

# The Harvard College Economist

## Thank You

This edition of *The Harvard College Economist* would not have been possible without the help of a number of people. First, we would like to thank Professors Richard Caves and John McHale and Andrei Shleifer for their continued guidance. Professor McHale will be missed by his students as well as by *The Harvard College Economist* staff when he leaves Harvard for Queens College this coming fall.

Additionally, we would like to thank: Anita Mortimer, Patty Boudrot and the entire Economics department staff in Littauer, Susan Cooke, Kate Quick, Gail Oskin, Claudia Goldin, Caroline Hoxby, Richard Freeman, and Benjamin Friedman.

Thanks to Michael Ames and Puritan Press for taking on the task of printing our journal.

Finally, we are indebted to Andrew Lamb, an old friend who gave selflessly in his assistance with the layout of the journal.

### THE HARVARD COLLEGE ECONOMIST

*Editor-In-Chief*

Matthew Rosenberg

*President*

Adam Taub

*Secretary*

Brooke Wagner

*Treasurer*

Emily Oster

*Faculty Advisor*

Professor Richard Caves

*Faculty Advisor*

Professor John McHale

*Technical Consultant*

Andrew Lamb

*Copy Editor*

Megan Todd

*Contributors:* Gernot Wagner, M. Marit Rehavi, Mwashuma Nyatta, Radoslav Raykov, Radim Rimanek, David Witkin.

## Table of Contents

Editor's Note.....	4
An Interview with Secretary Lawrence Summers.....	5
An Interview with Professor Andrei Shleifer.....	9
Greening the Capital National Income Accounts by Gernot Wagner.....	16
Keeping Up with The Joneses by Radoslav Raykov.....	27
An Analysis of Returns to Risk Arbitrage in Europe by David Witkin.....	42
The Tinkerer's Tale by Mwashuma Nyatta.....	70
Gibrat's Law for US Cities: A Test by Radim Rimanek.....	80

*The Harvard College Economist*  
email: [hce@hcs.harvard.edu](mailto:hce@hcs.harvard.edu)  
fax: (617) 495-7730 c/o *The Harvard College Economist*

*Please visit us at [www.hcs.harvard.edu/~hce](http://www.hcs.harvard.edu/~hce).*

# The Harvard College Economist

## Editors' Note

Welcome to the latest issue of *The Harvard College Economist*. *The Harvard College Economist* has existed for years, but in 1999 the magazine disappeared without a trace. We are reviving the journal to offer a forum for undergraduates to publish research, to exchange ideas and to read superior economic writing.

Unlike other economics journals, however, *The Harvard College Economist* is written entirely by undergraduates. Submissions to this inaugural issue were by recommendation of faculty in the Economics Department. Most of the published papers were written for an economics course at Harvard.

What results is a compilation of fine economics papers that are accessible to all students at the undergraduate level.

In keeping with the tradition of *The Harvard College Economist* of old, this issue features two interviews with prominent economists. We were fortunate enough to speak with Professor Andrei Shleifer on market efficiency and with Harvard's new President and former Secretary of the Treasury Lawrence Summers on international financial crises. We thank them both for generously agreeing to be interviewed.

As we look toward our next issue, we plan to open submissions to all economics students. For students who are not graduating this spring, please consider submitting for next fall. Send all submissions to [hce@hcs.harvard.edu](mailto:hce@hcs.harvard.edu). For more information regarding details of submissions, please visit us at [www.hcs.harvard.edu/~hce](http://www.hcs.harvard.edu/~hce). Also, we encourage all students interested in any aspect of the publication to e-mail us.

We would like to thank all those who submitted articles to the journal. All of the papers were excellent, but unfortunately, we do not have space to publish more than a few papers.

Finally, our hope is that the journal becomes a staple of the Harvard undergraduate community. Please do not hesitate to offer suggestions for future issues.

Sincerely,

Matthew Rosenberg  
Editor-In-Chief

Adam Taub  
President

# On International Bailouts: An Interview with Secretary Lawrence Summers

**The Harvard College Economist:** *During the Mexican and Asian crises, the IMF demanded that effected countries raise interest rates and curb their budget deficits. Some have argued that this exacerbated the crises. Do you believe that pursuing austere macroeconomic policies in the midst of crisis is an appropriate policy response to a loss of investor confidence?*

**Secretary Lawrence Summers:** Battlefield medicine is never perfect, but I think that overall the IMF did a very good job in confronting the evolving Asian financial crisis. When there is large scale capital withdrawal from a country, there is always a very difficult tradeoff between, on the one hand, the imperative of restoring confidence in the currency and retaining capital within the country and, on the other, the imperative of providing liquidity to a strained financial system. The judgement made in the effected countries was a two way point: getting control of the collapsing currency was prerequisite to any economic success, and that judgement was

a broadly correct one. Whether with the benefit of hindsight the precise parameters of the policy response could have been managed differently is a different question. There was, as has been widely recognized, an initial impulse towards fiscal consolidation, which as the seriousness of the problem became clear was reversed with the IMF's support, and that was certainly appropriate.

**HCE:** *What do you think of the Bush administration's decision to be more distanced from the international financial system than the Clinton administration? Do you foresee any repercussions in the Bush team's opposition to bailouts and its stance, in general, toward the world financial system?*

**LS:** I, as a matter of policy, am not commenting on the specifics of the new administration's actions or my successor's actions. I have enormous positive regard for Paul O'Neill who I've known for many years. In general, I think it is very important to global financial stability that the United States

## The Harvard College Economist

remains engaged and prepared to take an activist posture when there are financial crises. Confidence is the other side of moral hazard. The experience of the late 1920s and early 1930s speak in a very strong way to the consequences for the global financial system of abandoning countries facing very substantial financial difficulties, and the length and extent of Latin America's decline following the 1982 Mexican debt crisis reinforces that message. It is very important that the United States be prepared to work with the international financial institution in support of financial responsibility globally.

**HCE:** *You have been reported as saying that the most recent business cycle has been different in character from its recent predecessors. Can you explain? What are the implications for the likely speed of recovery from this recent economic downturn?*

**LS:** I think that there is a characteristic pattern of post war American business cycles in which the economy accelerates, inflation picks up, demand rises, the Federal Reserve steps on the break, and eventually the economy slows down. That was what happened along with supply shocks in 1990, in 1982, in 1975,

in 1971, in 1958, and so forth. The pattern of this business cycle appears to be one that is more characteristic of the pre-World War II period where the business cycle was more of a credit cycle. By that I mean it was a phenomenon of opportunity, investment, increases in asset prices, greater investment, rising productivity, but ultimately excess capacity, declining asset prices, reduced investment and financial strain. That, I think, is the more characteristic pattern of pre-World War II business cycle and is somewhat the pattern that we are now facing, particularly in the technology sector. What does this mean for the future is very difficult to know. We certainly have more understanding of business cycle phenomena and stronger tools in terms of monetary and fiscal policy than we did prior to the Second World War. But there are also risks that we have to recognize in the current situation.

**HCE:** *With its low national saving rate and the high current deficit, some commentators worry that the US is headed toward a "hard landing." If an emerging market economy had a current account deficit of 4.5 percent of GDP, we would probably be advising them to take steps to bring national saving in line with national*

*investment. Is the US also vulnerable to a turnaround in capital flows and a sharp depreciation of the dollar given the size of its deficit? Should US policy makers heed the advice they routinely given to the governments of emerging economies?*

**LS:** There's no question that current account deficit has to be a concern both now and in prospect. The current account deficit represents the difference between national savings and national investment, and it would be much better for national savings to increase than for investment to decrease. That's why fiscal discipline that increases public savings is so important, and that's why finding ways to induce Americans to save more at the personal level is important as well. It's also important to a healthy adjustment process that we expand exports to the maximum extent possible rather than compressing imports. That's why the United States has traditionally laid such emphasis on the importance of a rapidly growing, increasingly open, global economy.

**HCE:** *The IMF gives loans to countries subject to strict conditions and is often accused of micromanaging borrower countries, as in the much-*

*maligned case of the Indonesian clove monopoly. Do you think that it is worthwhile to demand such structural reforms in the midst of a crisis?*

**LS:** I think it's very difficult to generalize, as I've said, battlefield medicine is never perfect, but often structural issues can take on great significance symbolically or can take on great significance because of their implications for investor confidence, and they can take on real significance for how wisely and well limited resources provided to governments are used. For all those reasons, there will be occasions when structural issues do have to figure in IMF programs. In general, I made clear during the time that I was Secretary of the Treasury that I believe it is appropriate for conditionality to be increasingly focused and that there has been some tendency over the years for programs to be Christmas trees.

**HCE:** *The international community had been advising emerging market economies to liberalize their international capital accounts. Subsequent sharp reversals in capital flows have had a devastating impact on a growing list of countries, most recently on Turkey. As James Tobin famously*

## The Harvard College Economist

*quipped, "It takes a lot of Harberger triangles to fill one Okun gap" [meaning it takes a lot of market distortions to match the losses experienced in a recession]. Given the deep recessions we have observed in the wake of the emerging market financial crises, has your opinion on the wisdom of liberal international capital markets for emerging market economies changed?*

**LS:** Ironically, if you look at most of the emerging market crises, there were important policy non-neutralities in favor of short-term capital. In Thailand there was an offshore banking facility directed just at attracting short-term flows. South Korea's capital flows were directed at long-term capital flows, but not short-term capital flows. In Brazil and in Russia, government debt was channeled towards the objective of being an instrument that was attractive to hedge funds. In Mexico, Tesebonos were designed to represent a shortening and indexing of the liability structure, so just as in the environmental area, it's appropriate before moving to put taxes on pollution to avoid subsidies to pollution. My judgement in this area is that there

is an enormous amount that can be gained by reducing policy distortions in favor of short-term capital relative to long-term capital.

**HCE:** *Having been both an academic and a politician, have you found it difficult to transition between the two? Did you ever feel you were compromising your academic views when you were acting as a politician?*

**LS:** I never ran for office, so I was never a politician. I certainly did act in a political environment and part of acting in a political environment is reaching compromise, and, in that sense, I did not have to compromise on the content of my research, and I did have to compromise on policy actions that I was involved in, but I didn't think of compromise as representing a loss of integrity. Yes, the academic and the political environments are different. One learns, if one's going to succeed, to adapt to changing conditions, but I found my time in the Harvard Economics Department enormously satisfying. In a somewhat different way, I found my time in Washington very satisfying, and I'm very much looking forward to my return to Harvard.

## **On Market Efficiency: An Interview with Professor Andrei Shleifer**

**The Harvard College Economist:** *You are well-known for your view that the market is on the whole inefficient. What prevents arbitrageurs from entering the market and taking advantage of any mispricings?*

**Professor Andrei Shleifer:** In thinking about arbitrage, it's very important to recognize that it is an economic activity that has its rewards but it also has its risks. It is also very important to realize that arbitrageurs are investing other people's money. In principle, there are returns from market inefficiencies, but they require a fair amount of risk bearing and a fair amount of patience. There are people out there who manage to take advantage of the anomalies, but they tend not to be the standard institutional investors because the standard institutional investors do not have the patience or the risk bearing capacity to do this, and this is basically why the anomalies exist.

**HCE:** *Along those lines, Burt Malkiel, author of *Random Walk Down Wall Street*, is a major proponent of the efficient market hypothesis. He says*

*that not even institutional money managers are beating the market. In a recent exchange between you and Malkiel in the *Wall Street Journal*, you argued that this was not strong enough evidence for the market to be efficient. Why isn't it?*

**AS:** The question about institutional money managers is whether or not they are operating subject to constraints that prevent them from taking aggressive positions against market inefficiencies. To think about institutional money managers, you must recognize several things. Number one: they are evaluated by the investors who give them the funds to manage at a very high frequency, either once a quarter or once a year. The sponsors pull their money away from the managers who underperform the market by a substantial margin. That makes institutional money managers extraordinarily risk averse and cautious investors when it comes to taking positions against inefficiencies. Number two: inefficiencies often take a long time to work out. Once you take into consideration the institutional context of the professional money management activity, it's not something

## The Harvard College Economist

surprising that you don't find the best investors among the institutional investors.

**HCE:** *So if we accept your conclusion that the market is inefficient, is there anything that could occur, either through new players entering the market or additional regulations, to make the market efficient? And if so, what would be the advantages and disadvantages of having this efficient market?*

**AS:** This is a loaded question so let me take it in parts. With respect to new players entering the market, I think that over time as trading has become cheaper and as knowledge about valuations of securities has become more widely spread, markets have become more efficient. Another side of this is that as trading has become cheaper through the internet and other sources, there have been a lot of new very unsophisticated investors entering the market as well, who think they can never lose money in stocks. So overall I think the market is becoming more efficient through increased participation, especially by arbitrageurs with large pools of capital, but that doesn't mean that it has become completely efficient.

The question of regulation is very complicated. Overall, I am not an advocate of the position that government regulation of trading in securities markets has been a strong force toward improved efficiency. As I said in my Wall Street Journal article, government regulation often makes things worse. The down tick rule, where people are not allowed to sell short on the down tick, has made arbitrage more expensive, for example. So one always has to be very careful recommending government regulations, because more often than not they make things worse rather than better.

**HCE:** *And would we benefit from an efficient market?*

**AS:** I think overall the answer is yes. One could argue that there were significant benefits from the price bubble in internet stocks, because lots of people have started companies, and there was a tremendous amount of innovation associated with the stock market bubble, which might not have occurred were it not for the bubble and were it not for the incentives of all these potential entrepreneurs to go public at a profit. There are other instances in history in which stock price bubbles have been associated with rapid economic expansion.

sions and rapid increases in technological innovation, for instance, railroads or electricity, so there are some potential benefits of bubbles. But in order to believe in these benefits, you really need to believe that the rate of technological progress that would have occurred without them is below what is socially desirable. Overall, I think the world is probably better with more efficient markets than with less efficient markets.

**HCE:** *The story of Long-term Capital Management provides a clear example of an asset management firm that lost everything. They bet on spreads that widened in the short run, and they were unable to cover margin calls. Does such a visible example like LTCM deter other arbitrageurs from entering the market?*

**AS:** The example of LTCM has had a chilling effect on the market for hedge funds and arbitrage funds. But it's not because it discouraged arbitrageurs, but rather because it discouraged ultimate investors from trusting their money to arbitrage funds. A lot of investors have become scared following the LTCM experience of participating in arbitrage activity or entrusting their money to arbitrage funds. There probably has been

a two year lull in the market. At this moment, a lot of this money is coming back into the market.

**HCE:** *Over the past year, the market has been very volatile. There's been an incredible drop in the market. Were we in a price bubble? If so, are we still in one?*

**AS:** The evidence over the last two or three years is quite compelling on the proposition that we were in a price bubble, at least in the NASDAQ, an internet price bubble. The valuations that were achieved by some of the high technology stocks in 1999 and the beginning of 2000 were certainly virtually impossible to account for in any kind of a rational valuation model. Of course there were people who said this is all rational and we're living in a new world, but it's important to recall that every time there's a price bubble there are some people who say that it's all rational and we're living in a new world. The collapse of the internet and the NASDAQ prices has been very clear corroboration of the proposition that we were in a price bubble. If you look at the valuations of equities today, they are not nearly as outrageously expensive as they were at the peak of the bubble in early 2000,

## The Harvard College Economist

but they are still expensive. So are we still in a bubble? Probably not. But is there any reason to be optimistic about equities in terms of the next three to five years? I think that the answer to that is also no.

**HCE:** *Is it possible for the prices to fully adjust without the economy falling into a deep recession?*

**AS:** We have certainly seen instances in the United States and in other countries of stock prices adjusting quite considerably without the economy falling into a recession. A good piece of evidence we have seen was 1987, when the market experienced the largest crash in history with prices falling by over 20 percent in one day with there being no recession to follow it. It's very hard to say whether the stock prices falling and trillions of dollars of wealth being erased is going to be followed by continued good economic performance. I don't think we know that yet. But there are certainly instances of significant reductions in stock prices without a recession.

**HCE:** *I'd like to ask you briefly about value investing. In valuing a company, analysts often calculate a company's fundamental value. Is it*

*possible to truly calculate this value? And if it is, to what degree does the fundamental value affect the price of the security?*

**AS:** One cannot calculate fundamental value exactly, because in order to calculate fundamental value or even expected fundamental value one needs to make forecasts for earnings and dividends of the company as well as know the appropriate discount rates. So the calculation of fundamental values inevitably involves guesses or estimates of future earnings and appropriate discount rates. What one can do, however, is to try to relate this forecast to existing valuations and to ask the question: are existing valuations plausible or reasonable given the likely path of future earnings of the company? And one can look at some securities and say, "Look, it is extremely unlikely given what we know from history that the current valuation is reasonable given the forecast of future earnings and the estimate of fundamental value." So one of the ideas of value investing is precisely to identify securities where valuations are too low relative to some reasonable estimates of fundamental value.

**HCE:** *According to data, value stocks*

*vastly outperform glamour stocks. How much of this phenomenon can be explained by psychological biases? And if it can't, why don't more people practice value investing?*

**AS:** A good chunk of undervaluation of value stocks probably has to do with the fact that market participants do not find them attractive or interesting. The companies whose stock market prices turn them into value stocks tend to be companies that have been consistently poor performers. One probably cannot tell one's friends proudly about having invested in them. These kinds of psychological factors—avoidance of bad companies, avoidance of companies with long records of poor performance—are probably a fundamental psychological factor behind the attractiveness of value stocks as long term investments. That is to say, because they are beaten down so much they represent good investment value. Now, why don't more people invest more in value stocks? As I've mentioned earlier, from the point of view of institutional investors, extreme value strategies are unattractive because a) they require taking some risk of things going badly along the path and b) they may require a long time horizon to realize attractive returns.

And so, we see that even though many institutional investors call themselves value investors, at best they have a value bias, which is to say a small tilt towards value stocks as opposed to a significant commitment towards value. This is probably one of the reasons why value has continued to be a good investment strategy, namely that institutional investors have generally avoided taking significant value positions.

**HCE:** *If people realize the effectiveness of a value investing strategy, would the excess returns disappear the same way that phenomenon such as the January effect disappeared once it was discovered and became well-known?*

**AS:** If enough people realize that the value strategy works for them, and if they could tolerate the risks, and if they could have a long enough horizon, obviously value strategies are going to become less attractive. This is definitely correct. On the other hand, value strategies have been publicized and known to the investing public probably for about eighty years, since Graham and Dodd started writing about investing in stocks with high dividend yields. And despite the fact that we've seen an

## The Harvard College Economist

eighty year history of good performance of value strategies and eighty years of statistical data pointing to the attractiveness of value strategies, we have not seen massive convergence on value investing. It is probably more accurate to say that we see some trend toward appreciating the wisdom of these strategies over time, but it is certainly not significant enough to eliminate the advantages, at least it hasn't been significant enough to eliminate the advantages so far.

**HCE:** *Do you believe that it is possible to time the market?*

**AS:** When you ask whether it is possible to time the market, there are two very separate versions of it. If the question is: is it possible to predict what the market will do in the next six months or in the next year?, the answer is almost certainly that it is very, very difficult, nearly impossible. If the question is: is it possible to "time the market" in the long run?, I think the answer is closer to yes. For example, there was virtually universal agreement among all people who took statistics and history seriously in the years 99-00 that the expected future return on the market was going to be very, very low. Whether you believe

there was a fundamental reason for stock market values to be as high as they were in early 2000, or whether you thought that there was a bubble, I can't think of anybody who expected high stock market returns at that period of time. Timing the market in the sense of avoiding heavy exposure to stocks at that time would have emerged as an attractive strategy. That doesn't mean that people could anticipate a sixty percent decline in the value of NASDAQ over the course of one year, but I think that one could time the market at that time in the sense of knowing that stocks were not a good deal. Timing the market starting from the situation as we see today is much more difficult because stocks are quite expensive, but they are certainly not insanely expensive as they were a year and half ago.

