizing the sum of squares \( u \) and \( v \).

The question of whether \( E \) is independent of \( s \) deserves consideration. Ours is strictly a short-run model in which departments optimize their sectioning decisions based on the available teaching staff and projected enrollment. They react to enrollment shifts when they show up as overcrowded courses. Such events may affect departmental forecasts of future enrollments, but that does not introduce simultaneous-equation bias into cross-sectional estimation. The number of faculty can also vary with enrollment, since deans may be expected to add or subtract faculty billets in some relation to student numbers. But like enrollment forecasts, faculty numbers are appropriately viewed as predetermined in our short-run cross-sectional model.

The enrollment forecast error is embedded in the first term; hence there is no ambiguity about the direction in which its sum of squares should be minimized. The same cannot be said for the decision rule error, as there is no reason to associate it with any particular term in the equation. Errors distributed through an equation require special estimation techniques, but, since \( v \) accounts for only a small part of the overall error variance, we can ignore this complication.

### 3.1.1. The error transform.

The following new notation will simplify matters as we demonstrate how \( u \) is transformed through the "max" function. Let

\[
y_i = 2 \log \frac{E_i}{\eta_{0i}}, \text{ as observed}
\]

\[
Y_i = y_i + u_i, \text{ the department's forecast}
\]

\[
w_{ij} = \begin{cases} 1, & \frac{2s_j}{(S_j - A_j)} \max \left[ \log \frac{S_j - A_j}{\eta_{ij} F_j}, 0 \right] \\ \frac{-s_j}{(S_j - A_j)} \max \left[ \log \frac{S_j - A_j}{\eta_{ij} F_j}, 0 \right] & \end{cases}
\]

(7)

(Note that \( y_i \) and \( x_i \) can be computed from the data but \( Y_i \) is not observable.) Dropping the subscripts \( i \) and \( j \) temporarily, the structural equation becomes:

\[
\max[y + u - \alpha z, 0] = \lambda' w + \nu
\]

(8)

where \( \lambda' \) is defined as \( \lambda_{ACS,LL}, \lambda_{DTL,LL}, \lambda_{DTL,QQ} \) and both sides contain unobservable random variables. Next we define adjusted values of \( Y \) and \( y \):

\[
Y_a = Y - \alpha' z
\]

\[
y_a = y - \alpha_0' z
\]

where \( \alpha_0 \) is the true value of \( \alpha \). Using the new subscript \( m \) as a shorthand for the "max" function and substituting the definition of \( Y \) and \( y \) into the expression for \( Y_m \) yields:

\[
Y_m = \max \{ Y_m, 0 \} = \max \{ y_a + u - (\alpha - \alpha_0) z, 0 \}
\]

\[
y_m = \max \{ y_m, 0 \}
\]

For present purposes, the argument within the max function can be represented simply as \( Y_a = y_a + u^* \), where \( u^* = u + (\alpha - \alpha_0) z \) varies randomly from observation to observation. The mean of \( u^* \) for a given sample is non-zero due to the error in estimating \( \alpha \); however, for asymptotically unbiased \( \alpha \)-estimates the mean over many samples will approach zero. There is no way of calculating this error for a given sample, so we shall use \( u^* = u \) from this point forward. Therefore, \( Y_m = \max \{ y_m, u^*, 0 \} \).

Now define another random variable, where \( y_m \) is known but \( Y_m \) is stochastic:

\[
\xi = Y_m - y_m
\]

Substituting \( Y_m \) for the left side of Equation (8), subtracting \( y_m \) from both sides, and substituting \( \xi \) for \( Y_m - y_m \) recasts the structural equation to:

\[
y_m = \lambda' w - \xi + \nu
\]

which can be rewritten as

\[
y = \alpha' z + \lambda' w - \xi + \nu, \text{ if } y_m > 0
\]

\[
0 = \lambda' w - \xi + \nu, \text{ if } y_m \leq 0
\]

(9)

Suppose we redefine \( y \) in terms of trial values of \( \alpha \) (denoted by \( \alpha_{-1} \)). Then we could estimate the unknown parameter vectors \( (\alpha, \lambda) \), conditional on \( \alpha_{-1} \), by minimizing the sum of squared errors of Equation (9) and iterate until \( \alpha \) converges to \( \alpha_{-1} \). Convergence of the iteration seemed problematic, however, so we recast Equation (9) as follows:

\[
y = \alpha' z + \lambda' w - \xi + \nu, \text{ if } y_m > 0
\]

\[
\alpha_{-1}' z = \alpha' z + \lambda' w - \xi + \nu, \text{ if } y_m \leq 0
\]

The two equations are equivalent if \( \alpha_{-1} = \alpha \) but when \( \alpha_{-1} \neq \alpha \) the second line of the revised version produces a powerful force for convergence. Finally, we adopt the following variable redefinitions to simplify the notation when we come to minimize the sum of squared errors:

\[
y' = \alpha' z + \lambda' w - \xi + \nu, \text{ where}
\]

\[
y' = y \text{ if } y_m > 0
\]

\[
y' = \alpha_{-1}' z \text{ if } y_m \leq 0
\]

(10)

We shall use Equation (10) to estimate the coef-
ficients, but it is not equivalent to Equation (9) when it comes to estimating goodness of fit and the standard errors of $\alpha$ and $\beta$. While the sum of squared errors is the same for the two equations, only Equation (9) produces the appropriate sample variances and covariances for $x$ and $y$. Hence Equation (9) is the equation of choice for calculating goodness of fit and the standard errors.

3.1.2. The conditional mean and variance of $\xi$. Simply minimizing the sum of squares of Equation (9) or Equation (10) would not produce consistent and asymptotically efficient parameter estimates. The error in $\nu$ presents no problem, but $\xi$ violates OLS assumptions. The expectation of $\xi$ is not necessarily zero: in fact, it depends on the observed $y$ and $z$. To solve the resulting bias problem, we derive an explicit formula for the conditional mean and subtract it from the equation. The dispersion of $\xi$ also depends on $y$ and $z$, so we derive the conditional variance and adjust the equation for heteroscedasticity.

Figure 2 illustrates the elements that go into calculations for the conditional expectation of $\xi$, for the case where $y_0 = 0$ and $z = 0$. The expectation is calculated by integrating $Y_m|Y_m$ over the real line. Since $Y_m > 0$ for half of its probability density and zero elsewhere, the expectation of $Y_m$ is positive. The conditional expectation will be even larger when $y_m > 0$.

