Volume XLI, No. 2 2015

The Journal of Economics (MVEA)

ARTICLES

(Continued on back)

(Continued from front)

BOOK REVIEWS

Volume XLI, No. 2 2015

The Journal of Economics (MVEA)

The Journal of Economics (MVEA) *ISSN 0361-6576*

The Journal of Economics (MVEA) is published bi-annually by the Missouri Valley Economic Association and is printed in Cedar Falls, Iowa. It is indexed by the *EconLit, e-JEL, and JEL on CD, Sociological Abstracts*, and *Academic Search*. Abstracts can be found on *RePEc*. In addition, the Missouri Valley Economic Association has entered into an electronic licensing relationship with EBSCO Publishing. The full text of *The Journal of Economics* (MVEA) can be found on EBSCO Publishing's databases. Articles presented at the annual meeting of the MVEA have a submission fee of $25. All other submissions are $50. Both are subject to the usual refereeing process. Submit three hard copies of manuscripts to David R. Hakes, Editor, *The Journal of Economics* (MVEA), Department of Economics, University of Northern Iowa, Cedar Falls, IA 50614 or e-mail hakes@uni.edu. Authors may, with no submission fee, submit a comment with regard to an article already published in the Journal or a short note. These should be narrowly focused pieces, not in excess of 2,000 words, which make some particularly salient point(s).

Memberships in the Association are on a calendar year basis and include the *Journal* for that year. Individual memberships are $40 per year. Institutional memberships are $100 per year. Address correspondence regarding memberships, changes in addresses, and other business of the Association (except regarding the *Journal*) to Fred Abraham, Executive Secretary and Treasurer, MVEA, University of Northern Iowa, Department of Economics, Cedar Falls, IA 50614-0129. The subscription price of the *Journal* for non-members is $32 per year in the U.S. and $44 outside the U.S. For library rates and other issues regarding the *Journal* (except editorial matters), contact Cullen Goenner, Business Editor, University of North Dakota; Dept. of Economics; 93 Centennial Drive; Grand Forks, ND 58202 or e-mail cullen.goenner@und.edu.

The editors wish to acknowledge the kind cooperation of the University of Northern Iowa and the University of North Dakota. We also thank Beverly Barber for preparing the manuscript for this issue.

Officers and Executive Committee
Missouri Valley Economic Association
The Year 2014

RICHARD ANDERSON, *Lindenwood University*. . . . . President
MICHAEL SEEBORG, *Illinois Wesleyan Univ.* . . . . . President-Elect
JANICE WEAVER, *Drake University*. . . . . Vice-President
LINDA GHENT, *Eastern Illinois Univ*. . . . . Past President & Director
OSCAR FLORES, *Minnesota State University - Moorhead* . . . . . Past President & Director
PHIL MILLER, *Minnesota State University-Mankato* . . . . . Past VP & Director
JENNIFER BALL*, Washburn University*. . . . . Director
CHRIS DOUGLAS, *Univ. Of Michigan-Flint* . . . . . Director
FRED J. ABRAHAM, *Univ. Of Northern Iowa*. . . Executive Secretary
FRED J. ABRAHAM, *Univ. Of Northern Iowa*. . . . . Treasurer
ED BLACKBURNE, *Sam Houston State U.* Director of Public Affairs
SCOTT FAUSTI, *South Dakota State Univ.* Chairman, Editorial Board

Editorial Board

Scott Fausti, Chair
Cullen Goenner
Robert Stanley Herren
David R. Hakes

Institutional Members

Embry-Riddle Aeronautical University
Southern Illinois University – Edwardsville
University of Missouri – Columbia
University of Missouri – Kansas City
University of Northern Iowa
University of Oklahoma

Presidents of MVEA

1963-64 PAUL MERZ
St. Louis University
1964-65 WILLIAM CASSIDY
Compliance Officer, NLRB
1965-66 RUSSELL BAUDER
Univ. of Missouri-Columbia
1966-67 KERMIT WATKINS
William Jewell College
1967-68 ARTHUR MEYERS
St. Louis University
1968-69 JOHN MOORE
Lindenwood College
1969-70 WALTER JOHNSON
Univ. of Missouri-Columbia
1970-71 DAVID LAGES
Southwest Missouri State U.
1971-72 JOSEPH MCKENNA
Univ. of Missouri-St. Louis
1972-73 CHARLES DRAKE
Central Missouri State Univ.
1973-74 JEROME SCHWIER
St. Louis
1974-76 PAUL JUNK
Washburn University
1976-77 ANN SCHWIER
Southern Illinois-Edwardsville
1977-78 LEONARD WHITE
U. of Arkansas-Fayetteville
1978-79 STEVEN LIN
Southern Illinois-Edwardsville
1979-80 JERALD BARNARD
University of Iowa
1980-81 STANLEY JOHNSON
Univ. of Missouri-Columbia
1981-82 S. CHARLES MAURICE
Texas A & M University
1982-83 WALTER MULLENDORE
Univ. of Texas-Arlington
1983-84 A.J. KONDONASSIS
University of Oklahoma
1984-85 COLDWELL DANIEL III
Memphis State University
1985-86 W. ROBERT BRAZELTON
U. of Missouri-Kansas City
1986-87 LAWRENCE F. ZIEGLER
Univ. of Texas-Arlington
1987-88 DONALD L. BUMPASS
Oklahoma State University
1988-89 WALLACE PETERSON
University of Nebraska
1989-90 MICHAEL EDGMAND
Oklahoma State University
1990-91 KAREN VORST
U. of Missouri-Kansas City
1991-92 DENNIS JOHNSON
University of South Dakota
1992-93 DAVID E.R. GAY
U. of Arkansas-Fayetteville
1993-94 JOHN A. TATOM
St. Louis Federal Reserve
1994-95 DON SCHILLING
Univ. of Missouri-Columbia
1995-96 FRED J. ABRAHAM
University of Northern Iowa
1996-97 RONALD L. MOOMAW
Oklahoma State University
1997-98 PAULINE FOX
Southeast Missouri State U.
1998-99 CLAUDIA PECK-HEATH
University of Kentucky
1999-00 VAMAN RAO
Western Illinois University
2000-01 RONALD D. GILBERT
Texas Tech University
2001-02 JAMES R. GALE
Michigan Tech University
2002-03 MICHAEL STELLERN
Rockhurst University
2003-04 WILLADEE WEHMEYER
MidAmerica Nazarene
2004 GARY LATANICH
Arkansas State University
2004-05 WILLIAM GERDES
Clarke College
2005-06 GALE BLALOCK
University of Evansville
2006-07 DENNIS PETRUSKA
Youngstown State University
2007-08 ROBERT STANLEY HERREN
North Dakota State Univ.
2008-09 DANIEL BIEDERMAN
University of North Dakota
2009-10 JANNETT HIGHFILL
Bradley University
2010-11 SCOTT FAUSTI
South Dakota State Univ.
2011-12 OSCAR FLORES
Univ. of Minnesota-Moorhead
2012-13 LINDA GHENT
Eastern Illinois University

Table of Contents

Volume XLI, No. 2, 2015

Articles

The Journal of Economics, XLI, No. 2, 2015

BOOK REVIEWS

An Analysis of Earnings Differentials between College-Educated Chinese Immigrants and US Natives: Who Has the Advantage?

Lu Liao and Michael Seeborg*

ABSTRACT. This paper uses 2012 American Community Survey (ACS) data to examine the relative earnings performance of full-time employed college-educated Chinese immigrants compared to full-time employed college-educated natives. The college-educated Chinese immigrant population has nearly a $7,000 unadjusted earnings advantage over natives. We show that this advantage is primarily due to differences in human capital endowments between the two groups. For example, college-educated Chinese immigrants are more likely to have PhD degrees and to choose majors that lead to higher paying occupations than college-educated natives. When we control for human capital and demographic differences, the Chinese immigrant earnings advantage becomes a small disadvantage. We also find that Chinese that came as youth, and have been in the U.S. for many years, have a significant earnings advantage over other Chinese immigrants. (J11, J31, J61)

I. Introduction

Chinese immigration is an important source of U.S. high skilled labor and the flow of Chinese immigrants is increasing rapidly, with Chinese immigration surging from about 366 thousand in 1980 to about 1.881 million in 2010 (McCabe, 2012). Given the recent surge of Chinese students in American universities, it is reasonable to expect high skill Chinese immigration to continue.

Chinese immigrants differ in important ways from the native population. For example, they are much more likely than natives to have advanced degrees and to cluster into niche occupations. American Community Survey (ACS) data for 2011 show that college-educated Chinese immigrants are overrepresented in finance, computer science and mathematics, architecture, engineering, life sciences and physical

*Corresponding author Michael Seeborg is an economics professor at Illinois Wesleyan University and Lu Liao is a student at Illinois Wesleyan University. This paper was presented at the 2014 Annual Meetings of the Missouri Valley Economic Association. We would especially like to thank our discussant, Michael Zimmer, for insightful comments and suggestions.

sciences (Seeborg and Maynard, 2014). These occupations tend to pay well and require quantitative and analytical skills.

Offsetting the advantages that Chinese immigrants have in educational attainment and career choice are a set of disadvantages. First, many Chinese immigrants have some degree of English language deficiency (McCabe, 2012). Second, the skills acquired in China are often not completely transferrable to the U.S. job market. Finally, it takes time for immigrants to develop effective networks and fully assimilate into the U.S. labor market. Because of these offsetting advantages and disadvantages, it is not clear whether Chinese immigrants would have earnings above or below native earnings.

The purpose of this paper is to use recent data from the 2012 American Community Survey (ACS) to examine the relative earnings performance of full-time employed college-educated Chinese immigrants compared to full-time employed college-educated natives. The paper's focus on the relative earnings performance of Chinese immigrants is relevant because of their increasing presence in the U.S. labor force, especially in important high skill professional occupations. It is important for both Chinese immigrants and the economy that compensation is fair and reflective of human capital.

A. RELATED LITERATURE

A number of empirical studies find that proficiency in English is a major determinant of immigrant economic assimilation and that deficiencies in English often cause significant earnings disadvantage (Chiswick, 1991; Chiswick and Miller, 1998; Sandford and Seeborg, 2003). McCabe (2012) finds that Chinese immigrants are less likely than the total immigrant population to speak English at home and that 62 percent of Chinese immigrants report limited English proficiency.

Friedberg (2000) and Haley and Taengnio (2011) find that because human capital acquired in countries of origin is only partially transferrable to the destination country, economic assimilation is hindered. Kaushal (2011) and Tong (2010) show that immigrants who receive most of their formal education in their countries of origin have significant earnings disadvantages compared to immigrants who receive more of their education in the destination countries. Kaushal (2011) focuses on professional job markets and shows that immigrant scientists and engineers who receive more education in the United States earn

significantly more income than immigrant scientists and engineers who do not. Finally, Zeng and Xie (2004) focus exclusively on Asian immigrants and find that those who receive more education in their country of origin are at a disadvantage to those who received more of their education in the United States.

It takes time for immigrants to develop connections and adapt to new institutions in their host countries. Given the significant institutional and cultural differences between the United States and China, Chinese immigrants should have more difficulty adjusting to the U.S. labor market than immigrants, say, from an English speaking country with market based institutions (Haley and Taengnoi, 2011). However, assimilation challenges may be lower for recent cohorts of Chinese immigrants because China is embracing market based institutions and encouraging English language training in Chinese schools and Universities. Because of changes like these in China, more recent Chinese immigrants may have human capital endowments that are more transferrable to the U.S. labor market than earlier arrivals that were raised with non-market institutions and often did not have access to English language instruction in school.

The paper proceeds with a description of the 2012 ACS database and then describes the two empirical models used to predict the earnings of a large sample of native and Chinese college graduates. The first model estimates separate earnings functions for natives and Chinese immigrants using a set of standard human capital and demographic variables. The second model is an expanded earnings function for Chinese immigrants only that adds immigrant specific variables to the estimation. These variables include citizenship status, English language ability, number of years in the United States, and age at the time of immigration.

II. Data and Empirical Model

We use the 2012 American Community Survey (Ruggles, et. al., 2010) to examine the experience of 5,232 full-time employed college-educated Chinese immigrants relative to 268,320 full-time employed college-educated natives. The sample includes Chinese immigrants and native born Americans who:

- are between 21 and 65 years old
- work at least 36 hours per week
- record at least 48 weeks of work during the previous year, and

- have at least a bachelor's degree.

The empirical analysis proceeds in four steps. First, we define key variables and present descriptive statistics. Second, we run separate OLS earnings functions for natives and Chinese immigrants that include both human capital and demographic variables.

LnWAGE = f (Degree Level, Field of Study, Demographic) (1)

Third, we use the native earnings function to estimate what Chinese immigrants would earn if they had their own characteristics but were paid according to the native reward structure. This is done by applying the mean Chinese immigrant characteristics to the native earnings function (i.e., regression coefficients) and estimating what Chinese earnings would be if they were rewarded according to the native earnings function. Then, the Chinese earnings estimate is compared to their actual earnings. Parity between Chinese immigrants and natives is reached if Chinese actual earnings are equal to Chinese earnings estimated from the native earnings function. On the other hand, if actual Chinese earnings are less than estimated earnings, Chinese immigrants have not reached earnings parity with natives.

Fourth, we run a regression for Chinese immigrants that add immigrant specific variables to Equation (1).

LnWAGE = f (Degree Level, Field of Study, Demographic, Immigrant Specific) (2)

Immigrant specific variables only apply to the immigrant population and include such things as citizenship status, years in the United States, and age of immigration. The results could suggest reasons why some Chinese immigrants have earnings advantages over other Chinese immigrants.

III. Results

A. DESCRIPTIVE STATISTICS AND VARIABLE DEFINITIONS

Table 1 presents descriptive statistics for the native and Chinese immigrant sample and Table 2 defines the variables that are to be used in the regression analysis. Of primary interest are the 2012 annual wages

and salaries. This is the total annual wage and salary income reported by the respondent for the 12 months prior to the survey date. Since the ACS survey was conducted in 2012, the annual income would be partly from 2011 and partly from 2012. It consists of wages, salaries, commissions, cash bonuses, tips, and other money income received from an employer, but does not include payments-in-kind or reimbursements for business expenses (Ruggles, et. al., 2010).

TABLE 1–Descriptive Statistics

	Natives	**Chinese**
Sample Size	277,862	5,381
Average 2012 Wages	$81,211	$88,155
Fields of Bachelor degree (percent)		
Mathematics & Statistics	1.3%	3.3%
Computer Science	2.9%	9.4%
Engineering	8.7%	25.1%
Physical Science	2.9%	8.7%
Biology & Life Sciences	4.8%	8.6%
Medical Science	6.5%	4.3%
Psychology	4.6%	1.5%
Agriculture	2.3%	1.2%
Social Science	9.7%	5.5%
Business	23.2%	19.8%
Education	11.4%	2.1%
Communication	4.5%	1.1%
Fine Arts	3.6%	2.6%
Liberal Arts and Humanities	7.8%	5.2%
Other	5.8%	1.6%
Total	100%	100%
Degree Level (percent)		
Bachelor	63.4%	37.2%
Master	25.8%	36.9%

TABLE 1–Descriptive Statistics

	Natives	**Chinese**
Degree Level (percent) - continued		
Professional	7.2%	5.7%
PhD	3.7%	20.3%
Total	100%	100%
Demographic Variables		
Age	43.7 years	43.1 years
Uhrswork	45.6 hours	43.8 hours
Female	46.1%	46.1%
Married	66.0%	72.0%
Nchild	0.84	0.86
Immigrant Specific Variables		
Not Citizen		33%
Bad English		4.2%
YEAR OF IMMIGRATION		
Arrived after 2007		8.8%
Arrived between 2002-2007		10.4%
Arrived before 2002		80.8%
Total		100%
AGE OF ARRIVAL IN U.S.		
Immigrated before Age 15		20.8%
Immigrated Age 15 - 17		5.6%
Immigrated Age 18 - 30		53.7%
Immigrated after Age 30		19.9%
Total		100%

The mean wage and salary earnings of Chinese immigrants exceed the mean earnings of their native counterparts by nearly $7,000 ($88,155 vs. $81,211). However, this unadjusted earnings advantage could be the result of Chinese immigrant human capital advantages. Indeed, Table 1

shows that Chinese immigrants are much more likely to have advanced degrees. For example, 20.3% of Chinese college-graduate immigrants have PhDs compared to 3.7% of college-graduate natives, and Chinese immigrants are more likely to hold masters' degrees (36.9 % vs. 25.8 %). Also, Chinese immigrants in our sample have very different distributions across bachelor degree fields with overrepresentation in mathematics and statistics, computer science, engineering, physical science, biology and other life sciences. These differences in degree level and area of study could explain much of the gross earnings advantage that Chinese have over natives.

Table 1 shows that Chinese immigrants and natives have similar demographic characteristics. For example, both groups have similar average age, number of children and percent female. However, Chinese immigrants in our sample were slightly more likely to be married than natives (72% vs. 66%).

Several variables apply only to immigrants, such as citizenship status, English language skills, year of immigration, and age at immigration. Table 1 shows that about 33 percent of the immigrants in our sample were not citizens, 4.2 percent indicated poor English language skills, 19 percent arrived in the United States after 2002, and 20.8 percent arrived before they were 18 years old.

Given the significant changes underway in China, such as rapid movement toward market institutions and English language instruction in schools, the year of immigration is an especially important immigrant specific variable. It is likely that more recent immigrants from China have better English language skills and greater familiarity with market-based institutions than earlier immigrants had when they immigrated. Our sample consists of both recent arrivals and immigrants who have been in the United States for many years (Table 1).

Immigrant age of arrival in the United States is also important since those that came as children likely received most of their formal education in the United States, and immigrants that arrived as adults likely received most of their formal education in China. While the ACS data set does not allow us to determine precisely where the immigrant received education, a good proxy is the variable "age at immigration." Table 1 shows that while the majority of immigrants in our sample arrived between the ages of 18 and 30, there are significant numbers who arrived under age 18 and over age 30.

TABLE 2–Variables and Descriptions

Variable	Description
Dependent	
Ln Wage	Natural log of real wage and salary income
Independent	
Master	1 if master's degree
Professional	1 if professional school degree
PhD	1 if doctor of philosophy degree
Agriculture	1 if bachelor degree of agriculture
Communication	1 if bachelor degree of communication
Computer Science	1 if bachelor degree of computer science
Education	1 if bachelor degree of education
Engineering	1 if bachelor degree of engineering
Biology & Life Sciences	1 if bachelor degree of biology & life sciences
Mathematics & Statistics	1 if bachelor degree of mathematics & statistics
Physical Science	1 if bachelor degree of physical science
Psychology	1 if bachelor degree of psychology
Social Science	1 if bachelor degree of social science
Fine Arts	1 if bachelor degree of fine arts
Medical Science	1 if bachelor degree of medical science
Business	1 if bachelor degree of business
Other	1 if bachelor degree of other fields
Age	Age in years at last birthday
Age Square	Age squared
Uhrswork	Usual hours worked per week
Female	1 if female
Married	1 if married
Married Female	Married * Female interaction
Nchild	Number of own children in household
Not Citizen	1 if without US citizenship
Bad English	1 if does not speak English well
Arrived after 2007	1 if immigrated after 2007
Arrived 2002-2007	1 if immigrated between 2002-2007
Immigrate<15	1 if less than 14 years old when arrived in U.S.
Immigrate 15-17	1 if 15-17 years old when arrived in U.S.
Immigrate 18-30	1 if 18-30 years old when arrived in U.S.

B. REGRESSION COMPARISONS OF CHINESE IMMIGRANTS AND NATIVES

We start by running separate but identical regressions: one for the sample of full-time employed college-graduate natives and the second for the sample of full-time employed college-graduate Chinese immigrants. Variable definitions are given in Table 2. The dependent variable, wage and salary income (LnWAGE), is determined by the level of education, undergraduate major, and a set of personal demographic variables. Three dummy variables are included for level of education (Master, Professional, and PhD). The omitted category consists of bachelor's degree holders who did not complete advanced degrees. Since having a bachelor's degree is the reference group (omitted category), the coefficients to the dummy variables for more advanced degree levels (Master, Professional and PhD) should be interpreted as the advantages that respondents with these advanced degrees have relative to those whose terminal degree is a bachelor's degree. The model also includes dummy variables for fourteen undergraduate areas of study, with the reference group being liberal arts and humanities majors.

Personal demographic dummy variables include whether the respondent is female (Female), being married (Married) and the interaction between these two variables (Female*Married). The interaction term is included to test the hypothesis that being female and married has a negative effect on earnings because of the possibility of increased childcare responsibilities that are often assumed by married women. Finally, age (Age, Age Square) and number of children (NChild) are included to control for the effects of experience and family responsibilities on earnings. Age and age squared are included to capture the nonlinear relationship between earnings and experience (Borjas, 2016, p. 271). Since earnings are expected to increase, but at a decreasing rate with age over the life cycle, we expect a positive sign to the "Age" coefficient and a negative sign to the "Age Square" coefficient.

The results are presented in Table 3. The column labeled "Natives" is for the natives only sample and the column labeled "Chinese: Model 1" is for the Chinese sample. The results are consistent with expectations. Nearly all of the coefficients are statistically significant with the expected sign, except for a few of the "field of study" coefficients. Like natives, Chinese immigrants seem to benefit from having an advanced degree relative to the bachelor's degree.

TABLE 3–Regression Results
(Robust Standard Errors in Parentheses)

Variable	Natives	Chinese (Model 1)	Chinese (Model 2)
Constant	8.350***	8.011***	8.218***
	(0.020)	(0.184)	(0.193)
Degree Level			
Master	0.165***	0.184***	0.216***
	(0.003)	(0.021)	(0.021)
Professional	0.500***	0.453***	0.437***
	(0.006)	(0.052)	(0.050)
PhD	0.309***	0.216***	0.287***
	(0.007)	(0.027)	(0.028)
Field of Study			
Agriculture	-0.092***	-0.036	-0.028
	(0.010)	(0.075)	(0.069)
Communication	0.061***	0.114	0.050
	(0.007)	(0.102)	(0.097)
Computer Science	0.253***	0.427***	0.359***
	(0.008)	(0.050)	(0.049)
Education	-0.153***	-0.140**	-0.113*
	(.005)	(0.069)	(0.064)
Engineering	0.296***	0.345***	0.293***
	(0.006)	(0.047)	(0.046)
Biology & Life Sciences	0.122***	0.110**	0.077
	(0.007)	(0.056)	(0.053)
Mathematics & Statistics	0.206***	0.317***	0.266***
	(0.012)	(0.060)	(0.058)
Physical Science	0.160***	0.208***	0.178***
	(0.008)	(0.054)	(0.051)
Psychology	-0.027***	0.220***	0.096
	(0.007)	(0.070)	(0.066)
Social Science	0.093***	0.227***	0.169***
	(0.006)	(0.062)	(0.060)
Fine Arts	-0.093***	0.071	0.081
	(0.008)	(0.072)	(0.067)
Medical Science	0.200***	0.246***	0.193***
	(0.006)	(0.062)	(0.060)
Business	0.158***	0.172***	0.099**
	(0.005)	(0.048)	(0.049)

TABLE 3–Regression Results
(Robust Standard Errors in Parentheses)

Variable	Natives	Chinese (Model 1)	Chinese (Model 2)
Field of Study (continued)			
Other	-0.092***	0.083	0.007
	(0.006)	(0.084)	(0.079)
Demographic			
Age	0.082***	0.105***	0.83***
	(0.001)	(0.008)	(0.008)
Age Square	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)
Uhrswork	0.012***	0.008***	0.007***
	(0.000)	(0.001)	(0.001)
Female	-0.096***	-0.032	-0.037
	(0.004)	(0.035)	(0.033)
Married	0.195***	0.136***	0.152***
	(0.004)	(0.033)	(0.031)
Married Female	-0.144***	-0.123***	-0.120***
	(0.005)	(0.040)	(0.038)
Nchild	0.019***	0.021*	0.007
	(0.001)	(0.011)	(0.011)
Immigrant Specific			
Not Citizen			-0.077***
			(0.023)
Bad English			-0.417***
			(0.046)
Arrived after 2007			-0.296***
			(0.041)
Arrived 2002-2007			-0.063**
			(0.030)
Immigrate<15			0.273***
			(0.038)
Immigrate 15-17			0.182***
			(0.044)
Immigrate 18-30			0.133***
			(0.026)
Sample Size	268,320	5,232	5,232
Adjusted R Square	0.269	0.195	0.266

Note: ***significant at .01 level, **significant at .05 level, *significant at .10 level

Particularly notable, is that Chinese immigrants, like natives, receive very large returns from professional degrees (e.g., law degrees and MDs) with a 45% premium over the bachelor's degree. The return from PhD's, while significant, is lower for Chinese than for natives (21.6% vs. 30.9%).