**HCE:** *So, then what is your long run forecast?*

**AS:** I think the long run forecast is, a ten year forecast, continues to be low returns by historical standards.

**HCE:** *Moving back to the behavioral aspect of the stock market, people talk about irrational exuberance. If it weren't for this idea of irrational*

*exuberance, that individual investors feel they know better than the market, would there be any market for stocks at all?*

**AS:** Of course. We have seen financial markets in the United States and other countries operate for long periods of time quite successfully without having extreme stock market bubbles. The U.S. financial markets have seen the extreme valuations that you saw in the year 2000 since the late 1960s. This is a thirty year period of perfectly good functioning of the U.S. financial market without such bubbles. And before the 1960s, markets saw a bubble like that only in 1929. I think that it is true that certain things only occur in periods of extremely high valuations, but that doesn't mean that markets can't exist, that companies can't go public, or that other things don't take place.

**HCE:** *If you had to single out the most common mistakes that investors make when they play the market, what would they be? If you informed investors of their mistakes, would there be a turnaround?*

**AS:** Particularly in the last five or ten years, we've seen an emerging group of investors who've been convinced by the idea that stocks cannot go down, or even if they do go down, in the long run they're a fantastic investment. This is a very big mistake, especially when one starts looking from a level of very high valuations. Perhaps the best lesson that an investor can learn is that stocks can go down, and they do go down a lot, especially if we start from the world such as the one we're in today. It is not at all clear that even over a ten year horizon stocks are such a good investment.

# The Political Economy of Greening the National Income Accounts

*Gernot Wagner\**

## Abstract

National income figures currently in use are highly misleading. With some notable exceptions such as owner-occupied housing and government services, income calculations are limited to market activities. However, our economy reaches far beyond its traditional market-borders. Some negative economic effects on our environment are so large that they pose serious threats to our societal and economic well-being. Green accounting can significantly aid in policy decisions by assigning values to both the depreciation of natural resources and ecological services.

Because of political disagreement, however, official green accounting efforts in the United States have been on hold for the past half a decade. Whether and how to use the actual accounts in policy decisions will undoubtedly be a political issue, but the debate about creating the numbers in the first place should be moved from the political arena back to the Bureau of Economic Analysis (BEA) and the economics community.

## I. Introduction

While the United Nations and many other countries are actively exploring green accounting, official efforts in the United States have come to a virtual halt. In 1994, Congress commissioned a high-level study reviewing the Bureau of Economic Analysis's (BEA) work on its Integrated System of Environmental and Economic Accounts (ISEEA), but since its publication in 1999 has ignored the results and continued to bar BEA from

its efforts. The debate surrounding green accounting in the U.S. should be freed from the partisan struggle in Washington and moved to the hands of the scientific community. Green accounting efforts should seriously focus on including ecological services as well as mere resource depreciation, but in line with the recommendations of Nordhaus and Kokkelenberg (1999), the existing national income and product accounts framework should be used as a point of

\*Undergraduate Researcher, Program for Technology and Economic Policy, John F. Kennedy School of Government, 79 JFK Street, Cambridge, MA 02138, gwagner@fas.harvard.edu. I would like to thank Dale W. Jorgenson for guidance and support throughout my research and helpful comments on this essay. Darby Jack, Sheila Jasanoff, John B. Loomis, and Robert N. Stavins also provided helpful comments on an earlier draft. © AERE Newsletter, May 2001. Published with the permission of the Association of Environmental and Resource Economists.

departure, and the creation of official accounts should employ an incremental approach, giving first priority to areas where the necessary research is available.

## **II. History of Green Accounting in the US**

In 1989, Repetto et al. (1989) drew considerable attention to the shortcomings of economic indicators in a report entitled *Wasting Assets*. Using Indonesia as a case study, it concluded that the country's reported economic growth throughout the 1970s and 1980s would be cut in half if GDP calculations were modified to a so-called "Net" Domestic Product, taking timber, oil, and soil depletion into account. This World Resource Institute study was by no means the first to point out major shortcomings in national income measurements, but it sparked a considerable debate about "green accounting"—integrating environmental and economic accounts.<sup>1</sup> At that time, the UN Statistical Office was working in cooperation with several other international institutes and expert groups to amend its System of National Accounts (SNA) for publication in 1993, the first major revision in almost three decades. The 1993 revision did not include guidelines for an

integrated system of environmental and economic accounts, but in the appendix suggested the use of so-called "satellite accounts" for environmental statistics. Such a system leaves the core economic accounts untouched while providing some environmental information in a compatible, yet not fully integrated, set of supplementary statistics (SNA 1993: chapter XXI). In addition, the UN (1993) published a handbook on the system of integrated environmental and economic accounting (SEEA), with the goal of guiding national accountants in their efforts to create physical and monetary environmental statistics.

BEA is implementing the recommendations presented in the SNA with some modifications in its National Income and Product Accounts (NIPA). Even before the publication of the UN handbook, it had started to compile its own version of Integrated Environmental and Economic Satellite Accounts (IEESA) in 1992, following a recommendation of the Council of Economic Advisers under President Bush.<sup>2</sup> Two years later, BEA (1994) presented initial findings in its monthly *Survey of Current Business*. The issue included an overview of its efforts, as well as guidelines for the future and preliminary estimates for mineral resources. With the

# The Harvard College Economist

1995 Appropriations Bill, however, the 103<sup>rd</sup> Congress stopped BEA's work on all environmental accounts. In response to this order, the Department of Commerce, of which BEA is a part, asked the National Academy of Sciences to review the current state of environmental accounting in the US (US Congress, 1994). In 1999, a commission of leading environmental economists and national income statisticians published *Nature's Numbers*, which analyzed the state of green accounting in the US and prospects for future work (Nordhaus and Kokkelenberg, 1999). The study concluded that extensive research is still needed for developing a consistent set of accounts, but that "the rationale for augmented accounts is solidly grounded in mainstream economic analysis" and that BEA's work on IEESA is the best place for this research. So far, however, Congress has given BEA no mandate to continue its work on environmental accounts.

### III. Politics of Green Accounting

Throughout the last decade, the issue of green accounting has been highly polarized. The political debate splits along party lines: generally speaking, Democrats have favored environmental

accounting, while Republicans have opposed it. In *Earth in the Balance*, Vice President Al Gore (1992) called for a full integration of environmental and economic accounts, even though he acknowledged that national accountants did not believe this was feasible given the state of knowledge at the time.

President Clinton also addressed the importance of environmental accounting in his 1993 Earth Day address. Following the publication of the National Academy of Sciences study *Nature's Numbers*, the 2000 Economic Report of the President included an entire chapter on "Making Markets work for the Environment," which contains a discussion entitled "Taking Account of the Environment." In this section, the President's economic advisors stressed the lack of "a coherent framework for accounting for environmental quality and natural resource use in tandem with market economic activity" and referred to the recommendations presented in *Nature's Numbers*. They urged the creation of supplementary accounts for "assets and production activities associated with natural resources and the environment." The conclusion specifically states that integrated environmental and economic accounts do not merely contribute to the understanding of societal wel-

fare or human development but to the nation's economic development ("Economic Report of the President," 2000). Consistent with the platform of the Clinton administration, congressional Democrats have also been in favor of green accounting.

One notable exception in this support, however, was the decision of the 103<sup>rd</sup> Congress to stop funding for BEA's green accounting activities. Alan B. Mollohan, a Democratic House Representative from West Virginia and a longstanding member of the Appropriations Subcommittee responsible for allocating funds to the departments of Commerce, Justice and State, sponsored the bill (US Congress, 1994). This is not surprising, considering that he represented a district in West Virginia's coal country. The National Mining Association indicated it did not actively lobby on the issue, but it did express support of Representative Mollohan's action.<sup>3</sup> Mining corporations are afraid of being marginalized by a reform of the SNA and that instead of producing tangible wealth, their activities would be viewed as depleting one form of (natural) wealth and creating other forms of (human-made) wealth—a potential blow to their public image. Despite calls from economists and national income accountants

negating this claim, it was used as one of the arguments in support of Congress's decision to stop BEA's activities, which overshadowed the fact that the stop had its scientific merits. Similarly to the development of NIPA in the early parts of the last century, green accounting is a highly contentious field with many methodological problems. Moving the discussion from BEA to an independent scientific panel was an important step to reexamine the underlying issues. However, now that the findings have been published, Congress should again give BEA the opportunity to continue its research.

Congress commissioned the National Academy of Sciences study when stopping BEA's efforts, but the publication of *Nature's Numbers* has been ignored.<sup>4</sup> This is not surprising to political scientists, who would be hard pressed to cite an example of when an NAS report triggered direct political action. Members of Congress answer to their constituents, not to scientific panels, unless it is politically expedient. The main ideological argument on the side of Republican lawmakers is that integrating environmental statistics into economic data would give what some might consider 'environmentalist ideas' too much weight in economic decision

## The Harvard College Economist

making. Instead of being based purely on intangible values and quality of life factors that are virtually impossible to measure, an integration of environmental and economic accounts would give these ideas real dollar values, providing quantifiable data and therefore rendering them comparable to other policy decisions.

Ironically, Republicans are in favor of benefit cost analyses while Democrats generally oppose it. However, this division is reversed when it comes to taking account of ecological services. One example is the debate of mechanical treatment versus controlled burning in forest management.<sup>5</sup> Republicans, backed by the forest industry, favor mechanical treatment, while Democrats, catering to greens, prefer controlled burnings. However, the two parties support their positions with entirely different sets of facts. The Republican arguments for mechanical treatment are largely backed by hard economic data such as the number of jobs created in the logging industry. Democratic arguments for controlled burns, however, are to a large extent based on “soft” data such as increased ecological value. Assigning a dollar value to the ecological services a healthy forest provides, would put the argument for burn-

ing on equal footing with the economic arguments favored by Republicans. It is not surprising that the GOP would therefore oppose any such attempts to integrate environmental and economic accounts.

The debate of whether or not to go ahead with green accounting in the US, however, should rise above party politics. As Nordhaus and Kokkelenberg (1999) concludes, “a set of comprehensive accounts would illuminate a wide variety of issues concerning the economic state of the nation” (158). Many theoretical and methodological issues have yet to be resolved, but this should be reason for an expansive research agenda, not for a stop of BEA’s efforts.

The public debate in green accounting is currently dominated by the emerging field of ecological economics. Costanza et al. (1997a) attempted to evaluate the entire globe’s environmental services drew considerable public attention. There is definitely some merit in this work for the environmental community, but such numbers have very limited value for economic policy in the US.<sup>6</sup> Rather than estimating total “Green GDP,” analysts who aim to influence policy should focus on specific environmental values in the form of satellite ac-

counts. Research efforts should focus on smaller aspects of green accounting that correspond to clear policy questions. The debate surrounding prescribed burning versus mechanical treatment, for instance, could be considerably improved by means of U.S. forest satellite accounts. Similarly, such accounts could aid in the policy analysis of the Forest Service's Roadless Area Conservation initiative.

#### **IV. Application to Forestry**

*Nature's Numbers* also identified the refinement of timber value estimates as a next logical step in the implementation of the IEESA, and offered concrete suggestions. In its previous estimates, BEA used a shortcut approach similar to Repetto *et al.*'s method in *Wasting Assets* which is theoretically incorrect and, in practice, can vastly overestimate actual timber values.<sup>7</sup> Nordhaus *et al.* propose an alternative method for timber evaluation based on an unpublished manuscript which has been further developed by Jeffrey Vincent (1999). *Nature's Numbers* erroneously states that the present discounted value (PDV) method, which calculates the current value  $V(t)$  as the sum of discounted future income streams, harvest rate  $q(t)$  times stump-

age value  $p_s(t)$ , to yield

$$V(t) = \sum_{t=0}^T \frac{p_s(t)q(t)}{(1+r)^t}$$

is only theoretically correct for the case of "timber mining" of old-growth forests (Nordhaus and Kokkelenberg, 1999, 137–8.). The study then suggests the use of Vincent's method for managed second-growth forests. However, Vincent's approach is derived from this PDV method. Assuming  $T$  equal infinity, the PDV method correctly depicts the timber-value for any forest type—regardless of whether it is an old-growth or managed second-growth forest. Vincent's method still has the significant advantage that under certain conditions, it is easier to use current data, rather than future projections, which are required for the PDV approach. Nevertheless, the quality of physical forest data in the U.S. does not allow the Vincent method to be used to calculate precise results due to its particular sensitivity to fluctuations in forest area estimates. It is easier to obtain future projections of quantity and price paths than accurate current area data. Considering these caveats, it is relatively straightforward to calculate a theoretically correct set of timber values using

# The Harvard College Economist

the standard PDV approach in conjunction with projections based on already existing physical timber models (Wagner 2001).

## V. Accounting for Ecological Services

Timber-values, however, are only part of the picture. Non-timber values arguably constitute a much larger fraction of the total forest value. Considering the functions of watershed protection, prevention of soil erosion and the provision of wildlife habitats, as well as use and existence values, estimates of non-timber values are of crucial importance for policy evaluations, as in the case of prescribed burnings versus mechanical treatment, or the Roadless Area Conservation initiative. While the theory for including timber values is relatively well developed, the rationale behind including other ecological values has been less investigated. Even the question of whether or not to account for these services is a contentious issue. The ecological economics argument is based on the idea that we should focus on total societal welfare instead of economic welfare, and that human society is an integral part of the natural world.<sup>8</sup> This criticism of neo-classical economics has its validity, but the answers provided are

still inherently flawed and impractical. Our national income accounts have served us well in economic policy decisions, and currently practical solutions should focus on them as a point of departure.

In particular, one ought to start with the definition of what it is we are trying to measure. Martin Weitzman accomplished this for the case of the depreciation of natural resources in 1976.<sup>9</sup> Weitzman's paper lends theoretical credibility to the creation of timber accounts, but it does not do the same for comprehensive forest accounts. A theoretical argument would have to follow the notion of expanding the production function itself to include ecological services and provide a credible relationship between timber extraction and decreases in services the forest is able to provide.<sup>10</sup> Aggregated to a global scale, this calculation would evoke the same misleading notion criticized in Costanza et al.'s *Nature* article. This fallacy, however, can be avoided by applying the concept to a considerably smaller level, in this case, U.S. forests.

In contrast to timber accounts, which are theoretically well grounded, comprehensive forest satellite accounts would involve large uncertainties inherent in non-market valuation techniques.

Nevertheless, attempts are being made to overcome these barriers.<sup>11</sup> In this regard, green accounting can also have a positive impact on the traditional non-market evaluation literature since it uses the same valuation methods.

## **VI. Conclusion**

*Nature's Numbers* makes clear that timber accounts illustrate an area appropriate for immediate implementation, even though the specific procedure cited should be questioned. It would be premature to fully integrate ecological services into a comprehensive set of forest accounts, but more research in this area would undoubtedly be of high value to both green accounting efforts and, to a lesser extent, also to the non-market valuation literature. There are clear economic policy arguments for integrating environmental and economic accounts. The fact that some tough theoretical questions must still be addressed should not deter from tackling the issue, but should rather act as a call for more funding and research. Whether and how to use the actual accounts in policy decisions will undoubtedly be a political issue, but the debate about creating the numbers in the first place should be moved from the political arena back to BEA and the economics community.

Instructing BEA to resume work on a domestic green accounting system, in accordance with the National Academy recommendations, would be a first necessary step to do so.

## **Endnotes**

<sup>1</sup> For earlier critiques of the national income accounts, see Nordhaus and Tobin (1972) and Eisner (1988), among others.

<sup>2</sup> Earlier calls for environmental accounting in the U.S. reach back even further. The Ford Administration was the first to call for environmental accounting to track capital investment expenditures on pollution abatement, an initiative supported by President Carter (Nordhaus and Kokkelenberg (1999), 154).

<sup>3</sup> According to Connie Holmes, Vice President of Policy of the National Mining Association, formerly known as the National Coal Association (phone conversation on January 30, 2001).

<sup>4</sup> After the 103<sup>rd</sup> Congress halted BEA's work, every successive house report to the annual Appropriations Bill included a passage barring BEA from its work on the IEESA.

<sup>5</sup> See Bennett (2000) and Udall (2000) for a more complete discussion of this issue.

# The Harvard College Economist

<sup>6</sup> See Bockstael et al. (2000) for a comprehensive critique of Costanza et al.'s work, as well as Daily et al. (1999) for further explorations of the underlying issues.

<sup>7</sup> Under idealized conditions of a single, managed forest stand in perfect rotation sustaining its annual harvest at a constant rate into perpetuity, the Repetto method underestimates the actual timber value, since it ignores future timber growth and undervalues all tree stands up to the optimal rotation age. Since in most countries only a fraction of the total forest area is optimally managed and old-growth forests are considerably overvalued, in practice, the Repetto method overestimates the actual timber value.

<sup>8</sup> Very accessible overviews of this issue can be found in *You can't eat GNP: economics as if ecology mattered* by Eric Davidson (2000) as well as in Jane Jacobs (2000). Costanza et al. (1997b) contains a comprehensive discussion of the ecological economics perspective.

<sup>9</sup> Weitzman (1976) proves that including the net-depletion of subsoil assets in a comprehensive Green NNP measure sets it equal to the current sustainability equivalent of production, a notion which could equally well be

applied to the depreciation of timber.

<sup>10</sup> See Heal (1998) for a comprehensive treatment of this subject.

<sup>11</sup> Compare Adger et al. (1995) who created comprehensive forest accounts for Mexico.

## References

- Adger, W. Neil et al. 1995. "Total economic value of forests in Mexico." *Ambio* 24 (5), August.
- Bennett, R. S. 2000. "Q: Should Congress Halt Commercial Logging in the National Forests?; No: If Timber Harvesting is Permitted, there Will be Less Need for Controlled Burns." News World Communications, Inc. 26 June.
- Bockstael, Nancy E. et al. 2000. "On Measuring Economic Values for Nature." *Environmental Science and Technology* 34 (8): 1384–1389.
- Bureau of Economic Analysis. 1994. "Integrated Economic and Environmental Satellite Accounts." *Survey of Current Business* April.
- Costanza, Robert et al. 1997a. "The value of the world's ecosystem services and natural capital." *Nature* 387, 15 May: 253–260.
- Costanza, Robert et al. 1997b. *An introduction to ecological econom-*

- ics. International Society for Ecological Economics. Boca Raton, Fla.: St. Lucie Press: St. Lucie Press.
- Daily, Gretchen C. et al. 1999. "The value of nature and the nature of value." Beijer International Institute of Ecological Economics, Beijer Discussion Paper Series No. 126. "Economic Report of the President." 2000. Washington: U.S. Government Printing Office.
- Eisner, Robert. 1988. "Extended accounts for national income and product." *Journal of Economic Literature*. (December) 26: 1611–1694.
- Heal, Geoffrey. 1998. *Valuing the Future: Economic Theory and Sustainability*. Columbia University Press.
- Gore, Al. 1992. *Earth in the balance*. London: Earthscan Publications Ltd.
- Nordhaus, William D. and Edward C. Kokkelenberg (Eds.). 1999. *Nature's numbers: expanding the national economic accounts to include the environment*. Washington, D.C.: National Academy Press.
- Nordhaus, William D. and James Tobin. 1972. "Is growth obsolete?" *Economic Growth*, 50<sup>th</sup> anniversary colloquium V. New York: Columbia University Press for the National Bureau of Economic Research.
- Repetto, Robert et al. 1989. *Wasting Assets: natural resources in the national income accounts*. Washington: World Resource Institute.
- System of National Accounts 1993* (SNA). 1993. Commission of the European Communities (Eurostat), International Monetary Fund (IMF), Organization of Economic Cooperation and Development (OECD), United Nations (U.N.), and World Bank. Brussels/Luxembourg, New York, Paris, Washington DC.
- Vincent, Jeffrey R. 1999. "Net Accumulation of Timber Resources." *Review of Income and Wealth Series* 45, Number 2: 251-262.
- Wagner, Gernot. 2001. "U.S. Timber Accounts, 1957–1997." Working Paper, Environmental Science and Public Policy 91r. Cambridge, Massachusetts: Harvard University. January 23.
- Weitzman, Martin L. 1976. "On the Welfare Significance of National Product in a Dynamic Economy." *Quarterly Journal of Economics*, 90 (1), February: 156–162.

## The Harvard College Economist

Udall, S. L. 2000. "Let's Begin to Manage Our Forests." *The Arizona Republic*. 6 July.

U.N. 1993. *Integrated Environmental and Economic Accounting*.

U.N. New York, NY

U.S. Congress. 1994. House Report Accompanying HR 4603, U.S. Department of Commerce, FY 1995, Public Law 103-317. Washington, D.C.