The conditional expectation of $\xi$ is calculated as follows:

$$
E[\xi|y_0] = E[Y_m|y_0] - y_m = \int_0^\infty f(Y_m|y_0)dY_m - y_m
$$

$$
= \int_{-\infty}^\infty (y_u + u_x)f(u_x)du_x - y_m
$$

$$
= y_x \int_{-\infty}^\infty f(u_x)du_x + \int_{-\infty}^\infty u_xf(u_x)du_x - y_m
$$

$$
= E_{-\infty}^{\infty}[u_x] + y_x(p^* - \lambda m)
$$

(11)

where

$$p^* = \Pr[Y_m(\geq 0|y_u) = 1 - \int_{-\infty}^{0} f(Y_m|y_u)dY_m
$$

$$= 1 - \int_{-\infty}^{y_u} f(u_x)du_x = 1 - \int_{-\infty}^{u_x} f(u_x)du_x
$$

The conditional variance of $\xi$ is calculated in a similar fashion:

$$\text{var}[\xi|y_0] = E[(\xi - E[\xi|y_0])^2]
$$

$$= \int_{-\infty}^{\infty} \{Y_m(y_m) - E_{-\infty}^{\infty}[y_0] - y_0p^* + y_m\}^2 f(Y_m|y_0)dY_m
$$

$$= k_1 \int_{-\infty}^{\infty} f(u_x)du_x + \int_{-\infty}^{\infty} (u_x - k_2)^2 f(u_x)du_x
$$

$$= k_1(1 - p^*) + \int_{-\infty}^{\infty} (u_x - 2k_2u_x + k_2^2) f(u_x)du_x
$$

$$= k_1(1 - p^*) + k_2^2 - 2k_2 E_{-\infty}^{\infty}[u_x] + E_{-\infty}^{\infty}[u_x^2]
$$

(12)

where $k_1 = E_{-\infty}^{\infty}[u_x] + y_0p^*$ and $k_2 = k_1 - y_m$.

Figure 3 shows the conditional mean and variance as functions of $y_m$ under the assumption that $\text{var}[u] = 1$. The conditional mean reaches a maximum at $y_m = 0$ and is zero at the extremes. The conditional variance starts from zero at small values of $y_m$, where all the conditional density of $Y_m$ is to the left of zero, and goes to $\text{var}[u]$ as $y_m$ gets large.

3.2. The Estimator

Parameters are estimated by minimizing the sum of squares $Y + \nu = y^0 - \alpha z - \lambda w$, obtained from Equation (10). However, we must add $E[\xi|y_0]$ to both sides so that the expectation of the error is zero. First, though, we shall simplify our notation by combining the parameters and regressors on the right-hand side of Equation (10) into single vectors. Let

$$\beta' = [\alpha', \lambda']$$

$$x' = [z', w']$$

Then the sum of squares becomes:

$$\sum_i (\xi_i + \nu_i + E[\xi|y_0])^2 = \sum_i (y_i - \beta'x_i + E[\xi|y_0])^2
$$

(13)

Unconstrained minimization of Equation (13) produces a trivial solution point. To see why, refer back to Equation (10) and suppose, for example, that an element of $\alpha$ (element $q$, say) becomes very large.
This drives $y_a$ negative for all the data points: the large $\alpha$ dominates the solution and produces a perfect correlation between $\alpha_a$ and $\alpha_{a-1}$ as $\alpha_a = -\alpha_{a-1}$ and the other $\alpha$s go to zero in the limit. A similar problem arises in connection with Equation (9), so something must be done to forfend it.

3.3. Constraints on the Alphas

The trivial solution can be avoided by constraining $\alpha$ based on the data for normal class size ($\eta_n$). The $\alpha$s can be viewed as shifting the crossover point away from its nominal value, $\eta_n$. The values of $\eta_n$ were obtained by surveying department chairs, so it is reasonable to believe they hold on average over all the data points. A stronger, but still reasonable, assumption is that the survey values hold on average over each set of dummy variables; that is, over the structured and unstructured data points, over the data points for the five disciplinary domains, and (where applicable) over the data points for schools. To implement this assumption we partition the $\alpha$ and $z$ vectors into sub-vectors for structure (str), domain (dom), and school (sch) and define $\bar{z}$ as the vector of sample means for the dummy variables. Then we can write the three constraints as:

$$
0 = \alpha_{\text{str}} \bar{z}_{\text{str}}
$$

$$
0 = \alpha_{\text{dom}} \bar{z}_{\text{dom}}
$$

$$
0 = \alpha_{\text{sch}} \bar{z}_{\text{sch}}.
$$

(14)

Our task now is to minimize the sum of squared errors subject to Equation (14).

The first step is to eliminate one variable per constraint. Without loss of generality, we chose the first variable in each set: that is, unstructured, humanities, and school-1. Now we transform each of the remaining dummy variables according to

$$
z_{k} = \bar{z}_k - \frac{\bar{z}_1}{\bar{z}_1} z_{1}, \text{ for } k = 2,\ldots
$$

(15)

for each of the constraints. The resulting $z$-vectors, which are one element shorter than the originals, are used in the expression for minimizing the sum of squares. Once the minimum has been obtained, the estimates for the excluded coefficients and their variances are calculated by:

$$
\alpha_i = -\frac{1}{z_1} \alpha_z
$$

$$
\sigma^2_a = \frac{1}{n} \alpha_A \Sigma \alpha
$$

(16)

where $\alpha^*$ and $\Sigma^*$ are the estimates of the coefficients in the reduced set for the constraint, and their variance-covariance matrix, respectively.

3.3.1. Two-stage least squares. The objective now is to minimize the constrained sum of squared errors in a way that has desirable statistical properties. We proceed in two stages: first obtaining consistent estimates of $\beta$, $\theta$, and $\text{var}[v]$, and then improving the efficiency of the $\beta$ estimates by using the values for $\theta$ and $\text{var}[v]$ obtained in the first stage to stabilize the error variances via a generalized least-squares transformation.

The Stage-1 estimator is obtained directly from Equation (13):

$$
\text{SSE}_1 = \min_{\beta, \theta} \left( \sum_i (y_i - \beta' x_i + \text{E}[\xi_i | y_i, \theta])^2 \right)
$$

$$
\text{SSE}_1 = \sum_i \text{var}[\xi_i | y_i, \theta]
$$

$$
\text{var}[v] = \frac{\text{SSE}_1}{n - \text{Length}([\beta] - 1)}
$$

(17)

where the notation for the conditional mean and variance has been expanded to show the dependence on $\theta$. The number of degrees of freedom shown in the denominator of the expression for $\text{var}[v]$ reflects the number of parameters estimated to get $\text{SSE}_1$ (i.e. $\beta$ and $\theta$). The Stage-1 estimation requires a two-fold iteration: (1) an outer loop for alternative values of $\theta$, and (2) an inner loop in which successive values of $\alpha_{-1}$ are used to evaluate $y^*$ and $y_a$ until $\beta_{n-1} = \beta$. To find $\theta$, we adopted the so-called “golden section” procedure: the user inputs an admissible range of values and the algorithm optimizes the search within this range (Gellert et al., 1975, p. 642). Both the golden section method and the successive substi-
utions for \( \alpha \) work well, with convergence in a manageable number of steps to values independent of the starting position. The estimates are well-behaved, as demonstrated by their relative consistency from dataset to dataset and over fine-structure variations in the model’s specification.

