# Size matters. The relevance and Hicksian surplus of preferred college class size

Philipp Mandel[a], Bernd Süssmuth[a,b,*]

[a] *University of Leipzig, Germany*
[b] *CESifo Munich, Germany*

## ARTICLE INFO

## ABSTRACT

The contribution of this paper is twofold. First, we examine the impact of class size on student evaluations of instructor performance using a sample of approximately 1400 economics classes held at the University of Munich from Fall 1998 to Summer 2007. We offer confirmatory evidence for the recent finding of a large, highly significant, and nonlinear negative impact of class size on student evaluations of instructor effectiveness that is robust to the inclusion of course and instructor fixed effects. Beyond that, we run a survey based on the contingent valuation method and a representative sample of all Munich students of management science to quantify the welfare surplus of preferred class size. We find the average monetary value students ascribe to their preferred class size to lie between 5 and 300 Euros per semester and student. In an upper bound scenario, implied Hicksian surpluses can reach values of close to 500 Euros per semester and student.

© 2011 Published by Elsevier Ltd.

## 1. Introduction

Although there has been made some recent progress (Bedard & Kuhn, 2008; Kokkelenberg, Dillon & Christy, 2008; Westerlund, 2008), college courses still represent a relatively novel laboratory from which to infer class size effects. We subscribe to the view of Bedard and Kuhn (2008) that summarizes the abundant wealth of literature on class size, referring to the comprehensive reviews by Hanushek (2003) and Krueger (2003), in two central insights: (i) results can depend considerably on econometric specification and (ii) the profession has not yet reached a consensus estimate of the impact of class size on student performance. Bedard and Kuhn (2008) are the first to show

that insight (i) does, in contrast to results for test-based outcomes at both primary/secondary and college level, not apply to the result of a large negative impact of class size on instructor effectiveness as measured by college-level course evaluations. In the literature analyzing introductory and intermediate college economics courses, there is rather mixed evidence for the relationship between class size and student performance: little or no evidence is found by Saunders (1980) and Kennedy and Siegfried (1997) using scores on the U.S. TUCE (Test of Understanding College Economics) exam, while Lopus and Maxwell (1995) and Kennedy and Siegfried (1997) find a positive relationship using scores on the TUCE III exam. Finally, Arias and Walker (2004) and Kokkelenberg, Dillon, and Christy (2008) relying on student exam points and grades at public universities in the United States find a negative relationship. All these student test score-based studies are to some extent subject to measurement error, instructors' discretion over grades, attrition between courses, and several other deficiencies (Bedard & Kuhn, 2008, p. 254).

---

* Corresponding author at: Institute for Empirical Research in Economics (IEW), Econometrics, University of Leipzig, Grimmaische Str. 12, D-04109 Leipzig, Germany.

*E-mail addresses:* mandel@wifa.uni-leipzig.de (P. Mandel), suessmuth@wifa.uni-leipzig.de (B. Süssmuth).

A negative relationship between class size and instructional student evaluations is found in the studies by Bedard and Kuhn (2008) and Westerlund (2008) for a U.S. and a Scandinavian university, respectively. Both studies do not suffer from the possibility that results may confound the effects of class size and instructor quality as is the case for precursory work like McConnell and Sosin (1984), DeCanio (1986), and Siegfried and Walstad (1990).

The aim of the present paper is twofold. First, we analyze whether the findings of the large, highly significant, and nonlinear negative impact of class size on student evaluations of instructor effectiveness reported in the seminal study by Bedard and Kuhn (2008) can be replicated using data from a university outside the United States. Our sample consisting of 1438 economics classes (on 129 different topics) held by 299 instructors at the University of Munich from Fall 1998 to Summer 2007 exceeds the one of Bedard and Kuhn who studied 655 courses offered by 64 instructors between Fall 1997 and Spring 2004 at the University of California, Santa Barbara (UCSB). The econometric approach for the most part of our Section 2 adheres to their methodology. We find profound confirmatory evidence for the UCSB findings that we check – in contrast, e.g., to Westerlund (2008) – also for robustness to the inclusion of course and instructor fixed effects. The latter is particularly important considering problems of unobserved heterogeneity, for example, the possibility that the best instructors might have been systematically assigned to larger courses by department chairs.

