Volume 10, Number 2 **ISSN 1533-3604**

JOURNAL OF ECONOMICS AND ECONOMIC EDUCATION RESEARCH

An official Journal of the
Academy of Economics and Economic Education

Editor: Larry R. Dale
Arkansas State University

The *Journal of Economics and Economic Education Research* is owned and published by the DreamCatchers Group, LLC, and printed by Whitney Press, Inc. Editorial Content is controlled by the Allied Academies, a non-profit association of scholars, whose purpose is to support and encourage research and the sharing and exchange of ideas and insights throughout the world

Whitney Press, Inc.

Printed by Whitney Press, Inc.
PO Box 1064, Cullowhee, NC 28723
www.whitneypress.com

All authors must execute a publication permission agreement taking sole responsibility for the information in the manuscript. The DreamCatchers Group, LLC is not responsible for the content of any individual manuscripts. Any omissions or errors are the sole responsibility of the individual authors..

The *Journal of Economics and Economic Education Research* is owned and published by the DreamCatchers Group, LLC, PO Box 2689, Cullowhee, NC 28723, USA. Those interested in subscribing to the Journal, advertising in the Journal, submitting manuscripts to the Journal, or otherwise communicating with the Journal, should contact the Executive Director of the Allied Academies at info@.alliedacademies.org.

EDITORIAL REVIEW BOARD

EDITORIAL REVIEW BOARD

JOURNAL OF ECONOMICS AND ECONOMIC EDUCATION RESEARCH

CONTENTS

LETTER FROM THE EDITOR

We are extremely pleased to present this issue of the *Journal of Economics and Economic Education Research*, an official publication of the Academy of Economics and Economic Education Research, dedicated to the study, research and dissemination of information pertinent to the improvement of methodologies and effective teaching in the discipline of economics with a special emphasis on the process of economic education. The editorial board is composed primarily of directors of councils and centers for economic education affiliated with the National Council on Economic Education. This journal attempts to bridge the gap between the theoretical discipline of economics and the applied excellence relative to the teaching arts. The Academy is an affiliate of the Allied Academies, Inc., a non profit association of scholars whose purpose is to encourage and support the advancement and exchange of knowledge, understanding and teaching throughout the world.

The Editorial Board considers two types of manuscripts for publication. First is empirical research related to the discipline of economics. The other is research oriented toward effective teaching methods and technologies in economics designed for grades kindergarten through twelve. These manuscripts are blind reviewed by the Editorial Board members with only the top programs in each category selected for publication, with an acceptance rate of less than 25%.

We are inviting papers for future editions of the *Journal for Economics and Economic Education Research* and encourage you to submit your manuscripts according to the guidelines found on the Allied Academies webpage at www.alliedacademies.org.

Dr. Larry R. Dale

www.alliedacademies.org

ECONOMICS EDUCATION ARTICLES

EXPLAINING THE INTERBANK LOAN CRISIS OF 2008: A TEACHING NOTE

John Robert Stinespring, University of Tampa
Brian T. Kench, University of Tampa

ABSTRACT

This teaching note examines one aspect of the financial crisis: the month-long turmoil in the interbank loan market, which began with the bankruptcy of Lehman Brothers on September 15, 2008 and ended with the U.S. Treasury Department's $250 billion bank recapitalization on October 14, 2008. We model the crisis as a prisoner's dilemma in which banks with excess reserves play a game of Loan or No Loan with each other. The initial game results in a Nash equilibrium where each bank chooses No Loan and both are worse off. The second game incorporates the U.S. government's recapitalization program. It is argued that the program solves an asymmetric information problem that revives lending in the interbank loan market.

INTRODUCTION

One unintended consequence of the financial crisis of 2008 is a heightened student interest in economics. Having recently given a student-requested afternoon symposium on the topic to a standing-room-only crowd (a new experience for these economists), we find ourselves in a teaching moment – where students are getting a glimpse at the importance of economics in their lives, and demanding a reasoned explanation of the crisis. The intent of this note is to supply a response to our students' demand.