The estimate of \( \text{var}[\nu] \) turned out to be negative in a few of our test runs—although not in the large-sample runs for the fully-specified model. This did not really surprise us, given that the calculation involves the difference between the sums of squares of two noisy variables. The error of estimate for \( \text{var}[\nu] \) (i.e. \( \text{var}[\nu] - E[\text{var}[\nu]] \)), is the sum of the deviations between the two terms in the second line of Equation (17) and their expectations. The deviation for the first term, \( \text{SSE}_i - E[\text{SSE}_i] \), follows the familiar chi-square distribution. The second term’s deviation is the difference between the sum of the variances of \( \xi \) as calculated from Equation (17) using the estimated value of \( \theta \) and the sum of the true variances, based on the true (but unknown) value of \( \theta \). The two deviations are not independent, since the error in estimating \( \theta \) affects both. Moreover, the terms in each of the two sums that comprise the second term are not independent, since the single \( \theta \)-estimate affects all of them simultaneously. It is easy to imagine situations where the error in \( \theta \) produces estimates of \( \text{var}[\xi] \) which, when summed, exceed \( \text{SSE} \). When that happens, we simply set the estimated \( \text{var}[\nu] \) to a small positive default (currently 0.01) and proceeded with Stage II. Fortunately, the problem does not arise frequently and the Stage-II results are not sensitive to the choice of default in any case.

The Stage-II procedure uses the Stage-I estimates of \( \theta \) and \( \text{var}[\nu] \) to improve the efficiency of the \( \beta \) estimates by stabilizing the residual error variance:

\[
\text{SSE}_{II} = \min_{\beta} \sum_i \left( \frac{y_i' - \beta'x_i + E[\xi_i y_i', \theta]}{\sqrt{\text{var}[\xi_i y_i', \theta] + \text{var}[\nu]}} \right)^2
\]

(18)

This time the minimization has only one loop, which iterates on successive values of \( \alpha_{-1} \).

3.3.2. Statistical properties. Equation (18) involves the sum of squares of independent and identically distributed random variables with zero mean, so for large samples SSE is distributed approximately as chi-square with \( n - \text{Length}[\beta] - 2 \) degrees of freedom. Hence our estimation procedure can be characterized as approximately minimum chi-square, a type of generalized least-squares. (We inject a cautionary note because the higher-order derivatives of SSE are not continuous.) Minimum chi-square estimators are known to be consistent and asymptotically efficient (Rao, 1973, p. 352). The fact that the parameters enter non-linearly, which requires the minimum to be obtained by iteration, is immaterial.

The traditional \( R^2 \) represents the contribution of the explanatory variables to explaining \( y \). But \( y \) has no special standing in our model, and in any case it contains embedded parameters. Hence the traditional \( R^2 \) is not appropriate as a measure of fit. Since our model is in effect an identity with an error appended, it seems logical to take \( SSY + SSR \) as the fitted sum of squares, and to normalize this by \( SSY + SSR + SSE \). In other words:

\[
\text{Fit} = \frac{SSY + SSR}{SSY + SSR + SSE}, \quad \text{where}
\]

\[
SSY + SSR = \sum_i \left( \frac{y_i' + \beta'x_i + E[\xi_i y_i', \theta]}{\sqrt{\text{var}[\xi_i y_i', \theta] + \text{var}[\nu]}} \right)^2
\]

(19)

and SSE is obtained from Equation (18).

The fitting procedure produces the usual least-squares estimates of the standard errors of \( \alpha \) and \( \beta \), but they are conditional on \( \theta \) and \( \text{var}[\nu] \). Calculating the unconditional standard errors requires evaluation of the information matrix—the matrix of second-order partial derivatives of \( SSE \) with respect to the full parameter set. The second-order derivatives are not continuous, however, due to the “max” functions embedded in SSE. The resulting difficulties have frustrated our efforts to calculate the unconditional standard errors, so we report only the conditional ones for \( \alpha \) and \( \beta \). Doubtless the unconditional standard errors would be somewhat larger.

3.3.3. Significance tests for asymmetry. The asymmetry of \( uACS \) represents an important hypothesis about departmental behavior. Therefore, we sought a significance test for whether adding the “max” function (which produces the asymmetry) improves the model’s fit.

If the symmetric model is correct, the structural equation becomes:

\[
2 \log \left( \frac{E}{\eta_{0}S} \right) = \beta'x_i + 2u_i + \nu
\]

(20)

Minimizing the sum of squares with respect to \( \beta \) produces an estimate of \( \text{var}[u + \nu] \) with \( df_s = n - \text{Length}[\beta] \) degrees of freedom. If the asymmetric model is correct, \( \text{var}[u + \nu] \) is estimated by \( \text{var}[u + \nu] = \text{var}[\nu] + 4\theta'\log \hat{E} \) with \( df_s = n - \text{Length}[\beta] - 1 \) degrees of freedom. The ACS-asymmetric model is significantly better than the symmetric one when

\[
F' \left( \frac{df_s, df_k}{\text{var}[u + \nu]} \right) = \frac{\text{var}[u + \nu]}{\text{var}[u + \nu]}
\]

(21)

is larger than would be expected by chance. The numerator and denominator of Equation (21) are not independent, so \( F' \) does not follow the classic F-dis-
tribution. The F-tables do provide a conservative test of the hypothesis that the asymmetric model fits better than the symmetric one, however, since the positive correlation between numerator and denominator will reduce the computed F*. Therefore, we take significant F*-values as evidence favoring the asymmetric model.

Significance tests for asymmetry in μDTL can be obtained by forming the variance ratio for the alternative symmetric models. This refinement is unnecessary when the quadratic term itself is insignificant, however, since a monotonic function cannot be symmetric. Neither the DTL-symmetric variance-ratio test nor the test for asymmetric μACS is affected by the conditional-standard-errors problem, although both tests only hold asymptotically.

4. RESULTS

We ran the model for the liberal arts colleges, the research universities, and the two types combined. The core data are based on an analysis of course loads and course enrollments at each school for an entire academic year. We supplemented these data with a set of basic curricular measures derived from analyses of the transcripts of graduating seniors and from detailed interviews with the chairs of nearly all departments in the four liberal arts colleges and of the arts and sciences departments in the two research universities. We also benefited from a follow-up mail survey to the department chairs that provided additional information about the normal class sizes for particular teaching methods and styles.

Table 1 presents the figures for normal class size. For comparison, we also show the average class sizes as computed from our database, which generally are less than the normal class sizes. The normal class sizes for lectures with breakouts are taken to be open-ended, as described in Massy and Zemsky (1994). The teaching methodology definitions and their relation to the survey are described in the same reference.