Returns across areas of study show a similar pattern for Chinese immigrants and natives. For both groups the coefficients should be interpreted in reference to the omitted area of study, which is liberal arts and humanities. Of particular note are the relatively high returns for Chinese immigrants who have bachelors' degrees in computer science (42.7%), engineering (34.5%) and mathematics and statistics (31.7%). The high returns for these majors are important because they are disciplines that also attract a large number of Chinese immigrants as shown in Table 1.

The effect of the demographic control variables on earnings is very similar when comparing the native and Chinese immigrant equations. For both, earnings increase, but at a decreasing rate, with age; and marriage seems to increase earnings for men, but not for women. One notable difference between natives and Chinese immigrants is that being a woman is a statistically significant predictor for natives but not for Chinese immigrants. Therefore, Chinese immigrant women do not seem to be at a significant disadvantage relative to their male counterparts, ceteris paribus.[1]

C. SIMULATIONS

Next, we use the native earnings function to estimate what Chinese immigrants would earn if they had their own human capital and demographic characteristics but were rewarded according to the native earnings function. We do this by multiplying the coefficient estimates for natives in Table 3 by the average value of each Chinese characteristic from Table 1. For example, Table 1 shows that 36.9 percent of Chinese immigrants in our sample hold masters degrees. Therefore, we multiply 0.369 times the coefficient to MASTER in the native regression equation to determine the contribution of MASTER to LnWAGE. This process is repeated for all other variables and LnWAGE is estimated. When this is done, the estimated earnings for Chinese immigrants are 13.9 percent greater than their actual earnings. This means that, when demographics and human capital characteristics are taken into account, we cannot

conclude that the college-educated Chinese immigrants in our sample have reached parity with employed native college graduates. Indeed, what started out as a large unadjusted average wage advantage for Chinese immigrants over natives (Table 1), turns out to be an estimated 2.9 percent advantage for natives. Thus, after controlling for human capital and demographic differences, Chinese immigrants are at an earnings disadvantage to natives.

We next use the simulation results to partition the effect of native/Chinese differences in degree levels, field of study and demographic characteristics on earnings differentials. Table 4 shows that most of the gross unadjusted earnings advantage of Chinese over natives is attributable to degree level differences and differences in field of study. This is because Chinese immigrants are much more likely to have advanced degrees than natives and to pursue higher paying undergraduate fields of study.

TABLE 4–Summary of Simulation Results

	Simulated Percent Change	Chinese - Native Earnings Differentials
Average Wage Difference (without controls)		11%
Effect of Chinese/Native Degree Level Differences	6.2%	
Effect of Chinese/Native Field of Study Differences	8.8%	
Effect of Chinese/Native Demographic Differences	-1.1%	
Average Wage Difference (with all controls)		-2.9%

The first part of the decomposition estimates earnings from the native earnings function by changing the degree level distribution from the native's distribution to the Chinese distribution and computing the change in earnings. Natives clearly would have significantly higher

earnings if they had the Chinese distribution. In fact, the difference in degree level distributions accounts for 6.2 percentage points of the 11 percent earnings advantage that the Chinese immigrants in our sample have over natives.

The second part of the decomposition estimates earnings from the native earnings function by changing the field of study distribution from the native distribution to the Chinese distribution and then observing the change in earnings. This change causes a significant increase in the estimated earnings. Again, natives would have much higher earnings if they had the Chinese distribution of field of study. The difference in area of study distribution accounts for 8.8 percentage points of the 11 percent unadjusted earnings advantage that Chinese immigrants have over natives. Considered together, the advantages that Chinese immigrants have in degree level and area of study alone explain more than the unadjusted Chinese earnings advantage. This means that after we control for level of study and field of study the earnings advantage of Chinese immigrants becomes an earnings disadvantage.

The third part of the decomposition controls for differences in demographic characteristics between natives and Chinese immigrants. Again we multiply the averages of the Chinese demographic characteristics to the appropriate native coefficients. When we do this we see that natives would have slightly lower earnings (-1.1 percentage points) if they had the Chinese demographic characteristics.

When the three components of the decomposition are combined in Table 4, the estimated earnings disadvantage of the Chinese immigrants is -2.9 percent. Thus, what initially appeared to be a strong Chinese earnings advantage over natives is actually a small Chinese disadvantage after controlling for human capital endowments and demographic characteristics.

D. CHINESE MODEL WITH IMMIGRANT SPECIFIC VARIABLES

Finally, we explore the effect that immigration specific variables have on the earnings of Chinese immigrants by running the model for the Chinese immigrant sample that adds immigrant specific characteristics, such as citizenship status, years in the United States, age at immigration and English language proficiency. This model is labeled "Chinese Model 2" in Table 3. Chinese Model 2 differs from Chinese Model 1 that was discussed in the previous section in that it includes immigrant specific

variables in addition to the variables included in Chinese Model 1. This model includes all four groups of variables: degree level, field of study, demographic, and immigrant specific. Chinese Model 2 results are presented in the last column of Table 3.

Including the immigration specific variables increases the overall explanatory power of the Chinese immigrant regression considerably, with the adjusted R squared increasing from 0.19 to 0.26. The coefficient to "Not Citizen," for example, shows that not having citizenship decreases earnings by about 7.7 percent. In addition, Chinese immigrants who arrive at an earlier age have an earnings advantage over those who arrive when they are older, and immigrants who have been in the country for more years have an earnings advantage over those who arrive more recently. This pattern is consistent with expectations that are derived from human capital theory and are also consistent with findings in the immigration literature (Borjas, 1994, Chiswick, 1978).

Some of the immigrant specific variable effects are quite large. For example, those who immigrate before they reach 15 years have a 27.3 percent earnings advantage relative to those who immigrate when they are over 30 years. This result is consistent with prior research by Friedberg (2000) and Sandford and Seeborg (2003) that find that age of immigration is inversely related to earnings. A plausible explanation for the inverse relationship between age at immigration and earnings is that young immigrants receive more formal education in the United States and consequently do not encounter difficulties in transferring skills from the country of origin.

Also, the number of years in the United States is a significant predictor of earnings. For example, Table 3 shows that immigrants who arrived in the United States after 2007 are at a 29.6 percent earnings disadvantage compared to immigrants who arrived before 2002, and that those who do not have citizenship have an earnings disadvantage of 7.7 percent compared to immigrants who have citizenship. Also, the results show that reporting poor English language skills reduce earnings by 41.7 percent, which is a consistent with findings in previous research (Chiswick 1990; Chiswick and Miller 1998).

IV. Conclusions and Discussion

While college-educated Chinese immigrants enjoy a substantial unadjusted earnings advantage over natives (Table 1), the regression

analysis and simulations show that differences in degree level and area of study distributions explain much of this unadjusted earnings advantage of Chinese immigrants. For example, we find that Chinese immigrant overrepresentation among advanced degree holders and their overrepresentation in the higher paying areas of study contribute to their higher earnings. Once degree level and area of study are controlled for in the regression-based simulation, the Chinese earnings advantage becomes a disadvantage that is only slightly decreased by taking into account demographic differences between Chinese immigrants and natives (Table 4).

The analysis of the college-graduate Chinese immigrant earnings relative to native college graduates finds that Chinese earnings trail native earnings by about 2.9 percent once degree level, area of study, and demographic characteristics are taken into account. Therefore, we can conclude that our sample of college-educated Chinese immigrants is reasonably close to parity with their native counterparts.

While we do find that college-graduate Chinese immigrants as a group have nearly achieved income parity with natives, there are still large differences in earnings performance within the Chinese immigrant group. These differences are in part the result of immigrant specific variables like language skills, citizenship status, year of immigration, and age at immigration. The results of the regression that includes these immigrant specific variables (Chinese Model 2 in Table 3) show an earnings advantage for Chinese immigrants who immigrate when they are young, have been in the United States for many years, are U.S. citizens, and speak English well. Future research could partition the sample of immigrants along these dimensions and conduct simulations to see how these subgroups fare relative to the native comparison group. It is likely that this research would find that respondents with immigrant specific advantages would have significant earnings advantages over their native counterparts while those with immigrant specific disadvantages would have significant disadvantages.[2]

Fortunately, recent arrivals have the opportunity to increase their earnings over time as they acquire more U.S. specific human capital and improve language skills. Also, immigrants often attain citizenship as they continue to reside in the United States. Thus, there is reason to be very optimistic about the prospects for Chinese immigrants as a group. The disadvantages that often afflict new immigrant arrivals tend to disappear over time as immigrants acquire U.S. specific human capital and

citizenship rights.

We acknowledge that some Chinese earnings advantage may be the result of positive selection from the potential immigrant pool (Borjas, 1987), and that there also is likely selective emigration back to China of some immigrants who do not have a favorable labor market experience in the United States. However, from a U.S. labor market policy perspective, what matters is whether high skilled Chinese immigrants are productive labor market participants and whether their skills are being rewarded at the same rates as natives' skills. Our results suggest that college-educated Chinese immigrants have nearly reached parity with their native counterparts.

References

Benham, Lee. 1977. "Benefits of Women's Education within Marriage." *Journal of Political Economy* 82(2): S57-S71.

Borjas, George. J. 1994. "The Economics of Immigration." *Journal of Economic Literature* 32(4): 1667-1717.

Borjas, George J. 2016. Labor Economics. New York: McGraw Hill.

Borjas, George J. 1987. "Self-Selection and the Earnings of Immigrants." *American Economic Review* 77(4): 531-553.

Chiswick, Barry R. 1978. "The Effect of Americanization on the Earnings of Foreign-born Men." *Journal of Political Economy* 86(5): 897-921.

Chiswick, Barry R. 1991. "Speaking, Reading, and Earnings among Low-Skilled Immigrants." *Journal of Labor Economics*. 9(2): 149-170.

Chiswick, Barry R., and Paul W. Miller. 1998. "Language Skills and Earnings among Legalized Aliens." *Journal of Population Economics*. 12(1): 63-89.

Compton Janice and Pollak, Robert A. 2007. "Why Are Power Couples Increasingly Concentrated in Large Metropolitan Areas." *Journal of Labor Economics* 25(4): 475-512.

Friedberg, Rachel M. 2000. "You Can't Take It with You? Immigrant Assimilation and the Portability of Human Capital." *Journal of Labor Economics* 18 (2): 221-251.

Haley, Ryan M., and Sarinda Taengnoi. 2011. "The Skill Transferability of High-Skilled US Immigrants." *Applied Economics Letters* 18: 633-636.

Kaushal, Neeraj. 2011. "Earning Trajectories of Highly Educated Immigrants: Does Place of Education Matter?" *Industrial and Labor Relations Review* 64(2): 323-340.

Maynard, Paige and Michael Seeborg. 2014. "The Effect of Occupational Choice on the Earnings of College Educated Immigrants from China, India, the Philippines and Mexico," *The Journal of Economics (MVEA)*, 35(2): 19-35.

McCabe, Kristen. January 18, 2012. "Chinese Immigrants in the United States," Migration Information Source: Online Journal of the Migration Policy Institute, (http://migrationpolicy.org/article/chinese-immigrants-united-states).

Ruggles, Steven, J.T. Alexander, Katie Genadek, R. Goeken, Matthew B. Schroeder, and Sobek Maria. 2010. Integrated Public Use Microdata Series: Version 5.0

[Machine-readable database]. Minneapolis: University of Minnesota.

Sandford, Jeremy and Michael Seeborg. 2003. "The Effects of Ethnic Capital and the Age of Arrival on the Standard of Living of Young Immigrants." *The Journal of Economics*, 29 (1), 27-48.

Tong, Yuying. 2010. "Place of Education, Gender Disparity, and Assimilation of Immigrant Scientists and Engineers Earnings." *Social Science Research* 39:610-626.

Zeng, Zhen. and Yu Xie. 2004. "Asian-Americans' Earnings Disadvantage Reexamined: The Role of Place of Education." *Journal of Sociology* 109 (5): 1075-1108.

Endnotes

1. While not the focus of this research, future research could focus on the relative earnings advantage of Chinese women over native women. One possible explanation is that the favorable earnings performance of Chinese women is that college-graduate female Chinese immigrants are likely to marry college-educated spouses (Benham, 1977) and these "power couples" are more likely to locate in large metropolitan areas that offer good employment prospects to both husband and wife (Benham, 1977; and Compton and Pollak, 2007).
2. Another line of future research that further explores immigrant specific determinants of Chinese earnings could take advantage of the fact that Chinese immigrants come from places of origin that differ greatly in level of economic development, market institutions, and governance systems. This research could identify the place of origin of Chinese immigrants and run separate analysis of native/Chinese earnings differentials for each group. The results could yield interesting insights into how place of origin effects the transferability of human capital to the U.S. For example, Chinese immigrants who grew up in Hong Kong or Taiwan have likely acquired more U.S. specific human capital than Chinese immigrants who grew up in Mainland China. The level of economic development, legal institutions and economic institutions in Hong Kong and Taiwan are more similar to U.S. institutions than those in Mainland China. This suggests that immigrants from Hong Kong and Taiwan should have higher returns on their early human capital investments than immigrants from Mainland China who may find less transferability of their human capital (Friedberg, 2000).

A General Concentration Index for Multiproduct Firms with Differentiated Products

Christopher Bailey and Samuel Raisanen*

ABSTRACT. Economists commonly use the Herfindahl-Hirschman Index (*HHI*) to measure the competitiveness of markets. However, the *HHI* has two significant problems. First, defining the market boundaries can be arbitrary. Second, the products included in the defined market are effectively assumed to be homogeneous. This paper proposes a differentiated product concentration index (*DPCI*) that reduces these problems by weighing substitutes differentially. The *DPCI* mitigates the problematic effects of expanding a market definition to change the perceived degree of competition. The *DPCI* has many desirable properties and collapses to the *HHI* when products are homogeneous. (D43, L11, L40)

I. Introduction

The measurement of market power and the determination of relevant market boundaries have long been of concern to economists both in general research and for antitrust purposes. Various concentration indices, including the Herfindahl-Hirschman Index (*HHI*), have been used indicate the degree of competition in markets and the degree of firms' pricing power. Yet the *HHI* presents a significant problem. Using it effectively assumes that all the products included in the defined market are homogeneous. A related issue in using a concentration index like the *HHI* lies in defining the market to measure. The choice of placing a boundary around the appropriate market can be challenging and arbitrary. This paper proposes a new concentration index, the differentiated product concentration index, a more flexible index than the *HHI* when firms compete within related, but not perfectly substitutable, markets.

As an illustration, consider the market for carbonated soft drinks. Within this market there are bottled soft drinks, canned soft drinks, and soft drink syrups sold in bulk. These consist of, for example, colas and non-colas, and sugared soft drinks and diet soft drinks. There are other

*Christopher Bailey, Department of Economics, Central Michigan University, Mt. Pleasant, MI 48859. E-mail: baile1c@cmich.edu.
Samuel Raisanen, Department of Economics, Central Michigan University, Mt. Pleasant, MI 48859. E-mail: raisa1sr@cmich.edu.

substitutes as well: non-carbonated fruit drinks, iced-tea drinks, energy drinks and, perhaps more distantly, water. Which of these potentially substitutable beverages should be included in a concentration index with soft drinks?

Perhaps the relevant market should be more narrowly defined than all soft drinks. Bottled diet lemon-lime carbonated drinks may be sufficiently different from sugared cola syrups that they should be in separate markets. On the other hand, suppose the market is indeed defined as carbonated soft drinks and only carbonated soft drinks. The *HHI* then ignores the constraining power that fruit juice, water, iced-tea drinks, and energy drinks have on the prices of soft drinks. The market could instead be widely aggregated so that these latter substitutes are included. However, using the *HHI* then implicitly assigns the more distant substitutes the same price constraining power as the close substitutes of soft drinks.

At its heart, the choice of market boundaries for the *HHI* involves the notion of substitutability. If products are perfectly substitutable, they certainly should be included in the market definition for a concentration index. If products are perfectly not substitutable, they certainly should not be included in the concentration index. The issue of inclusion or exclusion of products in a concentration index arises not in the black and white cases of perfect substitutes or perfect non-substitutes, but in the gray areas of imperfect substitutes. As products are typically differentiated, most are imperfectly substitutable with one another, so the ambiguity of the gray areas frequently arises.

Depending on the chosen market boundaries, a number of different concentration index values could be calculated for a soft drink market. Still, the question remains as to which would best describe the market. Stigler (1982, p. 9) hints at the dilemma of defining markets by asserting that "the concentration ratios will be awesome or trivial," depending on who wins an argument to define a market. The *HHI* faces a puzzle in defining the relevant market and, therefore, in portraying the amount of competition in the market. A concentration index is needed that can effectively include all substitutes so that the choice of market boundaries is not such a critical issue.

Different products may have different capabilities to constrain the prices of its substitutes. Stigler and Sherwin (1985, p. 562), in tackling the problem of determining which products (or submarkets) ought to be placed together in the same (overall) market, noted that separate "[sub]markets can show every level of interdependence from absolute homogeneity to complete independence." It would be useful to have a

concentration index that provided a middle ground between the absolute homogeneity of products and their complete inclusion in an index (thereby weighing the products fully) and the complete independence of products and their complete exclusion from an index (thereby not weighing the products at all). This would allow the level of competition to be gauged more effectively.

A step in the direction of a new concentration index was undertaken by Lijesen (2004), who derived a Herfindahl-Hirschman type index with regard to product quality differences. The general finding was that higher quality products should be weighted more in a concentration index than lower quality products. Cesari (2000) used "transitional probabilities" to model customers switching between firms and suggested an index that multiplied customer loyalty by market share to form a "Fidelity Index" to measure the degree of non-competition in a market. Bailey and Taylor (2009) developed an index for differentiated single product firms. The present paper extends the Bailey and Taylor article.

We introduce an alternative to the *HHI,* the differentiated product concentration index (*DPCI*). The *DPCI* can be useful in much the same settings as the *HHI* has been useful, but the *DPCI* allows for differentiated accounting of differentiated products while alleviating problems associated with defining market boundaries.

Interest in alternate measures to the classical Herfindahl-Hirschman Index can be found in numerous recent economic papers across economic subfields. Recent papers in financial economics such as Wen-Yu (2013), Amidu-Wolfe (2013), and Berger, et. al. (2009) use alternate *HHI* measures based on deposits and loans in analyzing competition between banks. In energy economics, Hellmer and Werell (2009) uses an alternate *HHI* measure, developed by Melnik, et al. (2008) based on the two largest firms market shares, to examine market power in the Nordic electricity markets. In agricultural economics, Qin and Zhang (2012) use an alternate *HHI* measure to calculate degree of income specialization. They use multiple categorizations of related income categories to derive multiple *HHI* measures. Examples of the need for alternate *HHI* measures are certainly not limited to these few areas. In section VI, we present examples of potential usage of our alternate measure in the areas of industrial organization, trade, and sports economics.

II. Differentiated Product Concentration Index

Consider the *HHI* with the market divided into *N* submarkets and multiple firms dividing the submarkets, where a submarket is a grouping of products that have a very high degree of substitutability between each other relative to their substitutability to products outside the submarket. The *HHI* value for firm-f can be expressed in the unusual form given by

$$C_f^{HHI} = \left[\sum_{n=1}^{N} \sigma_n s_{nf}\right](\sigma_1 s_{1f}) + \left[\sum_{n=1}^{N} \sigma_n s_{nf}\right](\sigma_2 s_{2f}) + \ldots + \left[\sum_{n=1}^{N} \sigma_n s_{nf}\right](\sigma_N s_{Nf}) \quad (1)$$

In equation 1, the fraction that a submarket has of the overall market is called a "submarket share" and is denoted by σ_n for submarket-*n*. The fraction that a given firm's products have of the relevant submarket is a "firm share" or "product share" and is denoted by s_{nf} for firm-*f*'s products in submarket-*n*. Thus, $\sigma_n s_{nf}$ is the market share that firm-*f*'s products in submarket-*n* have as a fraction of the overall market. Firm *f*'s total market share is then

$$S_f = \sigma_1 s_{1f} + \sigma_2 s_{2f} + \ldots + \sigma_N s_{Nf}.$$

The index number, C_f^{HHI}, is then firm *f*'s total market share squared. (This paper takes market shares σ and s to have the range 0 to 1, not 0 to 100, as is sometimes done.)

With *F* firms in a market, the *HHI* value for the overall market is,

$$HHI = \sum_{f=1}^{F} C_f^{HHI} = \sum_{f=1}^{F}\left[\sum_{m=1}^{N}\left[(\sigma_m s_{mf}) \cdot \sum_{n=1}^{N}(\sigma_n s_{nf})\right]\right] \quad (2)$$

As shall be seen below, the *DPCI* value for the overall market will be

$$DPCI = \sum_{f=1}^{F}\left[\sum_{m=1}^{N}\left[(\sigma_m s_{mf})\cdot\sum_{n=1}^{N}\frac{\sigma_n s_{nf} b_{mn}}{B_m}\right]\right] \quad (3)$$

The differences between the *HHI* and *DPCI* lie in the presence of the terms b_{mn} and B_m. These terms represent respectively, the substitutability of products between submarkets *m* and *n*, and the fraction of the overall market in which submarket *m* products effectively compete. The *HHI* takes the substitutability between all products in the defined market to be 1 (perfect) and the fraction of the market in which all products effectively compete as 1 (the entire market). However, as differentiated products are not perfectly substitutable, they do not compete with the same effectiveness against the entire market. Note, however, if all products in the market are, in fact, perfectly substitutable, then $b_{mn} = B_m = 1$ for all *m* and *n*, and the *DPCI* collapses to the *HHI*. This is indicated in Property 1 below.

A user of the *DPCI* has some flexibility in how to compute the substitutable parameter b_{mn} depending on the context. For example, in an anti-trust or industrial organization context, the substitution parameter could be taken as a function of the cross-price elasticity between the goods in two related markets. Alternatively, per the Department of Justice merger guidelines for the Hypothetical Monopolist Test, the percentage of sales lost to products in each competing market is a measure of substitutability. Other potential methods of computing substitutability parameters are considered in Section VI.