In this paper, we examine one aspect of the financial crisis: the month-long turmoil in the interbank loan market. The crisis began with the bankruptcy of Lehman Brothers on September 15, 2008 and ended with the U.S. Treasury Department's $250 billion bank recapitalization on October 14, 2008. At the industry-wide level, the crisis was characterized by historic spikes in the federal

funds rate (hereafter *ffr*), a shrinking of the deposit multiplier, and the hoarding of excess reserves. At the individual bank level, the crisis was evidenced by unprecedented increases in the market for insurance against bank defaults on debt, known as credit default swaps. All this occurred despite massive injections of reserves by the Federal Reserve Bank (hereafter the Fed) and a 25 percent decrease in the target *ffr* from 200 basis points to 150.

We model the dysfunction as a prisoner's dilemma in which banks with excess reserves play a game of Loan or No Loan with each other. The initial game results in a Nash equilibrium where each bank chooses No Loan and both are worse off. The second game examines government intervention to revive interbank lending through the lens of an asymmetric information problem. The recapitalization program by the U.S. Treasury department (hereafter the Treasury) is discussed within this context. Two immediate benefits arise from this exercise. First, the model illustrates the key insight of John Nash that rational behavior at the individual level can lead to irrational outcomes in the aggregate. This outcome is particularly prevalent in financial markets where transactions depend significantly upon trust between participants. Second, students get a simple framework in which to analyze proposed government solutions to financial crises. Because financial crises follow similar patterns, this model can be applied to other time periods – such as the U.S. banking crisis during the Great Depression – and other countries – such as the 1992 financial crisis in Sweden. This approach should have broad appeal to economists because no finance or advanced economic theory is required and much empirical evidence supports the results.

BACKGROUND

To understand the economic theory behind the interbank loan crisis, first consider the usual monetary policy prescriptions for a tight credit market. To stimulate lending, the Fed injects reserves into the banking system to lower the *ffr*. The *ffr* is the interest rate banks charge one another for overnight lending of excess reserves – that is, reserves above what banks are required to hold against their deposits. These excess reserves are crucial during credit contractions as they provide insurance for banks against deposit withdrawals and declines in the value of bank assets. The *ffr* changes as banks supply and demand excess reserves from each other. A lower *ffr* reduces borrowing costs for banks that demand reserves while the injection of reserves expands the supply available. The Fed intervenes in the market every day by buying and selling treasury notes from and to banks to

maintain the Fed's target *ffr*. The Fed's effectiveness is evidenced by the small deviations of the daily high *ffr* from the Fed's target *ffr* which averaged 32 basis points for the six years between November 1, 2002 and September 14, 2008.[1]

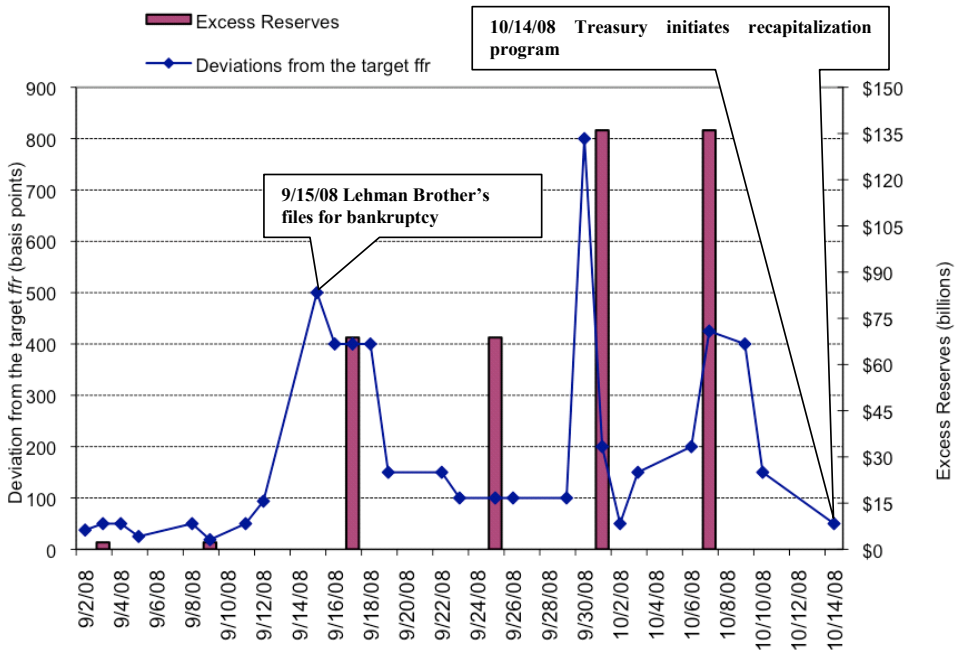
In addition to lowering the *ffr* and increasing the supply of available funds for the interbank loan market, an intervention by the Fed has a multiplier effect by which an increase in one bank's reserves, *R*, can lead to multiple deposits being created throughout the banking system.[2] Under normal credit conditions, the injection of reserves into a bank, call it Bank A, induces it to lend to its customers. As customers spend these loans, the money returns as deposits in another bank, say Bank B. Bank B uses the deposits to generate its own loans which lead to more deposits throughout the banking system. The deposit multiplier, α, measures the total increase in deposits created by each $1 injection of reserves. Although neither bank would claim to create money, each one unintentionally, as if lead by an *invisible hand*, contributes to the creation of deposits equal to the amount $\alpha * R$.