Table 2 provides an overall picture of the model’s performance. Degree-of-fit ranges from 38% to 44%, which is satisfactory given that we are working with cross-sectional data. The coefficients are readily interpretable and satisfy our sign expectations in most cases. The larger coefficients exceed their standard errors by a sufficient multiple that the conditionality problem discussed earlier is not likely to be material. Generally the Stage II results are more stable and interpretable than the Stage I results (not shown), as would be expected from the underlying statistical theory.

The error variance can be partitioned into components representing enrollment-forecast error (theta) and optimization error (var[υ]), as described earlier. Table 2 shows that forecast error represents 88.3% of the total variance for the liberal arts colleges and 72.1% for the research universities—which leaves about 12% and 28% for the effects of decision-process error. That the research universities seem to optimize less tightly than the more homogeneous and collegial liberal arts colleges is quite believable. We find the error variance estimates to be reasonable, and their consistency buttress our overall confidence in the model.

4.1. The Structure of Utility

Figure 4 presents graphs showing the shapes of the ACS- and DTL-utility functions for the liberal arts colleges and research universities. The left-hand panels support our hypothesis of average-class-size utility asymmetry, since the estimated function drops off more sharply above than below the normal class size (NCS). The ACS-linear coefficient would have to grow substantially before the utility function’s asymmetry would begin to erode. Indeed, the null hypothesis of μACS symmetry is rejected with high confidence: the F-values were 10.1 for the liberal arts colleges, 6.4 for the research universities, and 7.5 for the combined run. The graphs do not differ much by institutional type, although the coefficient for the universities is about half again as large as the one for the colleges and the difference is about 2.5 standard deviations. The positive ACS-linear coefficients imply class-size monitoring, as discussed earlier.

Unlike average-class-size utility, the term for departmental teaching load (DTL) does not demonstrate asymmetry. The departmental teaching load utility curves are nearly straight lines, as can be seen in the right-hand panels of the figure. The DTL-quadratic coefficients are not significant statistically and they hardly affect the shape of the graph.

The statistical results strongly support our hypothesis about the effect of curricular structure on normal

<table>
<thead>
<tr>
<th>Table 1. Base normal class size and average actual class size, by school type and teaching method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Lectures with breakout</td>
</tr>
<tr>
<td>Lectures without breakout</td>
</tr>
<tr>
<td>Discussion classes</td>
</tr>
<tr>
<td>Seminars</td>
</tr>
</tbody>
</table>
A Departmental Utility Model

Table 2. DTL-utility model

<table>
<thead>
<tr>
<th></th>
<th>ALL</th>
<th>LAC</th>
<th>RU</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Utility weights</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACS linear</td>
<td>0.082</td>
<td>0.064</td>
<td>0.091</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.010)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>DTL quadratic</td>
<td>−0.305</td>
<td>−0.038</td>
<td>0.303</td>
</tr>
<tr>
<td></td>
<td>(0.442)</td>
<td>(0.367)</td>
<td>(1.420)</td>
</tr>
<tr>
<td>DTL linear</td>
<td>−0.574</td>
<td>−0.134</td>
<td>−1.093</td>
</tr>
<tr>
<td></td>
<td>(0.121)</td>
<td>(0.134)</td>
<td>(2.023)</td>
</tr>
<tr>
<td><strong>Degree-of-structure</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Little structure</td>
<td>0.026</td>
<td>0.036</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.007)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>More structure</td>
<td>−0.036</td>
<td>−0.056</td>
<td>−0.017</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.010)</td>
<td>(0.012)</td>
</tr>
<tr>
<td><strong>Domain</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Humanities</td>
<td>0.013</td>
<td>0.007</td>
<td>0.018</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.012)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>Language</td>
<td>−0.060</td>
<td>−0.048</td>
<td>−0.080</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.015)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Mathematics</td>
<td>0.015</td>
<td>0.111</td>
<td>−0.060</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.036)</td>
<td>(0.039)</td>
</tr>
<tr>
<td>Science</td>
<td>−0.120</td>
<td>−0.145</td>
<td>−0.065</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.019)</td>
<td>(0.031)</td>
</tr>
<tr>
<td>Social science</td>
<td>0.110</td>
<td>0.111</td>
<td>0.113</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.013)</td>
<td>(0.018)</td>
</tr>
<tr>
<td><strong>Institution</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lac1</td>
<td>−0.030</td>
<td>−0.019</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.017)</td>
<td>—</td>
</tr>
<tr>
<td>lac2</td>
<td>0.111</td>
<td>0.126</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.013)</td>
<td>—</td>
</tr>
<tr>
<td>lac3</td>
<td>−0.205</td>
<td>−0.176</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.013)</td>
<td>—</td>
</tr>
<tr>
<td>lac4</td>
<td>0.056</td>
<td>0.074</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.013)</td>
<td>—</td>
</tr>
<tr>
<td>ru1</td>
<td>0.010</td>
<td>—</td>
<td>−0.001</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>—</td>
<td>(0.011)</td>
</tr>
<tr>
<td>ru2</td>
<td>0.018</td>
<td>—</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>—</td>
<td>(0.009)</td>
</tr>
<tr>
<td><strong>Variances</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>theta (%)</td>
<td>8.1</td>
<td>7.1</td>
<td>8.8</td>
</tr>
<tr>
<td>var[σ]</td>
<td>0.134</td>
<td>0.075</td>
<td>0.177</td>
</tr>
<tr>
<td><strong>Degree-of-fit (%)</strong></td>
<td>42.1</td>
<td>38.0</td>
<td>44.1</td>
</tr>
<tr>
<td>Asymmetry F</td>
<td>7.5</td>
<td>10.1</td>
<td>6.4</td>
</tr>
<tr>
<td>Sample size</td>
<td>3491</td>
<td>1531</td>
<td>1960</td>
</tr>
</tbody>
</table>

Note: The standard errors of the coefficients are in parentheses.

class size: structure decreases the anticipated NCS by a statistically significant amount for both the liberal arts colleges and the research universities. The LACs seem more sensitive to structure than the RUs, perhaps because the former are more likely to make sectioning adjustments in response to enrollment fluctuations. The coefficients for domain and school also are material: the adjusted NCS-values for science and language courses are below average by statistically significant amounts while those for the social sciences are above average, and some of the schools demonstrate distinctive variations as well.

Table 3 probes for possible interactions among the ACS-shift effect, by deleting successive blocks of variables. The coefficients are quite stable, indicating an absence of interaction. The table also shows that deleting the structure variables improves the degree of fit—especially in the case of the liberal arts colleges. This may seem counter-intuitive, since deleting variables usually decreases $R^2$, but there is a logical explanation. By adjusting normal class size, the NCS-shift variables can reduce the amount of utility variation seen in the data, thus reducing our degree-of-fit measure. For the liberal arts colleges, the larger "explanatory power" of the runs with the structure variables deleted is spurious, since some of the utility variation in this run represents a failure to take structure into account. The effect also arises in connection with the research universities, but it is quite small.