Using the *DPCI*, a firm may have multiple products in a submarket, but a given product is considered to be in only one submarket.[1] The *DPCI* is flexible with regard to the fineness of submarket division; a submarket may be a single product.[2] An overall "market" is the grouping of submarkets that have non-negligible substitutability with the core, or target, submarket(s).

The substitutability of submarkets is given by the parameters b_{mn}, which have a range from 0 (no substitutability) to 1 (perfect substitutability). The subscripts *m* and *n* each count the submarkets up to the total of *N* submarkets. A substitutability parameter for which $m \neq n$ is a "between submarket" substitutability parameter. It designates the substitutability between submarket-*m* products and submarket-*n* products. The parameter b_{12} indicates the relative degree to which ostensible buyers of submarket-2 products would consider buying products from submarket-1, and b_{21} indicates the relative degree to which ostensible

buyers of submarket-1 products would consider buying products from submarket-2.[3] Put another way, b_{mn} indicates how far products in submarket m can reach into submarket n.[4] A substitutability parameter for which $m = n$ is a "within submarket" substitutability parameter, b_{mm}. All products in a given submarket are deemed to have the same substitutability.[5]

These substitutability parameters are (1) used to scale the results when a firm's total market share is multiplied together (as in an *HHI*-type calculation) and (2) to determine effective market sizes. In the terms of Cesari (2000), the substitutability parameters are related to the probabilities of customers moving from a given submarket in period-1 to another submarket in period-2 in a two period model.

Regarding point (1), the substitutability parameters serve to differentiate submarkets (and products). The larger a parameter value, the less differentiated the submarkets, and the more that a multiplication of market share leads to a *HHI*-like result. The smaller the substitutability parameter value, the more differentiated the submarkets, making the result less like the familiar *HHI* outcome.

Regarding point (2), the *DPCI* uses the concept of "effective market size," as denoted by B_m, to show the size of the market in which a product in submarket-m effectively competes. The effective market size incorporates the notion that a product in submarket-m may compete intensely with other products in its own submarket and only to a lesser degree with products in other submarkets. This effective market size is the effective share of the overall market that ostensible buyers in a given submarket actually choose from. If all products in the overall market are equally substitutable ($b_{mn} = b$ for all m and n), the effective market size is 1 for each product. If products in submarket-m are essentially not substitutable with other products in the overall market, and submarket-m is very small, the effective market size for products in submarket-m is near 0.

ally, the effective market size is defined as $B_m = \sum_{n=1}^{N} \sigma_n b_{mn}$. In determining B_m, each submarket share (σ_n) is multiplied by the substitutability of submarket-m with submarket-n (b_{mn}); the resulting sum is the effective market size for products in submarket-m. Larger substitutability parameter values cause the effective market size to rise. If market segments are more substitutable, products effectively compete in a larger fraction of the overall market. The effective market size can be viewed as a weighted average substitutability of all market products for those products in submarket-m.

The market share that a firm's product has of its effective market is

$(\sigma_n s_{nf})/(B_m)$. Whenever substitutability is unequal for some products (i.e., if $b_{mn} \neq b$ for some m and n), the effective market size, B_m, will be less than 1. The presence of the B_m parameters tends to inflate market shares, which by itself would indicate greater market power. Because a product might not truly compete in the entire market, but only a portion of it, its true market power in its effective market may be greater than its nominal market power as indicated by its nominal overall market share. Ceteris paribus, the smaller the effective market size that a product competes in, the greater the market power for its producer within its (smaller) effective market, or niche.

Consider now the *DPCI* value for an arbitrary firm. Given N submarkets, for firm-f this is,

$$C_f^{DPCI} = \sum_{m=1}^{N}\left[(\sigma_m s_{mf}) \cdot \sum_{n=1}^{N} \frac{\sigma_n s_{nf} b_{mn}}{B_m}\right] \tag{4}$$

With F firms in a market, the concentration index value for the overall market is $DPCI = \sum_{f=1}^{F} C_f^{DPCI}$. This can be written as,

$$DPCI = \sum_{f=1}^{F} C_f^{DPCI} = \sum_{f=1}^{F}\left[\sum_{m=1}^{N}\left[(\sigma_m s_{mf}) \cdot \sum_{n=1}^{N} \frac{\sigma_n s_{nf} b_{mn}}{B_m}\right]\right] \tag{5}$$

The index can be constructed and thought of as follows. First, find each product's market share, which is $\sigma_n s_{nf}$. Second, inflate this market share by dividing it by the effective market size in which the product competes, to reflect the product's market power in its effective market. This is $(\sigma_n s_{nf})/(B_m)$. Third, for the firm, multiply each of its effective market shares by each of the firm's (actual) market shares, $\sigma_m s_{mf}$, and by the substitutability parameter, b_{mn}, linking the submarkets m and n. Then sum the market share multiplications. This generates the firm's concentration index value. Total the index values for all firms to find the market concentration index value.

The *DPCI* value for a firm consists of the submarket shares of the firm multiplied by an increasing function of the submarket shares of the firm. That is, for a firm, each $(\sigma_m s_{mf})$ is multiplied by each $(\sigma_n s_{nf} b_{mn})/(B_m)$. As is usual for a concentration index, a larger index value for the *DPCI* denotes a more concentrated market, implying a less competitive market.

III. Market Delineation and Distant Substitutes

Although the *DPCI* can include all non-negligible substitutes for the targeted submarket products, market boundary issues may still arise with its use. But a strength of the *DPCI* is that the omission of distant substitutes has less marginal impact on the *DPCI* value than similar omissions have on *HHI* values. Thus, the *DPCI* is less sensitive to market boundary definitions than is the *HHI*. This is illustrated in examples 1 and 2 below.

A. EXAMPLE 1

Consider a case of three related submarkets, X, Y, and Z. Let each submarket be of equal size and be inhabited by a single (different) firm. Suppose submarkets X and Y are moderately substitutable, with between submarket substitutability values of $b_{XY} = b_{YX} = .5$. Also suppose submarket Z is only distantly substitutable with either X or Y, with between submarket substitutability values of $b_{XZ} = b_{ZX} = .1$, and $b_{YZ} = b_{ZY} = .2$. (Within submarket substitutability values are $b_{XX} = b_{YY} = b_{ZZ} = 1$.) If just submarkets X and Y are considered to comprise the overall market, the *HHI* value is .5. But if all three submarkets are considered to be the relevant market, the *HHI* value is .333. The analogous *DPCI* index values are .667 for just submarkets X and Y, and .660 for all three submarkets. For the *HHI*, the inclusion or exclusion of submarket Z is important in determining the index value, while for the *DPCI*, the inclusion or exclusion of the distantly substitutable submarket Z makes little difference.

B. EXAMPLE 2

Consider four related submarkets, A, B, C, and D, where each submarket is of equal size and each is inhabited by two equal-sized (and unique) firms. Let submarkets A and B contain near substitutes, with $b_{AB} = b_{BA} = .8$. Next, let submarket C contain moderate substitutes for A and B, with $b_{AC} = b_{CA} = .5$ and $b_{BC} = b_{CB} = .7$. Finally, let submarket D contain more distant substitutes, with $b_{AD} = b_{DA} = .2$, $b_{BD} = b_{DB} = .3$, $b_{CD} = b_{DC} = .4$. (And with $b_{AA} = b_{BB} = b_{CC} = b_{DD} = 1$.) The *HHI* values are for a market of just submarkets A and B, .25; for submarkets A, B, and C combined, .167; and for all four submarkets, .125. The analogous *DPCI* values are for just submarkets A and B, .278; for submarkets A, B, and C combined, .215; and for all four submarkets, .209. Again, the inclusion or exclusion

of the moderate substitutes in submarket C and the more distant substitutes in submarket D have more dramatic impacts on the *HHI* values than on the *DPCI* values.

As the *HHI* treats distant substitutes in the same way that it treats very close substitutes, the marginal decreases in index values as the market definition expands tend to be greater for the *HHI* than for the *DPCI*. Halting the progressive expansion of the defined market at any point tends to be less critical in defining the level of competition for the *DPCI* than for the *HHI*.

IV. *DPCI* Properties Shared with the *HHI*

Below are properties the *DPCI* shares with the *HHI*. Due to space constraints, formal proofs of all properties are available at http://sites.google.com/site/raisanenecon/home/research. Properties 1-3 demonstrate that the *DPCI* is defined on the same zero to one range as *HHI* and that the two are identical measures if the products in the submarkets are perfectly substitutable.

Property 1: If all products are equally substitutable, a firm's concentration index value is the square of its overall market share, and the *DPCI* is equivalent to the *HHI*.
Property 2: The maximum value for the *DPCI* is 1.
Property 3: The infimum value for the *DPCI* is 0.

V. Product Concentration Index Properties of the DPCI

Encaoua and Jacquemin (1980) list a number of desirable properties for a homogeneous product concentration index. For the homogeneous case, the *DPCI* shares the properties of the *HHI*. For the differentiated product case, three of the Encaoua and Jacquemin properties are extended here for the *DPCI*. The first is that a transfer of share from a smaller market share to a larger market share should increase concentration. This is presented in section V.A. The second is that a merger of market shares should increase concentration. This is presented in section V.B. The third is that a greater number of equal sized market shares should lower concentration. This is presented in section V.C. Together these properties show the *DPCI* has the desirable properties of a product concentration index.

A. MARKET SHARE TRANSFER

Properties 4 and 5 indicate that a transfer of share from a smaller market share firm to a larger market share firm increases market concentration as measured by the *DPCI*. This is true both within a given submarket and for changing concentrations between submarkets.

> Property 4: When product share is transferred within the same submarket, from a firm with a smaller product share to a firm with a larger product share, the market concentration index value rises, given equal product shares in other submarkets.[6]
> Property 5: Assume constant between submarket substitutability, constant within submarket substitutability, submarkets of equal size, and equal firms shares within submarkets. Then the market concentration index value falls whenever there is an increase in the number of firms in a more concentrated submarket along with a simultaneous decrease in the number of firms in a less concentrated submarket, such that the submarket concentration rankings do not reverse.

B. MERGERS

Mergers increase concentration by reducing the number of firms operating in a market and increasing a firm's market share. Properties 6 and 7 show that mergers within a submarket and mergers of firms in different submarkets both increase *DPCI*. Property 8 shows the effect of merging two submarkets into one.

> Property 6: When two firms in the same submarkets merge, the market concentration index value rises.
> Property 7: If there is substitutability between their submarkets, when two firms without a common submarket merge, the *DPCI* value rises.

Together Properties 6 and 7 imply that if two firms merge that have at least one common submarket or if there is some substitutability between their pre-merger submarkets, the market *DPCI* value rises following the merger.

> Property 8: Given constant between submarket substitutability, merging two submarkets into one is equivalent to raising the between

submarket substitutability parameter values linking the two submarkets to the level of the within submarket substitutability parameter values for the two submarkets.

C. NUMBER OF FIRMS OR SUBMARKETS

A greater number of equal sized firms lowers concentration. Properties 9-11 show that increasing the number of identical firms decreases *DPCI* in the cases of firms that operate in different submarkets, firms that operate in all submarkets, and firms that operate in overlapping submarkets.

Property 9: Given a submarket containing just single submarket firms, when the number of equal-sized firms in the submarket rises, the submarket and market concentration index values fall.
Property 10: Given firms of equal size operating in all submarkets with identical submarket shares, when the number of firms rises, the *DPCI* falls.
Property 11: Given submarkets of equal size, equal submarket substitutability, and constant within market substitutability, if some firms are not in all markets, as the number of submarkets rises, the *DPCI* falls.

VI. Implementing the DPCI

A key step in using the *DPCI* is to estimate suitable substitutability parameter values. If the substitutability parameters are thought to be like probabilities as in Cesari (2000), surveys could be used to track consumers' probability of switching between sellers and products. Huettner (2002) remarks that the characteristics which make products closer substitutes include product prices more closely tracking each other and a cross price elasticity nearer unity. Hence, real price (or quantity) correlations or cross price/own price elasticity ratios are suitable candidates for substitutability parameters. In other contexts, different substitutability measures could be uses.

A. DPCI IN INDUSTRIAL ORGANIZATION

There are clear applications of the *DPCI* within industrial organization.

As early as Hausman, Leonard, and Zona (1992), shortcomings of the Herfindahl-Hirschman Index as applied to anti-trust policy for firms competing with differentiated products were identified. Firms often compete in multiple markets in related industries. Apple Inc., for instance, produces desktop and laptop computers, tablets, smart phones, digital music players, and competes in digital content markets. A concentration index that views Apple within just a single market will produce a distorted view of the company. Using cross-price elasticity data to construct the submarket substitutability parameters as the ratio of cross-price to own-price elasticities, results in a much more complete view of Apple as a company and the competition they face rather than just within a given market.

As applied to anti-trust policy, the *DPCI* is less sensitive to market definitions than is the *HHI* and can be easier to analyse in the context of horizontal merger guidelines. Section 4 of the 2010 Department of Justice and Federal Trade Commission Horizontal Merger Guidelines, outlines the Hypothetic Monopoly Test for determining a market's definition.

The Agencies employ the hypothetical monopolist test to evaluate whether groups of products in candidate markets are sufficiently broad to constitute relevant antitrust markets. The Agencies use the hypothetical monopolist test to identify a set of products that are reasonably interchangeable with a product sold by one of the merging firms.

Once the *HHI* is applied the products included in the market definition are deemed perfectly interchangeable. Use of the *DPCI* avoids this complication and is less likely to overstate the competitiveness of a given grouping of markets.

B. DPCI IN INTERNATIONAL TRADE

One method of examining the impact of trade liberalization on markets is to examine the effect on the *HHI* of opening up international markets for a given industry. The result of local firms facing an increase in international competition can be seen in the decline in industry *HHI*s as the market definition changes from within-state to an international region. A firm that was a local monopoly can be placed into a competitive market by lowering trade barriers. This approach can be seen in the Hernandez and Torero (2013) paper examining local and regional market

concentration *HHI*s on pricing behaviour. Treating local and foreign markets as perfectly substitutable, as the *HHI* does in this context, overstates the impact that trade liberalization is likely to have. Our *DPCI* measure allows for a more measured approach to estimating the impact of trade liberalization.

Consider the case of two countries with market sizes of 0.75 and 0.25 respectively, each with a monopoly firm in a given industry. Before liberalization, each country has an *HHI* and *DPCI* of 1. Combining the markets initially yields two firms with market shares 0.75 and 0.25 respectively and thus a market *HHI* of 0.625. If we recognize the two markets as not perfectly substitutable, perhaps due to a consumer preference for local goods, a substitutability parameter can be estimated. Actual post-liberalization trade flows provide strong guidance on what the substitutability parameter should be. Taking $b_{mn}=0.5$ for $m \neq n$ in this case yields a post-liberalization *DPCI* of 0.743. The substitutability parameter can track the degree of competition provided. The *DPCI* takes this into account whereas the *HHI* cannot.

C. DPCI IN SPORTS ECONOMICS

In sports economics, *HHI* is used to estimate competitive balance between teams in a league. Teams do not always play schedules of equal difficulty, but the *HHI* treats all games as equal. Owen, et al. (2005) and Humphreys (2002) examine the benefits and drawbacks of using the *HHI* measure in this context. Our *DPCI* measure provides a clean method for adjusting the competitive balance measure. The traditional competitive balance measure uses the square of wins by a team over games played in the league as the team concentration index and the sum of these team-win concentrations as the *HHI* with higher values denoting less competitive balance. In larger leagues teams play an unbalanced schedule, playing fewer teams outside their division/conference and the rest of their games against teams within their division/conference. By examining how each conference fairs against each other conference, the substitutability parameter of inter-conference games can be computed and used in calculating the *DPCI*. If all divisions are equally skilled, the inter-conference records are 0.500, and the substitutability parameter can be taken as unity yielding equivalence between the *HHI* and *DPCI* measures. As conferences differ, inter-conference games should be treated as less of a substitute than intra-conference games. In this case,

the substitutability parameters decline and the *DPCI* increases as the league becomes less competitive, whereas the *HHI* would stay constant.

VII. Conclusions

This paper has proposed and examined a new concentration index, with the objective of introducing a more comprehensive and flexible alternative to the *HHI*. Products in the defined market are no longer considered equally substitutable with each other, as they are when using the *HHI*. The *DPCI* instead allows the degree of substitutability between products to vary. Thus, for products differentiated by characteristics or geography, the *DPCI* has a better ability to account for a differentiated ability to constrain market power than the traditional *HHI*.

Further, by using the *DPCI* rather than the *HHI*, researchers do not face the same burden in determining market boundaries. In principle, all non-negligible substitute products are included in the defined market when using the *DPCI*, which can solve the market boundary problem. Yet the problem of market boundaries may still remain, in practice. Crucially, however, the delineation of market boundaries tends to be less critical in determining concentration index values when the *DPCI* is used rather than the *HHI*. Thus, much of the market definition problem associated with the *HHI* is avoided.

As demonstrated in section V, the *DPCI* has the properties that economists like for a concentration index. In particular, the index possesses the two important characteristics that Encaoua and Jacquemin (1980) give that concentration indices should exhibit in homogeneous (equally substitutable) markets. First, the transfer of market share from a smaller firm to a larger firm should not decrease the concentration index value. Second, when firms merge, the concentration index value should not decrease. The first characteristic is given in Property 4; the second characteristic is given in Property 6. Finally, for homogeneous products, the *DPCI* has the same properties as the *HHI*.

The *DPCI* is essentially an expansion of the *HHI*. Therefore, many of the problems and strengths associated with the *HHI* are also associated with the *DPCI*. For example, a concentration index does not precisely measure market power, whether in antitrust use or general research use. It is instead a useful tool in indicating market power. Hence, one can always construct a case in which a market more completely dominated by

a single firm is more competitive than a less concentrated market. The *DPCI* appears no more likely than the *HHI* to fall afoul of such hazards, however.

References

Amidu, M. and Wolfe, S. 2013. "Does bank competition and diversification lead to greater stability? Evidence from emerging markets." *Review of Development Finance* 3(3), 152–166.

Bailey, C. and Taylor, J. 2009. "A concentration index for differentiated products: the case of religious competition." *Applied Economics* 41(14), 1745–1759.

Berger, A. N.,et. al. 2009. "Bank competition and financial stability." *Journal of Financial Services Research* 35(2), 99-118.

Cesari, R. 2000. "A generalized measure of competition." *Applied Economics Letters* 7(7), 479–481.

Dalkir, S. and Warren-Boulton, F. 1999. "Prices, market definition, and the effects of merger: Staples-Office Depot (1997)." *The Antitrust Revolution: Economics, Competition, and Policy* (3rd Ed.) Oxford University Press, New York, pp.143–164.

Encaoua, D. and Jacquemin, A. 1980. "Degree of monopoly, indices of concentration and threat of entry." *International Economic Review* 21(1), 87 – 105.

Farrell, J. and Shapiro, C. 2010. "Antitrust evaluation of horizontal mergers: An economic alternative to market definition." *The BE Journal of Theoretical Economics* 10(1).

Ginevicius, R. and Cibra, S. 2009. "Additive Measurement of Market Concentration." *Journal of Business Economics and Management* 10(3), 191-198.

Hausman, J.A., et. al. 1991. "A Proposed Method for Analyzing Competition Among Differentiate Products." *Antitrust Law Journal* 60(3), 889– 900.

Hellmer, S. and Wårell, L. 2009. "On the evaluation of market power and market dominance--The Nordic electricity market." *Energy Policy* 37(8), 3235-3241.

Hernandez, M. A. and Torero, M. 2013. "Market concentration and pricing behavior in the fertilizer industry: a global approach." *Agricultural Economics* 44(6), 723-734.

Huettner, D. 2002. "Product market definition in antitrust cases when products are close substitutes or close complements." *Antitrust Bulletin* 47 (2002): 133.

Humphreys, B.R. 2002. "Alternative Measures of Competitive Balance in Sports Leagues." *Journal of Sports Economics* 3(2), 133-148.

Lijesen, M. 2004. "Adjusting the Herfindahl index for close substitutes: an application to pricing in civil aviation." *Transportation Research Part E: Logistics and Transportation Review* 40(2), 123-134.

McElroy, F. W. 1996. "Alternatives to the U.S. Antitrust Agency approach to market definition." *Review of Industrial Organization* 11(4), 511-532.

Melnik, A., et. al. 2008. "Assessing market dominance." *Journal of Economic Behavior & Organization* 68(1), 63-72.

Owen, P. D., et. al. 2007. "Measuring competitive balance in professional team sports using the Herfindahl-Hirschman index." *Review of Industrial Organization* 31(4), 289-302.

Pulak, M. and Neha, J. 2012. "Mergers, Acquisitions and Export Competitiveness:

Experience of Indian Manufacturing Sector." *Journal of Competitiveness* 4(1), 3–19.

Qin, Y. and Zhang, X. 2012. "Road to Specialization in Agricultural Production: Evidence from Rural China." International Association of Agricultural Economists Discussion Paper, October.

Stigler, G. 1982. "The economists and the problem of monopoly." *The American Economic Review* 72(2), 1-11.

Stigler, G. J. and Sherwin, R. A. 1985. "The Extent of the Market." *Journal of Law & Economics* 28, 555-585.

US Department of Justice and Federal Trade Commission. 2010. "Horizontal merger guidelines."
http://www.ftc.gov/sites/default/files/attachments/merger-review/100819hmg.pdf

Wen, S. Y. and Yu, J. 2013. "Banking Stability, Market Structure and Financial System in Emerging Countries." *Journal of Applied Finance & Banking* 3(3), 1-13.

Endnotes

1. Identical products could be sold by different final sellers, who may inhabit different submarkets, as determined by the relevant substitutabilities. Products would then be differentiated by locations or sellers' characteristics, rather than by direct product characteristics. For example, differentiation by seller characteristic was argued in the aborted Staples-Office Depot merger, see Dalkir and Warren-Boulton (1999). With this modification, a final seller's product sold at a given location is in only one submarket.
2. A mathematical standard could be set for defining a separate submarket. This could be a minimum substitutability that all products in the submarket must have for all other products in the submarket. The notion of a submarket can be essentially eliminated by letting every product be a monopoly product in its own submarket.
3. It can be expected that b_{mn} and b_{nm} generally rise and fall together.
4. Substitutability can vary depending on prices. A substitutability parameter value may then be with respect to some given price, for example, the competitive price or the actual price.
5. Internally, each submarket is treated as an *HHI* market is, all products are equally substitutable.
6. Property 4 causes the within submarket concentration index value to be minimized when the product shares in the submarket are equal.

A Closer Look at the Relationship between Grades and Teacher Evaluation: The Role of Grade Knowledge

Tin-Chun Lin*

ABSTRACT. Economic behavior theory was used to frame an investigation of the role of grade knowledge in student ratings of professors. A two-group experiment was conducted in which one group had midterm exams and thus received grade feedback before completing student evaluations of teaching (SET), while the other group did not have a midterm exam and thus had no specific grade knowledge before completing the SET. Both groups had a final exam and received exam feedback after the SET was administered. Results revealed that only in the midterm condition were grades significantly associated with SET. That is, SET is more strongly related to grades when students have had clear grade feedback prior to SET administration, implying that both students and professors engage in economic behavior and that a reciprocal relationship exists between students and professors. Moreover, two implied debatable issues on this topic are briefly discussed in the conclusion. (A20; A22; C30)

I. Introduction

The identification of precise methods for assessing a teacher's teaching quality (or performance) is a perennially important issue in higher education because teaching quality is a primary factor in student performance (e.g., De Paola, 2009; and Lin, 2010). Thus far, student evaluation of teaching (SET) is the assessment system most commonly and widely used by administrators in almost all U.S. universities and colleges. Nevertheless, a question remains: *Do grades significantly affect student behavior in rating professors*? Indeed, this question has been broadly discussed and investigated in several previous studies (e.g., Seiver, 1983; Nelson and Lynch, 1984; Krautmann and Sander, 1999; Clayson, 2004; Isely and Singh, 2005; McPherson, 2006; Langbein, 2008; and Matos-Diaz and Ragan, 2010). Most of these studies have concluded that students' expected grades exert a positive and significant

School of Business and Economics, Indiana University–Northwest, 3400 Broadway, Gary, IN 46408. I wish to thank the editor, and two anonymous referees for very helpful discussion and advice. Additional correspondence address: Phone : (219) 980-6634; Fax: (219) 980-6916; E-mail: tinlin@iun.edu.

effect on SET.