Secondly, university-rating agencies, students, and tuition paying parents frequently claim to place a high weight on small classes and an implied less anonymous and more personal learning environment. However, as these preferred, i.e., most satisfying, class sizes at the college level are neither guaranteed nor does there exist a direct market (both holding, in particular, for public schools in continental Europe), class size has the notion of an intangible. We make a first attempt to quantify the implied welfare surplus of this non-marketed intangible using a survey based on the contingent valuation method (CVM) for a representative sample of all students enrolled in management science (*Betriebswirtschaftslehre*) at the two universities in Munich. We find that the monetary value that students ascribe to the preservation of the status quo class size lies between 5 and 300 Euros per semester per student over the range of class sizes. As is common practice in the CVM framework, we derive our estimates from stated willingness-to-pay and willingess-to-accept responses in the counterfactual, though realistic, scenario of a merger of the two departments. To the best of our knowledge, no study of college-level class size has used such a CVM approach.

## 2. Class size and instructor effectiveness

### 2.1. Student evaluation data

The data for this study include nearly all economics classes offered at the University of Munich (Ludwig Maximilian University, henceforth: LMU) from Fall 1998 (*Wintersemester 1998/1999*) to Summer 2007 (*Som-*
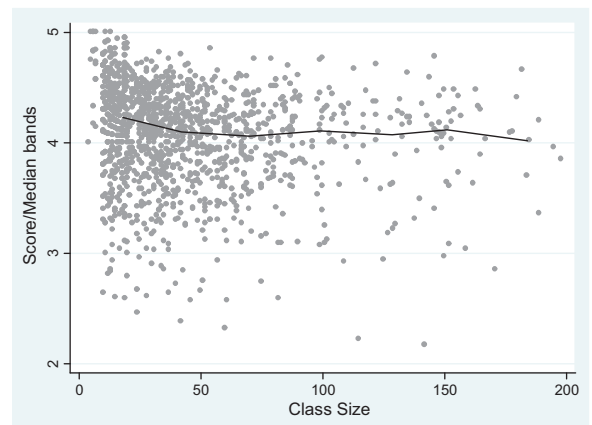


**Fig. 1.** Mean student evaluations across class size and median bands.

*mersemester 2007*)[1]: during this period of 18 semesters, 1438 economics classes on 129 different topics were offered by 299 instructors. Due to the fact that only in about 3% of considered classes, i.e., for 43 classes, class size exceeds 200 participants and in order to avoid placing too much weight at the upper extremity of the distribution (Fig. 1), we restrict our analysis to classes with ≤200 students.[2] Our data include information about class size, the semester (*Wintersemester*, *Sommersemester*), the year that each course was offered, the level of the class (lower division, upper division), whether or not the course is a program requirement, the instructor, and the average evaluation score. Summary statistics for all variables are reported in Table 1. We follow Bedard and Kuhn (2008) by using a variety of class size specifications to explore the relationship between class size and student evaluations of instructor effectiveness. In particular, we will use linear, quadratic, cubic, and – going beyond Bedard and Kuhn – also splined and fourth order polynomial specifications for class size, as well as categorical class size indicators to allow for the flexible estimation of any nonlinearity in the relationship between class size and student evaluations of instructor effectiveness.

The evaluation data are published and made available, corresponding to the natural unit of observation (Bedard & Kuhn, 2008, p. 255), in the form of student evaluation scores aggregated to class means:

$$E_{tci} = \frac{\sum_{j=1}^{R_{tci}} e_{tcij}}{R_{tci}}, \tag{1}$$

where $e$ denotes individual student evaluation scores, $E$ is the average class evaluation score, $R$ is the number of evaluation responses, $t$ denotes year ($t = 1998/1999, \ldots, 2007$), $c$ denotes course, and $i$ denotes instructor. The number of

---

[1] Disclosure of an instructor's rating results is not mandatory. However, the resulting attrition is less than 2%. A subsample of this data set is also used and described in Süssmuth (2006).

[2] Results for considering all class sizes can be found online in an extended version of our study: http://www.wifa.uni-leipzig.de/iew/size.pdf.