The variations in degree-of-fit also indicate that the liberal arts colleges are more heterogeneous over domains and institution-by-institution than are the
research universities. This may be due to the smaller number of schools in the RU set, or it may reflect the homogenizing power of being highly visible in the faculty marketplace.

4.2. Alternative Models

Table 4 compares the DTL-fixed and DTL-free parameter with those for the base model. The fixed model fits slightly better than the free model for the RUs but not for the LACs. This is consistent with the base-model curves presented in Figure 4, where the steeper slope for the RUs indicates a more powerful utility effect than for the LACs. The base model fits best for both institutional types, as would be expected given that it is the most flexible. However, there is not much difference in the degree of fit. The NCS-shift coefficients, theta, and var[τ] do not vary much across models either.

The DTL-fixed coefficients represent the marginal utility or shadow price of an extra unit of teaching load, and they are negative as expected. That the coefficient is significantly negative for the RUs but not for the LACs is consistent with the differences in goodness-of-fit and the slopes observed for the base model. The alternative-model results reinforce the model's stability and confirm that RU departments place a greater utility on low teaching load—and hence, by implication, faculty discretionary time—than do their liberal arts counterparts. While the base model indicated this quite clearly, we had harbored some concern about collinearity between the DTL-quadratic and the DTL-linear variables. The DTL-free and DTL-fixed models eliminate the collinearity, and the consistency of the remaining coefficients across models buttresses our confidence that the specification is robust.

4.3. Domain-Specific Results

The collinearity between the DTL-linear and DTL-quadratic variables did prove troublesome when we tested the model on domain-specific data. Hence we report only the DTL-fixed results, in Table 5. All the ACS-linear coefficients except language indicate class-size monitoring, with humanities and social science being strongly significant—apparently faculty in these domains feel more pressure to fill their classes. All the DTL-fixed coefficients are negative, with the effect being most pronounced for language and science. That the science coefficient shows strong teaching-load aversion is not surprising given the pressure to publish in the scientific disciplines. We were surprised at the size of the language coefficient in relation to that for the humanities, and also that the ACS-linear coefficient for languages is negative (though not significant). The result may be due to collinearity, or to variations in the mix of LAC and RU courses in the two datasets since school type was not controlled in the domain-specific runs. All domains except language exhibit statistically significant class-size monitoring effects.

The curricular structure effect is consistent across models, except in the case of the humanities where the coefficients are essentially zero. It appears that there are too few structured humanities courses to allow us to measure a meaningful effect. Language courses exhibit a powerful structure effect, which is not surprising, with social science courses coming in second. Once again, there are statistically significant and potentially interesting school effects. For
Table 3. Utility model, sets of NCS-shift variables deleted

<table>
<thead>
<tr>
<th>Utility weights</th>
<th>LAC</th>
<th>LAC</th>
<th>LAC</th>
<th>RU</th>
<th>RU</th>
<th>RU</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACS linear</td>
<td>0.064</td>
<td>0.098</td>
<td>0.071</td>
<td>0.091</td>
<td>0.097</td>
<td>0.106</td>
</tr>
<tr>
<td>(0.010)</td>
<td>(0.010)</td>
<td>(0.010)</td>
<td>(0.011)</td>
<td>(0.011)</td>
<td>(0.011)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>DTL quadratic</td>
<td>-0.038</td>
<td>0.074</td>
<td>-0.254</td>
<td>0.303</td>
<td>0.394</td>
<td>-0.079</td>
</tr>
<tr>
<td>(0.367)</td>
<td>(0.370)</td>
<td>(0.350)</td>
<td>(1.420)</td>
<td>(1.400)</td>
<td>(1.380)</td>
<td>(1.380)</td>
</tr>
<tr>
<td>DTL linear</td>
<td>-0.134</td>
<td>-0.161</td>
<td>-0.029</td>
<td>-1.093</td>
<td>-1.030</td>
<td>-0.723</td>
</tr>
<tr>
<td>(0.134)</td>
<td>(0.133)</td>
<td>(0.128)</td>
<td>(0.203)</td>
<td>(0.194)</td>
<td>(0.184)</td>
<td>(0.184)</td>
</tr>
<tr>
<td>Degree-of-structure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Little structure</td>
<td>0.036</td>
<td>—</td>
<td>0.037</td>
<td>0.013</td>
<td>—</td>
<td>0.013</td>
</tr>
<tr>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.008)</td>
<td>(0.008)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>More structure</td>
<td>-0.056</td>
<td>—</td>
<td>-0.057</td>
<td>-0.017</td>
<td>—</td>
<td>-0.017</td>
</tr>
<tr>
<td>(0.010)</td>
<td>(0.010)</td>
<td>(0.012)</td>
<td>(0.012)</td>
<td>(0.012)</td>
<td>(0.012)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>Domain</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Humanities</td>
<td>0.007</td>
<td>0.012</td>
<td>—</td>
<td>0.018</td>
<td>0.016</td>
<td>—</td>
</tr>
<tr>
<td>(0.012)</td>
<td>(0.012)</td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>Language</td>
<td>-0.048</td>
<td>-0.019</td>
<td>—</td>
<td>-0.080</td>
<td>-0.045</td>
<td>—</td>
</tr>
<tr>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.017)</td>
<td>(0.017)</td>
<td>(0.017)</td>
<td>(0.017)</td>
<td>(0.017)</td>
</tr>
<tr>
<td>Mathematics</td>
<td>0.111</td>
<td>0.068</td>
<td>—</td>
<td>-0.060</td>
<td>-0.065</td>
<td>—</td>
</tr>
<tr>
<td>(0.036)</td>
<td>(0.034)</td>
<td>(0.039)</td>
<td>(0.037)</td>
<td>(0.037)</td>
<td>(0.037)</td>
<td>(0.037)</td>
</tr>
<tr>
<td>Science</td>
<td>-0.145</td>
<td>-0.195</td>
<td>—</td>
<td>-0.065</td>
<td>-0.136</td>
<td>—</td>
</tr>
<tr>
<td>(0.019)</td>
<td>(0.018)</td>
<td>(0.031)</td>
<td>(0.029)</td>
<td>(0.029)</td>
<td>(0.029)</td>
<td>(0.029)</td>
</tr>
<tr>
<td>Social science</td>
<td>0.111</td>
<td>0.121</td>
<td>—</td>
<td>0.113</td>
<td>0.110</td>
<td>—</td>
</tr>
<tr>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.018)</td>
<td>(0.018)</td>
<td>(0.018)</td>
<td>(0.018)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>Institution</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lac1</td>
<td>-0.019</td>
<td>-0.021</td>
<td>-0.017</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>(0.017)</td>
<td>(0.016)</td>
<td>(0.016)</td>
<td>(0.016)</td>
<td>(0.016)</td>
<td>(0.016)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>lac2</td>
<td>0.126</td>
<td>0.124</td>
<td>0.110</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>lac3</td>
<td>-0.176</td>
<td>-0.166</td>
<td>-0.156</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>lac4</td>
<td>0.074</td>
<td>0.069</td>
<td>0.068</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>ru1</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>-0.001</td>
<td>-0.002</td>
<td>-0.004</td>
</tr>
<tr>
<td>(0.011)</td>
<td>(0.011)</td>
<td>(0.011)</td>
<td>(0.011)</td>
<td>(0.011)</td>
<td>(0.011)</td>
<td>(0.011)</td>
</tr>
<tr>
<td>ru2</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.001</td>
<td>0.001</td>
<td>0.003</td>
</tr>
<tr>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>Variances</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>theta (%)</td>
<td>7.1</td>
<td>6.3</td>
<td>7.3</td>
<td>8.8</td>
<td>7.5</td>
<td>7.5</td>
</tr>
<tr>
<td>var[r]</td>
<td>0.075</td>
<td>0.075</td>
<td>0.068</td>
<td>0.177</td>
<td>0.159</td>
<td>0.162</td>
</tr>
<tr>
<td>Degree-of-fit (%)</td>
<td>38.0</td>
<td>50.4</td>
<td>38.0</td>
<td>44.1</td>
<td>45.3</td>
<td>44.3</td>
</tr>
<tr>
<td>Asymmetry F</td>
<td>10.1</td>
<td>12.2</td>
<td>10.7</td>
<td>6.4</td>
<td>8.1</td>
<td>8.1</td>
</tr>
</tbody>
</table>