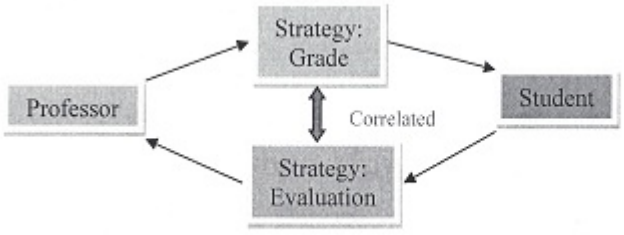
The positive relationship between student evaluations of teaching and student expected grades (or effective grades) rests on two factors: (1) students tend to offer more favorable evaluations in exchange for good grades, or (2) excellent instructors improve student skills to the point that they obtain good grades. Thus, if the positive effect is due to the first factor, both students and professors are engaging in economic behavior and a reciprocal relationship exists between students and professors. To verify that a positive relationship may be due to the first factor, it is necessary to use the theory of economic behavior as a basic framework when investigating the role of grade knowledge in student rating of professors. If specific grade knowledge positively and significantly affects students' behavior in rating professors, then a reciprocal relationship does exist between students and professors.

The reciprocal relationship has been broadly discussed in the business education literature (e.g., Clayson, 2004; Clayson, Frost, and Sheffet, 2006; and Lin, 2009a). The reciprocal relationship model based upon economic behavior theory shows that both groups will choose their best strategies to maximize their payoffs. In this paper we used the theory of economic behavior to look into the relationship between professors and students. Professors give their students grades while students also give their professors grades (i.e., student evaluation of teaching). In order to receive better evaluations from students, it is possible that some professors might adopt strategies (e.g., lowering the grading standard, creating easier exams, giving students extra bonuses [e.g., an attendance bonus], curving students' grades, avoiding some harder teaching materials that should be taught, being very nice to students, etc.) to ensure their students' approval. Strategies adopted by these professors may likely influence students' rating behavior. One of the most influential factors (or strategies) is students' grades, because grades may directly affect students' feelings (or emotions) and reflect their satisfaction with their professors. Clayson, Frost, and Sheffet (2006) explained the phenomenon as the reciprocal relationship between professors and students, in which students reward professors who give them good grades and punish professors who give them bad grades.

In 2009, Lin applied a static game of complete information to address the economic behaviors manifested between professors and students, and developed a model for the reciprocal relationship between professors and students. In the model, he created a production function of education

product constructed by both the professor and the student jointly and simultaneously. Each student and each professor has his/her own payoff function. His theoretical analysis showed that both the professor and the student will choose his/her best strategy to maximize his/her payoff and determine the Nash equilibrium. Consequently, professors' evaluations and students' grades are positively and endogenously correlated (see Figure 1). His theoretical evidence suggests that students' grades are one of the primary factors influencing students' rating behavior.

Figure 1–The Reciprocal Relationship between Professor and Student



Therefore, for this study we developed two research questions: (1) if students do not have grade knowledge before completing SETs, do course grades significantly affect student behavior in rating professors? And (2) If students have clear grade knowledge before completing SETs, do course grades significantly affect student behavior in rating professors?

To investigate these two research questions, we designed an experiment in which students were separated into two different groups: (1) no midterm exams, only one final exam; and (2) both midterm and final exams. In the first group, students had no grade knowledge before filling out the SETs, while students in the second group had clear grade knowledge (i.e., midterm exam grades) before filling out the SETs. We acknowledge that giving only a final exam may not be a good instructional method and do not promote it, but this strategy may be a better way to investigate the role of grade knowledge in student rating of professors.

This paper is organized as follows. First, we clarify several critical points regarding previous studies and briefly review some selected previous studies. Second, we describe our experiment design and data source. Third, we report the development of econometric models based upon our hypotheses. Fourth, we report our empirical results and provide a detailed discussion regarding the results. Fifth, we briefly discuss the

limitations of this research. Finally, conclusions may be found in the final section.

II. Clarification of Critical Points Regarding Previous Studies

Several critical points from previous studies should be clarified before moving on to the next section. Most previous studies have used self-reported grade expectations (i.e., expected grades) as a proxy for students' final grades (e.g., Seiver, 1983; Nelson and Lynch; Krautman and Sander, 1999; Isely and Singh; McPherson, 2006; Langbein, 2008; Matos-Diaz and Ragan, 2010). The reasons are understandable—student evaluations of teaching are anonymous, which means that it is impossible to match students' final grades with their SETs, and a student's expected grade may reflect grades on one or more midterm exams, problem set scores, term paper grades, class presentations, and/or attendance policies. Further, the expected grade also reflects anticipated information about the final exam and its grading.

To learn more about students' expectations for their final grades we conducted a preliminary study (via a survey)[1] which found that a student's expected grade covers two effects: (1) the student's midterm average grade; and (2) the student's satisfaction with the instructor's grading policy. For example, some instructors upwardly curve students' grades on the final exam. Other instructors give students several midterm exams but may drop the worst one or two scores. Such grading policies definitely benefit students and in turn impact their expectations about grades and hence their rating behavior. A student who is satisfied with an instructor's grading policy may be more likely to expect a higher grade because the grading policy appears to be favorable to his/her grade.

Based upon the survey evidence from our preliminary study, SET could relate to students' satisfaction with an instructor's grading policy. Therefore, it is reasonable to assume that the positive and significant effect between students' expected grades and SET scores may partially (or mostly) come from students' satisfaction with an instructor's grading policy rather than from "grades". To verify whether or not our assumption is correct, the effect of students' satisfaction with a grading policy needs to be extracted. Hence, in this study one question was added to a survey to ascertain students' satisfaction with the grading policy. If

the effect of students' satisfaction with a grading policy is expressed in a positive and significant manner on evaluations, this verifies that our assumption is correct. In addition, individual actual grades (i.e., midterm/final average grades) rather than expected grades were adopted.

It should be noted that while most researchers have used expected grades, some researchers have used actual grades. For example, Lanbein (2008) used both actual and expected grades for her investigation. However, the actual grade she used was the average actual grade in the class rather than individual student's actual grade.

Moreover, some researchers, such as Seiver (1983) and Nelson and Lynch (1984), have adopted a simultaneous framework to investigate this issue. Their basic argument in favor of simultaneity runs something like the following: instructors may have an "expected overall evaluation" that may be influenced by their grading policies and thus may influence students' "expected grades" and reported SET score. Readers may think that it seems far-fetched that an instructor's SET expectation can influence the students' expected grades. Indeed, it is not far-fetched. For example, if an instructor expects positive SET scores from students, the instructor may adopt a grading policy that is favorable to students' grades, such as curved, providing attendance bonus, dropping the worst one or two midterm-exam scores, etc. Due to a favorable grading policy, students may expect higher final course grades, which in turn may influence their SET scores. The instructor's "expected evaluation" is not observed but the class or student's "expected grade" is. The use of "expected grade" in their studies was required in order to run a simultaneous framework.

Additionally, expected grades were not adopted in this study; rather, the effects of midterm average grade and student's satisfaction with the grading policy were emphasized. This raises another question. In the second group, when the instructor gives students both midterm exams and a final exam, can students' final course grades and SET be simultaneously determined since midterm average grades may affect their emotions and in turn influence their rating behavior? That is, some (or many) students may use the SET to reward or exact revenge on their professors. Simply speaking, using midterm average grades, students can figure out their potential final grades for a course. Therefore, grades and SET are no longer simultaneously determined in this case. Thus, single-equation estimation was used for the second group. However, for the first group without midterm exams, students' final course grades and SETs

were simultaneously determined. Hence, a simultaneous framework was adopted for the first group. Some may be skeptical that these two variables (i.e., grades and SETs) can exert a simultaneous impact on one another because the prevailing thought is that a time sequence is generally in effect. The reason these two items are regarded as simultaneously determined is because students do not know their course grades when they fill in the SETs; similarly, the professor does not know the results of the SETs when he/she assigns students grades for the course. Unlike the group with midterm exams, students know their midterm average grades when they fill in the SETs so that they can forecast their course grades, which in turn may affect their rating behavior. Therefore, the effects for the group without midterm exams may be regarded as simultaneous even if technically they are not so.

Furthermore, the emphasis in the "typical SET study" was on cross-class and cross-instructor differences. The objective for the "typical SET study" was to understand why SETs differ across instructors and how grading policies affect that outcome. However, the macro data being used may fail to reflect individual differences across students. Therefore, an alternative course of action was selected for this study. Individual micro data were adopted and only one instructor was chosen in order to better understand why SETs differ across students given the same instructor and grading policy.

In addition to those critical points discussed above, some previous studies provided valuable findings that need to be pointed out.

1. Prior to 1983, almost all empirical work focused on single-equation estimation. Seiver (1983) adopted both a single-equation model and a simultaneous equation model. As a result, in his two-equation model estimated with two-stage least squares (2SLS), he did not suggest that students' expected grades related to their overall evaluations of teachers, but he found a statistically significant relationship between students' expected grades and SET in the ordinary least squares model.
2. Nelson and Lynch (1984) tested three hypotheses and concluded that: (1) the student evaluation process may create grade inflation (i.e., the easier the grading, the higher the teaching evaluations); (2) faculty who have been experiencing falling real incomes from teaching will adopt easier grading policies; and (3) the grade-evaluation relationship should be estimated using a simultaneous equations

model rather than a single equation model.

3. Krautmann and Sander (1999) considered grades as an endogenous determinant of SET. Therefore, they used the Instrumental Variable (IV) approach to solve the endogeneity problem. Findings indicated that grades positively affect student evaluations, which implies that faculty may intend to "buy" higher evaluations by lowering their grading standards. In addition, they pointed out an important conclusion: grade inflation may hurt the quality of higher education and weaken the signaling role of educational credentials in screening workers.
4. Clayson (2004) found that the reciprocity relationship is most critical in validating student rating behavior.
5. Isely and Singh (2005) showed that if an instructor for a particular course had some classes in which students expected higher grades, s/he would receive a more favorable average SET in these classes. In addition, they created another alternative variable—relative expected grade (i.e., expected grade relative to the incoming GPA of students). They concluded that the expected grade relative to incoming students' GPA provided more explanatory power.
6. McPherson (2006) tested for endogeneity and controlled for unobserved heterogeneity. He concluded that expected grades and SET are positively and significantly related, implying that instructors may "buy" higher evaluations by giving students higher grades. In his empirical work, he also found that an instructor's level of experience and class size were significant determinants of SET scores in principal classes.
7. Lanbein (2008) revealed that actual grades are positively and significantly related to SET scores, controlling for expected grade and fixed effects for both faculty and courses. After discussing the implications of this issue, she concluded that the SET has become a flawed measure of teaching quality and grades are a flawed signal of future job performance.
8. Matos-Diaz and Ragan (2010) applied risk aversion to the SET issue. They hypothesized that students are risk-averse, preferring less variability in grades. Thus, students' rating behavior will not only depend on their expected grades but also on the variance in expected grades (i.e., distribution of expected grades) because variance provides information on grading policy. Their empirical work showed that student ratings and expected grades are positively and

significantly related, but student ratings and variance in expected grades are negatively and significantly related. Their findings imply that faculty may be able to improve their SET scores by reducing the distribution of expected grades and increasing the mean for expected grades.

In short, the main differences between this study and previous studies may be briefly summarized as follows: (1) individual actual grades were adopted rather than expected grades; (2) a new variable—students' satisfaction with an instructor's grading policy—was developed to extract from the effect of expected grades; (3) cross-student data (individual micro data) were employed rather than cross-class data (macro data); and (4) an experiment was created by separating students into two groups—with and without midterm exams.

III. Experiment Design and Data Source

3.1. Experiment Design

In this experiment, four undergraduate Introduction to Microeconomics classes during the fall 2011 and spring 2012 semesters were chosen for the study. The group without midterm exams was enrolled in the fall 2011 semester class, while the group with midterm exams was enrolled in the spring 2012 semester class. That is, students in the group without midterm exams were given only one final exam, while students in the group with midterm exams were given three exams (i.e., two midterm exams and one final exam). Therefore, students' final grades in the group without midterm exams depended on the final exam, while students' final grades in the group with midterm exams were based on the average of these three exam scores.[2] There were 85 student participants in the fall 2011 semester and 90 student participants in the spring 2012 semester. There was no sufficient statistical evidence of a difference in the means for GPA (i.e., grade point average) between the fall 2011 and spring 2012 groups based on a two-tailed test at the 0.05 level of significance, implying that the quality of students in these two groups was not significantly different.

To conduct this experiment, the following factors needed to be held constant:

(1) *Teacher's instructional style and teaching materials.* Only one teacher was chosen to ensure consistent instructional style and teaching materials.
(2) *Incentive to attend class.* Students were given complete freedom to make their own choices about attending class. Thus, there were no mandatory attendance policies, no attendance bonus, and no quizzes.
(3) *Quality of classroom.* These two different groups (four different sections in two semesters) met in the same classroom to ensure the same classroom quality.
(4) *Same exam for each group and each section.* The same final exam was created for these two different groups (four different sections in two semesters), and the same midterm exams were also developed for two different sections in the second group (i.e., with midterm exam group, spring 2012). The final exam was comprehensive. All exams were collected when students turned in their answers and no student was allowed to use a cellphone/iphone. This practice decreases the probability that an exam will become public and makes it difficult for students to obtain information from a previous year's exams—the best and only ways to minimize the probability that students will gain information from a previous year's exams.

3.2. Data Source

The data used in this study included non-self-reported and self-reported. To avoid data collection problems or potential bias and to ensure that student data were confidential and anonymous, another instructor from a different institution was chosen.

The following four variables led to non-self-reported data:

(1) *Exam scores.* Both groups' exam scores were recorded. These exam scores can be used as a proxy for a student's grade performance. Students' grades used here were original grades without a curve.
(2) *Attendance record.* Daily attendance was taken by the instructor throughout both the fall 2011 and spring 2012 semesters. Students were aware that daily attendance was being taken but they were told by the instructor that it was just for record-keeping purposes and would not affect their final grades due to a lack of mandatory attendance policies. We understand that awareness of attendance monitoring could cause students to treat the variable as something

that might ultimately affect their grades. However, it is the only way to ensure correct attendance records. We believe that significant errors in data would result from asking students to self-report their attendance records. In addition, student's attendance may be used as a proxy for student's interest in the class. If a student is interested in the class, he/she will attend it more often. Alternatively, this variable may also be a proxy for conscientiousness—for example, even if a student has zero interest in the class topic. Moreover, it should be pointed out that the attendance might be a function of how good the teacher is, not just an exogenous measure of interest or conscientiousness.

(3) *Grade Point Average (GPA).* A student's grade point average (GPA) can be used to proxy his/her quality because GPA, regardless of a student's major, is a measure of a student's motivation and scholarly ability. Each student's GPA was provided by the office of the registrar.

(4) *Student's age.* Each student's age was supplied by the administration office. The office provided each student's birth date so that each student's age could be figured out. This variable can be used to proxy a student's maturity. In general, it is assumed that a student's maturity and age are positively related.

To collect the self-reported data, we created questionnaires for both semesters—these surveys were custom-designed for this experiment. Additionally, it was a challenge to match non-self-reported data with self-reported data due to the student data being confidential and anonymous. Thus, we adopted a strategy to keep the student data confidential and anonymous and also to avoid any possibilities of negativity. The strategy is described below.

In both semesters, on the day of the final exam, a proctor handed out the questionnaire to each student a few minutes before the exam began. Students were told that their ratings would be completely confidential and anonymous, so students understood and expected that the instructor would never see their individual SET scores. After all students finished the survey, the proctor went to each student to collect the response by herself, row by row and seat by seat. That is, students did not submit or pass their responses forward to the proctor, and there was no identification of each student on each response. After the proctor collected all responses from students, the proctor put the students'

responses in a big envelope and sealed it, and then the proctor handed out the final exam to each student. Students were told to place their answer sheets on their tables (but cover their answers) and then leave the classroom. After all students left, the proctor collected each student's answer sheet by herself and followed the same order as she had in collecting students' rating responses, row by row and seat by seat. Note that there was no seat number, but the proctor remembered the order in which she had collected students' responses. More importantly, when the instructor graded students' final exam, the instructor maintained the order created by the proctor and did not sort their exam answer sheets alphabetically; the proctor did not provide a sealed envelope containing students' rating responses to the instructor until the instructor had posted students' final grades online and submitted them to the office of the registrar. That is, the proctor gave students' exam answer sheets to the instructor first; then after the instructor posted students' grades online, she gave the sealed envelope to the instructor. We then picked up the sealed envelope and other information needed for the study from the instructor. This procedure ensured that students' final grades absolutely were not influenced by their rating responses.

The following nine variables led to self-reported data.

(1) *Student's efforts.* Two variables indicated this factor: (1) Frequency of studying for this class. There were five choices for this question: *1* = I study 1–5 hours before the test; *2* = I study 6–10 hours before the test; *3* = I study 11–15 hours before the test; *4* = I study 16–20 hours before the test; *5* = I study more than 20 hours before the test. (2) Frequency of practicing the study-guide before the exam. On a weekly basis, students were provided a study-guide with answers. There were five choices for this question: *1* = I never use the study-guide; *2* = I practice only once before the test; *3* = I practice 2 times before the test; *4* = I practice 3 times before the test; *5* = I practice more than 3 times before the test.

(2) *Student's work hours per week.* This variable can serve as a proxy for a student's opportunity cost of studying for the course. If a student works more, his/her opportunity cost will be higher. Students were asked to write down the total number of hours worked per week.

(3) *Student's math background.* Students were asked whether or not they had finished college algebra and calculus classes. This was a dummy variable so "yes" was set as 1 and "no" as 0. This variable was

considered because a math background is needed to learn economics since economics is more mathematical than other business and social sciences classes.

(4) *Depth of understanding of the lecture.* Students were asked: *how much do you usually understand the lecture in the class?* There were five choices for this question: *1* = Below 30%; *2* = 30 – 49%; *3* = 50 – 69%; *4* = 70 – 89%; *5* = Over 90%.

(5) *Professor's instruction skill.* Students were asked: *Do you agree that the instructor well organizes the lecture?*

(6) *Professor's communication skill.* Students were asked: *Do you agree that the instructor's speech and communication are clear and understandable?*

(7) *Professor's efforts.* Two questions were asked: (1) *Do you agree that the instructor is well prepared for the class?* (2) *Do you agree that the instructor is enthusiastic about teaching?*

(8) *Overall evaluation.* Students were asked: *Overall, I would rate the quality of this instructor as excellent.*

(9) *Student's satisfaction with grading policy*. Students were asked: *Overall, I am satisfied with the instructor's grading policy.*

The response options are the same for questions 5–9, which were five choices: *1* = No, I strongly disagree; *2* = No, I disagree but not strongly; *3* = Undecided; *4* = Yes, I agree but not strongly; *5* = Yes, I strongly agree.

In addition, a short summary of the data sources is presented in Table 1. There are three sources: instructor-supplied secondary data, university-supplied secondary data, and student-supplied primary data.

TABLE 1–Summary of the Data Source

Non-Self-Reported Data	Self-Reported Data
Instructor-supplied Secondary Data:	Student-supplied Primary Data:
Exam Scores	*Professor's instruction skill*
Attendance record	*Professor's communication skill*
University-supplied Secondary Data:	*Professor's efforts*
Student GPA	*Overall Evaluation*
Student age	*Student's efforts*
	Student's work hours per week
	Student's math background
	Depth of understanding of the lecture
	Student's satisfaction with grading policy

3.3 Descriptive Statistics

Table 2 reports means and standard deviations for the variables used in this study, t statistic, and K-S statistic. In addition, the reliability (i.e., Cronbach's alpha) of exams was measured. The Cronbach's alpha for the group with midterm exams was 0.87, which is high and indicates strong internal consistency among these exams. We were not able to compute Cronbach's alpha for the group without midterm exams, as there was only one exam.

In addition, two important points need to be mentioned before we present the results of the regressions.

(1) Satisfaction with grading policy was much lower (on average) when a midterm was not given, yet the overall evaluation was much higher (on average) when a midterm was not given. This is a very important finding, and we will provide a detailed discussion in a later section after presenting the empirical results.
(2) The average grade on the first midterm was 69, implying that many of the students failed the first midterm. Similarly, the final course grade was a C on average. Based upon the average GPA of 2.8, many of these students might receive a lower grade than they expected. It seems that the instructor may be an unusually "hard" teacher. Moreover, the teacher received worse evaluations from students when a midterm was given. That might be because a midterm was given and because it was graded hard. When these two reasons

coexist, students' rating behavior may be significantly related to grade feedback—the lower the grades, the worse the evaluations. That is, a reciprocal relationship exists between students and teachers—students reward teachers who give them good grades and punish teachers who give them poor grades. We will provide a detailed econometric investigation in the following section to prove our belief.

TABLE 2–Mean, Standard Deviation, t-Statistic, and K-S Statistic

Variables	Mean and Standard Deviation		Two-Tailed t Test	K-S Test K-S Statistic	
	(1) Fall 2011	(2) Spring 2012	(3) t Statistic	(4) Fall 2011	(5) Spring 2012
Overall Evaluation	4.08 (0.82)	3.56 (1.30)	3.23 (0.002)	0.295 (<0.010)	0.267 (<0.010)
First exam (scores)		69.08 (14.98)			0.067 (>0.150)
Second exam (score)		78.92 (12.96)			0.080 (>0.150)
Final exam (scores)	71.81 (12.99)	74.71 (13.08)	-1.47 (0.143)	0.095 (0.057)	0.085 (0.103)
Final course grade (scores)	71.81 (12.99)	74.03 (12.20)	-1.16 (0.247)	0.095 (0.057)	0.068 (>0.150)
Work hours per week	28.74 (14.29)	29.03 (13.77)	-0.14 (0.891)	0.194 (<0.010)	0.195 (<0.010)
Dummy variable-algebra	0.71 (0.44)	0.69 (0.47)	0.24 (0.808)	0.445 (<0.010)	0.437 (<0.010)
Dummy variable-calculus	0.35 (0.48)	0.36 (0.47)	-0.04 (0.971)	0.416 (<0.010)	0.414 (0.010)
Grade Point Average (GPA)	2.88 (0.49)	2.82 (0.49)	0.88 (0.379)	0.099 (0.044)	0.126 (<0.010)
Depth of understanding of the lecture	3.35 (01.08)	3.34 (1.08)	0.05 (0.959)	0.291 (0.010)	0.294 (<0.010)
Number of attendance (whole)	27.04 (3.50)	26.62 (3.97)	0.73 (0.466)	0.214 (<0.010)	0.216 (<0.010)

TABLE 2–Mean, Standard Deviation, t-Statistic, and K-S Statistic

Variables	Mean and Standard Deviation		Two-Tailed t Test	K-S Test K-S Statistic	
	(1) Fall 2011	(2) Spring 2012	(3) t Statistic	(4) Fall 2011	(5) Spring 2012
(continued)					
Number of attendance (midterm)		18.12 (2.21)			0.254 (<0.010)
Frequency of studying for the class	3.01 (1.14)	2.47 (1.13)	3.17 (0.002)	0.177 (<0.010)	0.315 (<0.010)
Frequency of practicing the study guide	3.52 (1.06)	3.67 (1.12)	-0.90 (0.369)	0.181 (<0.010)	0.183 (<0.010)
Well organized the lecture	4.29 (1.06)	4.30 (1.04)	-0.04 (0.970)	0.348 (<0.010)	0.349 (<0.010)
Well prepared for the class	4.62 (0.67)	4.61 (0.70)	0.12 (0.905)	0.430 (<0.010)	0.433 (<0.010)
Speech clear and understandable	3.44 (1.17)	3.47 (1.10)	-0.18 (0.856)	0.238 (<0.010)	0.252 (<0.010)
Enthusiastic about teaching	4.47 (0.85)	4.46 (0.84)	0.12 (0.907)	0.403 (<0.010)	0.387 (<0.010)
Student age	24.66 (5.18)	23.14 (4.33)	2.09 (0.038)	0.209 (<0.010)	0.193 (<0.010)
Satisfied with the grading policy	2.91 (0.63)	3.92 (0.71)	-10.06 (0.000)	0.359 (<0.010)	0.344 (<0.010)

Note: Number in parentheses in Columns (1) and (2) is standard deviation, while in Columns (3) - (5) is *p*-value.