Now imagine bankers are overcome with fear. Fear that commercial and real estate loans will default. Fear that depositors will withdraw their savings. Fear that fellow banks won't repay interbank loans. Fear that banks are hiding credit risk from the market with off-balance sheet securities such as mortgage-backed securities and credit default swaps.[3] Even fear that their own assets (perhaps some of which are backed by subprime mortgages) will become worthless. If banks perceive the situation to be severe enough, they will stop providing interbank and customer loans (they may even call in loans) and hoard excess reserves to bolster their balance sheets. Because of these sources of fear, the *ffr* rises and the deposit multiplier decreases. Thus, the very actions banks take to protect themselves will reduce the deposit base on which they depend. This is what occurred during the month following the bankruptcy of Lehman Brothers on September 15, 2008. Figure 1 illustrates.

With deviations from the target *ffr* on the left-side vertical axis, excess reserves on the right-side vertical axis, and the date on the horizontal axis, Figure 1 shows a spike in the average daily deviation from the target *ffr* of 193 basis points. The deviations from the *ffr* exceeded 100 basis points on 13 of the days, 200 basis points on 7 of the days, 400 basis points on 3 of the days and reached a peak of 800 basis points on September 30, 2008 when the *ffr* spiked to 10%. On that day, banks were so wary of lending excess reserves to each other *overnight* that they required a lending premium 400% higher than the Fed's target rate.[4] Over this period the deposit multiplier fell from its prior six-year average of 7.24 to 2.88 for the week of September 15.[5] The credit default swap market saw similar spikes for individual

banks over this period. The price of insurance against a bond default from Morgan Stanley rose approximately 500 basis points, while the price for Goldman Sachs and Citigroup rose approximately 200 and 100, respectively (Bloomberg data). The fact that these events all occurred while the Fed injected massive amounts of liquidity into the banking system – evidenced by the unprecedented jump in weekly excess reserves from a six-year average of $1.8 billion to $69 and then $133 billion within the month – indicates the impotence of monetary policy and magnitude of the crisis.

Figure 1: Deviations from the *ffr target* & Excess Reserves



Source: www.newyorkfed.org

LOAN OR NO LOAN: A GAME THEORETIC EXPLANATION

Such market dysfunction was well articulated by the Nobel-prize winning mathematician John Nash. The key insight of Nash was that rational behavior at the individual level might lead to irrational outcomes in aggregate. The interbank loan crisis exemplifies such outcomes and can be understood by means of a game called Loan or No Loan played over a one-month period between two banks, Bank A and Bank B. The payoffs in Figure 2 are the expected profits for each bank. Bank A chooses a row and Bank B chooses a column. The number to the left of the comma gives the payoff to Bank A while the number to the right gives the payoff to Bank B.[6]