Note: The standard errors of the coefficients are in parentheses.

example, LAC3 tends to favors small classes, especially in science, whereas LAC2 favors larger ones. Finally, we note that the forecast error variance (theta) is smaller for science and math than for the other domains.

4.4. Departmental Optimization Under Uncertainty

The model as derived in Equations (1), (2) and (3) predicts that the observed negative marginal utility of teaching load will drive departments toward underallocating section assignments. That is, departmental utility could be maximized somewhere on the right side of the ACS-utility curve, where the curve is downward-sloping because the class size is greater than its norm. However, at the end of our report on first results we observed that the frequency distribution of the ratio of actual to normal class size is skewed toward ratios less than one (Massy and Zemsky, 1994, Figure 6). We speculated that the tendency toward classes smaller than their norms represents "the next turn of the academic ratchet". That statement still seems accurate, but now we can begin to describe the mechanism involved.

The model as derived represents a "certainty-equivalent" (CE) formulation: the department applies its utility-maximization decision rule Equation (3) based on the enrollment-forecast point estimate without taking the probable forecast variance into account. Suppose, however, that the department seeks to zero its "expected marginal utility" (EMU) in light of the enrollment forecast. The required correction for the effect of enrollment forecast error on expected marginal utility turns out to be identical to the conditional
### Table 4. Free, constrained, and utility models

<table>
<thead>
<tr>
<th></th>
<th>LAC Utility</th>
<th>LAC Fixed</th>
<th>LAC Free</th>
<th>RU Utility</th>
<th>RU Fixed</th>
<th>RU Free</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Utility weights</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACS linear</td>
<td>0.064</td>
<td>0.051</td>
<td>0.070</td>
<td>0.091</td>
<td>0.099</td>
<td>0.125</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.009)</td>
<td>(0.006)</td>
<td>(0.011)</td>
<td>(0.010)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>DTL quadratic</td>
<td>-0.038</td>
<td>—</td>
<td>—</td>
<td>0.303</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(0.367)</td>
<td></td>
<td></td>
<td>(1.420)</td>
<td></td>
<td>—</td>
</tr>
<tr>
<td>DTL linear</td>
<td>-0.134</td>
<td>—</td>
<td>—</td>
<td>-1.093</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(0.134)</td>
<td></td>
<td></td>
<td>(0.203)</td>
<td></td>
<td>—</td>
</tr>
<tr>
<td>DTL fixed</td>
<td>—</td>
<td>-0.012</td>
<td>—</td>
<td>—</td>
<td>-0.015</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>-0.004</td>
<td>—</td>
<td></td>
<td>-0.003</td>
<td>—</td>
</tr>
<tr>
<td><strong>Degree-of-structure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Little structure</td>
<td>0.036</td>
<td>0.034</td>
<td>0.036</td>
<td>0.013</td>
<td>0.015</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>More structure</td>
<td>-0.056</td>
<td>-0.052</td>
<td>-0.055</td>
<td>-0.017</td>
<td>-0.020</td>
<td>-0.018</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.010)</td>
<td>(0.010)</td>
<td>(0.012)</td>
<td>(0.012)</td>
<td>(0.012)</td>
</tr>
<tr>
<td><strong>Domain</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Humanities</td>
<td>0.007</td>
<td>0.004</td>
<td>0.006</td>
<td>0.018</td>
<td>0.005</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.012)</td>
<td>(0.012)</td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>Language</td>
<td>-0.048</td>
<td>-0.039</td>
<td>-0.038</td>
<td>-0.080</td>
<td>-0.025</td>
<td>-0.010</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.015)</td>
<td>(0.014)</td>
<td>(0.017)</td>
<td>(0.015)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>Mathematics</td>
<td>0.111</td>
<td>0.098</td>
<td>0.114</td>
<td>-0.060</td>
<td>-0.095</td>
<td>-0.061</td>
</tr>
<tr>
<td></td>
<td>(0.036)</td>
<td>(0.034)</td>
<td>(0.035)</td>
<td>(0.039)</td>
<td>(0.038)</td>
<td>(0.038)</td>
</tr>
<tr>
<td>Science</td>
<td>-0.145</td>
<td>-0.144</td>
<td>-0.149</td>
<td>-0.065</td>
<td>-0.068</td>
<td>-0.098</td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.019)</td>
<td>(0.019)</td>
<td>(0.031)</td>
<td>(0.031)</td>
<td>(0.031)</td>
</tr>
<tr>
<td>Social science</td>
<td>0.111</td>
<td>0.107</td>
<td>0.106</td>
<td>0.113</td>
<td>0.075</td>
<td>0.063</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.018)</td>
<td>(0.017)</td>
<td>(0.017)</td>
</tr>
<tr>
<td><strong>Institution</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lac1</td>
<td>-0.019</td>
<td>-0.015</td>
<td>-0.017</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.016)</td>
<td>(0.016)</td>
<td></td>
<td></td>
<td>—</td>
</tr>
<tr>
<td>lac2</td>
<td>0.126</td>
<td>0.118</td>
<td>0.121</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.013)</td>
<td></td>
<td></td>
<td>—</td>
</tr>
<tr>
<td>lac3</td>
<td>-0.176</td>
<td>-0.170</td>
<td>-0.166</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.013)</td>
<td></td>
<td></td>
<td>—</td>
</tr>
<tr>
<td>lac4</td>
<td>0.074</td>
<td>0.073</td>
<td>0.069</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.013)</td>
<td>(0.013)</td>
<td></td>
<td></td>
<td>—</td>
</tr>
<tr>
<td>ru1</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>-0.001</td>
<td>-0.002</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.011)</td>
<td>(0.011)</td>
<td></td>
<td></td>
<td>—</td>
</tr>
<tr>
<td>ru2</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.001</td>
<td>0.002</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.009)</td>
<td></td>
<td></td>
<td>—</td>
</tr>
<tr>
<td><strong>Variances</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>theta (%)</td>
<td>7.1</td>
<td>7.1</td>
<td>6.8</td>
<td>8.8</td>
<td>7.8</td>
<td>7.5</td>
</tr>
<tr>
<td>var[ν]</td>
<td>0.075</td>
<td>0.068</td>
<td>0.072</td>
<td>0.177</td>
<td>0.159</td>
<td>0.164</td>
</tr>
<tr>
<td>Degree-of-fit (%)</td>
<td>38.0</td>
<td>37.5</td>
<td>38.1</td>
<td>44.1</td>
<td>43.2</td>
<td>43.0</td>
</tr>
<tr>
<td>Asymmetry F</td>
<td>10.1</td>
<td>10.4</td>
<td>10.9</td>
<td>6.4</td>
<td>7.8</td>
<td>8.1</td>
</tr>
</tbody>
</table>