IV. Econometric Models

In light of the research questions and given the data available for this study, we developed two testable hypotheses. Based upon these two hypotheses, we created econometric models to investigate this issue.

Hypothesis 1: *Without midterm exams (i.e., only one final exam), students' grades will not be significantly associated with SETs.*

To investigate Hypothesis 1, a simultaneous-equation model is required. Here, we used the Two-Stage Least Squares (2SLS) procedure to correct for simultaneous questions and to obtain unique estimates that were consistent and asymptotically efficient. Thus, in the first stage:

$$OEV = a_0 + a_1 WOR + a_2 WPR + a_3 ENU + a_4 DEP + a_5 ATD + \varepsilon_1 \text{ , and} \quad (1)$$

$$FGD = b_0 + b_1 FRS + b_2 FRP + b_3 ALG + b_4 CAL + b_5 GPA + b_6 ATD + b_7 WHR + \varepsilon_2 \text{ ,} \quad (2)$$

where *OEV* = overall evaluation; *WOR* = well organized the lecture; *WPR* = well prepared for the class; *ENU* = enthusiastic about teaching; *DEP* = depth of understanding of the lecture; *ATD* = total number of attended classes; *FGD* = student's final grade; *FRS* = frequency of studying for the class; *FRP* = frequency of practicing study guide; *ALG* = finished college algebra class; *CAL* = finished calculus class; *GPA* = grade point average; *WHR* = total work hours a week; and ε_1 , ε_2 = stochastic disturbance with a mean 0 and a variance σ^2 .

The results for Equations (1) and (2) are reported in Columns (1) and (2) of Table 3. $\hat{OEV}$ and $\hat{FGD}$ were saved—they are the predicted values of *OEV* and *FGD* as obtained from the reduced form estimates. The structural equations were estimated but *OEV* and *FGD* were replaced by $\hat{OEV}$ and $\hat{FGD}$. $\hat{OEV}$ and $\hat{FGD}$ are the instrumental variables (IV) here. Therefore, in the second stage the model for the professor's overall evaluation and student's final grade can be estimated in a linear form. The econometric models in the second stage can be expressed as follows.

$$OEV = \alpha_0 + \alpha_1 \hat{FGD} + \alpha_2 SPH + \alpha_3 AGE + \alpha_4 SAF + u_1 \text{ , and} \quad (3)$$

$$FGD = \beta_0 + \beta_1 \hat{OEV} + \beta_2 GPA + \beta_3 SPH + \beta_4 AGE + u_2 \quad (4)$$

where *SPH* = speech and communication are clear and understandable; *AGE* = student's age; *SAF* = student's satisfaction with the instructor's grading policy; and u_1 , u_2 = stochastic disturbance with a mean 0 and a variance σ^2.

In this formulation, the null hypothesis is that the parameters estimated by coefficients a_1 and β_1 are zero, while the alternative hypothesis is that the parameters are not zero.

Moreover, it should be noted that in the simultaneous equations model there is presumably an assumption of no correlation between the errors in both Equations (3) and (4).

Hypothesis 2: *With midterm exams, students' grades will be significantly associated with SETs*.

To investigate Hypothesis 2, a single-equation model is required. However, student's grade was an endogenous variable in the model[3]. When an endogeneity problem occurs, the Two-Stage Least Squares (2SLS) procedure is needed. Therefore, the student's grade was estimated in the first stage. The regression model was created in a linear form, such as:

$$GRD = d_0 + d_1 ATD + d_2 FRS + d_3 FRP + d_4 ALG + d_5 CAL + d_6 GPA + d_7 WHR + \mu_1 \quad (5)$$

where $GRD = GRD_F$ or GRD_M; GRD_F = final average grade (= mean of two midterm-exams and one final-exam scores); GRD_M = midterm average grade (= mean of two midterm-exam scores); $ATD = ATD_F$ or ATD_M; ATD_F = total number of attended classes in a semester; ATD_M = total number of attended classes in the midterm; and μ_1 = stochastic disturbance with a mean 0 and a variance σ^2.

The results for Equation (5) are reported in Columns (3) and (4) of Table 3. $\hat{GRD}$ was saved; the predicted value of GRD was obtained from the reduced form estimates. $\hat{GRD}$ was the instrumental variable (IV) here. Hence, in the second stage, the model for the professor's overall evaluation can be estimated in a linear form. The econometric model (Model 1) in the second stage can be expressed as follows.

$$OEV = \lambda_0 + \lambda_1 \hat{GRD} + \lambda_2 ENU + \lambda_3 WOR + \lambda_4 WPR + \lambda_5 DEP + \lambda_6 SPH + \lambda_7 AGE + \lambda_8 SAF + \tau_1 \quad (6)$$

where τ_1 = stochastic disturbance with a mean 0 and a variance σ^2.

TABLE 3–Determinants of *OEV, FGD, GRD$_F$*, and *GRD$_M$* in the First Stage

	Fall 2011 (No Midterm Exams)		Spring 2012 (With Midterm Exams)	
Explanatory Variables	OLS Explained Variable: *OEV* (1)	OLS Explained Variable: *FGD* (2)	OLS Explained Variable: *GRD$_F$* (3)	OLS Explained Variable: *GRD$_M$* (4)
Constant	0.814 (1.31)	-1.71 (-0.17)	11.82 (1.25)	9.78 (0.83)
WOR	0.395*** (3.94)			
WPR	0.116 (0.75)			
ENU	0.153* (1.70)			
DEP	0.082 (1.24)			
ATD	0.003 (0.16)	0.988*** (3.29)	1.05*** (4.04)	1.51*** (2.90)
FRS		4.558*** (4.82)	0.538 (0.57)	0.505 (0.47)
FRP		1.617 (1.60)	0.161 (0.17)	0.427 (0.40)
ALG		-0.323 (-0.14)	-0.636 (-0.29)	-1.067 (-0.43)
CAL		3.34 (1.57)	5.363** (2.50)	5.937** (2.46)
GPA		10.12*** (4.88)	12.50*** (6.13)	12.91*** (5.61)
WHR		-0.094 (-1.30)	-0.15* (-1.97)	-0.14* (-1.64)
R^2 $\overline{R^2}$	0.568 0.541	0.545 0.503	0.453 0.406	0.394 0.343
F-Statistic	20.81	13.17	9.69	7.63
Observations	85	85	90	90

Note: Number in parentheses is *t*-value; *OEV* = overall evaluation; *FGD* = student's final grade; *GRD$_F$* = final average grade; *GRDM* = midterm average grade; *WOR* = well organized the lecture; *WPR* = well prepared for the class; *ENU* = enthusiastic about teaching; *DEP* = deprth of understanding of the lecture; *FGD* = student's final grade; *FRS* = frequency of studying for the class; *FRP* = frequency of practicing study guide; *ALG* = finished college algebra class; *CAL* = finished calculus class; *GPA* = grade point average; *ATD* = total number of attended classes (*ATD$_F$* is for *GRD$_F$*; while *ATD$_M$* is for GRD$_M$); *WHR* = total work hours a week. ***p< .01; **p< .05; *p< .10

We also replace $\hat{FGD}$ in Equation (3) by $\hat{GRD}$, which would allow us to easily compare these two groups (without midterm and with midterm exams) due to the same explanatory variables. Thus, the econometric model (Model 2) in the second stage is shown below.

$$OEV = \gamma_0 + \gamma_1 \hat{GRD} + \gamma_2 SPH + \gamma_3 AGE + \gamma_4 SAF + \tau_2 \quad (7)$$

where τ_2 = stochastic disturbance with a mean 0 and a variance σ^2.

In this formulation, the null hypothesis is that the parameters estimated by coefficients λ_1 and γ_1 are zero, while the alternative hypothesis is that the parameters are not zero.

In addition, it should be pointed out that all of the error terms listed in the discussion of the econometric methods are based on an assumption of homoscedasticity rather than identical variances.

V. Results and Discussion

5.1. Results

Hypothesis 1

The results for Equations (3) and (4) are reported in Table 4. As Table 4 shows, student's final grade did not exert a statically significant effect on overall evaluation at any significant level in the all-students, Section I-students, and Section II-students groups.[4] Similarly, overall evaluation also did not exert a statistically significant effect on student's final grade at any significant level in the all-students, Section I-students, and Section II-students groups. These results imply that students' grades and SETs are not correlated. However, except for the Section II-students group, student's satisfaction with grading policy had a positive and statistically significant effect on overall evaluation at the 5% level in the all-students group and the Section I-students group, implying that student's satisfaction with grading policy and SET are correlated.

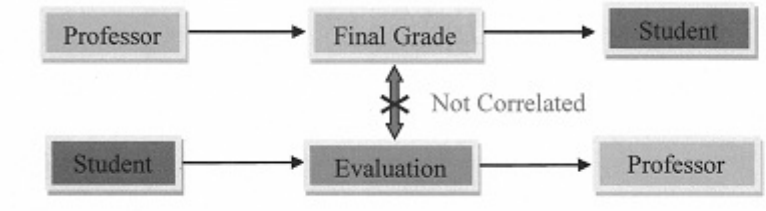
In short, Hypothesis 1 is supported. When students did not have grade knowledge prior to SET administration, their course grades were not significantly associated with SET. Figure 2 also explains this result.

TABLE 4–No Midterm Exams: Determinants of *OEV* and *FGD* in the Second Stage (Fall 2011)

	All Students		Section I		Section II	
Explanatory Variables	2SLS Explained Variable: *OEV* (1)	2SLS Explained Variable: *FGD* (2)	2SLS Explained Variable: *OEV* (3)	2SLS Explained Variable: *FGD* (4)	2SLS Explained Variable: *OEV* (5)	2SLS Explained Variable: *FGD* (6)
Constant	0.702 (0.98)	24.16** (2.21)	0.149 (0.13)	10.93 (0.66)	1.884* (1.93)	44.06** (2.54)
$\hat{FGD}$	0.012 (1.38)		0.011 (0.72)		0.010 (1.01)	
$\hat{OEV}$		3.184 (1.20)		3.789 (1.01)		0.039 (0.01)
GPA		7.682** (2.61)		10.756** (2.34)		5.996 (1.43)
SPH	0.230*** (3.11)	2.349* (1.75)	0.187 (1.35)	3.022 (1.54)	0.186** (2.28)	1.670 (0.85)
AGE	0.035** (2.43)	0.180 (0.71)	0.044* (1.73)	0.199 (0.51)	0.028* (1.75)	0.215 (0.63)
SAF	0.296** (2.37)		0.466** (2.19)		0.080 (0.52)	
$\frac{R^2}{R^2}$	0.334 (0.301	0.288 0.252	0.385 0.316	0.400 0.334	0.219 0.139	0.134 0.045
F-Statistic	10.04	8.08	5.63	6.01	2.73	1.51
Observations	85	85	41	41	44	44

Note: Number in parentheses is *t*-value; *OEV* = overall evaluation; *FGD* = student's final grade; *GPA* = grade point average; *SPH* = speech and communication are clear and understandable; *AGE* = student's age; and *SAF* = student's satisfaction with the instructor's grading policy.
***p< .01; **p< .05; *p< .10

Figure 2. Grade and Evaluation with out Midterm Exams



Hypothesis 2

The results for Equation (6) are reported in Table 5. As indicated there, both student's final average grade and midterm average grade exerted positive and statistically significant effects on overall evaluation at the 5% level in the all-students and at the 10% level in the Section I-students groups (but the effect is not statistically significant at any level in the Section II-students group).[5] Additionally, student's satisfaction with the grading policy exerted a positive and significant effect on overall evaluation at the 10% or 5% level in the all-students and Section I-students groups (while the effect is also not statistically significant at any level in the Section II-students group).

In addition, the results for Equation (7) are reported in Table 6. As Table 6 shows, both student's final average grade and midterm average grade exerted positive and statistically significant effects on overall evaluation at the 1% level in the all-students group and at the 5% level in the both Section I-students and Section II-students groups. Moreover, student's satisfaction with the grading policy exerted a positive and significant effect on overall evaluation at the 1% level in the all-students and Section I-students groups and at the 10% level in the Section II-students group.

Moreover, it should be noted that the coefficients of final and midterm average grades are much smaller than the other explanatory variables' coefficients (such as *ENU, WOR, WPR, DEP, WPR, DEP, SPH, and SAF*). This is because both grades are 100-point scale, while the other explanatory variables are 5-point scale. If we convert both grades to 5-point scale, the effect size of grades will become much bigger. For example, 0.021 (in Column 2 of Table 5) means that additional one point in midterm average grade is estimated to raise overall evaluation by approximately 0.021 points. If we convert it to 5-point scale, the effect size will become 0.42 (= (100/5) x 0.021), which means that additional twenty points in midterm average grade is estimated to raise overall evaluation by approximately 0.42 points.

Consequently, these results imply that grades and SETs are correlated, and that students' rating behavior may be affected by their satisfaction with the grading policy. In summary, Hypothesis 2 is supported. When students had grade knowledge prior to SET administration, their course grades were significantly associated with SET. Figure 3 also explains this result.

TABLE 5–With Midterm Exams: Determinants of *OEV* in the Second Stage (Spring 2012)–Model 1

	All Students		Section I		Section II	
Explanatory Variables	2SLS Explained Variable: *OEV* (1)	2SLS Explained Variable: *OEV* (2)	2SLS Explained Variable: *OEV* (3)	2SLS Explained Variable: *OEV* (4)	2SLS Explained Variable: *OEV* (5)	2SLS Explained Variable: *OEV* (6)
Constant	-4.280*** (-5.78)	-4.266*** (-5.63)	-4.345*** (-4.01)	-4.539*** (-4.02)	-5.568*** (-4.03)	-5.419*** -3.95)
$\hat{GRD}_F$	0.023** (2.50)		0.024* (1.79)		0.022 (1.51)	
$\hat{GRD}_M$		0.021** (2.36)		0.025* (1.88)		0.019 (1.37)
ENU	0.384*** (3.87)	0.388*** (3.89)	0.314** (2.09)	0.310** (2.07)	0.546*** (3.70)	0.550*** (3.66)
WOR	0.105 (0.99)	0.089 (0.83)	0.142 (1.03)	0.128 (0.93)	0.002 (0.01)	-0.023 (-0.13)
WPR	0.084 (0.54)	0.096 (0.61)	-0.004 (-0.02)	0.009 (0.04)	0.567* (2.00)	0.588** (2.07)
DEP	0.520*** (6.09)	0.522*** (6.08)	0.592*** (4.84)	0.586*** (4.80)	0.375*** (2.80)	0.389*** (2.93)
SPH	0.107 (1.39)	0.112 (1.45)	0.078 (0.70)	0.077 (0.68)	0.132 (1.18)	0.143 (1.27)
AGE	0.024 (1.56)	0.025 (1.63)	0.024 (0.93)	0.028 (1.06)	0.032 (1.61)	0.033 (1.62)
SAF	0.238* (1.91)	0.249** (2.01)	0.336* (1.79)	0.357** (1.96)	-0.024 (-0.12)	-0.031 (-0.16)
R^2 / $\bar{R}^2$	0.815 0.797	0.814 0.795	0.845 0.812	0.846 0.813	0.806 0.760	0.804 0.758
F-Statistic	44.64	44.20	25.85	26.08	17.66	17.42
Observations	90	90	47	47	43	43

Note: Number in parentheses is *t*-value; *OEV* = overall evaluation; GRD_F = *final average grade;* GRD_M = midterm average grade; *WOR* = well organized the lecture; *WPR* = well prepared for the class; *DEP* = depth of understanding of the lecture; *ENU* = enthusiastic about teaching; *SPH* = speech and communication are clear and understandable; *AGE* = student's age; and *SAF* = student's satisfaction with the instructor's grading.

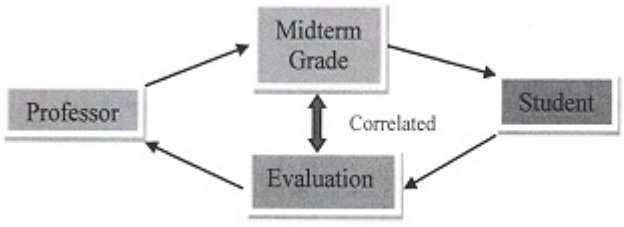
***p< .01; **p< .05; *p< .10

TABLE 6–With Midterm Exams: Determinants of *OEV* in the Second Stage (Spring 2012)–Model 2

	All Students		Section I		Section II	
Explanatory Variables	2SLS Explained Variable: *OEV* (1)	2SLS Explained Variable: *OEV* (2)	2SLS Explained Variable: *OEV* (3)	2SLS Explained Variable: *OEV* (4)	2SLS Explained Variable: *OEV* (5)	2SLS Explained Variable: *OEV* (6)
Constant	-4.637*** (-4.99)	-4.693*** (-4.99)	-4.827*** (-3.35)	-5.312*** (-3.61)	-4.131*** (-2.84)	-3.794** (-2.64)
$\hat{GRD}_F$	0.0423*** (3.53)		0.0418** (2.29)		0.0417** (2.29)	
$\hat{GRD}_M$		0.0420*** (3.54)		0.0471** (2.61)		0.0364** (2.05)
SPH	0.399*** (4.20)	0.404*** (4.27)	0.383*** (2.72)	0.374*** (2.69)	0.396*** (2.73)	0.419*** (2.88)
AGE	0.057*** (2.79)	0.060*** (2.92)	0.054 (1.57)	0.060* (1.76)	0.065** (2.28)	0.066** (2.27)
SAF	0.597*** (3.86)	0.605*** (3.94)	0.681*** (2.95)	0.690*** (3.11)	0.440* (1.76)	0.442* (1.74)
R^2 $\overline{R}^2$	0.618 0.601	0.619 0.601	0.647 0.614	0.659 0.626	0.530 0.480	0.518 0.467
F-Statistic	34.45	34.48	19.27	20.28	10.70	10.20
Observations	90	90	47	47	43	43

Note: Number in parentheses is *t*-value; *OEV* = overall evaluation; GRD_F = final average grade; GRD_M = midterm average grade; *SPH* = speech and communication are clear and understandable; *AGE* = student's age; and *SAF* = student's satisfaction with the instructor's grading.
***p< .01; **p< .05; *p< .10

Figure 3. Grade and Evaluation with Midterm Exams



5.2. Discussion

In addition to findings described above, we are left with one concern to clarify. Based upon the statistical evidence from the data used in this study, the professor's overall evaluations from the group without midterm exams were higher than those for the professor from the group with midterm exams.[6] This evidence might lead readers to argue that only giving one final exam would ensure better evaluations from students.

Additionally, finding that the SET is higher when there are fewer exams makes it easier for us to argue that students prefer less exam-work but does not necessarily indicate that we should argue that this grading procedure is somehow more fair, or led to better feedback, or for some other reason was highly desirable to students. More importantly, it also does not mean that we have support for the belief that fewer exams lead to better SET. To clarify this concern, we took a look at the correlation between SET and final course grade for these two groups. If the correlation was higher when students had grade feedback prior to SET administration, this may indicate that students knew their grades (i.e., midterm grades) prior to SET administration but not that having more exams alone would increase the correlation. To test the hypothesis, we used the one-tailed test (the upper tail test) by formulating the null (H_0) and alternative (H_a) hypotheses as below:

$$\begin{cases} H_0 : r_{12} - r_{11} \leq 0 \\ H_a : r_{12} - r_{11} > 0 \end{cases}$$

where r_{11} the population correlation between SET and final course grade for the group without midterms; and r_{12} the population correlation between SET and final course grade for the group with midterms.

We used "Fisher's *r*-to-*z* transformation" to transform correlations (r) into standard normal (z). As a result, the sample correlation between SET and final course grade for the group with midterms was 0.54, while the sample correlation between SET and final course grade for the group without midterms was 0.31. The *p*-value was 0.03, which is statistically significant at the 5% level, implying that the null hypothesis is rejected. Therefore, the answer is clear now—SET is more strongly related to grades when students have had clear grade feedback. That is, one of the main reasons for a lower SET when there are more exams is that students

in the group with midterms have had clear grade feedback prior to SET administration, and hence grade feedback affects their rating behavior.

Furthermore, on average, satisfaction with grading policy was much lower when midterm exams were not given to students. This might be because students would experience increased feelings of risk when only one exam determines their final course grades. In other words, a riskier grading procedure makes this grading policy less desirable to students. Nevertheless, based upon the evidence shown in Table 4, students' satisfaction with grading policy and overall evaluations are still positively correlated. This means that even though the grading policy is less desirable to students, students who were relatively more satisfied with the grading policy would still give their professor a better evaluation.

VI. Limitations

This research had two limitations—(1) experiment design; and (2) peer effect—that could possibly result in some potential errors. We illustrate these two limitations below.

6.1. Limitation of Experiment Design

Although we held constant important variables such as the teacher, classroom, exam, and double-blind protocol for collecting data in our experiment design, alternative designs could have been used to test the hypotheses while minimizing differences across the control and treatment groups. Our design led to some important differences across groups besides the treatment variable of two midterms: (1) the control group sections (no midterms) were all in the Fall 2011 semester classes, whereas the treatment group sections (with midterms) were all in the Spring 2012 semester classes; and (2) the class meeting time and number of times across sections could vary across groups. Of greatest concern was the treatment variable itself: the control group only had a final exam while the treatment group had two midterms and a final exam (equally weighted). This led to a considerable difference in grading policy and may have influenced students' rating behavior. Fortunately, we included "satisfaction with grading policy" as an explanatory variable in the regressions, making it much more strongly correlated with the Spring 2012 (treatment-group) semester courses (see Tables 4 and 6) and significantly higher in Spring 2012 (see Table 2). Furthermore, based

upon the one-tailed t-test at the 10% significant level, the final exam grade in the treatment-group was significantly higher because more frequent testing has been shown to have a positive impact on grades.