Figure 2
The Loan or No Loan Game

		Bank B: No Loan	Bank B: Loan
Bank A	No Loan	-10,-10	15,-15
	Loan	-15,15	10,10

If both banks choose Loan, the supply of interbank loans increases, the *ffr* falls, and deposits expand via the deposit multiplier process. With low borrowing costs and a sufficient number of deposits, lending is robust and generates an expected profit of $10. If both banks choose No Loan, interbank lending decreases, the *ffr* rises, and deposits contract. With high borrowing costs and fewer deposits created, banks turn their focus from profit expansion to capital protection. Lending is cut in favor of hoarding reserves and banks incur expected losses of $10 each. It

is clear from Table 1 that the socially optimal result – i.e., the Pareto optimal result – is for each bank to choose Loan.

Banks, however, are not in the business of making socially optimal decisions; they're in the business of making individually optimal decisions. Consider Bank A's decision making process. If Bank A thinks that Bank B will not provide interbank loans, Bank A's best response is to select No Loan. The logic is the following: If Bank A loans its excess reserves and Bank B hoards its reserves, the *ffr* remains high and deposit expansion is stymied. These results occur because each bank contributes to the deposit multiplier process, although no single bank can unilaterally affect α. With fewer reserves and little deposit growth, Bank A's likelihood of failure increases and its expected loss is \$15. If Bank A does fail, Bank B captures market share and may benefit from a fire-sale purchase of Bank A generating an expected profit of \$15.[7] Bank A doesn't like that scenario and so chooses No Loan. If Bank A thinks that Bank B will provide interbank loans, Bank A's best response is again No Loan. By loaning, Bank B takes on more risk without significantly affecting the *ffr* or total deposits in the banking system. If Bank B fails, Bank A may acquire its assets and gain market share thus providing an expected profit of \$15. In each case, regardless of what Bank B chooses, Bank A's best response – or in game theory terms, its *dominant strategy* – is to select No Loan. Because this game is symmetric, the same dominant strategy of No Loan applies to Bank B. These dominant strategies lead the banks away from the socially optimal result toward the No Loan equilibrium where both suffer a loss of \$10. The cell corresponding to No Loan, No Loan represents a *Nash equilibrium* in which neither player has an incentive to deviate unilaterally once at that equilibrium. Students will recognize the result of this Loan or No Loan game as a version of the classic prisoner's dilemma where all players are made worse-off. The only way to incentivize the players toward the social optimum is to change in the payoffs themselves.

FINANCIAL INNOVATION AND ASYMMETRIC INFORMATION

The symptoms of the interbank loan market crisis of 2008 are clear: 1) heightened fear led banks to charge overnight loan rates that greatly exceeded the target *ffr* and 2) banks rationally decided to hoard excess reserves in an effort to bolster their balance sheets. When combined, these symptoms caused a substantial decrease in the bank deposit multiplier, which in turn magnified the crisis. But what caused these symptoms? If we were to put our economic stethoscope on the pulse

of the crisis we would hear: asymmetric information, asymmetric information, asymmetric information.

Financial markets arise out of asymmetric information. Unlike product markets, financial markets facilitate the exchange of guarantees not goods. These guarantees are promises a borrower makes to repay a lender. Information is *asymmetric* because the borrower has much greater knowledge about his ability to repay than the lender. The greater the asymmetric information, the higher the degree of trust required among participants. Banks exist to assume the risk of asymmetric information by specializing in assessing loan risk and serving as intermediaries between borrowers and lenders. Banks profit from the difference between the interest rates they charge borrowers and the rates they pay depositors. But banking is a tricky business. Most bank assets (loans and securities) have long-term maturities while their liabilities (deposits) can be withdrawn at a moment's notice (Krugman 2009, p. 158). Because banks and their depositors assume this major liquidity risk, banks are required to meet regulations on the amount of reserves held against their deposits and the capital that must be held against their assets. These rules provide a cushion for banks against potential losses in the value of their assets. For many years, these strict regulations and long-proven methods of risk assessment made banks particularly adept at assessing interbank loans and minimizing asymmetric information.