Note: The standard errors of the coefficients are in parentheses.

The mean of $\xi$, derived in Equation (11). This produces an estimating equation identical that for the bias-adjusted certainty-equivalent formulation given by Equation (13). The utility-function parameter estimates are the same for the two models, though one may be able to distinguish between them by observing the actual class-size numbers.

Figure 5 illustrates the optimization for the number of course sections ($s$), using the RU results presented in Table 2 and assuming a department with only 10 teaching assignments and a discussion course with enrollment equal to the normal class size (i.e. $S = 10$, $y - \alpha'z = 0$). The upward-sloping curve (B) shows the right-hand side of Equation (8) as a function of $s$: the intercept equals $\lambda_{\text{ACSL}}$ and the slope depends on the two DTL terms. The lower of the two downward-sloping curves (A), labeled “certainty equivalent”, represents the left-hand side of Equation (8) minus the error term: $y^* = \max[2 \log(E[y^*]) - \alpha'z, 0]$. (We can ignore the fact that $s$ is integer for purposes of the illustration.) The higher upward-sloping curve (C), labeled “expected marginal utility”, adds the forecast error correction: $y^* + E[\eta|y^*]$. The certainty-equivalent optimum occurs where curve A intersects curve B. The solution requires $s < 1$ when $\lambda_{\text{ACSL}}$ is positive: actual class size will exceed normal class size.
### Table 5. Domain-specific fixed model: All institutions

<table>
<thead>
<tr>
<th></th>
<th>Humanities</th>
<th>Language</th>
<th>Math</th>
<th>Science</th>
<th>Social science</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Utility weights</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACS linear</td>
<td>0.122</td>
<td>-0.014</td>
<td>0.071</td>
<td>0.050</td>
<td>0.126</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.018)</td>
<td>(0.047)</td>
<td>(0.055)</td>
<td>(0.025)</td>
</tr>
<tr>
<td>DTL constrained</td>
<td>-0.004</td>
<td>-0.041</td>
<td>-0.006</td>
<td>-0.045</td>
<td>-0.008</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.008)</td>
<td>(0.014)</td>
<td>(0.041)</td>
<td>(0.013)</td>
</tr>
<tr>
<td><strong>Degree-of-structure</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Little structure</td>
<td>-0.004</td>
<td>0.030</td>
<td>0.041</td>
<td>0.028</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.013)</td>
<td>(0.064)</td>
<td>(0.036)</td>
<td>(0.015)</td>
</tr>
<tr>
<td>More structure</td>
<td>0.008</td>
<td>-0.052</td>
<td>-0.017</td>
<td>-0.020</td>
<td>-0.044</td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.023)</td>
<td>(0.027)</td>
<td>(0.026)</td>
<td>(0.023)</td>
</tr>
<tr>
<td><strong>Institution</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lac1</td>
<td>-0.044</td>
<td>0.021</td>
<td>-0.023</td>
<td>-0.033</td>
<td>-0.009</td>
</tr>
<tr>
<td></td>
<td>(0.075)</td>
<td>(0.044)</td>
<td>(0.158)</td>
<td>(0.108)</td>
<td>(0.055)</td>
</tr>
<tr>
<td>lac2</td>
<td>0.138</td>
<td>0.142</td>
<td>0.209</td>
<td>0.085</td>
<td>0.102</td>
</tr>
<tr>
<td></td>
<td>(0.041)</td>
<td>(0.045)</td>
<td>(0.088)</td>
<td>(0.060)</td>
<td>(0.053)</td>
</tr>
<tr>
<td>lac3</td>
<td>-0.142</td>
<td>-0.192</td>
<td>0.090</td>
<td>-0.741</td>
<td>-0.038</td>
</tr>
<tr>
<td></td>
<td>(0.043)</td>
<td>(0.046)</td>
<td>(0.105)</td>
<td>(0.086)</td>
<td>(0.048)</td>
</tr>
<tr>
<td>lac4</td>
<td>0.142</td>
<td>-0.046</td>
<td>-0.047</td>
<td>0.173</td>
<td>0.089</td>
</tr>
<tr>
<td></td>
<td>(0.036)</td>
<td>(0.048)</td>
<td>(0.152)</td>
<td>(0.070)</td>
<td>(0.045)</td>
</tr>
<tr>
<td>ru1</td>
<td>0.034</td>
<td>0.072</td>
<td>-0.242</td>
<td>-0.047</td>
<td>-0.095</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.026)</td>
<td>(0.073)</td>
<td>(0.053)</td>
<td>(0.032)</td>
</tr>
<tr>
<td>ru2</td>
<td>-0.076</td>
<td>-0.023</td>
<td>0.043</td>
<td>0.343</td>
<td>0.023</td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.021)</td>
<td>(0.054)</td>
<td>(0.066)</td>
<td>(0.027)</td>
</tr>
<tr>
<td><strong>Variance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>theta (%)</td>
<td>8.8</td>
<td>8.9</td>
<td>5.2</td>
<td>4.4</td>
<td>10.5</td>
</tr>
<tr>
<td>var[v]</td>
<td>0.079</td>
<td>0.064</td>
<td>0.091</td>
<td>0.177</td>
<td>0.156</td>
</tr>
<tr>
<td><strong>Degree-of-fit (%)</strong></td>
<td>52.4</td>
<td>35.1</td>
<td>42.1</td>
<td>37.9</td>
<td>41.4</td>
</tr>
<tr>
<td>Sample size</td>
<td>980</td>
<td>918</td>
<td>189</td>
<td>502</td>
<td>902</td>
</tr>
</tbody>
</table>

Note: The standard errors of the coefficients are in parentheses.