# A TWO-PLAYER GAME THEORETICAL MODEL OF INTERDEPENDENT CONSUMER CHOICE: KEEPING UP WITH THE JONESES

*Radoslav Raykov*

## Abstract

This paper proposes a theoretical model of a relatively simple social phenomenon: keeping up with the Joneses. Starting with an interdependent utility function that allows for social influences on consumer behavior such as snob and Veblen effects, we proceed to model consumer choice as a two-player game similar to the Cournot duopoly model in an intertemporal setting. The two main questions considered are the existence and the efficiency of a Nash equilibrium. We prove that under augmented Cobb-Douglas consumer preferences, there exists a single Nash equilibrium, and furthermore, that this equilibrium is inefficient. The economic implications of the Nash outcome include a particular form of saving myopia, which is evinced in higher spending and lower saving levels than predicted by standard economic theory, and a reduced saving rate of the poorer player under conditions of income inequality. The main argument is that this allocation of resources is inefficient, but not inconsistent with observed empirical patterns.

## I. Introduction

This paper attempts to analyze a game between two consumers in a modern industrial economy, each of whom has to choose a consumption level for the present and for the future in a two-period model, and each of whom is influenced by the choice of the other player. Given a progressive ideology, which dictates that “failure to consume in due quality and quantity becomes a mark of inferiority and demerit” (Veblen, 1899), there is little doubt that such an interaction is not only possible but also very

likely. An evolving body of literature, beginning with the pioneering work of Veblen, supports the notion that consumers might have utility indices that are not independent of each other, and that rivalry in consumption may be an important economic phenomenon with significant effects not accounted for by conventional economic theory. Here we attempt to quantify some aspects of “conspicuous consumption” and to develop a game-theoretical model of interdependent consumer choice based on a modified version of the Cournot

# The Harvard College Economist

duopoly model. Beginning with the construction of an interdependent utility function, we proceed to analyze how the choice of each player affects the consumption of the other and then investigate whether a stable level of individual consumption can be reached over time. Next, we compare the present consumption level of a player with interdependent utility to the standard microeconomic case in which agents optimize individually and consider the implications of the arising differences. In conclusion, we consider the efficiency of the outcome and comment on it as a plausible game theory explanation for current empirical data on consumption and savings.

## II. A Description of the Game

As long ago as 1899, the economist Thorstein Veblen suggested that individual consumption patterns might depend on more than the simple utility derived from the use of a number of goods consumed by a single individual. Even though it is not fully integrated in contemporary microeconomic analysis, the idea that individual spending decisions might have a social component has recently gained support from other sources as well and has been con-

sidered by economists like James Duesenberry, Harvey Liebenstein, and Robert Frank. In particular, Frank's latest book *Luxury Fever*<sup>1</sup> has focused particular attention on what we might term relative position or rivalry effects on consumer behavior. These suggest that an individual's utility is affected not only by the quantities of the goods consumed but also by the individual's relative standing with respect to the consumption of the rest of society. The existence of a psychological justification of this phenomenon had already been called to economists' attention by Duesenberry, who also proposed a utility function that would account for the influence of society on the individual. Along the same line, Frank has argued that much of the current spending boom in the United States can be explained in terms of a desire to emulate, or rival, the behavior of the richest agents in the economy, with patterns of high consumption spreading from the top to the bottom of the income distribution, driven by a desire to avoid negative social comparisons. Thus, any two consumers with similar interdependent utility functions can be expected to engage in a game in which each player observes the consumption of the other, increases his own consumption in order to enhance a posi-

tive social comparison and to avoid a negative one, and in turn induces the other player to increase the quantity of products consumed in order to remain competitive.

Thus, if interdependent consumer utility is indeed at the heart of interpersonal interactions of this type, its effects are by no means negligible and an attempt to model the behavior of competing consumers with the tools of game theory can substantially enhance our understanding of the matter.

### **III. Specification of the Game: Players and Strategies**

In order to simplify the analysis, we will approach this as a static game with complete information; that is, we will assume that quantities are chosen simultaneously and that both players' payoff functions are common knowledge. We will rely extensively on the methodology suggested by Cournot in analyzing duopoly systems, because both the behavior of a duopolist and that of an interdependent consumer involve (1) the maximization of a continuous payoff function and (2) choice of a best-response strategy from an infinite rather than finite strategy set. This will allow us to adopt a common framework for

analysis, which can be modified when necessary, while preserving some of the useful properties of the Cournot model. For example, instead of assuming that the outcome of the game is reached immediately after the players make a pair of simultaneous decisions, we could also view the game as a repeated one in which players don't know each other's utility functions, but merely accept their opponent's responses as given at each stage and then choose a best response. As in the Cournot model, such an interpretation is algebraically equivalent to the case with complete information, but possesses the advantage of being more realistic.

In order to provide a formal definition of the game, we need to specify the players, the strategies available to each of them, and their payoff functions. We begin by considering a model with only two players, each of whom faces a choice of how much to consume in the present, and each of whom is influenced by the choice of the other consumer. If we assume that consumption is measured in dollars and is continuously divisible, all possible choices of present consumption (consumption at time  $t$ , or simply  $C_t$ ) will comprise the interval from 0 to  $I$ , where  $I$  is the income of the consumer; for sim-

# The Harvard College Economist

plicity here we assume that income is constant and is received only in the current period, with no possibility for borrowing or lending. Since, unlike in other microeconomic problems, the choice the consumer faces here is not between the consumption of one good vs. all other goods, but between present and future consumption, we also need to take into account the influence of the additional variable of future consumption. At first, it appears that the introduction of an additional variable will greatly complicate our definition of each player's strategy sets, since now each player would be able to choose different quantities for two different variables. However, there exists a way to avoid this problem. If we assume that each player is rational and always attempts to maximize his payoff (in accordance with the standard propositions of most game theoretical models), the consumer's optimal choice between present and future consumption will always be on his budget constraint, since any other choice would be either unfeasible or not a utility maximum, given any reasonable utility function. Therefore, present and future consumption will always be related by means of the budget constraint if the consumer is rational, which allows us to conclude that under the assumption of rationality

the two strategy sets can be treated interchangeably since they are functionally related. In other words, if we signify present consumption as  $C_t$  and consumption in the future as  $C_{t+1}$ , once a player selects a strategy from the set {All possible  $C_t$ }, there will be one and only one strategy from the strategy set {All possible  $C_{t+1}$ } corresponding to it. With that in mind, we can write each player's strategy space as

$$S_i = \{\text{all feasible } C_t\} = [0, I],$$

with the understanding that we can translate our results in terms of future consumption when the analysis necessitates it. In summary, in this section we have specified the players of the game and their respective strategy spaces. The only remaining component that we need to complete the specification of the game is each player's payoff function.

## IV. Choice of the Payoff Function

Here we are faced with the difficult task of selecting an interdependent payoff function for each player, which is further complicated by the fact that there is no abundant literature on the question. As in the standard microeconomic case, we assume that the payoffs for each player are specified by a continuous utility function, which he is trying to

maximize subject to the constraint of his income. In his book *Income, Saving, and the Theory of Consumer Behavior*, Duesenberry<sup>2</sup> suggests that a suitable functional form of a utility function that takes social pressure on consumption into account would be

$$U_i = U [ C_i / \sum a_{ij} C_j ]$$

where  $C_i$  is the consumption of the  $i$ -th consumer,  $C_j$  is the consumption of the  $j$ -th consumer, and the term  $a_{ij}$  represents the relative weight placed by the  $i$ -th consumer on the consumption of the  $j$ -th. In other words, because the underlying idea behind interdependent utility is how one compares to the consumption standard of the rest of society, Duesenberry suggests that individual utility should be a function of the ratio of one's expenditures on consumption to the weighted summation of others' consumption. Even though this functional form has later been endorsed by authors like Easterlin in studies on economic growth and happiness,<sup>3</sup> it has two significant properties that make it very difficult to use in our case. First of all, the functional form suggested above is extremely general; it specifies an argument for the utility function but has little to say about the function itself, which could take virtually any form. The second problem is more serious and has to do with the

particular case we are considering, the case in which there are only two consumers. If our hypothetical "society" consists only of two individuals, then the sum over weighted social consumption is exactly the same as the weighted consumption of player  $j$ ; that is, the term  $C_i / (\sum a_{ij} C_j)$  becomes equivalent to  $C_i / (a_{ij} C_j)$ , which implies that if one of the players is not consuming anything, the other's utility function is not defined since  $C_j$  is equal to zero. While in the general case it is unlikely that the rest of society wouldn't consume anything, in the two-player case such a possibility cannot be excluded and a utility function of this type would create extreme complications for mathematical modeling. Therefore, we will need to construct a utility function with more desirable properties, which takes into account interdependence and is defined over all positive real numbers.

We begin with a set of assumptions that seem reasonable in the context of the game in order to construct such a payoff function. It seems reasonable to assume that

- (1) The  $i$ -th consumer has a diminishing marginal utility with respect to both present and future consumption.
- (2) The  $i$ -th consumer is risk-averse.
- (3) The share of a particular good's

## The Harvard College Economist

consumption in utility remains the same, or equivalently, that utility's elasticities with respect to present and future consumption are constant for each player.

(4) The  $j$ -th player's consumption affects  $i$ -th's utility negatively, or in other words, that  $U_i$  is a decreasing function of  $C_j$ .

(5) Players compare their present consumption, but are not astute enough to compare their consumption in the future. This might arise for a variety of reasons: even if current consumption is observed, the players might not know each other's income and therefore be uncertain about how much of it has been saved to be consumed in the future period of our two-period model. We have to add that this is very often the case in real-life situations, where information of this type is seldom available.

(6) Each player, if unaware of the other's consumption, or if given zero consumption on the part of the opponent player, would choose between goods as if he had independent preferences.

We know that if the consumer's utility is represented by a preference-independent utility function with convex indifference curves (when viewed from the point  $(0,0)$ ), the optimization process is particularly easy to model. There-

fore, it would probably be beneficial to consider the possibility of modifying a suitable utility function with convex indifference curves in such a way that it would be able to account for utility interdependence, while preserving the rest of its useful properties. In particular, let us consider a Cobb-Douglas utility function of the form

$$U_i = C_t^a C_{t+1}^{1-a}$$

It is easy to prove that its indifference curves are convex and that the function exhibits diminishing marginal returns to either variable when the values of the parameter  $a$  are set between 0 and 1. Moreover, the Cobb-Douglas utility function satisfies two more important properties, which we specified as desirable in propositions (2) and (3): it has constant elasticities with respect to consumption at present and in the future, corresponding to constant shares in utility from the consumption of each good, and it is consistent with risk-aversion. Let us now consider the remaining three propositions, properties (4), (5), and (6). If the players compare only their present consumption, and the consumption of each player enters the other's utility negatively, then the following expectation seems reasonable. Let us assume that player  $j$  has unexpectedly increased his present consumption (the only quan-

tity that player  $i$  can observe). Taking into account proposition (4), we can expect a fall in player  $i$ 's utility; that is, his previous optimum bundle, given an increase in rival consumption, would now correspond to a lower value of his utility index. In other words, we can interpret each increase in player  $j$ 's consumption as resulting in an outward shift in player  $i$ 's indifference map; for any given quantity of  $C_{t+1}$ , player  $i$  would now require more present consumption ( $C_t$ ) in order to attain his previous level of utility. Let us examine how we could model such a shift mathematically for a Cobb-Douglas utility function.

Since each indifference curve corresponds to a fixed level of utility, the indifference curves are essentially the same as the level curves of the utility function. Hence we can express an indifference curve with a utility level of  $U$  as follows, in terms of  $C_t$ :

$$C_t^a = \frac{U}{C_{t+1}^{1-a}}$$

or equivalently, as

$$C_t = U^{\frac{1}{a}} C_{t+1}^{\frac{a-1}{a}}$$

The rightward shift described in the previous paragraph, then, can be modeled quite easily by adding to the indifference curve equation a linear component, which would account for the

effect of player  $j$ 's consumption on the utility of player  $i$  and would be a function of  $C_{j,t}$ . Since we can expect that rival consumption would not reduce individual utility one for one, we will take a weighted form of player  $j$ 's consumption with a weight of  $k$ , where  $k$  is a positive number greater than 0 and smaller than 1. In this way we arrive at an interdependent indifference curve for the  $i$ -th player, which can be expressed as

$$C_t = U^{\frac{1}{a}} C_{t+1}^{\frac{a-1}{a}} + k C_{j,t}$$

In order to simplify the notation, from here on we will refer to the quantity  $C_{j,t}$  simply as  $C_j$  since we have already made the assumption that neither player takes the other player's future consumption into account. Also, all quantities without a subscript indicating the player will be understood to refer to the  $i$ -th player.

From here on, we can simply work in backward order in order to derive a utility function that would have interdependent indifference curves of the form suggested above. An implementation of this procedure yields the functional form

$$U_i = C_t - k C_j^a C_{t+1}^{1-a}$$

It can be shown that this function displays diminishing marginal returns to in-

# The Harvard College Economist

dividual consumption in either time period, preserves the constancy of elasticities with respect to individual consumption, depends negatively on the consumption of player  $j$ , and incorporates a comparison only between present consumption levels. These conditions satisfy the initial assumptions that we numbered from 1 to 5. Also, it becomes immediately evident that if  $C_j$  is set to 0, for example because of lack of information about the consumption of player  $j$ , the function immediately returns to its normal form with independent preferences, thereby also satisfying our last condition.

## V. The Specification of the Game Restated

After the lengthy discussion of the choice of appropriate functional form for the payoff functions, we are finally ready to complete our specification of the game. In summary, we have a game in which two rational agents, similar to two firms in the Cournot duopoly model, are trying to maximize their payoffs, which depend in part on their competitor's actions. As in the "pseudo-dynamic" interpretation of the Cournot model, players act sequentially, and neither knows his opponent's utility function. For each player, we assume a con-

tinuous strategy space ranging from 0 to  $I$  for present consumption, and a payoff function of the form

$$U_i = C_t - k C_j^a C_{t+1}^{1-a}$$

Unlike in the Cournot model, however, here we have a function in several variables in which the opposite player's consumption sifts individual preferences in favor of more consumption in the present; i.e., we have successfully modeled the behavior of our two players, but we cannot use the standard maximizing procedures used for one-variable functions. Instead, we will resort to the method of Lagrange multipliers in order to find out the optimum consumption bundle of each player, subject to an intertemporal budget constraint. For simplicity, we will assume that income can either be spent now or in the future and that there is no interest earned on savings. This does not fundamentally alter the nature of our argument but allows us to simplify the constraint, which would not change substantially even if interest were allowed. Thus we can write the intertemporal budget constraint as

$$C_t + C_{t+1} = I.$$

Because each player accepts the other's responses as given, the quantity  $C_j$  will be treated as a parameter; this will allow us to treat the utility function as a function of two variables and help

us further simplify the analysis. Thus our purpose amounts to solving the problem

$$\begin{aligned} \max U_i &= C_t - k C_{j,t}^a C_{t+1}^{1-a} \\ \text{s.t. } g &= C_t + C_{t+1} - I = 0 \end{aligned}$$

We can approach the problem by first taking the gradients of the function and the constraint, equating them with a factor of  $\lambda$ , and substituting the resulting expression into the intertemporal budget constraint. The expressions thus obtained for  $C_t$  and  $C_{t+1}$  will give us the optimum choice of each player, in which the other's consumption will also be present as a parameter. We begin by computing the gradient of the utility function:

$$\begin{aligned} \text{grad} U &= a C_t - k C_j^{a-1} C_{t+1}^{1-a} \mathbf{e}_1 + \\ &\quad (1-a) C_t - k C_j^a C_{t+1}^{-a} \mathbf{e}_2 \end{aligned}$$

where the boldface  $\mathbf{e}_1$  and  $\mathbf{e}_2$  signify the standard unit vectors. Next, we find the gradient of the budget constraint  $g$ , which is simply

$$\text{grad } g = \mathbf{1} \mathbf{e}_1 + \mathbf{1} \mathbf{e}_2$$

The system

$$\begin{aligned} \text{grad } U &= \lambda \text{grad } g \\ g &= 0 \end{aligned}$$

can be solved in three steps. First, we translate it into the equivalent form

$$\begin{aligned} a C_t - k C_j^{a-1} C_{t+1}^{1-a} &= \lambda \\ (1-a) C_t - k C_j^a C_{t+1}^{-a} &= \lambda \\ g = C_t + C_{t+1} - I &= 0 \end{aligned}$$

From the first two equations, we obtain the relationship

$$C_{t+1} = \frac{1-a}{a} C_t - k C_j$$

and finally, by substituting the above expression in the last remaining equation, we are able to arrive at the optimum level of present consumption for player  $i$ :

$$C_t^* = a I + (1-a) k C_j$$

By the very construction of our calculation, this choice maximizes player  $i$ 's utility for any consumption level  $C_j$  by the second player, subject to  $i$ 's budget constraint, no matter what player  $j$  does. Equivalently, we can say that the utility obtained by each player  $i$ , given a fixed strategy chosen by the other player, is greater than or equal to the utility that  $i$  would have obtained if he had played any other strategy  $C_{i,t}'$  than the one specified above. This tells us that according to Nash's 1950 definition<sup>4</sup>, the set  $\{C_{i,t}^*\}_i \{C_{j,t}^*\}_j$  is a pair of strategies that are tied to each player's best response, and which constitute an equilibrium from which neither side has an incentive to deviate. We have written this pair of best-response strategies in a slightly more complicated way than usual because we need to take one extra step in order to arrive at the Nash equilib-

# The Harvard College Economist

rium quantities. In the form in which we have expressed the pair of best response strategies so far, we know that they constitute a Nash equilibrium but we do not know the actual quantity of consumption corresponding to the equilibrium. In order to arrive at an expression that tells us the equilibrium amounts consumed, we have to take into account that in our case the expression  $\{C_{i,t}^*\}_i \{C_{j,t}^*\}_j$  corresponds to a system of best-response functions, which we can solve in a way similar to the Cournot model. If we take each best response function to have the form

$$C_i^* = a I_i + (1 - a) k C_j$$

then we can formulate the system of best response functions as

$$\begin{aligned} C_1 &= a I_1 + (1 - a) k C_2 \\ C_2 &= a I_2 + (1 - a) k C_1 \end{aligned}$$

from which we can solve for the Nash equilibrium quantities by substituting any of the equations into the other one. The Nash equilibrium consumption levels thus obtained are

$$\begin{aligned} C_1^* &= \frac{a I_1 + (1 - a) k I_2}{1 - (1 - a)^2 k^2} \\ C_2^* &= \frac{a I_2 + (1 - a) k I_1}{1 - (1 - a)^2 k^2} \end{aligned}$$

As we can see, the consumption outcome associated with the Nash equilibrium is symmetric and depends on the two consumers' incomes, on the rela-

tive weight placed on the other player's consumption, and on the elasticity of the utility function with respect to present spending, all of which are constant in our model and can be empirically determined. Such a result makes sense because it indicates that the budget constraint, as a function of income, has been taken into account, together with the second player's choice, which in turn depends on his income. We can be assured that the equilibrium quantities are positive, since their numerators are just weighted factors of both consumers' positive incomes, and since the denominators will always be positive for the range of  $a$  and  $k$  that we have specified. Also, the Nash equilibrium in this game is unique since there is only one pair of best-response strategies, corresponding to a system of best response functions which has only one solution.

## VI. Efficiency

An inevitable question that arises in all games involving Nash equilibria is whether a Nash equilibrium is efficient. Ever since the discovery of the Prisoner's Dilemma, game theorists have been aware that in many games the equilibria based on best-response strategies can result in various inefficiencies. An example of such a game is the problem

of the commons, and both Cournot's original version of duopolistic competition and Stackelberg's dynamic version of oligopoly exhibit this property. Therefore, it seems reasonable to ask whether the Nash equilibrium in the consumer interdependence game considered here is efficient.