Recent financial innovations changed all that. Instead of banks holding loans on their balance sheets, they began to package them into securities to be sold to off-balance sheet financial entities such as Structured Investment Vehicles (SIVs).[8] Though still affiliated with the originating banks, SIVs avoid capital-to-asset regulations because they fund their purchases with short-term debt (typically, asset-backed commercial paper) rather than deposits. Though the securities often consist of subprime mortgages, the affiliated banks provide credit lines to their SIVs to ensure a AAA credit rating.[9] Because banks have zero capital requirements for providing such credit lines, SIVs enable banks to originate more loans without a commensurate increase in required capital. Asymmetric information increases significantly because the credit risk of the loans remains with the bank and is obscured by the process of securitization. Assets of SIVs and other entities in the so-called "shadow banking system" totaled approximately $10.5 trillion in early 2007 compared to traditional banking sector assets of about $10 trillion.[10] Lehman Brothers was a major participant in the shadow banking system with "exposure to at least $2.6 billion in SIVs" by December 2007.[11] When Lehman failed, fear of

these off-balance sheet vehicles spiked, trust among banks evaporated, and the banking sector found itself within the prisoner's dilemma previously described.

GOVERNMENT INTERVENTION: THE TROUBLED ASSET RELIEF PROGRAM

One way to escape the prisoner's dilemma is for an outside agent to create a trust mechanism among the players. In terms of the payoff matrix, greater trust raises the expected profits of lending and, if done sufficiently, can move the Nash equilibrium to the social optimum of Loan, Loan. Governments have often played the role of outside agent in providing trust mechanisms for markets rife with asymmetric information, and with some success. Consider the used car market.

Sellers of used cars always know more than buyers of used cars. This form of asymmetric information tends to drive all the good, higher priced used cars out of the market. This is so because buyers only consider the lowest priced used car, and sellers have no *trustworthy* mechanism to reveal the quality of their good used car to a potential buyer. Thus, at the end of day, only lemons (or bad used cars) are left on the market.

Many state governments have "lemon laws," which provide consumers with a money back guarantee if their used car turns out to be a lemon.[12] The lemon law increases the level of trust in the marketplace, and it creates real value because it has the effect of bringing the good used cars back into the market.

Similar to the implementation of a lemon law, the Treasury created the $700 billion Troubled Asset Relief Program (TARP) in an effort to stabilize financial markets.[13] The first $250 billion of these funds, after much input was given by the economic community, was used to recapitalize qualifying banks through direct purchases of bank stock in two steps. First, $125 billion was provided on October 13 to the nine largest U.S. banks in return for bank stock. Though some of the nine banks did not want to partake in the program, Treasury Secretary Henry Paulson persuaded all to join telling them "[t]he system needs more money, and all of you will be better off if there's more capital in the system".[14] The second step requires all other banks to "apply" for a capital injection. After an arduous evaluation, the Treasury approves the applications of qualified banks.[15] This *de facto* seal of approval by the Treasury has become a trustworthy bank sorting mechanism that has decreased asymmetric information and produced an increased level of trust among approved banks.

In terms of the Loan or No Loan game, the TARP raises the lending bank's expectation of loan repayment among approved banks and thus the profitability of interbank loans.[16] Given the numerical values in Table 1, the government program only needs to raise expected profitability of interbank loans by $6 to change banks' decisions at the margin. Figure 3 illustrates.

Figure 3
The Loan or No Loan Game Redux

		Bank B: No Loan	Bank B: Loan
Bank A: No Loan		-10,-10	15,-9
Bank A: Loan		-9,15	16,16

In this version, if Bank A loans its excess reserves and Bank B hoards its reserves, the *ffr* remains high and deposit expansion remains stymied, but Bank A's likelihood of failure is lower because the perceived probability of repayment is higher. The bank's expected loss is only $9 rather than the original $15. If Bank A does fail, Bank B may still capture market share and benefit from a fire-sale purchase of Bank A leaving the expected profit of $15 unchanged. If both banks choose Loan, the supply of interbank loans increases, the *ffr* falls, deposits expand via the deposit multiplier process but the expected profits are higher by $6 with the higher perceived probability of repayment. Each bank has a dominant strategy of Loan and the Nash equilibrium in the game is Loan, Loan. By altering the institutional rules through the recapitalization program, the competitive Nash equilibrium result is now simultaneously the socially optimal result. It is important to note that the interbank loan market is revitalized not by the government guaranteeing repayment but merely changing loan profitability at the margin.