---

**Figure 5.** Certainty-equivalent and expected marginal utility decision-rule terms, and the right-hand side, for a department's utility maximization.

In this case, however, adding the error correction can shift the solution (now the intersection between curve A and curve C) to $s > 1$. The department guards against the heavy disutility of large classes by "buying insurance" in the form of extra sections. Ignoring the integer constraint, the illustration shows a certainty-equivalent solution of about $s = 0.92$ and an EMU solution of $s = 1.07$, which correspond to average class sizes of 9% larger and 7% smaller than the norm, respectively. Thus the "insurance" produces a reduction of 16% in average class size.

The illustration demonstrates the direction of the EMU-optimization effect and suggests that its magnitude may be material, but quantification would require simulating the optimization—with the integer constraint included—over the whole dataset. The simulation would have to take account of the fact that the dataset contains enrollment outcomes, not the enrollment forecasts upon which the departments are assumed to have based their calculations. Hence prediction would equal the expected number of sections, based on the probability distribution for what the forecast might have been given the actual enrollment: that is...
\[
E[s_i^*] = \int S^i E[ X_i^* f(E_i^*)] dE_i^*
\]

where \(E[S^i] \) is the predicted number of sections for the \(i\)th course in the dataset, \(E_i^* \) is the forecast enrollment, \(s_i^* \) is the simulated optimum number of sections based on \(E_i^* \), and \(f(E_i^*) \) is the probability density of \(E_i^* \) conditional on \(E_i^* \). This quantity can be calculated from the estimated model parameters, but the simulation remains a subject for future research.

5. CONCLUSION

This paper presents a utility model for academic department decision making and describes the structural specifications and estimation techniques needed to analyze it. The techniques produce stable results, and the model adds to the growing body of theory and empirical findings about utility maximization in higher education. The model confirms the class-size utility asymmetry predicted by the authors' academic ratchet theory, but shows that marginal utility with respect to teaching loads is always negative. Curricular structure and disciplinary domain shift normal class size in the expected directions. The alternative DTL-free and DTL-fixed models also produce reasonable results, but they do not out-perform the base model except in the domain-specific case where the fixed model mitigates collinearity. Future work will involve extending the model and estimator to include additional variables, as well as building the simulation described above.

Acknowledgements—The Academic Department Modeling Project was sponsored by the Pew Charitable Trusts. Joy Mundy’s invaluable assistance in perfecting and implementing the estimation algorithm and Ted Fu’s assistance during the later phases of the research are gratefully acknowledged.

REFERENCES


University of Pennsylvania, Philadelphia, PA.


APPENDIX

The relation between B- and C-departments is given by the following matrix equation which maps the vector of B-department sections, \(s_B = \{S_{B1}, S_{B2}, \ldots \} \), into the vector of C-department sections, \(s_C = \{S_{C1}, S_{C2}, \ldots \} \).

\[
s_C = M_B s_B
\]  

\(M_B \) is the “faculty mapping matrix” described in the text, and its elements \(m_{Bj} \) represents the fraction of sections in courses listed in C-department \(j \) that are taught by faculty in B-department \(k \). (We could map B-department auxiliary faculty into the C-departments, but since A is exogenous that refinement is unnecessary.) \(M_B \) equals the identity matrix if there is a one-to-one correspondence between the B- and C-departments; however, the addition of interdisciplinary C-departments makes the matrix rectangular. We adopted the convention that the first \(n_B \) rows of \(M_B \) represent the B-departments, which also are C-departments since they invariably are listed in the catalog. The remaining rows of \(M_B \) represent the interdisciplinary C-departments. The equations presented in the text use \(M_B = I \), but the empirical results do not depend on this assumption.
A Departmental Utility Model

The first step in deriving the general case is to ask what happens at the B-department level when the number of sections taught by a C-department changes. The faculty mapping matrix, $M_f$, tells us what proportion of B-department teaching assignments currently are used in each C-department, but it does not provide information on changes in the B-departments’ teaching loads in response to changes in the demand for teaching by C-departments. Therefore, we must introduce some additional structure.

Let $\rho_{ij}$ be the expected change in B-department $k$’s teaching load in response to a one-unit change in the teaching sections demanded by C-department $j$. In other words,

$$\rho_{ij} = \frac{\partial S_k}{\partial S_j}$$

The simplifying assumption that $M_f = I$ implies that $\rho_{ij}$ equals 1 when $j = k$ and 0 otherwise. However, we need an additional postulate to determine $\rho_{ij}$ in the general case.

Two possible formulations suggest themselves:

*Modal model:* an extra section in C-department $j$ increases the teaching load in the B-department that already contributes the most staffing to $j$; in other words, the modal B-department is responsible for the C-department’s incremental demand, with the other B-departments’ commitments being limited to their current teaching assignments. This means $\rho_{ij} = 1$ if $m_{jk}$ is the largest value in its row and $\rho_{ij} = 0$ otherwise.

*Proportional model:* an extra section in C-department $j$ increases the load in B-department $k$ in proportion to the existing contribution of $k$ to the staffing of $j$. In other words, all B-departments participate proportionally in the C-department’s incremental demand. This leads to the formula $\rho_{ij} = m_{jk}$.

Preliminary tests showed that for our data there is little difference between the two formulations. Therefore, in the interest of simplicity, we report results based on the modal model.

Utility maximization involves joint decision making between the B- and C-departments, since the former represent the faculty collegium with respect to discretionary time—and thus research and scholarship—and the latter are responsible for delivering educational services. Denoting B-departments by the subscript $k$ and C-departments by $j$, text Equation (2) becomes:

$$\frac{\partial U_j}{\partial S_j} = \frac{\partial u_{ACS_j}}{\partial S_j} + \frac{\partial S_j}{\partial S_i} \sum_k \frac{\partial u_{DTL_k}}{\partial S_j}$$

$$= \frac{\partial u_{ACS_j}}{\partial S_j} + \frac{\partial S_j}{\partial S_i} \sum_k \rho_{kj} \frac{\partial u_{DTL_k}}{\partial S_j}$$

$$= \frac{\partial u_{ACS_j}}{\partial S_j} + \frac{\partial S_j}{\partial S_i} \sum_k \rho_{kj} \frac{\partial u_{DTL_k}}{\partial S_j} = 0,$$

for all $i \in j$. (A2)

Carrying the extension forward to text Equation (3) yields:

$$0 = 2\lambda_{ACS,0} \max \left[ \log \frac{E_i}{\eta_0 S_i} - \alpha z_i, 0 \right] + \lambda_{ACSL}$$

$$-\lambda_{DTL,0} \sum_k 2\rho_{kj} S_j S_i - A_i \max \left[ \log \frac{S_i - A_i}{\eta_0 F_i}, 0 \right]$$

$$-\lambda_{DTL,0} \sum_k \rho_{kj} S_j S_i - A_i, \text{ for all } i \in j.$$ (A3)

The remaining text equations are transformed similarly.