There is a broad class of inefficiencies that we could conceive of if we translate the outcome of this two-person game to the society in general. First of all, we shall argue that competition between consumers creates a distortion in intertemporal choice, which is evinced in a bias in favor of present consumption, and correspondingly, in an incentive to save less to be able to live up to the expectations of society today. Upon reviewing the results of our model, it would be useful to compare whether the Nash equilibrium quantities that we derived are greater or smaller than what each player would have consumed individually in the absence of social pressure. This can be done quite easily if we return to the preference-independent intertemporal utility function that we introduced in the beginning and we repeat the Lagrangian maximization exercise subject to the same budget constraint. A replication of the procedure already described in section IV yields

an optimal quantity of

$$C_{i,t}^{**} = a I_{i,t}$$

whereby  $C^{**}$  we signify the quantity of consumption that each player would have chosen if his preferences were strictly independent of the preferences of the other player. If we simply rewrite our actual Nash equilibrium consumption quantities in the form

$$C_1^* = a I_1 + \frac{a(1-a)kI_2 + a(1-a)^2k^2I_1}{1 - (1-a)^2k^2}$$

$$C_2^* = a I_2 + \frac{a(1-a)kI_1 + a(1-a)^2k^2I_2}{1 - (1-a)^2k^2}$$

it is quite easy to see that

$$a I_1 + \frac{a(1-a)kI_2 + a(1-a)^2k^2I_1}{1 - (1-a)^2k^2} > a I_1$$

and

for values of  $a \in (0,1)$  and  $k \in (0,1)$ . In other words,

$$\begin{aligned} C_{1,t}^* &> C_{1,t}^{**} \\ C_{2,t}^* &> C_{2,t}^{**} \end{aligned}$$

and each consumer is consuming more in the present and saving less if he has interdependent preferences. So far, our mathematical results are in agreement with our intuition, which suggests that the more each player consumes, the more he induces rival consumption on the part of the second player.

We can imagine that a variety

## The Harvard College Economist

of people might be influenced by a similar kind of reasoning in their everyday lives without even realizing the powerful influence of society. What are the practical implications of our findings? The riddle of excess consumption, or “conspicuous consumption” existing solely for social purposes, has puzzled economists for more than a hundred years. In his pioneering work, *The Theory of the Leisure Class*, Thorstein Veblen<sup>5</sup> first called economists’ attention to the existence of a class of consumption that is not physically needed but is rather dedicated to satisfying the social demands placed on the individual. Because such consumption also creates distributional concerns, it is worthwhile to explore whether the reduced payoffs of all players are the only negative consequence of games of interdependent consumption. In one of the latest studies on excess spending, Robert Frank interprets the existence of spending patterns biased towards present consumption as an *allocative* inefficiency, which results in forgone human capital accumulation. In particular, Frank contends that investments vital for the health and the well-being of the population, such as replacement of municipal water supply systems containing toxic metals and higher pay for high school teachers, are being fore-

gone since “paying for luxury consumption has also meant having to curtail spending in the public sphere.”<sup>6</sup> Frank also provides a very striking illustration of a “duopolistic” competition between two consumers similar to the interaction discussed in our game-theory model. The beginning of his book recounts the “battle” between two of America’s top millionaires, one of whom equipped his yacht with more and more luxuries in an effort “to outdo a rival shipping magnate... whose own yacht, the 375-foot *Atlantis*, was designed by an architect whose explicit instructions were to make it 50 feet longer.”<sup>7</sup> One can imagine that the resources spent in equipping the two vessels could have been much more appropriately distributed if they had been directed to subsidizing housing for the homeless or the alleviation of child poverty. Therefore, there is clearly also a distributional aspect of the Nash equilibrium inefficiency, which our two-player model has not been able to capture effectively because it has assumed a society consisting of only two individuals. However, upon extending our analysis to the society as a whole, it becomes immediately evident that such game outcomes result in allocative inefficiencies with tangible implications.

## **VII. The Effect of Utility Interdependence on the Saving Rate**

So far we know that an individual who takes social norms of consumption into account will be biased toward consumption in the present, and he is likely to save less than he would have done if the decision on how much to spend had been dependent entirely on himself (and not on the predominant beliefs of what and how much it is prestigious to consume). It is important to note that in reaching this conclusion, we haven't made any initial assumptions about the incomes with which our two players are starting; our analysis is perfectly general and allows for any amount of income for either player. Therefore, it would be especially interesting to explore what happens if we introduce a degree of income inequality. In particular, we know that each player consumes more in the present than he would have chosen to do independently, but does this imply anything at all for the saving rates of people with different earnings? It might well be the case that while each player increases present consumption and lowers his saving rate, the new saving rates after the Nash equilibrium is reached might be the same for both rich

and poor people alike. To see whether this is indeed so, let us assume that one of the players has an income that is  $m$  times larger than that of his rival. Then we can simply express the second player's income as

$$I_2 = m I_1$$

Given that, we can ask ourselves the question when (for which values of  $m$ ) the saving rate of a poor social class would be lower than that of a relatively high-income group of the population; i.e. is it plausible to justify theoretically the empirical observation that the present consumption boom in the US is surprisingly fueled predominantly by middle and low-income households? In his book *Luxury Fever*, Robert Frank reports the curious example of an upper-middle class housewife who purchased a \$17,500 wristwatch for her husband's birthday<sup>8</sup>. To most conservative consumers, such expenditure will appear outrageous, but nevertheless such occurrences are by no means rare. Frank also reports that the American consumer debt to income ratio has substantially grown and is now close to unity, and that the demand for luxuries is now growing four times faster than overall spending. Given these overwhelming figures, it is quite unlikely that such an in-

# The Harvard College Economist

crease in demand could be attributed only to consumers in the top income bracket. Is it then possible to seek an explanation for the increasing share of consumption in income and the declining saving rate in the game we have just described? To find out what answer our model will give us, we need to look at the share of current spending in income, which is by definition equal to  $(1-s)$ , one minus the saving rate. How much richer do the rich have to be in order to induce a poorer individual or social class to lower its rate of saving? It turns out that the inequality

$$\frac{C_1^*}{I_1} > \frac{C_2^*}{I_2}$$

or equivalently

$$\frac{C_1^*}{I_1} > \frac{C_2^*}{mI_1}$$

can be reduced to the form

$$m^2 - 1 - a(1 - a)kI_1 > 0$$

whose only positive solution is

$$m > 1.$$

In other words, we have arrived at a remarkable result: as long as one player's income is greater than that of the other player, the poorer player will always have a lower saving rate. This is a striking finding not only because of its practical implications but also because of its consistency with observations made by

Robert Frank, who writes that disincentives to save come from competition.

## VIII. Conclusion

Our game-theoretical model of interdependent consumer choice has enabled us to reach several important conclusions about the behavior of utility-maximizing agents in a social setting with established consumption standards. We can formulate them as follows:

- (1) Two rational consumers with interdependent utility functions will each consume more in the present than if each of them was optimizing independently, and correspondingly, each of them will save less than standard microeconomic theory predicts.
- (2) The outcome of such rivalry in consumption, given no changes in income, is stable in time and neither consumer has an incentive to deviate from it.
- (3) The outcome is also inefficient in the sense that it distorts the optimal allocation of resources.
- (4) Under conditions of income inequality, the richer player will always have a higher saving rate than the poorer one, since the latter would be spending a disproportionate amount of his income trying to emulate the absolute consumption standard set by the rich.

In our view, this set of conclu-

sions can provide a plausible explanation for the comparatively low American saving rate, the current consumption and luxury spending boom in the US, and the increasing spending to income ratio observed in middle and lower-class households. Whether this explanation is correct remains to be tested empirically.

<sup>1</sup> Frank, 1999.

<sup>2</sup> Duesenberry, 1949, 32.

<sup>3</sup> Easterlin, 1973, 112.

<sup>4</sup> Gibbons, 1992, 8.

## References

- Duesenberry, James. *Income, Saving, and the Theory of Consumer Behavior*. Harvard University Press, 1949.
- Easterlin, Richard. "Does Economic Growth Improve the Human Lot?," in *Nations and Households in Economic Growth*, 1973.
- Frank, Robert. *Luxury Fever*. The Free Press, 1999.
- Gibbons, Robert. *Game Theory For Applied Economists*. Princeton University Press, 1992.
- Veblen, Thorstein. *The Theory of the Leisure Class*. New York, Modern Library, 1899.

# Crossing the Atlantic and Playing Deals: An Analysis of Returns to Risk Arbitrage in Europe

*David Witkin*

## Abstract

This paper evaluates the profitability of risk arbitrage in Europe by measuring returns to investing in 633 mergers from 1989 to 1999. Results indicate that European risk arbitrage generates small positive returns when transaction and borrowing costs are excluded and returns fall to levels not statistically different from zero when these costs are included. Results from investing in British deals are particularly impacted by transaction and borrowing costs relative to deals in the other countries. In a Capital Asset Pricing Model regression, European risk arbitrage returns are not statistically different from a risk-adjusted European market return, though cash deals have statistically significant positive alpha when transaction and borrowing costs are excluded, and stock deals have significant negative alpha when these costs are included. Returns are also analyzed in light of two recent papers focusing on risk arbitrage as an investment strategy in the United States. The first paper suggests that transaction and other practical costs have a statistically negative effect on risk arbitrage returns. European results would confirm this transaction cost theory. The second paper posits that transaction costs do not affect returns, but that “limited arbitrage” does: when the supply of dedicated risk arbitrage capital falls, subsequent returns rise, and vice versa. The European returns provide some indirect support for this theory. Finally, the results are compared to U.S. returns calculated in these two papers. European risk arbitrage appears to underperform significantly its U.S. counterpart.

## I. Introduction

The increase in European merger and acquisition activity since the mid-1990s has been well documented by the financial press and duly noted by American investment banks. They have devoted

substantially more resources and attention to the European M&A market in an attempt to capture a slice of the burgeoning business. Meanwhile, on the next flights over to London have come risk arbitrageurs, who make a living out

of betting whether announced takeovers will or will not be completed. By all accounts, the risk arbitrage market in Europe has grown rapidly in the last few years.<sup>1</sup> At the same time as money is flowing into Europe to play deals, however, there is a debate in the United States regarding the profitability of risk arbitrage in general. One side contends that it produces, in the words of one paper, “considerable” excess returns,<sup>2</sup> which Baker and Savasoglu (1999)<sup>3</sup> attribute to the existence of “limited arbitrage.” Arbitrageurs receive a premium (in the form of the “arbitrage spread”) for assuming the risk that an announced deal will fail, and the smaller this supply of arbitrage capital to provide “deal insurance,” the greater the premium existing arbitrageurs can demand. The opposing view is posed by Mitchell and Pulvino (2001),<sup>4</sup> who suggest that excess returns to risk arbitrage are not as large as Baker, or others, have documented.<sup>5</sup> According to Mitchell and Pulvino, transaction costs and other practical limitations cause excess risk arbitrage returns to fall from “considerable” to simply mediocre.

This paper uses the expanding merger market in Europe to test the conflicting views of risk arbitrage as an investment strategy. For the years 1989

through 1999, portfolios are constructed from deals announced in seven countries: the United Kingdom, France, Germany, Italy, the Netherlands, Spain and Switzerland.<sup>6</sup> A position in a deal is taken on the first trading day following the date of formal announcement of the offer. The position is held until the deal is either completed or withdrawn. If the deal is completed, a period of two weeks is assumed for the arbitrageur to receive payment from the acquirer for his shares. If the deal is withdrawn, the position is exited on the first trading day following the date of formal announcement of withdrawal.

The portfolios are run under different assumptions: weighting position sizes equally or by target market value; incorporating transaction costs and the costs of carry or rebate, either of these two costs, or no costs at all; and restricting the portfolio to only cash deals or only stock deals. Because seventy percent of the deals in the sample come from the UK, separate portfolios are also designed for UK and non-UK deals, to test for differences in deal returns between Britain and the rest of Europe. Portfolio returns are compared to a European risk-free interest rate (the “Eurorate”) to measure excess returns, and these excess returns are regressed

## The Harvard College Economist

in a Capital Asset Pricing Model against excess returns on a European market portfolio (the “Euroindex”) to determine if there are abnormal returns to risk arbitrage in Europe.<sup>7</sup> In the process, the effects of transaction costs and borrowing costs, two “practical constraints” which Mitchell and Pulvino claim significantly decrease returns to risk arbitrage (at least in the United States), can be observed.

Results indicate that risk arbitrage in Europe does make a profit with close to zero market beta, but that transaction and borrowing costs have a substantial negative effect on returns. With no costs, the mean return on a portfolio of all deals is about 0.9 percent per month (10.8 percent annually), but when costs are included, the return declines to 0.45 percent per month (5.4 percent annually). When compared to the risk-free interest rate, transaction and borrowing costs cause excess returns to go from slightly positive (2.8 percent per year) to slightly negative (-2.6 percent per year). Finally, in CAPM regressions against a European market portfolio, not even raw portfolios (i.e., not facing transaction or borrowing costs) generate alpha statistically greater than zero, but transaction and borrowing costs do cause alpha to decline from 0.21 per-

cent monthly (2.5 percent annually) to around -0.24 percent (-2.9 percent annually). Across categories, cash deals in the sample earn much higher returns than stock deals, and UK deals perform slightly better than non-UK deals when no costs are assumed, yet much worse than non-U.K. deals when transaction and borrowing costs are included. Clearly, the results provide strong support for Mitchell and Pulvino’s contention that, after practical costs are accounted for, returns to risk arbitrage are not considerable at all.

Also, even though the European results do not support Baker’s (or others’) findings of high excess risk arbitrage returns, they do lend evidence to his theory that arbitrage returns should decrease when arbitrage capital increases (this support is rather indirect, however, as extremely scarce data on risk arbitrage capital in Europe made a formal test impossible). In the late 1990s, much of the new arbitrage capital was directed at continental Europe, as risk arbitrage had already been relatively prevalent in the UK.<sup>8</sup> According to the Baker model, non-UK risk arbitrage returns should have decreased relative to UK returns in this time period. In this paper, a portfolio of UK deals (assuming no transaction or bor-

rowing costs) would have earned a mean annual return of 12.3 percent from 1991 to 1996, and 13.2 percent from 1997 to 1999. A portfolio of all non-U.K. deals (also assuming no transaction or borrowing costs), however, would have seen its mean annual return decrease from 10.4 percent over 1991-96, to 3.0 percent over 1997-99. Limited arbitrage may be playing a role in the declining European returns.

Section II of this paper illustrates the mechanics of risk arbitrage and gives examples of traditional arbitrage investments. Section III describes the dataset used to calculate the European risk arbitrage time series. Section IV explains the characteristics and assumptions of the various portfolios constructed to model European arbitrage returns. Section V presents and discusses the results, and Section VI concludes.

## **II. Types of Risk Arbitrage Investments**

There are two primary kinds of mergers, cash mergers and stock mergers. In a cash merger, the acquirer offers a fixed dollar amount of cash for every share of the target. Often, following the announcement of a merger for  $x$  dollars (or pounds or euros, as the case may be), the target will trade at a

discounted price  $\$(x-m)$ . The risk arbitrageur simply buys the stock at  $\$(x-m)$  and holds it until the merger goes through, at which point he will exchange his share for  $\$x$ , thus capturing the discount  $\$m$ . The arbitrageur is also entitled to any dividends the target firm pays during the period he holds the stock.

In stock deals, the acquirer offers shares of its own stock in exchange for each share of the target (for instance, in a cross-border deal announced and completed in 1999, the French pharmaceutical firm Rhone-Poulenc SA exchanged 0.75 of its shares for each share of a German rival, Hoechst AG). After the announcement of a stock deal, the target will usually trade at  $\$(rp-m)$ , where  $r$  is the number of acquirer shares offered per target share, and  $\$p$  is the share price of the acquirer. Stock deals are arbitrated by buying the target stock at  $\$(rp-m)$  and simultaneously selling short  $r$  shares of the acquirer at  $\$p$  for each share of the target owned. When the deal closes, and the arbitrageur receives  $r$  acquirer shares for each target share he owns, he returns the  $r$  shares to the lender and captures the discount  $\$m$ . In the meantime, the arbitrageur is entitled to dividends on the target, and he also receives interest (called the “re-

## The Harvard College Economist

bate”) on  $\$rp$ , the proceeds from the short sale.<sup>9</sup> His holdings will also decline, however, by  $r$  times any dividend payments per share that the acquirer makes.

The exchange ratio  $r$  is “fixed” in most stock deals, meaning the acquirer will pay the same number of shares for each target share whether  $\$p$  rises or falls (thus leaving target shareholders exposed to swings in the acquirer’s stock price). In “floating ratio” deals,  $r = \$v/\$p$ , where  $\$v$  is a specified “dollar value,” and  $\$p$  is determined in a pricing period specified in the terms of the merger. Target shareholders in floating-ratio deals are not exposed to decreases (or increases) in  $\$p$ , because as  $\$p$  falls (or rises) during the pricing period,  $r$  adjusts upward (or downward) to give the target shareholder more (or fewer) acquirer shares. Though floating-ratio stock deals are typically arbitrated by selling short acquirer shares during the pricing period, this paper, like Baker and Mitchell, uses a simplifying assumption and treats them as cash deals (as Baker notes, “The key is that the consideration does not depend on the level of [the acquirer’s] share price.”).<sup>10</sup>

Many deals, both in the US and Europe, feature more complex terms

than cash, fixed-ratio stock or floating-ratio stock considerations. For example, in a “collar” deal, the merger agreement specifies that  $r$  will change as the acquirer’s stock moves in and out of certain specified price ranges during a pricing period. Collar deals have certain option-like features that can be arbitrated using derivatives. In other mergers, acquirers will offer securities such as warrants, preferred stock, or bonds for the target shares. These deals and collar deals are excluded from the sample, as they are in Baker and Mitchell, due to the complexity involved in arbitrating them. As Mitchell notes, “determining the value of the ‘hedge’ in [these transactions] is [often] not possible since market values of hybrid securities are generally unavailable.”<sup>11</sup>

Baker and Mitchell exclude some types of deals, however, that this paper includes. In one of these types (to be called “mix” deals), the target receives both cash *and* stock from the acquirer. The arbitrage on this transaction is performed in the same manner as if the deal were 100 percent stock: by buying one share of the target and shorting  $r$  shares of the acquirer.<sup>12</sup> However, the value of the deal is now  $\$(rp+x)$ , where  $x$  is the cash portion of the offer. The arbitrageur buys the target stock

after announcement at  $\$(rp+x-m)$ , simultaneously sells short  $\$rp$  worth of acquirer shares, and hopes to collect  $x$  if the deal is completed, thus capturing the discount  $m$ .

Another deal type included in this paper is a “choice” deal, in which target shareholders are offered the choice of  $\$x$  or  $r$  for each share they own. “Choice” deals are arbitrated based on which alternative offers the larger spread. If  $\$rp > \$x$ , then the spread on the stock alternative ( $\$(rp-m)$ ) is larger than the spread on the cash alternative ( $\$(x-m)$ ), and thus arbitrating the merger as a stock deal is more attractive (gross of any transaction costs). The paper assumes that the arbitrageur chooses the alternative that offers the larger spread on the first trading day after the deal’s announcement, which is the day he sets up all his positions. Thus, if the stock alternative is more attractive on that day, he buys one target share and shorts  $r$  acquirer shares, holding this position until the deal closes, at which time he informs the acquirer he would like to receive stock instead of cash. If the cash alternative is higher on the day after announcement, the arbitrageur simply buys one target share and elects to receive cash when the deal closes.<sup>13</sup>

If a deal fails, the arbitrageur’s

exit strategy depends on whether the failed deal is a cash or stock deal (failed “mix” deals are exited in the same manner as stock deals, and failed “choice” deals are exited like either cash or stock deals, depending on the manner in which the arbitrageur initially chose to play the deal). If a cash deal fails, the arbitrageur simply sells his share in the target. The paper assumes this transaction takes place the day after the offer is formally withdrawn, ensuring the arbitrageur suffers the full (likely negative) effect of the announcement. If a stock deal fails, the arbitrageur sells his target stock and purchases  $r$  acquirer shares in the market in order to fulfill delivery of the  $r$  shares he borrowed and sold short when entering the position. These transactions are also assumed to take place the day after the deal is formally withdrawn.