Our design created a considerable difference in grading policy and may have influenced SET, but it allowed us to verify that both expected grades and final course grades were not the "grades" affecting SET, and still provided convincing enough evidence and a step in the right direction for explaining the causes of grade inflation—a controversial issue in higher education today. In a future study, we will consider other alternative designs. For example, in one design we could have a midterm SET administered before the midterm in one group and after in the other. Another design could involve students who know that only the professor will observe the SET results, thereby reducing the incentive to reciprocate since there are no consequences to the professor. Both of these designs would be a bit cleaner and enables the hypotheses to be tested without changing the grading schema across groups.

6.2. Limitation of the Peer Effect

A "*peer effect*" may be found in students' rating behavior. For example, if a student had a bad experience with the professor for some reason, this student might share negative comments about the professor with other students. Therefore, the rating behavior of students may likely be influenced by the student's negative comments, and hence lead to a bias. Similarly, former students' comments (either positive or negative) about the professor may also affect current students' rating behavior. This is because students could simply distribute information and comments to all other students. Although the "*peer effect*" may exist, it cannot be easily observed and measured, creating a significant challenge in data collection. Thus, to collect such data, we will have to develop a strategy that maintains the confidentiality and anonymity of student data and also avoids any possibility of negativity. These will be left to future research on this issue.

VII. Conclusion

In this paper we developed a two-group experiment—one group had midterm exams and hence received grade feedback before completing

student evaluations of teaching, while the other group did not have a midterm exam and thus had no specific grade knowledge before completing student evaluations of teaching. Both of these two groups had a final exam and received final course grade feedback after the SET was administered.

In light of the empirical results described previously, three major findings may be summarized as follows:

1. Student's final grade was not significantly correlated with professor's overall evaluation in the no-midterm exams group, while student's final and midterm average grades were positively and significantly associated with professor's overall evaluation in the midterm-exam group. That is, neither expected grades nor final grades affect students' rating behavior, because expected grades are not real grades and final grades are determined after students fill out the evaluation. In fact, midterm average grades (or midterm grades) may significantly affect students' behavior in rating professors.
2. Overall, student's satisfaction with a grading policy exerts a positive and statistically significant effect on professor's overall evaluation in both groups (without and with midterm exams). The evidence confirms our assumption that the positive and significant effect between students' expected grades and SET scores may partially (or mostly) come from students' satisfaction with an instructor's grading policy rather than from "grades".
3. According to the correlation test, student evaluation of teaching is more strongly related to grades when students have had clear grade feedback prior to SET administration, which explains why the professor received a lower evaluation when there were more exams in the group with midterms because grade feedback prior to SET administration would affect students' rating behavior.

While our main finding seems to be similar to the main finding from past studies—students' grades are positively associated with SETs—our study indeed neither replicates nor refutes the main findings of past studies done on class averages (not the individual level). This is because we used individual actual grades and cross-student data rather than expected grades and cross-class data, which were commonly used in previous studies. Thus, students' rating behavior actually is influenced by midterm grades rather than expected grades.

Furthermore, before we wrap up this paper, two important implied debatable issues regarding SETs need to be pointed out.

1. In addition to the issue of *halo error*[7] in student evaluation of teaching (Orsini, 1988; Clayson, 1989, 1999; Clayson and Haley, 1990; Simpson and Siguaw, 2000; Clayson and Sheffet, 2006; and Madden, Dillon, and Leak, 2010), this assessment system ignores the existence of economic behaviors between teachers and students. Both teachers and students are economic individuals so both will respond to each other via economic behaviors. Therefore, the best strategy for receiving good grades and feedback is "*collusion*". In other words, teaching evaluations may lead professors to intentionally inflate grades[8] in order to receive good comments (a form of "*cheating*") and thus foster "*collusion*". As Simpson and Siguaw (2000) reported, some (or many) faculty members may use "*halo effects*" to their advantage by managing student evaluation. Similarly, some faculty members could also use "*collusion*" to manipulate evaluations.
2. Obermiller, Fleenor, and Raven (2005) pointed out that the role of students and their relationship with professors are always complex. Their survey showed that students generally prefer the customer orientation. A recent online survey on campus also found that the majority of students perceive themselves as "*customers*" in school[9]. If students are customers, then professors become "*servers*" or "*sellers*". Then the question is: how can servers evaluate (or judge) their customers? (Only customers evaluate servers.) That is, professors (i.e., "*servers*") are not allowed to test their students (i.e., "*customers*") and thus cannot provide grades for their students (i.e., "*customers*"). Rather, if grades must be given to students, the only grade that "*servers*" are allowed to give "*customers*" must be A (or A+), because "customers are always right". Therefore, it is impossible for "*servers*" to fail "*customers*". Hence, Franz (1998) stated that this belief could lead higher education to emphasize entertainment and professors whose role is to delight students rather than truly teach them. On the other hand, if it is inappropriate to regard students as "*customers*", then professors should not be regarded as "*servers*" and a customer orientation is no longer relevant or appropriate.[10] For that reason, the question here is: how can students evaluate (or judge) professors? In other words, under the circumstances described here, students do not have the right to evaluate professors. Unfortunately, in a system focusing on student

evaluation of teaching, both professors and students evaluate one another at the same time, creating a "*paradox of roles*" between professors and students.

In summary, the main contribution of this study to economic education is our verification of an important fact: midterm grades (grade feedback prior to SET administration) actually affect students' rating behavior, implying that both students and professors do engage in economic behaviors and that a reciprocal relationship does exist between students and professors. Hence, grades (especially midterm grades) could have been frequently used as a strategy to influence students' rating behavior. The school authority should be aware of the importance of this fact and identify a better form of faculty performance assessment in order to avoid the existing reciprocal relationship between students and professors.[11]

References

Clayson, D. E. 1989. "Halo Effects in Student Evaluation of Faculty: A Question of Validity." Paper presented at "Positioning for the 1990s," *Proceedings of the Southern Marketing Association*, New Orleans, LA.

Clayson, D. E. 1999. "Students' Evaluation of Teaching Effectiveness: Some Implications of Stability." *Journal of Marketing Education*, 21(1): 68–75.

Clayson, D. E. 2004. "A Test of the Reciprocity Effect in the Student Evaluation of Instructors in Marketing Classes." *Marketing Education Review*, 14(2): 11–21.

Clayson, D. E., Frost, T. F., and Sheffet, M. J. 2006. "Grades and the Student Evaluation of Instruction: A Test of the Reciprocity Effect." *Academy of Management Learning and Education*, 5(1): 52–65.

Clayson, D. E., and Haley, D. A. 1990. "Student Evaluation in Marketing: What is Actually Being Measured?" *Journal of Marketing Education*, 12(3): 9–17.

Clayson, D. E., and Haley, D. A. 2011. "Are Students Telling us Truth? A Critical Look at the Student Evaluation of Teaching." *Marketing Education Review*, 21(2): 101–112.

Clayson, D. E., and Sheffet, M. J. 2006. "Personality and the Student Evaluation of Teaching." *Journal of Marketing Education*, 28(2): 149–160.

De Paola, Maria. 2009. "Does Teacher Quality Affect Student Performance? Evidence from an Italian University." *Bulletin of Economic Research*, 61(4): 353–357.

Franz, R. 1998. "Whatever You Do, Don't Treat Your Students Like Customers." *Journal of Management Education*, 22(1): 63–67.

Isely, P., and Singh, H. 2005. "Do Higher Grades Lead to Favorable Student Evaluations?" *Journal of Economic Education*, 36(1): 29–42.

Jones, E. B., and Jackson, J. D. 1999. "College Grades and Labor Market Rewards." *Journal of Human Resources*, 25(2): 253–266.

Krautmann, A. C., and Sander, W. 1999. "Grades and Student Evaluations of Teachers." *Economics of Education Review*, 18(1): 59–63.

Langbein, L. 2008. "Management by Results: Student Evaluation of Faculty Teaching and the Mis-measurement of Performance." *Economics of Education Review*, 27(4): 417–428.

Lichty, R. W., Vose, D. A., and Peterson, J. M. 1978. "The Economic Effects of Grade Inflation on Instructor Evaluation: An Alternative Interpretation." *Journal of Economic Education*, 10(1): 3–11.

Lin, T.-C. 2009a. "Application of a Static Game of Complete Information: Economic Behaviors of Professors and Students." *Economics Bulletin*, 29(3): 1683–1691.

Lin, T.-C. 2009b. "Implications of Grade Inflation: Knowledge Illusion and Economic Inefficiency in the Knowledge Market." *Economics Bulletin*, 29(3): 2321–2331.

Lin, T.-C. 2010. "Teacher Quality and Student Performance: The Case of Pennsylvania." *Applied Economics Letters*, 17(2): 191–195.

Madden, T. J., Dillon, W. R. Dillon, and Leak, R. L. 2010. "Students' Evaluation of Teaching: Concerns of Item Diagnosticity." *Journal of Marketing Education*, 32(3): 264–274.

Matos-Diaz, H., and Ragan, J. F. 2010. "Do Student Evaluations of Teaching Depend on the Distribution of Expected Grade?" *Education Economics*, 18(3): 317–330.

McPherson, M. A. 2006. "Determinants of How Students Evaluate Teachers." *Journal of Economic Education*, 37(1): 3–20.

Nelson, J. P. and Lynch, K. A. 1984. "Grade Inflation, Real Income, Simultaneity, and Teaching Evaluations." *Journal of Economic Education*, 15(1): 21–37.

Obermiller, C., Fleenor, P., and Raven, P. 2005. "Students as Customers or Products: Perceptions and Preferences of Faculty and Students." *Marketing Education Review*, 15(2): 27–36.

Orsini, J. L. 1988. "Halo Effects in Student Evaluations of Faculty: A Case Application." *Journal of Marketing Education*, 10(2): 38–45.

Sabot, R., and Wakeman-Linn, J. 1991. "Grade Inflation and Course Choice." *Journal of Economic Perspectives*, 5(1): 159–170.

Seiver, D. A. 1983. "Evaluations and Grades: A Simultaneous Framework." *Journal of Economic Education*, 14(3): 32–38.

Shapiro, C., and Stigliz, J. 1984. "Equilibrium Unemployment as a Worker Discipline Device." *American Economic Review*, 74(3): 433–444.

Simpson, P. M., and Siguaw, J. A. 2000. "Student Evaluations of Teaching: An Exploratory Study of the Faculty Response." *Journal of Marketing Education*, 22(3): 199–213.

Spense, M. 1973. "Job Market Signaling." *Quarterly Journal of Economics*, 88(3): 355–374.

Endnotes

1. A survey was conducted in spring 2008 to learn more about students' expectations for their final grades. This survey was distributed and collected sometime after the midterm exams but before the final exam. (Note: responses were anonymous.) Students were asked two questions and for one explanation: "*What grade for this class do you expect to receive? Why do you expect that grade? Express your*

reasons." The results showed that students' expected grades basically depended on their midterm average grades and a probable curve on final grades after the final exam (Note: the instructor's grading policy states: each exam weights one third of the final grade, and I reserve the right to curve up your final grades if the overall final average of the class is below 78%.) They also believed that they would do at least as well on their final exams as they did on their midterm exams, or even better.

2. The instructor's grading policy for fall 2011 and spring 2012 is described below. Fall 2011: (1) *Graded items*: the final exam is the only graded item for this course, and this item is graded on a 100-point scale. The final exam is comprehensive. There are no projects, no term papers, no mandatory attendance policy, no attendance bonus, and no quizzes. Participation in class discussions is encouraged but not part of the course grade. (2) *Grade scale*: following the standard straight scale (no curved). (3) *Weights of graded items*: the final exam is 100% of the course grade. The grading policy for spring 2012 is the same as the other group in fall 2011 except for graded items consisting of two midterm exams and one final comprehensive exam. Each exam is one-third of the course grade.
3. Based upon the Hausman specification test, the null hypothesis that student's grade is an exogenous variable is rejected, implying that it is an endogenous variable.
4. The main difference between the Section I-students and the Section-II students is that the Section I-students meet in the morning, while the Section II-students meet in the afternoon. We report their regressions separately instead of just creating a Section-II dummy variable, because it allows us to see how each explanatory variable impacts the explained variable in the regressions in each section.
5. The same reason as in Endnote 4 for why we report their regressions separately rather than just creating a Section-II dummy variable.
6. We used the one-tailed test. As a result, the mean for the overall evaluation was significantly higher in the group without midterm exams than in the group with midterm exams at the 1% level.
7. The halo error is "a mistake or bias that can occur in evaluating an individual's performance where they are consistently rated based on the evaluator's overall impression, rather than on their actual performance in various categories" (see http://www.businessdictionary.com.) According to evidence in previous studies (e.g., Orsini, 1988; Clayson, 1989, 1999; Clayson and Haley, 1990; Clayson and Sheffet, 2006; and Madden, Dillon, and Leak, 2010), a number of factors that determine student evaluations of teaching are not related to professors' actual teaching performance. For example, Clayson and Sheffet (2006) discovered that a professor's personality would significantly influence student behavior in rating professors. In addition, Clayson (1989) showed that, for marketing students, the halo error in student evaluation of teaching is highly related to expected grade.
8. Grade inflation has a number of negative impacts on both the labor and knowledge markets. Briefly, it can: (1) make education a less efficient signal (Spence, 1973); (2) lead universities/colleges toward the Giffen good case (Lichty, Vose, and Peterson, 1978); (3) exacerbate information asymmetry and raise monitoring costs (Shapiro and Stigliz, 1984); (4) lead to signals of comparative advantage to bias (Sabot and Wakeman-Linn, 1991); (5) create a biased signal in the labor market (Jones and Jackson, 1999); and (6) create knowledge illusion and economic inefficiency in the knowledge market (Lin, 2009b).
9. In March 2010, an online survey on campus was conducted. The question was: *"Do*

you have the perception that students are customers in school?" There were two choices: yes or no. In total, 1,016 students responded to this survey within a few days. A total of 54% of the sample chose "yes", while 46% chose "no".

10. The George Mason University Faculty Senate passed a resolution officially stating that it is inappropriate to regard students as customers. The Faculty Senate Statement: *"Corporate models" of education in which students are viewed as "customers" are not appropriate. Education is a unique activity in a democratic society that differs markedly from both business and government. Universities are absolutely essential in contemporary society as centers of free inquiry, free expression, open discovery, and dissent. Any attempt to force education into a corporatist mold devalues faculty, lowers academic standards, and harms both students and the institution itself"* (GMU Faculty Senate, September 2002).
11. In addition to the problem of the reciprocal relationship between students and professors and the halo error in student evaluations of teaching, students may not offer honest responses on evaluations. Clayson and Haley (2011) found that a majority of students admitted that an estimated 30% of their answers on evaluations were not true.

The Impact of Smoking Bans on Bar and Restaurant Values

Michael Allgrunn and Christopher C. Douglas*

ABSTRACT. The state of Florida implemented an indoor smoking ban in 2003 that exempted bars. Using a data set containing the sales price of bars and restaurants in Florida that spans 1999-2011, we utilize a difference-in-difference framework to examine whether or not a smoking ban impacts the value of bars and restaurants and in what direction. We find that the value of restaurants decreased following the smoking ban, while the value of bars increased. These results suggest that a smoking ban has a negative impact on a business' value. (D23, H75, I18)

I. Introduction

Since the introduction of the first indoor smoking ban in San Luis Obispo, California on August 2, 1990, indoor smoking bans have become a public policy staple in numerous states in the United States. According to the Americans for Nonsmokers Rights, twenty-four states have smoking bans that do not include exemptions for bars and restaurants. Only ten states do not impose any sort of a smoking ban.[1]

Given the widespread prevalence of smoking bans, it is of interest to examine the impact they have on bars and restaurants. In particular, we examine the impact that smoking bans have on the valuation of bars and restaurants that are covered by a smoking ban, with valuation operationalized as the sale price of the bar or restaurant when it changes hands. Alamar and Glantz (2004, henceforth "AG") argue that smoking bans increase the value of restaurants. They argue that revenues will either remain unchanged or rise, as consumers frequent restaurants more because they are smoke free. Costs should also fall as employees take

*Allgrunn: Department of Economics, University of South Dakota, 414 E. Clark Street, Vermillion SD 57069. Email: Mike.Allgrunn@usd.edu.
Douglas: Corresponding author, Department of Economics, University of Michigan-Flint, 303 E. Kearsley St., 220 French Hall, Flint, MI 48502. Email: ccdougla@umflint.edu.
We thank former University of Michigan-Flint student Tom Herron for providing outstanding research assistance as well as Heather Bednarek, participants at the 51st annual meeting of the Missouri Valley Economics Association, and two anonymous referees for helpful comments and suggestions. All remaining errors are our own.

fewer sick days and capital needs to be replaced less frequently due to reduced smoke damage. Indeed AG, using a data set containing the sales prices of restaurants from across the United States, find that a smoking ban increases a restaurant's value. Other studies, however, such as Pakko (2008a and 2008b) find the opposite effect. He finds that sales revenues fall as consumers who prefer an environment that allows smoking go to the smoke-free businesses less, and this is not compensated by nonsmoking consumers going to the businesses more. It seems likely that lower sales revenues would translate into a lower sales price when the business is sold.

We utilize data on the sales prices of 1,015 bars and restaurants in the state of Florida from 1999-2011 as collected by the Biz Comps database, which is the same data source as in AG. Florida implemented a smoking ban in 2003, but exempted bars from the ban. This presents a nice framework to investigate the impact of a smoking ban on a bar or restaurant's value using a difference-in-difference framework, which allows us to see the effect of the smoking ban on the treatment group, relative to the control group. If a smoking ban increases a business's value, we should find that the sales price of bars falls relative to restaurants, following the imposition of Florida's smoking ban. If, on the other hand, a smoking ban decreases a business's value, we should find the opposite effect, with the value of bars rising relative to restaurants, as bars were exempt from Florida's smoking ban.

We find the smoking ban reduced the value of restaurants and increased the value of bars. These results suggest that a segment of the public that goes to bars and restaurants has a preference for being able to smoke in these places, and the restaurant covered by a smoking ban suffered a loss of profitability by not being able to serve this segment of the market.

Section II surveys the previous literature that examines smoking bans. Section III describes the data. Section IV describes the econometric model. Section V presents the results, and Section VI concludes.

II. Previous Literature

Research on the economic impact of smoking bans address two general research questions. First, how does a bar or restaurant accommodate smokers and nonsmokers absent a smoking ban? Second, if a smoking ban is put into place, does the ban help or harm the bar or restaurant

covered by it?

Survey data is used to address the first research question. Boyes and Marlow (1996) surveyed 64 bars and restaurants in San Luis Obispo, California in 1992 to see how they accommodated smokers and nonsmokers absent a smoking ban. They found that 61% of bars and restaurants surveyed attempted to reduce smoke through smoking and nonsmoking sections, ventilation, the use of a patio, and/or other means. The authors also found that 57% of bars and restaurants felt no impact from the smoking ban, while 25% reported a negative impact and 17% reported a positive impact.

Dunham and Marlow (2000a) use a 1996 nationwide survey of 650 bars and 650 restaurants by the National Licensed Beverage Association to examine what determines the seating allocation in smoking versus nonsmoking sections. The authors find that smoking laws are not a statistically significant predictor of the fraction of seats in a bar or restaurant that are designated as nonsmoking. States with such laws already had fewer smokers and hence a larger demand for nonsmoking sections prior to the law. Instead, the authors found that the size of the smoking and nonsmoking sections depended on the incidence of smoking in the state where the business is located, whether the business was part of a national chain, and whether or not the business was a bar. The authors argue that consumer preferences shape the size of smoking and nonsmoking sections as businesses cater to these preferences. Dunham and Marlow (2004) found very similar results in a 2001 survey of 1,000 bars and restaurants in Wisconsin.

Biehl and Douglas (2011) surveyed 125 restaurants in Genesee County, Michigan. This survey was conducted prior to Michigan's smoking ban being implemented in order to see how bars and restaurants accommodated smokers and nonsmokers absent a ban. The authors found that restaurants accommodated nonsmokers simply by banning smoking, while bars accommodated nonsmokers by restricting smoking to the bar-area of the establishment or in a separate room. Just under half of bar owners were supportive of Michigan's proposed smoking ban, while nearly 70% of restaurant owners were supportive of the ban.

The second research question – whether a smoking ban is beneficial or detrimental to a bar or restaurant – has received more attention from researchers. Perhaps the first paper addressing this question is Glantz and Smith (1994). Using data spanning 1986-1993 that includes 15 California cities that had a smoking ban and 15 that did not, the authors

find that a smoking ban did not lower restaurant sales as a proportion of total retail sales for a given city. The authors argue this is evidence that a smoking ban does not adversely impact restaurants. This result was questioned by Evans (1997). He argues that Glanz and Smith (1994) ignored the fact that 14 out of the 15 cities with smoking bans also exempted free-standing bars, bar areas of restaurants and cocktail lounges, patios, and separately ventilated and enclosed smoking rooms from their smoking bans. Evans (1997) also alleges that Glantz and Smith (1994) misclassified 8 out of the 15 cities without smoking bans. When he addresses this issues, Evans (1997) finds that a smoking ban resulted in a 5% decline in restaurant sales.

In a series of papers, Pakko (2006, 2008a, and 2008b) finds that a smoking ban either harms a business (Pakko 2008a and 2008b) or has so many exemptions that it has no effect on a business (Pakko 2006). Pakko (2008a) uses monthly sales tax receipts from 2001-2007 for the city of Columbia, Missouri to investigate the impact of the city's smoking ban, which took effect January, 2007. The author found a 5% average loss in sales tax revenue following the smoking ban, despite there being no change in the sales tax rate at that time. Pakko (2008b) uses monthly net proceeds from Delaware's "racinos" (a combination of a casino and horse race track) from 1997-2005 to examine the impact of the state's 2002 smoking ban on these businesses. The author finds that a smoking ban resulted in an average revenue decline of 14.9% at the racinos, costing the state $33 million in foregone tax revenues annually.

Pakko (2006) uses quarterly sales tax data for Maryville, Missouri from 1998-2004 to investigate the city's smoking ban, which took effect on June 9, 2003. The author cites a study from the Missouri Department of Health and Senior Services (DHSS) that found that bar and restaurant sales increased, following the smoking ban. Pakko (2006) finds that this increase stems from the opening of a very popular Applebee's that occurred around the same time. Taking out this effect, the author finds no impact of the smoking ban on bar or restaurant sales. The likely explanation is that the city's smoking ban exempts stand-alone bars, any business that receives more than 60% of its revenue from alcohol sales, and in fact exempts seven establishments by name. At the time the ban was implemented, 70% of restaurants in Maryville were already smoke free.

Adams and Cotti (2007) used nationwide, county-level data from the Quarterly Census of Employment and Wages from 2001-2004 to

investigate the impact of a smoking ban on bar and restaurant employment. The authors find that in counties with an average smoking prevalence, bars see a 7.1% reduction in employment while restaurants have no statistically significant change in employment.

Alamar and Glantz ("AG," 2004) find that a smoking ban increases the sales price, or value of a restaurant when it is sold. Using a dataset of 608 restaurants across the United States spanning 1991-2002, the authors find that restaurants covered by a smoking ban are worth an average of 16% more than restaurants that are not. The authors operationalize a restaurant's value as the ratio of its price when it sold to its gross establishment sales, or *P/S*. In other words, a smoking ban is associated with a 16% higher *P/S* for a restaurant. They do conduct an analysis for bars and find that a smoking ban increases bar values as well. However, their sample includes only 78 bars and just 5 of them are covered by a smoking ban. As a result, the authors state that their results regarding bars "should only be considered preliminary" (p. 524).