Has the TARP been successful? As of now, the TARP has yielded mixed results. With respect to the interbank loan crisis, it has been successful in stabilizing the market and revitalizing lending. As Tim Bond of Barclays Capital said in mid-October referring to interbank loans, "compared with [late September], borrowing volumes are up by as much as ten times."[17] *The Economist* magazine stated that "American banks including JPMorgan Chase and Citigroup have, in [mid-October], made loans to European counterparts for up to three months. And Europe's biggest bank, HSBC, is lending billions to other banks."[18] The TARP has not, however, lead to a substantial increase in consumer or business lending. For this reason, many politicians and pundits have considered it a failure.[19]

CONCLUSION

Since the day the recapitalization plan was announced, interbank loan rates declined dramatically and lending increased significantly. Deviations from the target federal funds rate remained below 75 basis points over this period. The Treasury's aggressive intervention to make banks "better off" effectively dealt with the deficiency of trust among banks caused by increased asymmetric information and has stabilized the interbank loan market. It is hoped that when the broader financial market stabilizes, private markets will augment the level of trust in the financial system by providing better credit rating agencies or other yet-to-be-discovered entrepreneurial tools to effectively value assets in the financial marketplace – much like private certifications of used vehicles offered by private automobile dealers guarantee the quality of their vehicles. Until that time arises, governments may continue to play a role in providing stability in financial markets.

ENDNOTES

1 Data exclude the final day of each reserve period when the federal funds rate characteristically strays from its target.

2 By focusing on the deposit, rather than the *money*, multiplier we ignore changes in currency demand. For intermediate courses, instructors may wish to include this aspect of M1 and the money multiplier as it reinforces our story and introduces other crisis phenomena such as the "breaking of the buck" that occurred in money markets.

3 Credit default swaps provide a real-time estimate of a bank's financial condition. As of 2007, the market was estimated to be $62.2 trillion of which banks sold 44%

of the total. A good exposition can be found in "The Great Untangling," *The Economist*, November 6, 2008.

[4] These spikes occurred throughout OECD nations as evidenced by the overnight LIBOR (London Interbank Offer Rate) rising to 6.44 percent on September 16 and falling to 1.94 percent on October 16. By November 6, the rate stood at 0.33 percent (Bloomberg data).

[5] The deposit multiplier is calculated as Demand Deposits at Commercial Banks divided by Total Reserves. Both figures are provided weekly and are seasonally adjusted. *Board of Governors of the Federal Reserve System*.

[6] For a review of the basics of non-cooperative game theory see Frank and Bernanke (2009), Mankiw (2009), or Hubbard and O'Brien (2008).

[7] Consider JP Morgan's purchase of Bear Stearns, Bank of America's purchase of Merrill Lynch and Barclay's purchase of Lehman Brothers' assets.

[8] SIVs profit from these securities – typically, Collateralized Debt Obligations, CDOs – by slicing them into different risk-based *tranches* that are sold off to other investors. Because many of these investors are other banks themselves, the credit risk remains within the banking sector.

[9] These credit lines are also referred to as "liquidity backstops".

[10] "By early 2007, conduits, structured investment vehicles and similar entities that borrowed in the commercial paper market and bought longer-term asset-backed securities, held roughly $2.2 trillion in assets, according to the Fed's Geithner. Another $2.5 trillion in assets were financed overnight in the so-called repo market, Geithner said. Geithner also highlighted big brokerage firms, saying that their combined balance sheets held $4 trillion in assets in early 2007. Hedge funds held another $1.8 trillion, bringing the total value of asset in the "non-bank" financial system to $10.5 trillion, he added. That dwarfed the total assets of the five largest banks in the U.S., which held just over $6 trillion at the time, Geithner noted. The traditional banking system as a whole held about $10 trillion, he said" (June 20, 2008, "Brokers threatened by run on shadow bank system" *MarketWatch).*

[11] Ibid.