### **III. Data Description**

The original data set consisted of 1,137 merger and acquisition announcements reported by Securities Data Company (SDC) from January 1, 1989 through December 31, 1999 for the seven countries studied in this paper, and valued by SDC at \$100 million or greater at the time of the deal’s announcement. The SDC summaries provided the names of the target and the

## The Harvard College Economist

acquirer, the dates of deal announcement and completion (or withdrawal), and terms of the deal, including the mode of consideration paid by the acquirer (e.g. “Cash”, “Cash and Loan Notes”<sup>14</sup>, “Ordinary Shares”, or “Ordinary and Preferred Shares”), and the securities or assets of the target which were being acquired (e.g. “Ordinary Shares”, “Ordinary and Bearer Shares”, “Assets”, or “Ordinary Shares and Convertible Bonds”). Of these deals, 205 were excluded right away because the acquisition reported was for only part of the target company, or for shares other than ordinary shares, thus unlikely to be arbitrated by outside investors.

For deals announced in 1993 and following, dates and deal terms were double-checked using *Bloomberg*, whose data was more reliable and comprehensive than SDC’s but went back only to mid-1992. Thus, for deals announced beginning January 1, 1993, the “announcement,” “completion” and “withdrawal” dates and deal terms are those reported in *Bloomberg*, which are often reproductions of press releases or wire reports. For deals announced in 1989 through 1992, the dates and terms are those provided by SDC.

Using the company names provided in the original data set, 6-digit

*Datastream* codes were looked up manually for all targets, and for each acquirer in stock deals.<sup>15</sup> The codes were then entered into *Datastream* to download daily historical closing stock prices and trading volumes for the various dates the companies were involved in a deal. If no sensible *Datastream* quotes could be found for a stock, *Bloomberg* historical closing prices (dating back to June 1992; deals before this date had to be dropped) were used.

Of the 928 deals remaining following the initial exclusions from the SDC list, 295 were further cut, for six possible reasons. First, after checking on *Bloomberg*, the “announcement” reported by SDC was found to be of merger discussions or of intentions to launch an offer, not of a definitive agreement or bid. Second, the consideration being paid by the acquirer consisted of securities other than cash or ordinary shares, making the arbitrage too complex, or impossible, to simulate using ordinary share prices. Third, sensible stock price data, deal announcement, withdrawal or completion dates, or terms of the deal could not be found on *Datastream*, *Bloomberg* or *Lexis-Nexis*. Fourth, the target (in any deal) or acquirer (in deals with stock consid-

eration) had average trading volume below 5,000 shares a day for the 5 trading days following the merger announcement, rendering the entering of arbitrage positions almost impossible without incurring enormous indirect transaction costs. Fifth, the target had a very liquid American Depositary Receipt (ADR) which could be traded in lieu of its ordinary shares, thus enabling the deal to be arbitrated in US markets, as a US deal, instead of in Europe (LVMH's 1999 offer for Gucci NV offers an example of a large deal of this nature).<sup>16</sup> Finally, the target was financially distressed and being sold in a government-run auction, a situation which merger arbitrageurs tend to avoid because of inadequate transparency in the sale process.

Thus, the remaining sample contains 633 transactions. British targets account for around two-thirds of the sample. France has the second largest number of deals, with close to 100. Italy has 26 deals over the 11 years, while the other countries have 25 deals or less. The number of deals in a given country varies over time: for instance, Spain has no more than 4 deals in a given year until 1998, when it has 10.

The holding period for a transaction is defined as the number of days, including weekends, that the deal is held

by the portfolio. For completed deals, this period extends from the trading day after announcement, when the position is set up, to ten business days following the completion of the deal (or, for UK deals, ten business days following announcement that the offer is "unconditional in all respects"), when the arbitrageur receives payment from the acquirer. For failed deals, the holding period runs from the trading day following announcement of the deal, to the trading day following announced withdrawal of the deal.<sup>17</sup> UK deals have the shortest holding periods, around two and a half months, while French holding periods are a little over three months, and deals in other countries are held four to five months, on average.

## **IV. Construction of Portfolios**

### ***A. Basic Portfolio Activity and Initial Funding***

Using the sample of 633 mergers, mechanical European Risk Arbitrage Portfolios (ERAPs) are calculated for all deals, all cash deals and all stock deals ("mix" deals are placed in the stock deal category for return calculations, while "choice" deals are placed in either the cash or stock category, de-

## The Harvard College Economist

pending on the alternative chosen by the portfolio), then subjected to varying assumptions to measure the respective effects of different factors on risk arbitrage returns. This subsection describes the assumptions under which all portfolios operate.

The life of each ERAP begins when the first merger relevant to that portfolio is announced. For ERAPs of all deals or cash deals, this deal is Peel Holdings' offer of 340p cash per share of London Shop PLC on January 5, 1989. For stock portfolios, the first announced deal is Cadbury Schweppes' offer for Basset Foods PLC on February 2, 1989. Each ERAP is assumed to be seeded with enough money (in dollars) to cover two requirements: one, the posting of initial collateral to open margin accounts in each country where a deal is announced, until the point where the portfolio has generated enough value to post collateral from its existing profits; and two, the repayment of margin loans if deals break early in the portfolio's life, and the ERAP cannot square the loan using either the proceeds it expected from the arbitrage or funds accumulated from prior successful deals.

The ERAP invests in every announced deal at the closing price on the

first trading day after the merger announcement (this assumption explicitly ensures that the portfolio does not benefit from the initial spike in price usually experienced by the target immediately following the announcement), as reported by SDC or *Bloomberg*. In the London Shop example, this date is January 6<sup>th</sup>. The ERAP draws on its initial funding, goes into the spot foreign exchange market, and exchanges dollars for enough British pounds to open a margin account in London and borrow 330p (this collateral amount is perhaps 50 percent of the loan). The ERAP then takes out the loan, buys the share of London Shop, and waits for the next deal announcement. On January 6, the first day of the holding period for the London Shop deal, the ERAP is assumed to earn a return of zero percent, gross of any transaction or borrowing costs.

The process continues as each new deal is announced. For stock deals, the ERAP buys the target share with borrowed money, and simultaneously shorts  $r$  acquirer shares, keeping the short proceeds in a bank account in the acquirer's country, earning a rebate.<sup>18</sup> The interest rate on the loan for the long position (this rate is called the "cost of carry") is assumed to be the risk-free

rate in the target country. The assumed rebate is the risk-free rate in the acquirer's country, but decreases according to an assumed schedule (explained below) if the acquirer stock is illiquid.

Two basic types of portfolios are constructed, as in the Baker paper, assuming different methodologies for determining relative position sizes in each deal. In "value-weighted" portfolios, the weight given to a single deal is proportional to the target's market capitalization (in dollars) relative to the sum of the market caps of all the targets on that day. Each target market value is determined as of the close of market on the day following announcement of the offer for that target.<sup>19</sup> In "equal-weighted" portfolios, position sizes are equalized across deals. Both types of portfolios, however, are subject to a limit on a single deal's position size to 10 percent of the entire portfolio, based on a rule of thumb followed by most arbitrage funds, and cited by Mitchell, to insure that the portfolio does not suffer a devastating loss if one large deal fails.<sup>20</sup>

If a deal is completed, the margin loan is squared ten business days after the announcement of the completion, on the same day the ERAP receives payment for its shares.<sup>21</sup> If the deal breaks

and the merger offer is withdrawn, the loan is squared the day after the announcement of withdrawal, after the ERAP has unwound its position. In the first example, the London Shop deal closes successfully on February 1<sup>st</sup>, 1989, and the ERAP receives payment of 340p from Peel Holdings ten business days later, on February 15<sup>th</sup>. That day, it makes a payment to the bank of the principal, 330p, plus the accumulated interest of 3.63p (based on an annualized cost of carry of 10.4 percent, the British risk-free rate on January 6<sup>th</sup> when the money was borrowed), and keeps a profit of 6.37p. Had the deal broken, and London Shop fallen to 320p on February 15<sup>th</sup>, the ERAP would have had to sell its shares at 320 and come up with 13.63p on its own to square the loan. Its net loss on the deal would be 13.63p.<sup>22</sup> After the completion or withdrawal of a stock deal, the ERAP simply withdraws the short sale proceeds it had deposited in the rebate-earning account, and sets this aside for future deals, though if the stock deal in question has broken, these funds might be used by the broker to cover the short position.

### ***B.Raw Daily Returns***

In "raw" portfolios, transaction and borrowing costs are zero. The daily

## The Harvard College Economist

return on each cash deal is given by

$$R_c = (P_t^T + D_t^T - P_{t-1}^T) / (P_{t-1}^T),$$

where  $R_c$  is the daily cash deal return,  $P_t^T$  is the target's closing stock price on day  $t$ ,  $D_t^T$  is the dividend paid by the target on day  $t$ , and  $P_{t-1}^T$  is the target's closing stock price on day  $t-1$ . For stock deals, the daily return is given by

$$R_s = (P_t^T + D_t^T - P_{t-1}^T) / P_{t-1}^T - r(P_t^A + D_t^A - P_{t-1}^A) / P_{t-1}^A$$

where  $R_s$  is the daily stock deal return,  $r$  is the share exchange ratio,  $P_t^A$  is the acquirer's closing stock price on day  $t$ ,  $D_t^A$  is the dividend paid by the acquirer on day  $t$ , and  $P_{t-1}^A$  is the acquirer's stock price on day  $t-1$ .

During the ten-business-day period following a deal completion announcement, in which the ERAP awaits payment, the gross daily return is assumed to be zero, to reflect the fact that the arbitrageur is not vulnerable to swings in the target or acquirer stock price in the period between deal completion and payment.<sup>23</sup> *Ceteris paribus*, this feature lowers returns to European arbitrage relative to U.S. arbitrage, because the arbitrageur in a European deal must wait seven business days longer than an arbitrageur in a U.S. deal to reinvest that deal's proceeds in new deals earning positive average returns. In portfolios that assume a cost of carry, the Euro-

pean arbitrageur suffers even more relative to his American counterpart, because he must continue to pay the cost of carry (and receive a rebate, but the value of the former is usually larger than that of the latter) for seven extra business days until he can settle his loan.

### C. Portfolio Returns Including Transaction Costs

For institutional investors, according to active arbitrageurs, a good approximation of European direct transaction costs, including brokerage commissions, clearing and settlement fees, and government duties, is 50 basis points (0.5 percent) times the share price, no matter the country.<sup>24</sup> This is the baseline used for the ERAP portfolios that include transaction costs.

In addition to brokerage fees, transaction costs in the UK have an additional, significant component: the "stamp tax" (similar to the US transfer tax) of 50 basis points times the share price, for both purchases and sales of stock. Thus, in effect, total direct trading costs are 1 percent in the UK and 0.50 percent in the other European countries.<sup>25</sup>

Daily deal returns including transaction costs, then, are given by

$$R_x = (P_t^T + D_t^T - P_{t-1}^T - X_t^T - X_t^A) /$$

$$P_{t-1}^T - r(P_t^A + D_t^A - P_{t-1}^A) / P_{t-1}^A$$

where  $R_x$  denotes deal return with transaction costs,  $X_t^T$  represents transaction costs incurred buying target shares on day  $t$ , and  $X_t^A$  represents transaction costs incurred while selling short acquirer shares on day  $t$ . If it is a cash deal, of course, the right half of the right side of the equation equals zero.

The ERAP assumes that transaction costs are incurred in entering positions, exiting positions in failed deals, and changing hedge ratios (by shorting or buying back additional acquirer stock) in revised stock deals. These assumptions are similar to those of Mitchell's RAIM portfolio. The ERAP does not incur transaction costs, however, when reweighting the portfolio after a new deal is announced or an old one is completed. This is an assumption that biases returns upward. Mitchell and Pulvino avoid this problem in their RAIM portfolio by never reweighting: the portfolio is seeded with \$1 million in 1963 and cannot borrow money.<sup>26</sup> Even if an enormous new deal is announced and the RAIM portfolio has only a little available cash, only a small position will be taken in the deal, and no existing deals unwound or loan taken out to fund a larger position. If the portfolio is already fully invested, the deal is

skipped altogether, until an existing deal closes and cash is freed up.

Mitchell and Pulvino's assumption introduces its own bias, however, because the RAIM portfolio does not accurately reflect the cross-section of deals on a given day, but rather is restricted by Mitchell and Pulvino's choice of \$1 million with which to fund the RAIM.<sup>27</sup> The bias introduced in the ERAP by excluding transaction costs from reweighting is somewhat offset by its freedom to spread capital among positions as the deal universe expands or contracts, an alternative available to most funds in real life but which the RAIM is denied.

#### ***D. Portfolio Returns Including Borrowing Costs***

This paper uses "borrowing costs" to reflect a common theme in European risk arbitrage, and one reason risk arbitrage has not been as popular in Europe as in the US: the difficulty of borrowing shares of the acquirer to short in stock deals.<sup>28</sup> The "cost" of borrowing shares in Europe is often reflected in a significantly decreased rebate. In the U.S., as Baker and Mitchell correctly assume, the rebate is often the risk-free interest rate, sometimes higher. In Europe, however, rebate on all but

## The Harvard College Economist

Average Acquirer Volume for the 5 Days Following Merger Announcement	Rebate (%)
500,000 shares or above	rf [the risk-free rate in acquirer's country]
400-500,000 shares	(rf)-1
350-400,000 shares	(rf)-2
300-350,000 shares	(rf)-2.5
250-300,000 shares	(rf)-3.5
200-250,000 shares	0
150-200,000 shares	0
100-150,000 shares	-5
0-100,000 shares	-10

the largest and most liquid stocks can be two or three percentage points below the risk-free rate, sometimes zero percent, and in some cases negative, meaning the short seller actually pays a “fee” to borrow shares. In real life, these rebates or fees are negotiated privately between short sellers and lenders of shares, who often have repeated interaction with each other, so coming up with a scientific formula to approximate the costs is almost impossible. For this paper, the ERAPs “with borrowing costs” are assumed to face the following schedule of rebates based on a proxy for acquirer stock liquidity:<sup>29</sup>

Because rebates are paid on short proceeds, the calculation of net return on the overall deal must also in-

clude the cost of carry on the long position, which is assumed to be the risk-free rate in the target's country. Thus, the daily return equation for deals including borrowing costs is

$$R_b = (P_t^T + D_t^T - P_{t-1}^T - r_{f0}^*P_0^T + r_{e0}^*P_0^A) / P_{t-1}^T - r(P_t^A + D_t^A - P_{t-1}^A) / P_{t-1}^A,$$

where  $R_b$  represents daily deal return under borrowing costs,  $r_{f0}^*P_0^T$  is the daily cost of carry (calculated by multiplying the daily risk-free rate on  $t=0$ , the day of the target stock's purchase, by the price of the target at  $t=0$ ), and  $r_{e0}^*P_0^A$  represents the daily rebate, calculated by multiplying the daily rebate rate at  $t=0$  by the acquirer stock price at  $t=0$ .

Baker and Mitchell do not sub-

tract the cost of carry when calculating their gross returns, instead incurring it implicitly when calculating returns over the risk-free rate and performing regressions of these excess returns against the CAPM. Baker remarks that “when we net out the risk-free rate, it is arbitrage conditional on success: the position provides a positive and fixed profit if the merger is successfully completed and requires no money up front if the arbitrageur finances it with borrowed money.”<sup>30</sup> However, when testing portfolio returns in the CAPM, using an “arbitrage-conditional” risk-free rate seems somewhat misleading, as the risk-free rate used in the CAPM is not traditionally arbitrage-conditional. For this reason, this paper subtracts the risk-free rate (as the cost of carry) *before* the CAPM, in the calculation of actual portfolio returns, rather than actually in the CAPM. Consequently, for the ERAPs with borrowing costs, “returns over the risk-free rate” effectively have the risk-free rate being subtracted twice.

***E. Portfolio Returns with Transaction and Borrowing Costs***

The most realistic set of ERAPs faces both transaction and borrowing costs. In the return calculations for these

portfolios, the transaction costs are assumed to be incurred first, and are paid for by taking out a larger margin loan. Thus, the equation for daily deal returns under both costs is given by

$$R_z = [P_t^T + D_t^T - P_{t-1}^T - X_t^T - X_t^A - r_{f0} * (P_0^T + X_0^T + X_0^A) + (r_{e0} * P_0^A)] / P_{t-1}^T - r(P_t^A + D_t^A - P_{t-1}^A) / P_{t-1}^A$$

where  $R_z$  denotes daily deal return under both transaction and borrowing costs, and  $X_0^T$  and  $X_0^A$  represent the transaction costs incurred buying the target and shorting the acquirer on the day the position was initially set up. For instance, in the London Shop deal, the cost of buying one share is  $330p + 0.005 * 330 + 0.005 * 330 = 333.3p$ , because of the 1.65p brokerage commission and the 1.65p stamp tax. The arbitrageur takes out his loan for 333.3p instead of 330p, and thus pays more interest, for the same share of stock, than in the solely borrowing-cost portfolio. Not surprisingly, returns on ERAPs with transaction *and* borrowing costs included are always lower than those for identical ERAPs that face only transaction costs or only borrowing costs.

***F. Currency Hedging***

Implicit in the construction of these portfolios are two assumptions

## The Harvard College Economist

regarding foreign exchange activity: that currencies can be bought and sold at no cost, and that currency risk is perfectly hedged. In the majority of cases, currency hedging does not add or subtract significantly from deal returns.<sup>31</sup> Thus, most of the time, hedging foreign exchange risk does not significantly add or detract from overall returns, and so omitting this activity from arbitrage portfolio returns should not bias the results one way or another.

### V. Returns to European Risk Arbitrage

First, monthly returns for various portfolios are presented. Differences across stock and cash deals, value-and equal-weightings, and assumptions of portfolios assuming no costs, only transaction costs, only borrowing costs, or both types of costs, are analyzed. These returns are compared with US returns calculated by Baker and Mitchell. Finally, the returns are regressed in a Capital Asset Pricing Model to determine their performance relative to a European market portfolio. These “abnormal” returns are compared to results from Baker and Mitchell’s regressions of U.S. excess returns.

#### A. *All Deals, Cash Deals and Stock Deals, and The Effect of Transaction Costs*

Monthly returns for the ERAPs are calculated by compounding daily returns. Overall, returns indicate that European risk arbitrage generates positive returns with low standard deviation. The value-weighted raw portfolio of all deals earns a mean monthly return of 0.93 percent (implying a mean annual return, when monthly returns are compounded, of 11.7 percent), while the mean equal-weighted raw portfolio return is 0.88 percent (11.1 percent annually). In a paired-sample *t*-test, these returns are not significantly different ( $p=0.65$ ), indicating the choice of weighting method does not meaningfully affect returns.

These raw returns from Europe appear fairly similar to the mean monthly US returns of 1.25 percent (14.3 percent annualized) for a raw, value-weighted portfolio in the Baker paper,<sup>32</sup> and 0.93 percent for Mitchell’s VWRA portfolio, which incurs no costs but is also exempt from the 10 percent position size limit.

When transaction costs are included, the mean monthly returns to European arbitrage decrease to 0.76 percent (8.2 percent annualized) and

0.64 percent (7.4 percent annualized) for the value-weighted and equal-weighted ERAPs of all deals, respectively. These results are not very different from Mitchell's RAIM portfolio, which includes transaction costs plus other practical constraints, and which yields a mean monthly return of 0.75 percent (9.3 percent annualized). Two-tailed paired-sample *t*-tests of the null hypothesis that the returns with transaction costs are equal to raw returns, yield  $t = 5.61$  ( $p < 0.0001$ ) for the value-weighted ERAP and  $t = 5.49$  ( $p < 0.0001$ ) for the equal-weighted portfolio, indicating transaction costs have a statistically significant effect on returns. One-tailed tests of the null hypothesis that returns including transaction costs are greater than raw returns, also indicate significance at the  $p < 0.0001$  level, with  $t > 5$  for both the value- and equal-weighted portfolios. There is reason to reject the null hypothesis, indicating that transaction costs have a statistically significant negative effect on raw returns, a result supportive of the Mitchell and Pulvino theory.

Cash deals earn a much higher mean raw monthly return when compared to stock deals, 1.18 percent as opposed to 0.38 percent (robust for both value- and equal-weighted port-

folios), with slightly smaller standard deviations. The raw cash portfolio outperforms the raw stock portfolio in 87 of the sample's 132 months, or 66 percent of the time. Still, by looking at Baker's results, both cash and stock deals in Europe appear to underperform cash and stock deals in the US.