Our paper extends AG's in several respects. First, they conduct their main analysis excluding bars whereas our dataset includes data on 131 bars. Second, AG does not employ a before-and-after study to investigate how restaurant sales prices change following a smoking ban being implemented. Instead, what the authors find is that, ceteris paribus, P/S is larger for restaurants covered by a smoking ban. It could be that a smoking ban forced less profitable restaurants out-of-business. Or, smoking bans may be endogenous in that they are only implemented in areas with profitable restaurants or where few people smoke anyway, and thus are unlikely to result in financial hardship. In contrast, we utilize a difference-in-difference methodology, which allows us to see the differential effect before and after a policy change for the treatment group relative to the control group. In this case, we can see the ceteris paribus valuation of restaurants (our treatment group) before and after 2003, and see how that compares to the change in bar valuations (our control group) over the same time.[2]

Finally, AG divide sales price and seller's discretionary earnings (or *SDE*, which is roughly pre-tax profit, which is used as a control variable in their regression) by total restaurant sales to control for differences in restaurant sizes.[3] However, it cannot be determined if the positive and statistically significant estimated coefficient on the smoking ban dummy variable in their regression means that *P/S* increased because *P* increased or *S* decreased. This is problematic because the former and latter have

opposite implications for the effect a smoking ban has on a restaurant's value.[4] We instead include a quadratic term for *SDE* as a control variable. This allows *SDE* to have a diminishing impact on the sales price of a bar or restaurant, should *P* not increase in a linear fashion with *SDE*.

III. Data

Data for the sales price of 1,015 bars and restaurants in the state of Florida spanning the time period 1999-2011 were purchased from the Biz Comps database.[5] Biz Comps provides the sales prices of businesses across various SIC codes. The database is used by the real estate industry for the purposes of establishing "comparables" in real estate transactions.[6] This is also the source of data employed by AG.

We downloaded sales data for businesses with SIC code 5812 ("Eating Places") and SIC code 5813 ("Drinking Places") from the Biz Comps database. In addition to the sales data, the Biz Comps database also contains data on seller's discretionary earnings (SDE). SDE is defined by Bizcomps as a business' net profit before taxes and any compensation to the business owner is paid, plus amortization, depreciation, interest, and other non-cash and non-business related expenses. The SDE provided by BizComps assumes one owner.[7] Businesses with SIC code 5813 were classified as bars while businesses with SIC code 5812 were classified as restaurants in our data set. We classified individual restaurants as being either a family restaurant, an ethnic restaurant, a take-out restaurant, fast food, or coffee shop based on the name of the restaurant provided by Biz Comps. Given they have limited (if any) indoor seating, take-out restaurants were dropped from the analysis. In order to control for business cycle conditions that might impact the bar or restaurant's sales price, we also obtained data on the average weekly wage and unemployment rate during the month that the bar or restaurant was sold. We converted all prices to real terms using the consumer price index so that all prices are in constant December, 2011 dollars.

Table 1 presents summary statistics for all of the bars and restaurants in the data set. Summary statistics include the mean, standard deviation, minimum, and maximum values for all variables in the data set. An establishment sold for an average of just under $200 thousand with an average SDE of $126 thousand. "Income" (which is the average weekly

wage) is, on average, $808.52 while the average income growth rate, defined as the year-to-year percentage change in income, is 3.3%. In the data set, 13%, or 131 out of 1015, of the observations are bars as defined by the SIC code. In terms of restaurants, 21% are fast food restaurants while 30% are family restaurants. Note from the table that bar and restaurant transactions are fairly evenly spaced throughout the data set, given that the data spans 1999-2011. Twenty-five percent of the transactions took place prior to 2003, 51.7% of transactions took place after 2003, while 10% of transactions took place in 2003. Thus, the sale of these businesses are not clumped together before or after the smoking ban in 2003.

TABLE 1–Summary Statistics

Variable	Mean	Standard Deviation	Min	Max
Price (thousands of $)	$199.99	$168.67	$10.27	$1,457.10
SDE (thousands of $)	$126.89	$269.29	-$16.60	$6,915.79
Income	$808.52	$34.88	$661.72	$903.38
Income growth rate	3.30%	2.89%	-15.02%	11.53%
Unemployment rate	5.47%	2.14%	3.30%	11.40%
Bar	0.13	0.34	0	1
Fast Food	0.21	0.41	0	1
Family	0.30	0.46	0	1
Percentage of transactions				
Before 2003	25%			
In 2003	10%			
After 2003	51.7%			
Observations	1015			

IV. Econometric Model

As discussed in the previous section, the data set we utilize spans 1999-2011 and contains the sales price of bars and restaurants for the State of Florida for when these businesses changed hands. The smoking ban in Florida took effect in 2003 and provided an exemption for bars. These two characteristics of Florida's smoking ban allows us to employ a difference-in-difference framework in testing whether a smoking ban increases or decreases a bar or restaurant's value. Difference-in-difference estimation allows us to see the differential effect before and after a policy change for both treatment and control groups. In this case, we can see the ceteris paribus valuation of restaurants (our treatment group) before and after 2003, and see how that compares to the change in bar valuations (our control group) over the same time.

Define *ban* as a dummy variable equal to unity if the year is 2003 or beyond, and zero otherwise. Define *bar* as a dummy variable equal to unity if an observation in the data set is a bar (SIC 5813 classified business), and zero otherwise. Define *P* as the sale price of the bar or restaurant. The difference-in-difference estimation would then estimate the following equation for the state of Florida:

$$P = \beta_0 + \delta_0 ban + \delta_1 ban \times bar + \beta_1 bar + \beta_2 SDE + \beta_3 SDE^2 + \beta_4 avweeklywage + \beta_5 avweeklywagegrowthrate + \beta_6 umemployment + \beta_7 fastfood + \beta_8 family + \beta_9 time + u \quad (1)$$

The parameter of interest in equation (1) is δ_1, which is the difference-in-difference parameter that measures the impact of the smoking ban on bars relative to restaurants, following the policy change. Since bars are exempt from Florida's smoking ban, the AG result would predict $\delta_1<0$. That is, if a smoking ban increases a bar or restaurant's value, the fact that bars are exempt from the smoking ban should cause their value to decrease. On the other hand, if a smoking ban reduces a bar or restaurant's value, perhaps because the business can no longer cater to its smoking customers, then we would expect $\delta_1>0$. In other words, being exempt from a smoking ban would increase the bar's value.

Economic control variables in equation (1) include *averageweeklywage*, which is the average weekly wage in Florida during the month the bar or restaurant was sold, *avweeklywagegrowthrate* which

is the year-to-year growth rate of the average weekly wage, *unemployment*, which is the unemployment rate in Florida during the month the bar or restaurant was sold. A person's average weekly wage likely determines how much he or she can afford to go out to dine or drink in a given week. Likewise, a higher unemployment rate should lead to fewer people going out to eat or drink, which would depress bar or restaurant values. Following AG, for restaurants in the data set, *fastfood*, is a dummy variable taking the value of unity if the restaurant is a fast food restaurant, *family* is a dummy variable taking the value of unity if the establishment is a family restaurant. The dummy variable *time* consists of yearly dummy variables for all observations in the data set.

Similar to Graham (1997), it is apparent that there is an issue with equation (1) in that SDE and the other economic control variables are linearly correlated. A smoking ban is alleged to increase a bar or restaurant's value in-part because the impact it has on the bar or restaurant's profit. As argued by AG, additional customers attracted by the smoke-free policy result in additional revenues and hence profit for the bar or restaurant, and thus a higher sale price when the business is sold. If this is the case, then it would not make sense to estimate equation (1) as written. The estimated coefficient on the smoking ban dummy variable in equation (1) would be interpreted as the partial effect of a smoking ban on a bar or restaurant's sales price, holding profit (or *SDE*) constant. This is inconsistent with how a smoking ban is hypothesized to increase a bar or restaurant's value.

The same is true regarding the economic control variables in equation (1). The impact of economic variables such as the average weekly wage and unemployment rate will fall on a bar or restaurant's profit, as these economic control variables determine how often people go out to eat or drink, which then impacts the bar or restaurant's price when it is sold. For instance, higher weekly wage would induce people to go out to eat or drink more, which would increase a bar or restaurant's profit, which would then increase a bar or restaurant's price when it is sold. In equation (1), the interpretation of estimated coefficients on the economic control variables would be the partial effect of these variables on a bar or restaurant's sales price, holding profit (or SDE) constant. For example, the estimated coefficient on the average weekly wage would be the partial effect of the average weekly wage on a bar or restaurant's sales price, holding profit constant. This is inconsistent with how the average weekly wage would impact a bar or restaurant's sales price.

Following Graham (1997), we employ the following procedure to correct for this issue. First, we regress *SDE* on all the other control variables in equation (1):

$$SDE = \beta_0 + \beta_1 ban + \beta_2 bar + \beta_3 ban \times bar + \beta_4 avweeklywage + \beta_5 avweeklywagegrowthrate + \beta_6 umemployment + \beta_7 fastfood + \beta_8 family + \beta_9 time + u \quad (2)$$

Second, we gather the residuals from equation (2), defined as *SDEresid*. *SDEresid* are thus the bar or restaurant's profit not explained by the average weekly wage (or growth rate of), the unemployment rate, whether (in the case of restaurants) the restaurant is a fast food or family restaurant, or the passage of time. We then replace *SDE* and SDE^2 with *SDEresid* and $SDEresid^2$ in equation (1):

$$P = \beta_0 + \delta_0 ban + \delta_1 ban \times bar + \beta_1 bar + \beta_2 SDEresid + \beta_3 SDEresid^2 + \beta_4 avweeklywage + \beta_5 avweeklywagegrowthrate + \beta_6 umemployment + \beta_7 fastfood + \beta_8 family + \beta_9 time + u \quad (3)$$

The estimated coefficient on, say, the average weekly wage no longer has the interpretation of being the partial effect of the average weekly wage on the sale price of the bar or restaurant, holding profit (*SDE*) constant. The variable *SDEresid* is orthogonal to these other variables as a consequence of equation (2). That is, the interpretation of *SDEresid* is the bar or restaurant's profit not explained by whether or not the bar or restaurant is covered by a smoking ban, the average weekly wage rate and unemployment rate, and so forth.

V. Results

Table 2 presents the results of our difference-in-difference estimation given by equation (3). The difference-in-difference parameter, which is the estimated coefficient on *banxbar*, suggests that bars saw an average increase of $149 thousand in their sales price following the implementation of the smoking ban, relative to restaurants. The smoking ban resulted in a $76 thousand reduction in overall business value (as coefficient on the "ban" dummy variable is 75.65 and significant at the 1% level). Thus bars saw, on average, a $73 thousand increase in value following the smoking ban being enacted. Prior to the smoking ban, there

was no statistically significant difference in the sales price of bars compared to restaurants (the estimated coefficient on the *bar* dummy variable is insignificant).

TABLE 2–Difference-in-Difference Results

Variable	Coefficient
ban	-75.65** (18.01)
banxbar	149.02** (25.69)
bar	32.89 (18.02)
SDEresid	1.26** (0.2604)
$SDEresid^2$	-0.000203** (0.0000364)
weekly wage	0.426** (0.129)
weekly wage growth rate	-1.76 (1.47)
unemployment	-22.96** (3.56)
fast food	-8.51 (10.44)
family	14.76 (10.71)
time	17.95** (3.83)
constant	-71.20 (92.86)
R^2 = 0.4949 **N = 1,015**	

Notes: "**" denotes statistically significant at the 1% level, "*" denotes statistically significant at the 5% level. Heteroscedasticity robust standard errors in parenthesis.

Given that the smoking ban implemented in 2003 applied to only restaurants and not bars, these results are suggestive that smoking bans harm, rather than help, businesses covered by them. These results are similar to those in Pakko (2008a and 2008b) who finds that a smoking ban is associated with decreased sales tax revenue, stemming from a loss of sales revenues at businesses covered by a smoking ban. Though Pakko (2008a and 2008b) looks at sales revenue rather than the sales price of a business, it seems likely that lower sales revenue will translate into a lower sales price when the business is sold.

The results of equation (3) also suggest that a $1 thousand rise in SDE increases a bar or restaurant's sales price by $1.26 thousand, though this effect diminishes as the estimated coefficient on the square of SDE is negative and significant at the 1% level. A 1 percentage point increase in the unemployment rate reduces a bar or restaurant's value by approximately $23 thousand. The coefficient on the time trend variable ("time") suggests that each additional year results an increase of bar or restaurant values by about $18 thousand, holding all other variables in equation (3) constant.

These results stand in contrast to those found by AG. Their results suggest that a smoking ban increases the value of restaurants and bars (albeit with a very small sample size in the case of the latter). Our difference-in-difference estimation suggests that a smoking ban negatively impacts a restaurant's sales price. This can be seen from the negative and statistically significant estimated coefficient on the ban dummy variable (*ban*) as well as the positive and statistically significant difference-in-difference parameter. If a smoking ban increased a restaurant's sales price, then we would expect a positive coefficient on *ban* (that is, a smoking ban would be associated with an across-the-board increase in bar and restaurant values) and a negative coefficient on *banxbar* (being exempt from a smoking ban harms bars).

Our results are in line with studies that find that bars and restaurants are able to accommodate both smoking and nonsmoking customers privately absent a smoking ban. Dunham and Marlow (2000a and 2000b) find that consumer characteristics such as the percentage of smokers in the adult population and whether or not an business is a bar impact the size of the nonsmoking section in the bar or restaurant. Dunham and Marlow (2004) finds that the quantity of smoking and nonsmoking seats are related to the demographics of the consumers they serve. For instance, bars and restaurants catering to a white collar clientele and to

families with children tend to allocate a greater proportion of their seats to nonsmokers while bars and restaurants that hold liquor licenses allocate some of their seats to nonsmokers. Likewise, Biehl and Douglas (2011) find bars and restaurants take numerous steps to keep smokers away from nonsmokers, such as restricting smoking to a separate room away from nonsmokers.

These results, along with the results of the current paper, suggest that a segment of the population that goes to bars and restaurants has a taste for being able to smoke indoors, and bars and restaurants suffer if they are unable to cater towards this segment of the market. The rise in the value of bars in Florida that are exempt from the smoking ban suggests that an exemption from the ban gives these bars increased market power over this particular market segment, which gets reflected in a higher sale price.

It is true that on one hand, this differential effect would be absent if smoking was banned in both bars and restaurants as bars no longer act as substitutes for restaurants for smokers. On the other hand, other substitutes might become available. These might include simply eating or drinking at home or in bars which elect to pay fines rather than abide by the ban. Marlow (2010) found widespread noncompliance by bars after Ohio's smoking ban was implemented in 2007, as Ohio's smoking ban provided no exemption for any business open to the public. If a bar's customers are less likely to turn it in for noncompliance with the smoking ban than a restaurant's, a bar might simply flaunt the law and the differential effect might still be present.

VI. Conclusion

This paper utilized a dataset consisting of sales prices of bars and restaurants to investigate whether a smoking ban impacts a bar or restaurants valuation and in what direction. We find that a smoking ban reduces the value of bars and restaurants, as those that are covered by a smoking ban see their value fall by $75 thousand on average (Table 2). Yet, those that are exempt from smoking bans see their value increase by nearly $73 thousand, on average. This is consistent with the idea that a segment of the public has preferences towards indoor smoking and places suffer a loss of value when they cannot cater to this market segment.

Previous studies in the literature have found similar results. Boyes and Marlow (1996) found that smokers were much less likely to be

supportive of a smoking ban than nonsmokers. Pakko (2008a and 2008b) found a decrease in sales tax revenues in restaurants in Columbia, Missouri and racinos in Deleware after these businesses were covered by a smoking ban. Adams and Cotti (2007) found a statistically significant decrease in employment in bars after these businesses are covered by smoking bans.

It is important to note that this paper does not address the normative question as to whether or not smoking bans are good public policy. When conducting cost-benefit public policy analysis, it is important to correctly specify the costs and benefits of the policy. This paper reinforces the emerging consensus in the smoking ban literature that the impact of smoking bans on bars and restaurants should fall on the cost side of the ledger.

References

Adams, Scott, and Chad D. Cotti. 2007. "The Effect of Smoking Bans on Bars and Restaurants: An Analysis of Changes in Employment." *The B.E. Journal of Economic Analysis & Policy* 7 (1): Article 12.

Alamar, Benjamin C., and Stanton A. Glantz. 2004. "Smoke Free Ordinances Increase Restaurant Profit and Value." *Contemporary Economic Policy* 22 (4): 520-525.

Ashenfelter, Orley, David Card. 1985. "Using the Longitudinal Structure of Earnings to Estimate the Effects of Training Programs." *Review of Economics and Statistics* 67 (4): 648-660.

Biehl, Amelia, and Christopher C. Douglas. 2011. "How Do Private Markets Address Smoking Externalities?" *Journal of Economics (MVEA)* 37 (1): 39-57.

Boyes, William J., and Michael L. Marlow. 1996. "The Public Demand for Smoking Bans." *Public Choice* 88 (1): 57-67.

Dunham, John, and Michael L. Marlow. 2000a. "The Effects of Smoking Laws on Seating Allocations of Restaurants, Bars, and Taverns." *Economic Inquiry* 38 (1): 151-157.

Dunham, John, and Michael L. Marlow. 2000b. "Smoking Laws and Their Differential Effects on Restaurants, Bars, and Taverns." *Contemporary Economic Policy* 18 (1): 326-333.

Dunham, John, and Michael L. Marlow. 2004. "The Market for Accommodation: Determinants of Smoking Policies in Restaurants and Bars." *Eastern Economic Journal* 30 (3): 377-391.

Evans, Michael K. 1997. "A Review of 'The Effect of Ordinances Requiring Smoke-Free Restaurants on Restaurant Sales." working paper.

Glantz, Stanton A., and Lisa R.A. Smith. 1994. "The Effect of Ordinances Requiring Smoke-Free Restaurants on Restaurant Sales." *American Journal of Public Health* 84 (7): 1081-1085.

Graham, Michael H. 1997. "Factors Determining the Upper Limit of Giant Kelp, Macrocystis Pyrifera Agardh, Along the Monterey Peninsula, Central California,

USA." *Journal of Experimental Marine Biology and Ecology* 218 (1): 127-149.

Henderson, David R. 2007. "Smoking in Restaurants: Who Best to Set the House Rules?" *Econ Journal Watch* 4 (3): 284-291.

Marlow, Micheal L. 2010. "The Economic Losers from Smoking Bans," *Regulation*, Summer: 14-19.

Pakko, Michael R. 2008a. "The Economic Impact of a Smoking Ban in Columbia, Missouri: An Analysis of Sales Tax Data for the First Year." Federal Reserve Bank of St. Louis *Regional Economic Development* 4 (1): 30-40.

Pakko, Michael R. 2008b. "No Smoking at the Slot Machines: The Effect of a Smoke-Free Law on Delaware Gaming Revenues." *Applied Economics* 40 (14): 1769-1774.

Pakko, Michael R. 2006. "On the Economic Analysis of Smoking Bans." *Federal Reserve Bank of St. Louis Regional Economic Development* 2 (2): 115-130.

Endnotes

1. See, http://www.no-smoke.org/pdf/WRBLawsMap.pdf
2. See Ashenfelter and Card (1985) for one of the first uses of difference-in-difference methodology
3. See page 522 of Alamar and Glantz (2004)
4. See Henderson (2007) for a detailed critique of the AG methodology.
5. See, http://www.bvmarketdata.com/defaulttextonly.asp?f=BIZCOMPS%20-%20Main%20Street%20Business%20Private%20Transactions
6. The database contains the state the business was located in when it was sold, but not more specific locational data such as city or zip code.
7. See, http://bizcomps.com/definitons/

Book Reviews

Edited by Robert Herren

Endangering Prosperity: A Global View of the American School. ERIC A. HANUSHEK, PAUL E. PETERSON, AND LUDGER WOESSMANN. Washington, D.C.: Brookings Institution Press, 2013. Pp. xi, 147. $26.95.

If American education were to improve so that students would achieve the same skills as those currently achieved by Canadian students, by the end of the century, GDP per capita in the United States would be 20% higher than otherwise, the present value of such an increase is estimated to be around $77 trillion, which represents over five times the current United States GDP. The heart of the message in this well written, short book can be thus summarized: the opportunity costs of not improving United States K-12 education are staggering.

Eric A. Hanushek, Paul E. Peterson, and Ludger Woessmann (HPW), arrive at this conclusion by first arguing that education has a significant impact on the long-run economic growth of a nation, Chapter 2, and finding some estimates of the relationship between education and growth. In Chapters 3 and 4, they compare the math and reading skills achieved by American students to the skills achieved by students in other nations and find that American students do not compare well; this is true even when using the best performing American students as the basis for comparison. In Chapter 4, HPW run simulations of the impact of changes in skill achievement on United States GDP arriving at results stated above. In the last chapter HPW discuss several recent improvements in United States student achievement compared to what has been achieved elsewhere. They find such improvements unimpressive as well.

The topic under investigation is fraught with many thorny problems with which they deal parsimoniously. Overall, I am convinced by the logic of the arguments and their results. My problem with this book is in the policy recommendations and the obstacles HPW identify to improvements in American K-12 education. First, let me deal with their policy recommendations: there are not many clearly defined policy

recommendations in this book because their main goal is to add urgency to the call for structural reform in education (p. 15). HPW make it clear however, that it is imperative to establish policies that would enhance the overall quality of the teaching workforce (p. 14) and point out that decreases in class size and pay increases for teachers have not resulted in improvements in educational outcomes. Some of this should be intuitive for economists, even those not familiar with the literature on education: there is likely to be an optimal class size and class smaller than that optimal size will not add to the educational experience.

Second, according to HPW's account much of the blame for the current state of American education and the main obstacle for improvement are teachers and their unions, after all the school workforce "favors only those changes to the status quo that enhance their income or lighten the workload. They oppose changes in the organization and structure of the school system that would be likely to enhance learning opportunities of those for whom they are educationally responsible" (p.3).

Let me be clear, yes incentives matter, and there could be changes in teachers' contracts that would at the margin affect teachers' behavior and yes there could be changes in the organizational structure of schools that may improve teaching quality. However, there may be many factors which also affect student performance beyond the control of teachers. For example, as college teachers we know that a handful of disruptive students can affect the whole class performance, I wonder if different countries have different policies regarding such behavior and how to deal with it. Along the same lines, I wonder if there are significant differences in other school policies and curriculum that may affect student performance. Are there differences in the curriculum at education schools that may later impact teachers performance? In addition, early childhood has a significant impact on individuals' intellectual curiosity and readiness to learn; hence the nature of policies, if any, regarding daycare may impact later learning; for example, one of the countries that outperforms the United States in math is Finland, (page 40, Figure 3-1) which has a heavily subsidized day care system that some credit in part for their success in education (see http://www.npr.org/2014/03/08/287255411/what-the-u-s-can-learn-from-finland-where-school-starts-at-age-7). What happened to Head-Start? In short, I am not sure that teachers and their unions are the main obstacle to improvement.

Despite these concerns, I recommend this book highly. HPW identify a significant problem and contribute immensely by providing an estimate of the benefits from improving the skills of our young people. In addition, an economist looking for a new research agenda could find much inspiration in almost every page in this book.