[12] In many cases, private markets can solve the asymmetric information problem by developing for-profit solutions. For example, in the used car market, it is in the best

interests of Audi, BMW, and the like to offer their own binding inspection certification systems, which guarantee that the dealership will fix a problem with the vehicle for free for a specified period of time.

[13] The mafia is always a clear example of how institutions can change the incentive structure in a prisoner's dilemma game: members of the mafia end up "swimming with the fishes" in the long run if they rat out other members of the mafia in the short run. Because the former changes the incentive structure on the margin, members of the mafia are less likely to rat out other members of the mafia.

[14] Damian Paletta, Jon Hilsenrath and Deborah Solomon, "At Moment of Truth, U.S. Made Bankers Blink," *The Wall Street Journal*, October 15, 2008.

[15] Those banks deemed unqualified were "recommended" by the Treasury to "withdraw [their] application" and were persuaded to "find a buyer". To incentivize consolidation among banks, the government effectively subsidized the bank acquisitions by relaxing accounting rules to let banks benefit from the accumulated tax losses of banks they acquired. Fitzpatrick, Dan and Sidel, Robin "Federal Aid? Not for Us, Proclaim Some", *Wall Street Journal*, November 17, 2008, and Francis, Theo "How Uncle Sam is Reshaping Banking" *Business Week,* December 8, 2008.

[16] The Treasury's initial plan was to increase capital in the banks by purchasing their troubled (or "toxic") assets. The important difference between the programs is the information that they provide. Capital injections through equity stakes signal that the government deems the receiving bank to be of a particular quality. Purchases of troubled assets might only signal the poor asset management at the receiving bank. Thus the latter could raise uncertainty among banks and unintentionally lower the expected profits of interbank loans.

[17] "Thawing Out" *The Economist*, October 23, 2008.

[18] Ibid.

[19] The Wall Street Journal estimated that by December 30, 2008, the TARP program had a return of 4% since its inception, which corresponds to an annualized rate of 16%. Moore, Heidi, N. "Smartest Guy In the Room?" *Wall Street Journal*.

REFERENCES

Frank, R., & B. Bernanke, (2009). *Principles of Economics (Fourth Edition)*. McGraw Hill.

Hubbard, R.G., & A.P. O'Brien, (2008). *Economics (Second Edition)*. Pearson/Prentice Hall.

Krugman, P., (2009). *The Return of Depression Economics*. Norton.

Mankiw, N. G., (2009). *Principles of Economics (Fifth Edition)*. Cengage.

ECONOMICS ARTICLES

STANDARDS, PRACTICES, AND METHODS FOR THE USE OF ADMINISTRATIVE CLAIMS DATA

John Robst, University of South Florida
Roger Boothroyd, University of South Florida
Paul Stiles, University of South Florida

ABSTRACT

Administrative data are attractive sources of information in research and evaluation studies for numerous reasons including relatively low cost, and the availability of longitudinal information and large subject pools. While many professional organizations set standards for members, there exists a patchwork of practices for researchers to follow when performing research. The purpose of this paper is to outline standards and practices for researchers, and to discuss common analysis issues related to the proper use of administrative data. The discussion focuses on data from the two largest United States government-funded health care programs, Medicare and Medicaid. This focus is chosen due to the wide use of such data, and the sensitive nature of healthcare information.

INTRODUCTION

In a recent paper, Safran et al., (2007) discuss the increasing secondary use of health data for research and other purposes. The authors note that the "lack of coherent policies and standard good practices for secondary use of health data impedes efforts to transform the U.S. health care system" (p. 1). This paper seeks to contribute to this important discussion in two ways. First, a set of standards and practices for researchers to follow is proposed for the acquisition and proper use of administrative data. Second, the literature is reviewed that relates to specific shortcomings with administrative databases and methods to address the problems. The paper is geared towards students with an interest in health economics, but may also be useful to other students and established researchers given the increasing use of administrative data (both health-related and otherwise). The goal is to help

researchers use administrative data correctly so that policy makers can have greater confidence in findings, and consequently research can have a greater effect on public policy.