In a portfolio of cash deals, one would expect transaction costs to have a relatively smaller effect on raw returns for cash deals than for stock deals, because arbitraging cash deals entails incurring transaction costs only on the long side of the trade, whereas stock deals face transaction costs setting up both the long and short positions. This expectation is strongly confirmed by the results. The mean monthly return for a cash portfolio with transaction costs is 1.05 percent, a decrease of 13 basis points, or 12 percent of overall value, from the raw cash return. The stock portfolio with transaction costs sees its monthly returns cut by more than half, however, earning around 0.18 percent monthly, a decrease of 19 basis points from the raw return. These results are also robust for value- and equal-weightings.

### ***B. Returns Including Borrowing Costs***

Borrowing costs also appear to

## The Harvard College Economist

have a significantly negative effect on raw returns, although no comparable effect has been calculated for the U.S. with which to compare European results.<sup>33</sup> For the entire sample, monthly value-weighted returns with borrowing costs average 0.57 percent, a decrease of 36 basis points (39 percent) from the raw return. Equal-weighted monthly returns with borrowing costs average 0.48 percent, a decrease of 40 basis points (45 percent) from the raw return. Introducing borrowing costs should affect returns on cash deals about twice as much as stock deals, *ceteris paribus*, because both cash and stock deals charge the same cost of carry while only stock deals pay a rebate. Only if the acquirer shares in stock deals are illiquid, causing the rebate to be lower than the cost of carry, will stock deal returns suffer with borrowing costs. The expectation of greater cash deal effect is confirmed: borrowing costs reduce raw returns on cash deals by 34 basis points, and on stock deals by 15 basis points. The existence of a negative effect of borrowing costs on stock deals, however, suggests acquirer shares in Europe are somewhat hard to borrow and thus are charged a rebate below the risk-free rate.

For cash deals, the absolute dif-

ference (i.e., not percentage difference) between raw returns and returns with borrowing costs in a given month should be close to the weighted-average risk free rate for that month, because the borrowing cost *is* the risk-free rate.<sup>34</sup> Initially, results do not confirm this hypothesis. The absolute effect of borrowing costs in cash deals is 34 basis points per month, while the mean monthly Eurorate is 0.67 percent, or 67 basis points, and these two values are statistically different at  $p < 0.0001$ . A closer look at the data, however, finds two years in which there were significant periods (a total of 5 months) in which there were no cash deals at all: 1992 and 1993. During these periods, the portfolio had all of its capital in the bank, earning the Eurorate and not paying a cost of carry on any deals. Thus, for these periods, the effect of borrowing costs on the portfolio would be biased downward. Eliminating 1992 and 1993 from the analysis (both for the portfolio returns and for the Eurorate), the mean monthly absolute effect of borrowing costs is 86 basis points, and the average monthly Eurorate is 66 basis points. These results are statistically different only at the  $p = 0.15$  level.

A one-tailed test of the null hypothesis that raw returns are greater than

returns with borrowing costs, yields  $t > 4$  ( $p < 0.0001$ ) for both value- and equal-weighted portfolios of all deals, cash deals and stock deals, indicating borrowing costs also have a statistically significant negative impact on arbitrage returns.

### ***C. Returns Including Transaction and Borrowing Costs***

Returns with both transaction costs and borrowing costs included should be statistically lower than returns that include one or the other (or none), because an arbitrageur facing both costs must borrow more in order to pay transaction costs on both the long and short sides, raising total interest payments and decreasing overall returns. ERAP results confirm this intuition. For the entire sample, average monthly returns with both types of costs included are 0.46 percent using value-weighting, and 0.44 percent using equal-weighting. Compared to returns with only transaction costs, both the value- and equal-weighted portfolios of all deals including transaction and borrowing costs are significantly lower ( $t > 5$ ,  $p < 0.0001$ ). Compared to returns with only borrowing costs, returns with both costs included are also significantly lower.

ERAPs of only cash deals are also significantly affected when both costs are included. For cash deals facing transaction costs *and* borrowing costs, the mean monthly return is around 0.70 percent, significantly less than cash deal returns with only transaction costs ( $t = 17$ ,  $p < 0.0001$ ) or only borrowing costs ( $t = 11$ ,  $p < 0.0001$ ). The results for portfolios of stock deals show a similar, statistically significant negative effect of combining transaction and borrowing costs.

### ***D. UK Deals Versus Non-UK Deals***

Looking at returns of portfolios that include only British mergers and only non-British mergers, makes it clear that transaction costs and borrowing costs both affect UK deals much more than they affect non-UK deals. When no costs are assumed, UK deals outperform non-UK deals by 23 basis points, 0.90 percent per month to 0.67 percent. However, as soon as any costs are added, UK returns fall below non-UK returns. The transaction-cost effect, 46 basis points in the UK to 13 basis points outside the UK, is not surprising, given that UK trades, because of the stamp tax, cost twice as much as non-UK trades. The borrowing-cost

## The Harvard College Economist

effect is more surprising: borrowing costs decrease UK deal returns by 49 basis points per month, compared to 16 basis points for non-UK deals. The results suggest that acquirer shares tend to be much less liquid in the UK than elsewhere: This result makes sense when considering that the UK merger market is more mature than in continental Europe. In the UK, smaller companies are comfortable with making acquisitions, but in the rest of Europe, only very large, liquid acquirers have endeavored public takeovers.

### *E. Returns over the Risk-Free Rate*

With an average monthly Eurorate of 0.67 percent (8.3 percent annualized), the raw portfolios including all deals in the sample generate a mean value-weighted monthly return of 0.26 percent (3.3 percent annualized) and a mean equal-weighted monthly return of 0.21 percent (2.6 percent annualized) over the risk-free rate (these returns will hereafter be called “excess returns”). These excess returns are statistically greater than zero at one-tailed  $p < 0.0001$ , suggesting European risk arbitrage outperforms a riskless security when transaction and borrowing costs are not included.<sup>35</sup> However, the

contrasting performance of cash and stock deals is made clear by subtracting the risk-free rate. A raw portfolio of only cash deals generates a statistically significant *positive* mean monthly excess return, 0.51 percent (6.3 percent annualized), robust for value- and equal-weightings, while stock deals exhibit *negative* mean monthly excess returns of 0.30 percent (-3.5 percent annualized), also robust across weightings. This negative excess return is statistically less than zero at one-tailed  $p = 0.024$ .

When both transaction and borrowing costs are included, the ERAPs of all deals and cash deals have excess returns indistinguishable from zero (-0.23 percent and 0.03 percent, respectively), while stock deals underperform the Eurorate by a statistically significant 0.48 percent.

### *F. Regressions Versus a European Market Portfolio*

In a standard CAPM, monthly ERAP returns over the risk-free rate are regressed on the monthly return of the Euroindex over the risk-free rate to measure abnormal, risk-adjusted returns. Results, presented in Table 3, indicate that the only portfolios which would statistically outperform the Euroindex are the value- and equal-

weighted portfolios of cash deals assuming no costs ( $\alpha=0.49$  percent per month, or 6.0 percent annually, statistically significant at  $p<0.0001$ ), and the value- and equal-weighted portfolios of cash deals assuming only transaction costs ( $\alpha=0.36$  percent per month, significant at  $p=0.01$ ). Portfolios of all deals fail to generate returns statistically different from zero, while stock deals display statistically *negative* alpha in scenarios incorporating any type of practical cost.

### ***G. Discussion of Results***

Two features of the results are particularly surprising. The first is the overall poor risk-adjusted performance of European risk arbitrage relative to US arbitrage, even when practical costs are not taken into consideration. Possible determinants for this result include the longer waiting period to receive payment; greater reluctance of, or simply a lack of interest by, target company shareholders in Europe to sell their shares at a meaningful discount to the value they will receive if the deal is completed; and a deal failure rate that is greater in Europe than in the United States.

A “failed deal” is defined as a transaction in which an acquirer withdraws its bid without another, ultimately

successful bid already having been launched for the target. Thus, if there is a bidding war for a target company, and one bidder eventually drops out of the bidding while the other suitor successfully completes the acquisition, the withdrawn deal is not counted as a “failure,” because the returns to arbitraging such a situation are likely to be positive if any deal is consummated at all. If both (or all) suitors in a bidding war end up failing, and the target remains independent, all offers are counted as “failures.” As shown in Table 6, returns on failed deals range anywhere from  $-80$  percent to positive 32 percent, with returns for an individual deal calculated by compounding the daily returns for each day the deal was in progress, excluding any transaction or borrowing costs. Based on the target market capitalization, the average return on a failed European deal is  $-5.2$  percent, with British failed deals returning  $-16.5$  percent, and failed deals in continental Europe actually yielding a positive return of 2.2 percent.

Overall, the European deal failure rate (defined as number of failed deals per year divided by deals announced in that year) is 10.6 percent, with UK deals failing at a 10.2 percent rate. This is not drastically different from estimates of US failure rates, which

## The Harvard College Economist

range from 5 to 15 percent depending on year.<sup>36</sup> If losses on failed deals are to help explain differences in risk arbitrage returns across countries, then British risk arbitrage, with a failure rate similar to the other European countries but a loss per failed deal significantly greater than any other nation, should exhibit lower returns than non-UK risk arbitrage, practical costs notwithstanding. However, the evidence from the sample does not support the deal-failure theory: UK deals actually outperform non-UK deals by 23 basis points per month when practical costs are not included.

The second surprising feature of the results is the considerably poorer performance of stock deals relative to cash deals. Deal failure does not appear to be an explanation, as one can observe that failed cash deals in Europe actually suffer sharper losses than failed stock deals. The effect of target shareholders' unwillingness to sell their stock to arbitrageurs should not be exacerbated in stock deals, and in fact should be smaller, if target shareholders prefer to cash out after a takeover announcement rather than have the value of their stock tied to the acquirer's performance. The poor performance of stock deals remains a puzzle.

## VI. Conclusion

This paper, along with being the first to document returns to a risk arbitrage strategy in Europe, also sheds light on a continuing debate over risk arbitrage in general by examining the effects of transaction costs and other practical constraints on European arbitrage returns. Results support Mitchell and Pulvino's contention that transaction costs have a statistically significant negative effect on arbitrage returns, and that previous studies have overestimated risk arbitrage returns by not including these costs. Limited *ex-post* support is provided for Baker's theory of limited arbitrage, which suggests that returns to risk arbitrage will decrease as arbitrage capital increases, and vice versa. Looking at the European returns themselves, European risk arbitrage does not appear to be a very profitable strategy, especially when compared to a European risk-free rate and European market return. Either the risk arbitrageurs flocking to play European deals are seeking returns they will not realize, or they are extremely skilled at picking which deals to arbitrage and which to avoid. The near future, as European mergers continue to increase and capital markets in Europe become more mature, will doubtless bring further op-

portunities to re-evaluate risk arbitrage in Europe, specifically, and as an investment strategy in general.

## References

<sup>1</sup> According to conversations with Mitchell and Pulvino, almost every large risk arbitrage hedge fund has set up a London office in the last two or three years to invest in European deals. One practicing European arbitrageur estimates that capital dedicated just to arbitrage on the Continent (i.e., excluding the UK, where arbitrage has been prevalent for a longer period of time) has increased from \$2 billion in 1996 to \$30 billion in 2001.

<sup>2</sup> Jindra and Walkling (1999).

<sup>3</sup> Hereafter referred to as “Baker”, unless otherwise noted.

<sup>4</sup> Hereafter referred to as “Mitchell”, unless otherwise noted.

<sup>5</sup> These papers include Jindra and Walkling (1999), Dukes, Frohlich and Ma (1992), and Karolyi and Shannon (1998).

<sup>6</sup> Because the monetary cost of acquiring data for all European countries was prohibitively high, the sample had to be restricted to these seven countries. Thus, the results exclude returns for some countries where there has been a reasonable amount of M&A activity,

including Belgium, Sweden, Finland, Denmark and Portugal. The author, based on conversations with risk arbitrageurs who invested in deals throughout Europe over the period covered by this paper, does not believe that the omission of these countries introduces any substantial bias to the results.

<sup>7</sup> Both the Eurorate and Euroindex are calculated based on the geographic distribution of deals across countries in a given year. Thus, of the target companies of deals announced in year  $t$ , if 68% came from the UK, 15% from France, and 17% from Germany, the Eurorate for year  $t$  would be calculated using weights of 0.68 for the UK risk-free rate, 0.15 for the French rate, and 0.17 for the German rate, and the Euroindex return would have similar weightings for the FTSE 100, CAC 40 and Xetra DAX indices, respectively.

<sup>8</sup> Once again, based on conversations with risk arbitrage professionals.

<sup>9</sup> Both Baker and Mitchell assume that the rebate is always the risk-free rate. They admit this assumption is somewhat unrealistic, as rebates can be greater than the risk-free rate, if acquirer shares are extremely easy to borrow, or less than the risk-free rate if they are hard to borrow. In many European deals, the rebate can be close to zero or even nega-

## The Harvard College Economist

tive, implying that the short seller is in effect paying to borrow stock. This paper adjusts the rebate in its portfolio calculations, based on the expected degree of difficulty in borrowing acquirer shares, to make the results more real.

<sup>10</sup> Baker and Savasoglu, *Limited arbitrage in mergers and acquisitions*, p. 7.

<sup>11</sup> Mitchell and Pulvino, *Characteristics of Risk and Return in Risk Arbitrage*, p. 11.

<sup>12</sup> From conversations with Baker and Mitchell, the author believes they excluded “mix” deals because, in breaking down their samples into cash and stock deals to do subanalysis, it would be difficult to figure out where to place deals with both cash and stock. There appears to be no reason why excluding mix deals biases their results in any way, because their samples are quite large and the number of mix deals in the US was relatively quite small. Because the number of mergers in Europe (and thus candidates for inclusion in the sample) is much smaller, including mix deals was necessary to generate more robust results. For subanalysis purposes, this paper places mix deals into the category of stock deals, because the transaction costs and rebate discounts on the short position, regardless of the percentage

of the entire deal that the stock portion represents, are large enough to make a difference to the overall return, and thus carry an important characteristic of 100% stock deals.

<sup>13</sup> Obviously, in real life, one alternative may start out as more attractive but then become less and less so as the deal progresses. For instance, if an arbitrageur starts out treating the deal as a stock deal, and shorts the acquirer, and then the acquirer stock plummets to a point where the cash alternative is worth much more, he would cover his short and remain “naked long” the target. [Indeed, the return to choice deals is identical to that from being long the target plus long a put option to sell the target for the cash offer price, so the real-life arbitrageur would likely have used derivatives to hedge the deal instead of choosing one option over the other]. Fortunately, no deals in the sample suffered such dramatic shifts in the relative value of a cash vs. stock alternative. Thus the simplifying assumption of selecting the alternative based on one-day post-announcement value does not bias the results to any “choice” deal significantly upward or downward.

<sup>14</sup> Almost all British “cash” deals were really “cash and loan note” deals. The loan notes are highly liquid, however, so

a deal described by SDC as having terms of “355 pence a share in cash and loan notes” is considered to be a deal for 355p in cash.

<sup>15</sup> Extra care had to be taken to make sure the right security was selected. For instance, SDC reports that Credit Suisse Group took over Winterthur Schweizerische in 1997. However, a *Datastream* search for “Credit Suisse Group” turned up 10 different codes, each for a different class of stock or warrant, some delisted and some trading on exchanges other than the Swiss exchange. To get the right security, prices for all 10 codes were downloaded and compared to information from the SDC summary and Bloomberg articles to narrow down the candidates. In many cases, acquirer names changed from the time they were inputted into SDC. For example, the Swiss drug firms Sandoz and Ciba-Geigy merged in 1998. The code for Sandoz was easy enough to find, but Ciba-Geigy’s prices actually turned up under the code for Novartis, the name of the merged entity.

<sup>16</sup> On the other hand, if a target stock had an illiquid ADR but liquid local shares, it was included in the sample. For instance, both Vodafone and Mannesmann had shares traded on the New York Stock Exchange, but their

local shares were much more liquid than their ADRs. Consequently, in this stock deal, most arbitrageurs bought Mannesmann in Frankfurt for euros, and sold Vodafone short in London, receiving pounds.

<sup>17</sup> Deal announcement, completion and withdrawal dates are as reported by SDC (for 1989 through 1992) and *Bloomberg* (for 1993 through 1999). The assumed ten-day waiting period to receive payment after a deal is completed is based on conversations with professional arbitrageurs.

<sup>18</sup> Notice the ERAP does not use its short proceeds to pay for its long position. The two transactions, long and short, are completely separate. The ability to open a short account also assumes the ERAP has enough initial funding to post the amount of capital required by the broker in the acquirer’s country.

<sup>19</sup> Dollar-market-cap information for foreign stocks was downloaded from *Datastream*.

<sup>20</sup> Mitchell and Pulvino, *supra*, p.17.

<sup>21</sup> Payment in US mergers usually occurs within three business days following completion of the deal. In Europe, the payment time differs across countries and even across deals, but conversations with practicing arbitrageurs indicate that two weeks, or ten business

## The Harvard College Economist

days, is a good approximation, even in the UK. In deals with cash or stock elections, payment can take even longer to occur, because the actual election of consideration is sometimes not held until a week or so after the deal formally closes. Since there are not many deals of this type in the sample, the ten-business-day assumption is still used.

<sup>22</sup> All these calculations assume the portfolio includes borrowing costs. For the raw portfolios, and those assuming only transaction costs, the cost of carry would be zero. The ERAP would borrow 330p on January 6<sup>th</sup>, receive 340p on February 15<sup>th</sup>, repay the principal, and keep 10p as profit. Obviously, in the case of cash deals, portfolios with borrowing costs will always have lower returns than raw portfolios. In stock deals, portfolios with borrowing costs might actually earn higher returns than raw portfolios, if the cost of carry and rebate are equal (or negligibly different). The reason is this: as Baker notes, in most stock deals the arbitrageur is actually net short, because the value of the long position is a little less than the value of the short (hence the arbitrage). Thus the total interest earned on the short will be greater than the total interest paid on the long.

<sup>23</sup> In a large majority of European deals,

the target stock is not delisted until months, or even years, after the deal has been completed. For example, Paribas SA stock still trades on the Paris Bourse, albeit with very little volume, even though Paribas was formally acquired by BNP in August 1999. Even though the acquirer stock still trades and fluctuates, the arbitrageur does not care about its price if the deal is done because he is guaranteed delivery of the shares, and will be able to return them to the lender no matter what their value.

<sup>24</sup> Various reports attempt to enumerate transaction costs by country. For instance, BrainBank Finance (1997) provides a summary of all fixed transaction costs in markets worldwide. However, for the European markets in question, “brokerage commissions” are characterized as “negotiable” (except for the Italian Stock Exchange, which allows a maximum commission of 70 basis points). The 50bp assumption attempts to account for these “negotiable” commissions.

<sup>25</sup> For most stocks, then, these numbers mean that direct trading costs are higher in the Europe than in the U.S, even without the country-specific government fees. A stock that trades at a price of \$10 in New York will have a direct trading cost, according to Mitchell, of \$0.04

per share. Now assume that the stock trades, in dollars, on the Paris Bourse. Its direct trading cost will be 50 basis points times \$10, or \$0.05 a share. Thus, unless a stock trades under \$8, in which case its commission in Europe would be \$0.04, the cost of trading in Europe will be higher. Many thanks to Patrick Burke for these numbers.

<sup>26</sup> Mitchell's assumption that the RAIM does not borrow money is not trivial. Shleifer and Vishny make clear in "The Limits of Arbitrage" (1997) that ultimate success in any type of arbitrage depends on who supplies the arbitrageur's capital. If the arbitrageur's "initial funds" come from other people's money, as is the case for institutional investors and margin investors, an agency problem arises if prices initially diverge. The arbitrageur gets excited as he sees a greater profit to be made, but his investors and creditors, who "do not know or understand exactly what he is doing, will only observe him losing money . . . They may infer from this loss that the arbitrageur is not as competent as they previously thought, refuse to provide him with more capital, and even withdraw some of the capital, even though the expected return from the trade has increased." Their analysis eerily presages the downfall in the summer and fall of

1998 of large arbitrage (though not necessarily merger arbitrage) funds such as Long-Term Capital Management, Ellington Management and D.E. Shaw, which were forced to liquidate positions at large losses to meet margin requirements and fund redemptions when their positions moved too far against them.