OSCAR FLORES

Minnesota State University Moorhead

The Great Recession: Lessons for Central Bankers. JACOB BRAUDE AND OTHERS (EDITORS). Cambridge, MA: The MIT Press, 2013. Pp. xii, 380. $22.00.

This book is a collection of papers presented at an international conference on *Lessons from the World Financial Crisis.* The topic was the Great Recession (2008-2009). Sponsored by the Bank of Israel, the participants were mainly central bankers but included representatives of international financial institutions (such as the International Monetary Fund).

Papers are organized into four different sections: Part I) Monetary Policy in View of the Crisis; Part II) Macroprudential and Financial Policies; Part III) Capital Flows, Capital Controls, and Exchange Rate Policies; and Part IV) The Crisis and its Lessons: Some Case Studies. Each Part does not present a consensus position of conference participants but rather the views of individual central bankers concerning issues addressed in their particular papers.

Part I outlines monetary policy responses by central bankers from three different geographic locations affected by the financial crisis: the Euro Area, Turkey, and Chile. Hugh Pill and Frank Smets (from the European Central Bank) focus their attention on the Euro Area. One challenge that raised concerns about the effectiveness of monetary policy was the (zero) lower bound on nominal interest rates, something that Pill and Smets argue was less of an issue in the Euro area.

More attention was given to the introduction of unconventional monetary policy measures (or nonstandard measures – NSMs). These included massive asset purchase programs (many in unconventional markets) and liquidity intermediation (swapping illiquid assets for cash). The authors viewed these policy innovations favorably because they

appeared to both repair financial market transmission mechanisms and to mitigate the impact of financial market instability on the real economy. The success of these policies was due, in part, to a prior commitment to price stability on the part of central banks.

The papers by Alp and Elekdag (Turkey) and Claudio Soto (Chile) examine monetary policy responses in second-world economies. The period from 2002-2010 was one where the Central Bank of the Republic of Turkey (CBRT) implemented a policy of inflation targeting. Because the Great Recession occurred in this time frame, the authors were able to employ their econometric model of the Turkish economy to conduct counterfactual policy simulations. They found that the recession in Turkey would have been much worse had CBRT not previously implemented inflation-targeting along with a flexible exchange rate policy.

Chile's Gross Domestic Product (GDP) fell by 1.5 percent in the wake of the financial turmoil unleashed in 2008, but decline was short-lived. The economy was expanding by the last quarter of 2009 and advanced by 5.0 percent in 2010. Claudio Soto (Central Bank of Chile) attributes the short duration to aggressive policies undertaken by the Central Bank whose initial response was to provide liquidity, but later slashed sharply the overnight lending rate. When this rate reached its (zero) lower bound, the Bank (in mid-2009) unconventionally extended its lending up to six months at the low policy rate. Soto's policy simulations suggest that these actions averted an additional two percent fall in Chile's GDP.

One possible way to avoid repeating the Great Recession experience is to make policy changes. Conference participants uniformly endorsed the need for both enhanced monitoring and additional regulation of the financial system. In conventional rhetoric, they favor *macroprudential* policies that are the focus of the two papers in Part II.

Noting that developing the tools of macroprudential policy is a work in progress, Alberola, Trucharte, and Vega (all of the Bank of Spain) make a case for one particular tool: dynamic loan-loss provisioning. The Bank of Spain has been experimenting with this tool for some time; it involves varying bank loan loss provisions over the business cycle—raising them in the expansionary phase and reducing them in the contractionary phase. Ideally, this reduces the risk of a credit crisis and builds a buffer of bank reserves that makes the financial system more resilient if one occurs. While acknowledging the impossibility of

knowing the counterfactuals, the authors' empirical results lead them to believe that Spanish banks benefitted from the use of dynamic provisioning during the Great Recession.

In the second paper (Part II), Helene Schuberth (Central Bank of Austria) suggests that tax policies contributed to the risk-taking and financial imbalances observed in the recent financial crisis. Hence, Pigouvian tax policies should not be excluded from the macroprudential tool kit because they offer several advantages. Tax policies have an impact throughout the financial sector, thus reaching those institutions not typically affected by more conventional bank regulation. Moreover, as a complement to increased regulation, tax policy can serve as a viable instrument when the objective is to reduce (or eliminate) negative externalities that appear within the financial sector.

Capital flows during the Great Recession led to renewed debates about appropriate exchange rate policies (Part III). Countries recovering relatively quickly often experienced significant capital inflows that led to discussions of possible capital controls to mitigate some consequences. A paper by Jonathan Ostry (IMF) reviewed previous work on exchange rate policies as they relate to such capital inflows.

While Korea was the recipient of substantial capital inflows from 2000-2007, there was a sharp reversal during the financial crisis, and Korea suddenly experienced a significant capital outflow. Chung and Kim (both Bank of Korea) discuss both periods. They infer that flexible exchange rates and relatively free capital flows were not sufficient to insulate the country from the severe external shock that occurred during the Great Recession. A final paper by Capistran (Bank of America Merrill Lynch), and Cuadra and Ramos-Francia (both Bank of Mexico) employed a New-Keynesian Model to test different policy responses to such external shocks. According to their results, countries with stronger economic fundamentals were better positioned to effectively employ countercyclical monetary and fiscal policies during the Great Recession.

Gleaning lessons from the Great Depression experience is the concern in Part IV. Each of the four papers is a case study for an individual country (Australia, Norway, Israel, and Ireland). The first three experienced relatively mild recessions and recovered relatively quickly.

Ireland, the fourth, had a severe financial crisis accompanied by significant turbulence in the real sector. Browne and Kelly (both Central Bank of Ireland) focus on labor market distortions and a troublesome

monetary policy. The former resulted from a collective bargaining process in Ireland that promoted equalization of nominal wages across the traded goods (TG) and nontraded goods (NT) sectors, something that slower productivity growth in the NT sector did not warrant.

Ireland's monetary policy problem related to the Euro Area's common currency, and the ongoing convergence of relatively low-income euro countries to those in the core (Germany, France, and Netherlands). The low interest-rate monetary policy implemented by the European Central Bank was deemed useful for the core euro economies experiencing sluggish growth. It was not appropriate for the more rapidly growing economies in the periphery: Ireland, Greece, Spain, and Portugal. In Ireland, the monetary policy created a wedge between low real interest rates and the marginal product of capital. The resulting overinvestment led to a boom/bust cycle, especially in real property sector.

Study of the three mild recession-quick recovery economies (Australia, Norway, and Israel) might prove instructive for countries (such as the United States) wanting to avoid a repeat of the severe financial crisis of 2008-2009. Sound macroeconomic management was a common feature, including more fiscal discipline. The Australian government, for example, ran a small budgetary surplus, on average, in the 20 years preceding the crisis.

The importance of a credible monetary policy that was conservative and focused on price stability was also cited. Sound banking practices and the near absence of complex financial instruments were also prevalent in these countries. Despite the relative success, central bankers in the countries generally endorsed a need for improved macroprudential policies.

As suggested by the title, this is largely a book by central bankers for central bankers. It holds potential interest for a much larger audience precisely because it reflects the thinking of central bankers. It is reasonable to assume that the lessons they learned will serve as a cornerstone for monetary policy going forward.

WILLIAM D. GERDES

Clarke University

Global Interdependence, Decoupling and Recoupling. YIN-WONG CHEUNG AND FRANK WESTERMANN (EDITORS). Cambridge, MA: The MIT Press, 2013. Pp. vi, 308. $ 35.00

This tome is a collection of essays looking at the extent to which economies are "coupled" in the sense of having simultaneous (synchronized) business cycles, and why some countries differ from such (or are "decoupled"). Although the articles were initially presented at a 2011 Italian conference, several insights are useful now, especially with the Greek crisis at the time of this review's Summer 2015 writing. Cheung is at the City University of Hong Kong, while Westermann is at the University of Osnabrueck (Germany).

After the editors summarize the book's contents (Chapter 1), Lombardi et al. discuss several issues in tracking international business cycles. The researchers use a "vector error correction model" and monthly industrial production indexes for the United States, the euro area, and Japan for the 1980-2010 period to assess coupling among the three economies. They concluded that United States recessions had the most impact on the three, with United States recoveries less so because of different fiscal and monetary policy responses in the respective nations.

Cubadda et al. in Chapter 3 examine if within the European Union (EU) business cycle "synchronicity" has emerged. Looking 24 EU member states over the 1999-2011 period and their respective real GDP growth rates, the authors specify a VAR-type model. They determined that ten of the EU states had a common cycle. Another ten countries had cycles that moved with a lag to the former group. Finally, four countries – including Britain – had "idiosyncratic" cycles apart from the others.

Siklos in Chapter 4 divides the world into the United States, China, Japan, and six regional groups (including one for the EU). Using 1980-2010 panel data and a quantile regression model, he concludes that real GDP growth rates have become more synchronous over time, but that shocks can create short term "interruptions." Inflation-targeting monetary policies in many countries have abetted this apparent coupling. Siklos notes that China and the Central and East European countries were not as impacted by the 2008 financial crisis as were other nations.

Hirata et al. in Chapter 5 also take a world-wide perspective, but ask whether regionalization has had a greater impact than globalization on business cycles. They group 106 countries into seven geographic regions

(with NAFTA being one of the seven). They then decompose fluctuations in real GDP into a global factor, a regional factor, a country-specific factor, and a residual factor, using 1960-2010 data. Specifying and estimating a dynamic factor model, the economists conclude that the global factor, while important, has diminished over time, and the regional factor has gained in influence. The latter appears to have appeared because of growing intraregional trade (say through trade agreements like NAFTA and the EU), as well as growing regional financial linkages. Furthermore, the global factor is more significant in explaining business cycles in North America and Europe than elsewhere. Country-specific events are less important, the more "open" an economy is; however, the country-specific factor is the most important of the four for most nations.

Lugauer and Mark look at household saving in China. Based on other studies, the two authors argue that China's rapid economic growth has been due primarily to investment rather than exports. In turn, China's high saving rate has helped it finance its domestic investment spending. Using Chinese microdata, the researchers argue that cutbacks in government-guaranteed pensions and employment (as state owned enterprises were privatized) led to a rise in precautionary and retirement savings. Furthermore, China's one-child policy reduced family sizes, but also will create a higher dependency ratio as the number of retired relative to employed workers increases over time. They also compared saving behavior in China with saving behavior in the United States.

In Chapter 7, Pula and Peltonen examine South-South trade, focusing especially on Asia, noting that most South-South trade is intraregional trade in intermediate goods within "emerging" Asia. Using the Asian International Input-Output table, the economists determined the extent to which value added in emerging Asia depends on final demands in the United States, Japan and the EU (the G-3 countries). Looking at the 2000-2008 period, they conclude that G-3 final demand has had a declining impact on Asian value added; rather, Asian final demand is much more important. This suggests less "coupling" outside of Asia.

Chapter 8 by De La Cruz et al. examines the Mexican economy, looking at "vertical specialization" along supply chains. The paper provides a short history of the Maquiladora program (and the similar PITEX program). Using a Mexican input-output table, the authors estimate that average domestic value added in Mexican exports is around 30 percent. At the valued added low end are electronic goods, reflecting extensive imported components; automotive and clothing goods (among

others) are in the middle; and oil and other mining products at the high end of value added.

Berger and Nitsch in Chapter 9 look at trade imbalances within the EU, especially before and after the introduction of the euro. Using data from 1948 to 2008 for 15 (eventual) EU nations, the economists determined that euro-area countries saw increases in trade imbalances with other euro-area nations. They note in particular that Greece and other "small" economies ran high imbalances. The authors state that not having floating exchange rates increases the time needed to adjust to external accounts, and forces those adjustments through price changes.

Dooley (in Chapter 10) and Goldberg (in Chapter 11) both wrote "think pieces." Dooley notes that the 2008 crisis did not reduce economic activity as much in emerging markets compared to developed nations' markets. Within the EU, he states that Germany's trade surplus was "offset" by deficits in the EU's "peripheral" countries, but these latter countries (presumably including Greece) will be unable to finance these external deficits much into the future. Goldberg looks at the future of the US dollar, arguing that its "international role" should not change much in coming years.

The final (twelfth) chapter by Fujii looks at the "Penn effect," the argument that countries with high per capita incomes have high price levels. Fujii discusses problems with data sets, noting recent revisions to World Bank data. Using 1975-2009 panel data, he finds evidence of the Penn effect. The effect is more pronounced for OECD lands than for others. Controlling for nontradables, openness and exchange rate regimes does not eliminate the Penn effect, although countries with "rigid" exchange rate regimes appear to have higher price levels.

In all, this book is a very informative read, especially in light of current (2015) issues in the EU. The articles are very much empirical rather than theoretical pieces, using both econometric as well as input-output analyses. Several articles will be of interest to those studying the impacts of supply chains, as well as those examining regionalization. While the articles have little to say about Africa or Latin America (other than Mexico), they do collectively highlight issues in Asia, North America and Europe. I would encourage regional and international economists, and also economic geographers to look at this book.

MARK JELAVICH

Baker University (adjunct)

Keynes: Useful Economics for the World Economy. PETER TEMIN AND DAVID VINES. Cambridge, MA: The MIT Press, 2014. Pp. xiii, 117, $24.95.

We live in a world where John Maynard Keynes has largely influenced the education of policymakers and the content of the policies they crafted in the major economies. The premise by the end of the twentieth century became that to know Keynesian macroeconomic thought meant having an ability to devise the best policies to meet economic goals. In *Keynes: Useful Economics for the World Economy*, Peter Temin and David Vines use this aforementioned premise and a belief that Keynesian thought has been forgotten or neglected more recently in order to carve out an opportunity. An opportunity to teach the reader the basics of Keynesian thought, outline the intellectual insights of David Hume and Alfred Marshall that later informed Keynes, describe the episodic embrace and rejection of Keynes as the twentieth century proceeded, and apply Keynesian thought to economies after the 2008 financial crisis.

Their discussion begins with the statement that a country whose policymakers follow the tenets of Keynesian macroeconomic theory achieve economic growth. As proof of this they cite the fact that the United States economy from the late 1940s to the late 1960s and then again from the 1980s until 2008 represented twin golden ages of economic growth that were both fueled by Keynesian inspired policies. The economic problems experienced after the financial crisis have been caused, to some extent, by policymakers who may claim to follow Keynesian policies but contradict this by failing to consider the global consequences of their actions and the potential interactions among countries. Temin and Vines posit that the lackluster economic performance after the crisis could be resolved if policymakers remembered the global aspects of Keynesian macroeconomic thought.

The authors wrote this book not for economists but for policymakers who have some basic familiarity with economics. This text is organized into three parts. In the first three chapters the authors provide an overview of economic thought in a historical context by tracing back the intellectual heritage of Keynesian thought to David Hume and Alfred Marshall and the interactions (and frustrations) of Keynes first at Versailles at the conclusion of World War I and later with the Macmillan Committee after the 1929 stock market crash. This historical discussion highlights the revolutionary nature of the Keynesian thought in the early

twentieth century and emphasizes the global perspective that Keynes considered so important in economic theory. In the second part of the text covered by chapters 4-8 the reader is presented with the basics of Keynesian macroeconomic theory as developed by Keynes, the IS/LM Keynesian model developed by John Hicks, the Swan diagram developed by Trevor Swan, and the policy prescriptions suggested by Keynes in a post-war world. For a trained economist, these chapters provide insufficient detail to the otherwise complicated models and concepts they revel in. The authors explicitly state in the preface that this book is not intended for economists but instead for policymakers who can make the world better by understanding Keynesian theory. The text concludes in part three (chapters 9 and 10) with a discussion of the role of Keynes in the Bretton Woods agreement and the International Monetary Fund institutional structure. The intent of this discussion is that readers appreciate the influence Keynes had on the twentieth century and how the neglect of Keynes in policy has contributed to lackluster and unstable economic growth more recently.

For the non-economist, this text provides a valuable trifecta of history, economic theory, and policy. Although the professional economist reading the text may find the economic theory discussion frustratingly simplified, the historical discussion is something not typically encountered in macroeconomic theory textbooks and makes reading the text a worthwhile investment of the economist's time. Both the non-economist and economist however would have benefited from a significantly longer discussion of how policies and institutional reactions changed after the 2008 financial crisis. In just 10 pages the authors discuss the international paradox of thrift as it applied to economies in the closing decades of the twentieth century and during and immediately after the financial crisis. Consequently, frustration understandably emerges from the forward-thinking policymaker who has essentially worked through 90% of the text to be left with a rather short critique of policies adopted after the financial crisis and of what should be done in the future.

The text concludes with a hope that policymakers will be able to resolve international tensions more easily than was in the case in the interwar period by creating the conditions for sustainable economic prosperity in a global context. Yet, the reader with some basic understanding of economic performance before and after the crisis may be surprised of the lack of discussion in the text of the recent transition

from convergence to divergence in global growth rates. Leading up to the crisis, the major and minor economies in the world were largely in sync and growing strongly at roughly similar rates. Once the crisis occurred, the same major and minor economics declined together. This global business cycle and the synchronized movements of economics reflected, to some extent, the long-term vision of Keynes where global institutions coordinated growth and beggar-thy-neighbor policies were limited in their effect. Years after the crisis, a divergence has emerged with the United States growing three times faster than Germany and India growing annually at 8% while Russia faces the prospects of a deep recession. The return of country-wide business cycles signals a lack of coordination and an incentive for policy makers to treat trading partners with disdain if it means boosting domestic growth of Gross Domestic Product.

This transition from convergence to divergence in global growth rates bears mentioning when discussing the consequences of us all being Keynesians now. Is it the case that divergence in growth rates reflects policymakers forgetting Keynesian theory? And if the answer to that question is in the affirmative, doesn't the presence of this divergence pose problems for capital flows and international cooperation? Or is it the case that the return of divergence highlights the limits of Keynesian theory? And if the answer to that question is in the affirmative, could it be said that policymakers won't resolve problems by following Keynesian theory more closely but rather fuel the problems? Temin and Vines in this text artfully present one side of this debate. What is needed now is someone to take up the other side of this debate as effectively as Temin and Vines.

DR. TOM SCHEIDING

University of Hawai'i – West O'ahu

Making in America: From Innovation to Market. SUZANNE BERGER. Cambridge, MA.: MIT Press, 2013. Pp. xiv, 250. $24.95.

Suzanne Berger provides a detailed account of how innovation flows from ideas into production and the implications for the manufacturing sector in the United States and abroad. Her analysis is primarily based upon interviews and surveys of managers in U.S. manufacturing

establishments of all sizes across a variety of industries. She also included a select group of manufacturing firms from Europe and Asia. The interviews and surveys were conducted by MIT's Production in the Innovation Economy (PIE) research group. This group was composed of faculty across disciplines for the purpose of learning how manufacturing contributes to knowledge creation and economic growth.

One objective of the PIE research group was to determine why there has been slow productivity growth and industrial stagnation across many areas of U.S. manufacturing. The recent evolution of manufacturing has been influenced by structural change within the United States and competitive pressure stemming from globalization.

In the 1970s and 1980s, large, vertically integrated manufacturing firms in the United States often controlled all stages of production. This kept research and development, design, fabrication, prototyping, simulation, testing, demonstration, pilot production, packaging, test marketing, and full-scale commercialization activities in house. It was believed that complete control of each production stage improved efficiency and reduced risks associated with multiple suppliers. Manufacturing processes embody valuable knowledge that many firms do not want to lose to potential competitors. Patents may offer little protection if they reveal too many details. In many cases, trade secrets and tacit production knowledge can be more valuable than patent protection.

Over time, the benefits of vertical integration began to decline. Outside forces such as increased manufacturing capabilities from abroad (e.g. China) including lower wages, taxes and land prices began to entice U.S. firms to begin outsourcing some production activities.

Manufacturing firms began to restructure production processes and focus on activities where they had a comparative advantage. This led to greater specialization by identifying production activities that were most closely aligned with their "core competencies". Any remaining activities were spun-off or closed if they were not in the "core". Some residual activities not in the core were sold off and often became targets of mergers or takeovers by firms already engaged in similar activities. Essentially, there was a restructuring of firms moving towards buying up activities of other firms related to their core competencies and shedding activities that were related to the core competencies of acquiring firms. This restructuring was also a partial response to the demands imposed by the financial markets. Greater emphasis on "shareholder value" requires

large, integrated manufacturing firms to break-up. Many investors believe that conglomerate firms – those that are vertically and horizontally integrated – suffer from diseconomies of scale. Breaking them up should reduce long-run per-unit average costs, improving profitability and raise stock prices.

The restructuring of U.S. manufacturing eventually began to separate innovation from production activities. While this strategy can improve short-run efficiency and profitability, it can jeopardize future innovation. A significant amount of learning takes place on the factory floor as technicians bring problems back to engineers and designers. These face-to-face interactions can create tacit knowledge critical for future innovations and profits. Further, new jobs can be created when production stays close to R+D and design. New knowledge often leads to new and improved products enabling firms to enter new markets.

As the various stages of manufacturing have become fragmented and distributed across the domestic and global supply chain, many smaller and specialized firms have been created. The fragmentation of U.S. manufacturing has led to holes in the industrial ecosystem that is not conducive to supporting the innovation required for a dynamic economy.

Excluding regional economies such as the Silicon Valley, holes in the industrial ecosystem result from market failure or absence of complementary resources that diminish local capabilities. Smaller firms typically operate with only internally generated resources. Many smaller firms are isolated from complementary resources and have underexploited potential for expansion, job creation and profits. In some cases, smaller firms find it harder to get loans as local banks are absorbed by national basks with little understanding of local manufacturing.

Berger emphasizes that coordination failures and lack of public goods have contributed to the development of holes in the United States industrial ecosystem. Leadership is essential if the United States is to improve productivity and expand its industrial base. Leaders could be dynamic individuals or public authorities, emerge from industry, industry associations, trade schools, colleges and universities. Leaders must be able to see the connection between complementary resources and their impact on the innovative process. The role of leaders is to serve as a catalyst or convener. A convener can create new regional manufacturing capabilities and ultimately attract new industry and partners. The creation of public goods and shared resources is essential if the U.S. is to fill in the holes of the industrial ecosystem.

The PIE researchers emphasize that the United States has the resources to transform regional manufacturing economies into engines of growth. These resources need to be organized in a way that generates new knowledge and innovations that propel productivity and economic growth.

Knowledge spillovers occur when firms are in close proximity. Interdependencies among complementary activities, not narrow specialized clusters produce innovations. While specialization can improve short-run efficiency and profits, collaboration is essential for innovation and long-run regional and economic growth.

Berger provides interesting examples of innovation in specific manufacturing firms and how these firms have evolved in the United States and abroad. The value added of this book is to give readers a much richer picture of U.S. manufacturing than could be obtained from inside the walls of a university. This book would be an excellent complement to standard undergraduate or graduate texts in courses such as: industrial organization, regional economics, technological change, economic geography, economic growth and development, supply-chain management, organizational behavior and learning and strategic management.

JEFFREY G. WOODS
Indiana University-Purdue University Indianapolis

Index of Authors

Volume XLI, No. 2, 2015

Indexed by last name and first page of contribution

Index of Articles, by Author
Volume XLI (2015)

Index of Books Reviewed

Volume XLI (2015)

Plan Now to Attend the

Missouri Valley Economic Association

53rd Annual Meeting

At the Hyatt Regency St. Louis at the Arch
St. Louis, Missouri
October 27-29, 2016

For more information, contact:
Ken Brown, MVEA President
Missouri State University
Springfield, Missouri
KennethBrown@missouristate.edu