<sup>27</sup> As Mitchell and Pulvino admitted in conversation, a portfolio seeded with only \$1 million, even in the 1960s, would have to forgo or limit its positions in many deals. Curiously, when they assume seed money of \$10 million in their paper's sensitivity analysis, annual returns (with direct and indirect transaction costs) actually decrease from 10.6% to 6.9%. Perhaps lower returns to the larger fund were caused by a greater portion of that fund having to be held in cash to avoid incurring high indirect transaction costs by building large positions in illiquid stocks.

<sup>28</sup> In a well-known deal in 2000, Terra Networks of Spain acquired Lycos, a US internet portal trading on NASDAQ, in a stock transaction. Notwithstanding any deal-specific risk, the arbitrage spread remained astoundingly wide up to the date of deal closure, simply because arbitrageurs could not borrow Terra in Spain to hedge potential long positions in Lycos. This deal

## The Harvard College Economist

also illustrates why arbitrageurs almost never settle for being “naked long” the target if they cannot short the acquirer. Between May 2000, when the deal was announced, and October, when it closed, internet stocks plunged, and an investor who was naked long Lycos would have taken a significant hit. Had he been able to short Terra, however, his loss on Lycos would have been offset by a profit made on the short position, because Terra, also an internet portal, was pummeled as well.

<sup>29</sup> Once again, many thanks to Patrick Burke for guidance in creating this schedule. five-day post-announcement trading volume was used to approximate liquidity because it is in these five days that an arbitrageur will likely look to borrow shares and sell short. Of course, liquidity is not an exact proxy for relative difficulty of borrowing, but it is in fact a measure arbitrageurs consult for an initial impression of the likelihood of being able to borrow a stock at a reasonable price.

<sup>30</sup> Baker and Savasoglu, *supra*, p.9.

<sup>31</sup> Once again, based on conversations with professional arbitrageurs. The notion that currency hedging is unimportant with respect to overall returns does not apply to all international investments. Recent results from internationally-di-

versified mutual funds indicate that funds which did not hedge currency risk in 1999 and 2000 vastly underperformed those funds which did hedge foreign exchange risk. See Aaron Lucchetti, “Fund Managers Disagree on Value of Currency Hedging,” *Wall Street Journal*, February 2, 2001, p.C1.

<sup>32</sup> This result incorporates only the years 1989 through 1996, the year Baker’s sample ends. Returns on the entire Baker sample, representing 1978 through 1996, are around 1.82% per month, value-weighted. Many thanks to Malcolm Baker for providing monthly return data for his portfolio.

<sup>33</sup> The Mitchell RAIM portfolio deals with illiquidity in acquirer shares by excluding those deals altogether. This assumption is one of only many in the RAIM regarding “practical constraints.” Thus, the individual effect of “borrowing costs,” or acquirer illiquidity, cannot be determined.

<sup>34</sup> The two numbers (absolute effect of borrowing costs, and weighted-average risk-free rate) should not be *exactly* the same, because the formulas that determine them are different. The determination of borrowing costs reflects the actual geographic distribution of deals in a given month, assuming deal sizes are relatively stable across countries,

while the Eurorate is determined by the geographic distribution of deals in the previous calendar year. Thus, if month  $m$  falls in calendar year  $t$ , the Eurorate for  $m$  is dependent on the distribution of deals in year  $t-1$ , which is likely to be different than the actual distribution above.

<sup>35</sup> To check for robustness across choices of the risk-free rate,  $t$ -tests of the raw excess returns using the aver-

age UK risk-free rate were also performed. Excess returns in this case also proved to be statistically greater than zero.

<sup>36</sup> Estimates for the US are based on discussions with arbitrageurs. Baker and Mitchell attempted to predict deal failure using certain variables but did not actually provide results for the number of U.S. deals that failed in a given time period.

# A Tinkerer's Tale: Examining the Effect of Big Business on Inventors at the Turn of the 20<sup>th</sup> Century

*Mwashuma K. Nyatta*

## Abstract

Big Business came of age in America at the turn of the 20<sup>th</sup> century, having widespread effect not just on industry but also on wider society. Efficiency became a pervasive mantra and entrepreneurs met this demand for efficiency in innovative ways. In this paper, the American Inventor is presented as one of these entrepreneurs. Inventors have been shown to be very responsive to market conditions throughout economic history. In the Big Business era, this responsiveness took the form of a shift in types of inventions from final goods, to what the paper calls “process inventions.” Primary patent data shows that, during the period from 1870 to 1910, inventors produced an increasingly large number of inventions aimed specifically at increasing the efficiency of production processes. This paper also draws a connection between the observed rise in process inventions and the decline in inventor independence that occurred during the same period.

## I. Introduction

The image of the Yankee tinkerer is almost as ingrained in American self-perception as are hotdogs and democratic idealism. Portrayed as an important part of the engine that has driven American success and uniqueness over the centuries, the tinkerer is an icon familiar to most: the independent, practically-thinking, innovative individual, hammering away in his backyard, creating yet another wonder to make our lives easier. However, this image of the inventor to which we are so accustomed is one

based almost entirely on the popularized inventors of the 19<sup>th</sup> century – Edison, Fulton, Morse, Tesla – the list is endless. We rarely consider inventors in the periods after this “great inventor boom,” and little has been written about the role they have played in this country’s progress. One period of specific interest is the turn of the 20<sup>th</sup> century. This paper considers a forty-year window between 1870 and 1910 and analyzes the effects of the rise of big business on inventors. It illustrates that they were, like the inventors of old, an important,

albeit understated part of the “engine that drove American success,” though they occupied a different role than the more traditional inventors.

Part II of this paper will provide the historical context for the period in question, reviewing inventor trends, company trends in invention, and the rise of mass production and big business. Part III will deal with inventors and specifically their entrepreneurial characteristics, showing that they consistently respond to economic incentives. In Part IV, drawing from these two sections, this paper will put forward the hypothesis that the rise of big business and mass production led to a reduction in American society’s demand for new and novel items, instead shifting attention to the manner in which existing items were produced. “Efficiency” became a catchword and production processes were streamlined in order to produce the best goods for the largest number of people in the shortest time. The paper’s primary hypothesis is that inventors, conscious of the above-mentioned societal trends, exercised their entrepreneurial tendencies to produce more “process inventions,” to increase efficiency in production processes than they had previously. There was an overwhelming shift in the tinkerer’s role from someone that pro-

duced new goods to one who helped streamline processes for the mass-production of already existing products. The paper will support this hypothesis with primary data obtained from turn-of-the-century patent records that show a marked increase between 1870 and 1911 in process invention. In Part V, the paper will conjecture that the decline of inventor independence observed during this same period may have been related to this increase in process inventions. Part VI will provide a brief conclusion.

## **II. Efficiency and the Rise of Big Business**

The rapid spread of the railroad between 1870 and 1911 transformed America into a single economic space. It became easy to transport factors of production and finished goods throughout the country, creating, in the eyes of producers, a huge market waiting to be exploited. Mass production methods were perfected. One of the major ideas underlying mass production was the absence of “fitters” – workers whose job it was to file, shape, and otherwise refine various components, such as handgun parts, so that they would fit seamlessly into the finished product. Mass production methods attempted to

## The Harvard College Economist

eliminate fitters by manufacturing perfectly interchangeable parts, alike in every respect and crafted to fit together to ensure that only assembly was needed to produce the final product. Perfect interchangeability of parts therefore set the stage for assembly lines, the most famous of which is probably Henry Ford's, churning out hundreds of Model-T cars daily at his Highland Park Factory. This prolific production quickly became the norm in many industries and was titled "Fordism" in honor of the man who called most attention to it.

The Ford Motor Car Company is only one example of an important turn-of-the-century phenomenon in America, termed "the rise of big business," that spanned approximately the period from 1860 to 1920.<sup>1</sup> Embedded in the rise of big business was "The Great Merger Movement," which peaked between 1898 and 1902 and spawned several gigantic companies, such as Pittsburgh Plate Glass, General Electric, Nabisco, Dupont and Kodak, many of which still exist today. These businesses were characterized by multiple plants in multiple cities, a large scale of capital needs, particularly fixed capital, and complex operation. All these attributes of big business led to various complications and diseconomies of

scale, some of which are summarized as follows:

"Because of the scale and scope of their operations, the situation for big businesses was quite different . . . . The coming of the complex new technologies and the multi-site, multifunction companies had significant effects on the behavior of the firms involved. The many factories, mills, refineries, warehouses, blast furnaces, assembly lines, and distribution outlets represented enormous amounts of capital, so these firms experienced substantially higher constant costs than had their antebellum predecessors. This made it more costly to cease production when business turned bad .... Start-up costs were substantial, and market share might be lost to competitors during [a] slowdown."<sup>2</sup>

In other words, mass production methods, together with the technological and organizational complexities that surrounded them, required a focus on the functioning of big businesses to maintain their competitiveness in the market. The threat of competition and loss of market share compelled businesses to attempt to minimize their operating costs and maximize their production efficiently. The notion of efficiency played a crucial role in the interactions between big businesses and society and between big businesses and inventors specifically.

In addition to efficiency as mandated by the desire to be competitive, the notion of efficiency had also become a general societal mantra in the early 20<sup>th</sup>

century. Evidence for this idea is embodied in yet another doctrine, that of “Taylorism,” whose founder, Frederick W. Taylor, was obsessed with the “scientific management” of industry. One writer claims that “under [Taylor’s] leadership the movement took on the overtones of a great crusade. Taylor was the messiah, and his close followers were often referred to as disciples.” Moreover, “in a short time Taylorism became famous far and wide as the embodiment of the era’s love affair with the idea of efficiency.”<sup>3</sup> As supporters of Taylorism fervently preached efficiency, industrial leaders, foremen, and managers of big businesses grabbed the baton with zeal and urged improvements in their production processes not just because efficiency enhanced competitiveness but also because efficiency was trendy and “progressive.” Although many eventually concluded that Taylorism was a farce based on “a host of hidden arbitrary assumptions and subjective judgments,”<sup>4</sup> during the first decade of the 20<sup>th</sup> century, Taylor’s gospel of “efficiency, efficiency, efficiency!” was sacrosanct and as embedded in big business practice as mass production itself.

How is efficiency related to inventors? The next part of this paper will focus on inventors’ entrepreneurial char-

acteristics, demonstrating that inventors respond definitively to economic pressures. Combining inventors’ entrepreneurial spirit with big businesses’ emphasis on efficiency, many inventors responded to the rise of big business by changing their roles from groundbreaking pioneers of new final products to refiners in a corporate system that had replaced them as America’s unique and revolutionary economic trump-card.

### **III. The Inventor as Entrepreneur**

Thomas Edison is one of the world’s best-known inventors. Everyone, either directly or indirectly, benefits from the incandescent light bulb, Edison’s most famous invention. However, Edison’s achievements were not limited to the light bulb: over the course of his life, he received hundreds of patents for all sorts of gadgets, from stock tickers to congressional vote counters. He is a prime example of the quintessential inventor – the man dedicated to solving problems just because he has an inquisitive turn of mind and a desire to provide answers. This sort of inventor is, perhaps, the idealized “pure” inventor; he is “a man who is possessed or obsessed by the inventive faculty [and] invents because he cannot help himself.”<sup>5</sup>

## The Harvard College Economist

Though this is the stereotypical inventor that most people envision when they conceive of the process of invention, the actuality is more cynical. Even the most “pure” inventors, like Edison, could not eat their inventions (for the most part, anyway) and needed to sell their new products in order to feed themselves and their families. Edison learned this harsh lesson when one of his inventions was deemed unnecessary, thereafter redirecting “his attention to ideas which had possibilities of becoming commercially practicable.”<sup>6</sup> Another prolific and well-known inventor, Frederick L. Fuller, was also faced with economic constraints in his inventive activity. Regarding his employment by the National Cash Register Company, which had a large pool of funds available for his inventions and his salary, Fuller said, “I anticipated with great hopes the possibility of working where I was not hampered by lack of money.”<sup>7</sup> As succinctly phrased by one writer, “The most prevalent explanation of the cause of invention is the desire for economic reward.”<sup>8</sup> Clearly, inventors are constrained in their pursuit of solutions to life’s problems by economic demands.

Other studies confirm this picture of inventors as entrepreneurs. In a paper entitled “Schemes of Practical

Utility: Entrepreneurship and Innovation Among ‘Great Inventors’ in the United States, 1790-1865” (1993), authors Kenneth Sokoloff and Zorina Khan argue that all inventors, including the “Great Inventors,” responded to market conditions. Using primary patent data and biographical information, the paper supports its hypothesis rigorously, showing that all inventors in early industrial America directed their inventive activity to cater to market needs. Analyzing primary data, economist Jacob Schmookler arrives at the same conclusion: “inventive effort is responsive to economic pressures and opportunities.”<sup>9</sup> It is likely that inventors’ entrepreneurial tendency also existed during the period under consideration, indicating that inventive activity was sensitive to market conditions. There was greater specialization in invention as the 19<sup>th</sup> century progressed as documented by Lamoreaux and Sokoloff in their paper “Inventive Activity and the Market for Technology in the United States, 1840-1920” (1999). It also seems reasonable to postulate that turn-of-the-century inventors became more reliant on the successful sale of their patents or inventions than they were before. An inventor that specializes almost entirely in his craft has virtually no other sources

of income and thus relies exclusively on successful patent or product sale for sustenance. Products sell successfully only if there is demand for them; patents sell successfully only if the buyer of the patent perceives potential returns from sale of the product, which again relies on demand. Therefore, with the rise of big businesses in turn-of-the-century America, inventors were at least as dependent on market demand for their products as were inventors in early industrial America, mainly because these later inventors depended almost exclusively on patent-sales for income.

#### **IV. Patent Records and Process Inventions**

How did inventors' entrepreneurial characteristics play out in their interaction with the economic conditions that prevailed at the turn of the century? As mentioned in Part I of this paper, there was a rising tide of big businesses whose emphasis was on efficiency in production. Big businesses were quickly becoming the major players in the American economy; therefore, big business demand formed a major component of the total demand in the economy. Sensitive to market pressures, inventors were acutely aware of the returns to be exploited in meeting big businesses' de-

mand for efficiency in production. As a result, we would expect that at least some of them switched their efforts to inventing products that would meet this demand for efficiency. Investors achieved this by producing process inventions. These were mostly "microinventions" that did not introduce new goods into the market but rather improved the efficiency of existing production processes or presented new ways to produce existing products efficiently.

Data obtained from patent records supports the hypothesis that there was a notable shift in inventive activity from final products to production-related goods. This data was collected by randomly sampling one hundred patent records from each of three periods: 1870-1871, 1890-1891, and finally 1910-1911.<sup>10</sup> The inventions in each period were then divided into process inventions and other types of inventions. For simplicity, only the manufacturing sector was considered in determining which patents would count as process inventions. Therefore, for example, a new method for the efficient hoisting of product-parts would be included in process inventions, but a more efficient reaper-binder would not.<sup>11</sup> The hypothesis presented in this paper sug-

## The Harvard College Economist

gests a rise in the number of process inventions over the three periods, most notably in the 1910-1911 period, which fell almost a decade after “The Great Merger Movement” and well into the establishment of big businesses as America’s economic dictators.

As the hypothesis presented in this paper would predict, there is indeed a large jump from 1890 to 1910 in the number of patented process inventions. More specific results are tabulated below:

The included table shows that there was an increase in the number of process inventions from three percent in the first period, to four percent in the second period to 29 percent in the final period. In other words, the approximate proportion of total inventions devoted to production processes changed from small fractions to almost a third of the entire number of inventions. Whereas 29 percent of all inventions does not mean that most inventors started to produce process inventions, it does indicate that there was a marked shift in the number of inventors who did.

Viewing the above results in the context of the preceding material on big business and entrepreneurial inventors, it seems that inventors shifted to process inventions in response to the higher

demand for these inventions inspired by the notion that efficiency in production was paramount. Because this demand-shock occurred primarily in the immediate wake of “The Great Merger Movement,” as big businesses,

<b>Years</b>	<b>Total No. of Observations</b>	<b>No. of Process Inventions</b>
1870-1871	100	3
1890-1891	100	4
1910-1911	100	29

“Fordism” and “Taylorism” flourished, we would expect process inventions to be much more numerous in the years after 1902 than the years before. Indeed, this is the case. With the rise of big business and corporate structure, many inventors could no longer survive financially by inventing new gadgets and other final, consumer-oriented products—the hallmarks of the traditional tinkerer—that had previously earned them fame and economic reward. Instead, many inventors had to find a way to become a part of the new corporate regime that was quickly becoming the dominant paradigm of American society. One author states:

“Big business proved to be the seedbed of a new social and economic order. The new managerial class, governed by the engineering values of efficiency and systematic

approaches to problems, having first arisen to help create and then to serve the modern corporation, soon became the dominant element in an urban and then suburban civilization . . . Soon almost the entire society would come to be influenced by corporate ways of doing things.”<sup>12</sup>

As this “managerial class” became more economically prominent, inventors found themselves marginalized. They could not easily become part of management in business, especially since they had, particularly in the late 19<sup>th</sup> century, specialized in their craft of invention and ceased to take part in the active marketing and sale of their products.<sup>13</sup> Therefore, these inventors followed the economic trends to generate income by catering to the needs of this new corporate system and becoming cogs in a giant industrial wheel. Increasingly, inventors became merely process-inventors: producers of goods for factory-process consumption rather than final goods for consumers.

## **V. Inventor Independence**

This trend in process inventions noted in the period between 1870 and 1911 sheds some light on the reasons behind another trend in invention that occurred at about the same time – the decline in inventor independence. As the 20<sup>th</sup> century began, many inventors joined various companies, under whose

umbrella they continued their inventive work. Given the analysis on process inventions above, there are at least two reasons that the rise of big business and its effect on the types of inventions produced may relate to this loss of independence.

First, to create useful inventions, inventors needed to be familiar with the production processes for which they were inventing. The idea that outside inventors often provide the best solutions to a company’s problems, perhaps because of their unbiased perspective, is limited. To improve a complex production process, inventors must be intimately familiar with it. As one writer says, “When it comes to those things which are kind of peculiar to the nature of your business, where intimate knowledge of the day-by-day affairs are concerned, the outsider just cannot possibly know about that....”<sup>14</sup> Big business production processes were extremely complex and therefore required inventors with specific knowledge of their functioning. With the increasing degree of product homogeneity between businesses, one of the only ways to differentiate a product was by price, and the more efficient a production process, the cheaper a product. This suggests that firms would guard their production pro-

# The Harvard College Economist

cesses to maintain their competitive advantage. In addition, one would expect firms to retain inventors who had specialized knowledge of a firm's production process to prevent the loss of competitive advantage in the case of an inventor moving to another firm. One way to retain inventors at a particular firm and to directly control their inventive efforts would be to require inventors to sign an employment contract.

A second reason that reduced inventor independence relates to the rise in process-inventions is apparent when viewed from the inventor's perspective. Many big businesses established rudimentary invention departments to improve their products and foster the development of their production processes. These departments presented inventors with an opportunity for stable incomes and job security, which they would lack as tinkering vagabonds. Like the famous and prolific Frederick L. Fuller, who joined such a department in the National Cash Register Company,<sup>15</sup> we would expect many inventors to snap up the opportunities to have sufficient funds for invention and regular incomes. If the demand for efficiency led to the establishment of invention departments to foster process invention, the rise in process inventions and loss of inventor

independence are closely linked. Considering the reduction of inventor independence as a result of the rise of big business may be a useful first step in understanding this relationship.

## VI. Conclusion

This paper has dealt with inventors at the turn of the century and shown how the rise of big business, the increasing prominence of a "corporate class" and especially big businesses' demand for efficient production substantially affected inventors. Though partially driven by curiosity and the desire to solve problems, inventors were also economically constrained and forced to produce inventions that corresponded to market demand. Therefore, big business imposed market forces that led to an increase in process inventions – inventions whose purpose was to facilitate efficiency in production. The paper finally suggests that the loss of inventor independence at the turn of the century was linked to the rise of big businesses and process inventions.

Yankee tinkerers occupied a special place in the societal structure of the time and helped revolutionize the world by tinkering in their backyards. The turn of the century may have been the backdrop for a shift in emphasis from

mechanical genius to corporate genius. It is difficult to conceive many great inventors whose ideas were patented after 1910. The reasons for this could be numerous, including the institutionalization of invention and growth of research laboratories. This paper does not claim to provide any comprehensive answer to this puzzle. However, the paper does begin to provide ways to look at the history of inventors and particularly the intriguing and often neglected period in their history that occurred at the beginning of the 20<sup>th</sup> century.

<sup>1</sup> These dates are as estimated in Glenn Porter's book, *The Rise of Big Business* (1992).

<sup>2</sup> Porter, 11.

<sup>3</sup> *Ibid*, 106 and 107.

<sup>4</sup> *Ibid*, 107.

<sup>5</sup> Statement made by an inventor taken from "Hearings before House Committee on Patents, Oldfield Revision and Codification of the Patent Statutes, 1912, No. 3., 5.

<sup>6</sup> Patterson, *America's Greatest Inventors*, 129.

<sup>7</sup> Fuller, *My Half Century as an Inventor*, 128.

<sup>8</sup> Vaughan, *Economics of Our Patent System*, 15.

<sup>9</sup> Schmookler, "Inventors, Past and

Present," 18.

<sup>10</sup> For more on the sampling procedure, please see the Appendix, Section 1: Sampling (online).

<sup>11</sup> For more on how the inventions were classified, please see the Appendix, Section 2: Classifying the Inventions (online).

<sup>12</sup> Porter, 92.

<sup>13</sup> This specialization in invention is given thorough treatment in Naomi R. Lamoreaux and Kenneth L. Sokoloff's "Inventive Activity and the Market for Technology in the United States, 1840-1920" (1999).

<sup>14</sup> Folk, *Patents and Industrial Progress*, 159.

<sup>15</sup> From Fuller's autobiography, *My Half Century as an Inventor*.

<sup>16</sup> This is a statistical test used for two samples that are uncorrelated and that have unequal variance.

## References

- Centennial Celebration of the American Patent System, 1836-1936*, United States Government Printing Office. 1937.
- Folk, George E. *Patents and Industrial Progress*. Harper and Brothers. 1942.
- Fuller, Frederick L. *My Half Century as an Inventor*. 1938.

## The Harvard College Economist

- Hatfield, H. Stafford. *The Inventor and His World*. Kegan Paul, Trench, Trubner and Co. Ltd. 1933.
- Hearings Before House Committee on Patents, 1912. United States Government Printing Office, 1912.
- Hutchins, John G. B. "Recent Contributions to Business History: The United States." *Journal of Economic History*, Vol. 19, No. 1. (Mar., 1959), pp. 103-121.
- Lamoreaux, Naomi R. and Sokoloff, Kenneth L. "Inventive Activity and the Market for Technology in the United States, 1840-1920." NBER Working Paper 7107. May 1999.
- Patterson, John C. *America's Greatest Inventors*. Thomas Y. Crowell Co. 1943.
- Porter, Glenn. *The Rise of Big Business 1860-1920*. Harlan Davidson, Inc. 1992.
- Schmookler, Jacob. "Inventors Past and Present." *The Review of Economics and Statistics*, Volume 39, Issue 3. August, 1957.
- Sokoloff, Kenneth. "Inventive Activity in Early Industrial America: Evidence From Patent Records, 1790-1846." *The Journal of Economic History*. December 1988.
- Twain, Mark. *A Connecticut Yankee at King Arthur's Court*. Justin Kaplan, 1971.
- Vaughan, Floyd. *Economics of Our Patent System*. The Macmillan Company. 1925.
- Woodbury, Robert S. *Studies in the History of Machine Tools*. The Massachusetts Institute of Technology, 1972.

# Gibrat's Law for US Cities: A Test

*Radim Rimanek*

## Abstract

Many man-made and naturally occurring phenomena, including city sizes, are distributed with surprising accuracy according to power laws. Specifically, Zipf's law for cities implies that the number of cities with a population larger than  $P$  is proportional to  $1/P$ , a relationship confirmed empirically for different parts of the world. No generally accepted explanation for this relationship had been offered until Xavier Gabaix in 1999 showed a straightforward mathematical proof leading to Zipf's distribution for cities. His account, however, relies on a critical assumption that the growth of cities follows Gibrat's law. We examined city size data for a group of more than 50 of the largest US cities from 1970 until 1998 to prove or disprove the assumption of Gibrat growth. We found that there is a statistically significant difference in the variance of growth rates across city sizes. This finding constitutes a violation of Gibrat-compliant growth. Having disproved Gabaix's critical assumption, we conclude that his account of the emergence of Zipf's law for cities is questionable.

## I. Introduction

Zipf's law determines very tightly the admissible boundaries for models of local growth. (Gabaix, 1999). The implication of this power law is that the number of cities<sup>1</sup> with a population larger than  $P$  is proportional to  $1/P$ . This relationship has been confirmed empirically in many studies for different parts of the world. This is quite striking, given the numerous (relatively unsuccessful) attempts in the past decades by authors to explain this conspicuous relationship (see for example Krugman, 1996 for a summary of theories). Finally, it has been

shown conclusively by Xavier Gabaix (1999) that a plausible mathematical explanation exists for the appearance of Zipf's law for cities. However, his account relies on one crucial assumption that the growth of cities follows Gibrat's law<sup>2</sup> (see Sutton, 1997 for discussion of literature on Gibrat's law). Gabaix presents a straightforward mathematical proof that when the key condition of Gibrat's law growth holds, then the emergent distribution of such city growth process must necessarily conform to Zipf's law with a power exponent of one. Through analysis of data for US

## The Harvard College Economist

cities in the past 30 years, we find that one of the crucial assumptions of Gibrat's law, equal variance of growth rates across city sizes, is violated. Gibrat's law therefore doesn't seem to hold for the growth of US cities at least for recent decades.

Gabaix does not attempt to provide much empirical evidence for his assumption of Gibrat's law for the growth of cities. He claims that "more work is needed to establish this entirely, but it appears that empirical analyses seem to support Gibrat's law." He uses the work of Glaeser, Scheinkman, and Shleifer (1995) to allude to the equality of mean growth rates of cities across sizes. While that paper does examine city growth extensively, it does not aim to prove the equality of mean growth rates. In citing this paper, Gabaix was most likely referring to the part where the authors conclude that "population of larger cities grew slower, a finding which is not robust." While it is difficult to reconstruct from the paper the exact way in which this result was obtained, it is immediately obvious that the regression for this result included a number of control variables. It needs to be noted that such treatment makes the result useless for the purposes of proving Gibrat's law for cities. Gabaix's main point is that

Zipf's law pattern will necessarily emerge if cities' growth, *irrespective* of the source or components of growth, follows Gibrat's law. Therefore, any effort to prove that city growth is Gibrat, while stripping growth figures through the use of control variables, is misguided and unwarranted.

The other basic condition for Gibrat's law, equal variance across city sizes<sup>3</sup>, is taken care of in Gabaix's paper by referring to results stemming from a paper by Eaton and Eckstein (1997). Their analysis pertinent to our case is even more marginal than that abstracted from Shleifer et al.'s paper. The brief footnote that contains this result claims that for France and Japan the difference in the variances of log-growth rates of a group of large cities and a group of smaller cities is statistically insignificant. Apart from Gabaix's vague but crucial reference to some control variable for capital (which implies the same major problem as that of Shleifer et al.'s regression), there is little relevance of these results to our case. We aim to establish that US (as opposed to French or Japanese) city growth rates are Gibrat. This is a direct consequence of our effort to link the apparent conformity of US cities to Zipf's law as shown by Gabaix. Therefore, we will look at the mean

growth rate and variance of US cities across sizes to establish whether Gabaix's theory for the emergence of Zipf's law applies to the US or not.

## **II. Data Selection**

Several types of urban entities offer themselves for our analysis. In principle, the main three entities that could be useful to look at, and whose data can be feasibly obtained, are cities, metropolitan areas and counties. The last one has been studied the least and is of no interest to us for this paper. It is then up to debate whether cities or metropolitan areas should be used for the study of Zipf's law and Gibrat's law. Gabaix makes the point that "agglomerations" are better suited for such analysis. Rosen and Resnick (1980) show that Zipf's law actually holds better for more carefully constructed agglomerations.<sup>4</sup> The main concern that drives this argument is the fact that as cities grow, population tends to move to the suburbs even when people work in the city (Glaeser et al, 1995). Due to such expansion, we would ideally use perfectly drawn standard metropolitan statistical areas (SMSAs) for our analysis.

Unfortunately, difficulties arise with such approach. First, data for

SMSAs are available only for little over 20 years while data for cities have been available for around a century. Secondly, such data is very hard to work with as definitions of these areas often change, new areas emerge, several areas merge, and area boundaries are altered for statistical purposes. While working with this data, we would run the risk of not taking into account some of these changes and, instead, attributing them to intrinsic changes in the areas' population. On the other hand, not only do we have many decades of data for cities, such data also seems more consistent for intertemporal comparison. We concede that city boundaries also obviously change over time. However, these changes generally should, by the nature of cities, reflect the growth/decline of city population, as opposed to arbitrary annexation/separation/redefinition of areas for statistical purposes (as is the case for SMSAs). Hence city data should be absolutely comparable while raw SMSA data are not. While ideally we would look at SMSAs over a longer period of time to find whether their growth conforms to Gibrat's law, it should be safer to look at cities for now and leave the analysis of SMSAs for careful examination in the future.

As we have already implied, the

## The Harvard College Economist

primary measure of city size and growth in this paper is population. This comes naturally from the usual application of Zipf's law for cities on the size of cities in terms of population. Indeed, Gabaix's paper, from which we draw and which we aim to complement, uses city population as a measure of city size and growth. *The Statistical Abstract of the United States* and its supplement, *The County and City Data Book*, were used as data sources for the population of cities. *The County and City Data Book 1977* contains a ranked list of largest US cities for the year 1975. First 55 cities from this list were used as our city sample. One of them (Baton Rouge Metro, LA) had to be eliminated from the data set due to inconsistency with later data<sup>5</sup>. Population for Honolulu, HI for 1975 was also missing and was interpolated in such a manner as to minimize the made-up value's influence on the sample mean and variance.<sup>6</sup>

The year 1970 was chosen as a suitable (albeit somewhat arbitrary) starting date. It provides a sufficiently large data set while keeping the effect of cities moving in and out of the top 54 cities to a minimum.<sup>7</sup> The data sources give us city size for every five years from 1970 to 1980 and every two years thereafter until 1998. We thus have

twelve data points for each city for the time period. We then calculate average growth rates for these periods and, to make them comparable, normalize them by conversion to average annual growth rates.<sup>8</sup> This data is then used to calculate the mean annual growth rate and the variance of annual growth rates.

It is important to note here that the mean and especially the variance values calculated need not be unbiased estimators of the true (should we have annual data for all cities) mean and variance of growth rates of the sample. We can suspect that the annualization of two-year and five-year data might introduce a certain (unknown) bias in our results. However, as long as this bias is consistent throughout the analysis across all city sizes (and there is no reason to believe that is not the case), our results will be valid as far as statistical significance goes, even though the actual mean and variance data might be somewhat biased.<sup>9</sup>

To prove Gibrat's law as we defined it, we will examine whether the mean growth rate and the variance of growth rate over four different city sizes are the same. Ideally, we would take each city and compare its growth to the growth of the whole population of cities in the US (or at least those in the upper

Table 1 F-Test for Variances

<b>City Growth</b>	<b>Top 27</b>	<b>Bottom 27</b>
Mean	-0.01023	0.447151
Variance	2.204201	3.395664
Observations	297	297
df	296	296
F	0.649122	
<i>P(F&lt;=f) one-tail</i>	<i>0.000108</i>	
<i>F Critical one-tail</i>	<i>0.825705</i>	

Table 2 t-Test Assuming Unequal Variances

<b>City Growth</b>	<b>Top 27</b>	<b>Bottom 27</b>
Mean	-0.01023	0.447151
Variance	2.204201	3.395664
Observations	297	297
Hypothesized Mean Difference		0
df		566
t Stat		-3.33093
<i>P(T&lt;=t) one-tail</i>		<i>0.000461</i>
<i>t Critical one-tail</i>		<i>1.64755</i>
<i>P(T&lt;=t) two-tail</i>		<i>0.000922</i>
<i>t Critical two-tail</i>		<i>1.964163</i>

tail, as is noted by Gabaix [1999]). However, it is infeasible to gather sufficient data for such analysis. Therefore, we will choose an alternative approach that is alluded to briefly in Eaton and Eckstein's (1997) paper. We will divide the 54 cities in our data set into two groups, the top 27 cities and the bot-

tom 27 cities. We will then compare the means and variances of growth rates for the two groups. If at least one of the values is different between the two groups in a statistically significant manner, then we reject the notion that the cities examined grew in accordance with Gibrat's law.

# The Harvard College Economist

## III. Analysis and Results

In our first analysis, we arranged the 54 cities according to the size they had in 1970 and then divided them into two groups of 27, whose composition remained constant for all years in our analysis (this is despite the fact that cities indeed move in and out of the two groups over time). We then calculated the means and variances of growth rates for each group and compared them (Table 1 and 2). The mean growth for the group of largest cities is  $-0.01$  percent for the period and that of the group of smaller cities is  $0.45$  percent. A heteroskedastic<sup>10</sup> t-test reveals that we can safely reject the null hypothesis that the means are the same ( $p=0.0009$ , in its stronger, two-tail, form). This violates one condition of Gibrat's law. Variances of growth rates for largest and smaller cities are  $2.204$  and  $3.396$ , respectively. An F-test testing for no difference between the two variances robustly rejects the null hypothesis ( $p=0.0001$ ). We can thus conclude that there is a statistically very significant difference between the variances of the two groups of cities. This result violates the second condition of Gibrat growth.

One potentially important caveat exists for the above analysis. We know that as cities grow, they shift be-

tween the two groups. The setup above fails to account for this movement because it takes the ranking at 1970 and keeps it for the whole time period. Therefore, in years after 1970, we actually have cities that should belong to the first 27 cities but are actually in the second 27 cities and vice versa.<sup>11</sup> We can cope with this problem by readjusting the ranking (and thus grouping) of the cities for each time period. In other words, before we calculate the growth rate for each time period, we re-rank the cities according to then-current city size (e.g. for the 1980-82 period, we first sort the cities by 1980 size and only then get growth rates for cities in each of the two updated size groups). The moving-rank method ensures that we really have only growth data for largest cities in the first group and only data for smaller cities in the second group, as opposed to the somewhat arbitrary and time-inconsistent ranking by 1970 for the whole time span.

Interestingly, we get somewhat different results when we use this more sophisticated method (Table 3 and 4). The mean growth rate for the first group is somewhat larger than that for the second group. What is of much more interest to us is the fact that the difference between the mean growth rates for the

Table 3 F-Test for Variances		
<b>City Growth</b>	<b>Top 27</b>	<b>Bottom 27</b>
Mean	0.255629	0.181295
Variance	2.326751	3.375293
Observations	297	297
Df	296	296
F	0.689348	
<i>P(F&lt;=f) one-tail</i>	<i>0.000716</i>	
<i>F Critical one-tail</i>	<i>0.825705</i>	

Table 3 F-t-Test Assuming Unequal Variances		
<b>City Growth</b>	<b>Top 27</b>	<b>Bottom 27</b>
Mean	0.255629	0.181295
Variance	2.326751	3.375293
Observations	297	297
Hypothesized Mean Difference	0	
Df	573	
t Stat	0.536479	
<i>P(T&lt;=t) one-tail</i>	<i>0.295918</i>	
<i>t Critical one-tail</i>	<i>1.647518</i>	
<i>P(T&lt;=t) two-tail</i>	<i>0.591836</i>	
<i>t Critical two-tail</i>	<i>1.964113</i>	

two city groups is no longer statistically significant. A heteroskedastic<sup>12</sup> t-test comparing the means of the two samples reveals that we cannot reject the null hypothesis that they are not different (p=0.59, in its stronger, two-tail, form [and half that in the weaker one-tail form]). We thus lose the result that seemed so strong in the first (fixed-rank-

ing) version of our analysis. The mean-equality condition of Gibrat's law now seems to hold for US cities. On the other hand, our previous result for variance does not change. Variance is again larger for the group of smaller cities (3.38 vs. 2.33 for largest cities). This difference is again very highly significant (p=0.0007), allowing us to robustly re-

# The Harvard College Economist

ject the null hypothesis of no difference. The equal-variance condition of Gibrat's law once again does not hold.

## IV. Conclusion

Provided we had access to complete and comparable data, it would have been more illustrative to examine the growth of metropolitan areas (wider agglomerations) instead of cities. However, even cities fit Zipf distribution quite well and Gabaix's theory is fully applicable to them. Indeed, Gabaix himself uses the term "cities" throughout his paper. It seems therefore perfectly warranted to examine Gabaix's basic assumption of Gibrat grow on cities.

We have managed to seed some serious doubt on Gabaix's claim that the growth of cities seems to obey Gibrat's law. While our second, more sophisticated analysis did not show a statistically different mean for the two groups of cities, we obtained a very statistically significant difference in mean growth rates for the two groups. This result means that Gabaix's explanation for the emergence of Zipf's law distribution with a power exponent of one has to be taken with reservation. Without the condition of Gibrat law growth being satisfied, Gabaix's theory has little merit in explaining why cities in the US fit Zipf's

law with a power exponent of one. While it is possible that some population of cities (perhaps outside the US) grow through a Gibrat-compliant process, that does not seem to be the case for recent growth of cities in the US.

## Endnotes

<sup>1</sup> More generally, we should consider "urban areas" which include, depending on definition, cities, metropolitan areas, counties, etc. For most of this paper, we will use the term "cities" for clarity.

<sup>2</sup> We will work with an intuitive interpretation of Gibrat's law that says that city growth has, over time, a common mean and a common variance that are both independent of city size.

<sup>3</sup> This notion is equivalent to a homoscedastic distribution of growth rates across city sizes.

<sup>4</sup> I borrow this citation from Gabaix's paper.

<sup>5</sup> Apparently, the metro area was dropped from the statistics in favor of a much smaller city area.

<sup>6</sup> That effect cannot be completely eliminated since any value that is different from the (unknown) true value will have some effect on the mean and variance. However, this effect is so small due to the large data set that it can be neglected

without any concern. I felt that the detriment of this alteration would be smaller than that of eliminating Honolulu, HI from the data set.

<sup>7</sup> The number of cities that were in the top 54 in 1970 but were not in the top 54 by 1996 is only seven. This number indeed increases as we move the starting date back, making our analysis less and less accurate should we use an earlier starting date.

<sup>8</sup> We simply divide the growth rate for a period by the number of years in the period.

<sup>9</sup> A consultation with an econometrist will hopefully clarify this issue.

<sup>10</sup> As we will see shortly, the two samples have a statistically significant difference in variance.

<sup>11</sup> The effect of cities shifting in and out of the whole sample of 54 cities seems to be negligible, given the robustness of all results obtained in the previous and following setups. Indeed, as we already pointed out, there were only 7 such “misplaced” cities in 1998.

<sup>12</sup> Again, we will see shortly that the two samples still have a statistically significant difference in variance.

## References

*County and City Data Book 1997*

(Supplement to the Statistical Abstract of the United States), Bureau of the Census.

Eaton, J., Z. Eckstein, 1997, “Cities and Growth: Theory and Evidence from France and Japan,” *Regional Science and Urban Economics*, XXVII, 443-474.

Gabaix, Xavier, 1999, “Zipf’s Law for Cities: An Explanation,” *The Quarterly Journal of Economics*, August 1999, 739-767.

Glaeser, E., J. Scheinkman, and A. Shleifer, 1995, “Economic Growth in a Cross-Section of Cities,” *Journal of Monetary Economics*, XXXVI, 117-143.

Krugman, P., 1996, “Confronting the Urban Mystery,” *Journal of the Japanese and International Economies*, X, 399-418.

Rosen, K., M. Resnick, “The Size Distribution of Cities: An examination of the Pareto Law and Primacy,” *Journal of Urban Economics*, VIII, 165-186.

*Statistical Abstract of the United States 1990, 1994, 1996, 1998, 1999*, Bureau of the Census.

Sutton, J., 1997, “Gibrat’s Legacy,” *Journal of Economic Literature*, XXXV, 40-59.