Public health care programs in the U.S. such as Medicare and Medicaid finance health care for millions of people. The information collected as a result of health care delivery, enrolling members, and reimbursing for services is referred to as administrative data (Iezzoni, 1997). Despite widespread use for research purposes, there exist limited standards and practices for researchers to adhere to in using administrative data (Retchin & Ballard, 1998; Safran, et al., 2007). In addition, while undergraduate and graduate students in economics (and other social sciences) encounter a wide array of courses during their education, few academic programs teach students how to acquire and properly use data.

This paper focuses on data from the two largest government-funded health care programs, Medicare and Medicaid, but the issues discussed in this paper apply to all types of administrative records. The focus was chosen because of the sensitive nature and yet widespread use of such data, the increased vulnerability of the subjects, and the evolving U.S. federal regulatory landscape for healthcare information in general. Examples are discussed based on experiences during the lead author's five years at the Centers for Medicare & Medicaid Services (CMS), the government agency that oversees the programs.

ADVANTAGES AND DISADVANTAGES TO ADMINISTRATIVE DATA

First, let's review a few of the advantages and shortcomings of using administrative data for research. There are a number of advantages to administrative data (Iezzoni, 2002; Pandiani & Banks, 2003; Roos, Menec, & Currie, 2004; Roos et al., 2008). It is conceivable to study (almost) all individuals age 65 and above with Medicare enrollment and claims data. The use of population based data enables questions to be considered that could not be addressed with a sample. However, due to cost considerations and the sheer size of the databases, almost all studies use a sample. For example, as discussed in more detail later in the paper, much research uses a 5% sample of Medicare beneficiaries which is approximately 800,000 people. Despite being a small proportion of beneficiaries, the sample size remains substantial and limits concerns about the generalizability of results found in small sample studies. In addition, the large size also allows for adequate numbers of minorities for statistical analysis.

The records are not limited to specific types of setting (e.g., hospitals). Information can be longitudinal covering individuals and institutions across many years. Confidentiality can be maintained due to the large sample sizes. The data exist, and thus are relatively inexpensive to acquire compared to primary data collection, plus the low cost also allows for easy replication of previous studies. Survey attrition due to a loss of contact or refusal to participate is also minimized.

There are, however, many potential problems with the use of administrative data (Retchin & Ballard, 1998; Drake & McHugo, 2003). Such problems include a lack of information on the reliability or accuracy. Public use files may not be available for several years, reducing usefulness for current policy questions. Large samples can lead to statistically significant results that are not very meaningful, as even very small effects are precisely measured. Similarly, researchers may look for questions to fit the data, rather than forming questions and then looking for the appropriate data. Medicaid and Medicare enrollment and claims records include protected health information under the Health Insurance Portability and Accountability Act (HIPAA) and therefore require stringent privacy protection measures.

Due to such potential problems, users should adhere to standards on data use. While most professional organizations establish standards for members, there are no clear standards and practices for users of administrative data to follow. However, appropriate use is crucial in order to increase public confidence in the use of sensitive health care information for research purposes, and for federal agencies to continue to allow access to the data (Safran, et al., 2007).

THE RESEARCH PROTOCOL: DATA ADEQUACY AND ACQUISITION

Acquisition of administrative data typically begins with the development of a detailed research protocol. The protocol is assessed by the data owners to determine whether access should be granted to Medicaid or Medicare enrollment and claims records. A useful resource for researchers developing a protocol, although involving data for Canada, is provided at The Manitoba Centre for Health Policy (MCHP) web site (http://www.umanitoba.ca/centres/mchp/). Some of the information is specific to the MHCP mission and data on Manitoba residents, but much of the information applies to administrative data in general.

The protocol should detail the research questions and explain why they are important to the mission of the Medicare and/or Medicaid programs. Given the

inherent concern in releasing sensitive information, research questions need to be of sufficient interest to the data owners to warrant release of the data. The protocol must also identify the specific dataset(s) and justify that the source is appropriate for the proposed analysis. van Eijk, Krist, Avorn, Porsius, & de Boer (2001) created a checklist guidelines to determine whether available data are adequate to answer the research questions. Important considerations used to decide whether the available data are adequate to meet research needs include sample size, whether the claims contain sufficient detail for the study (e.g., diagnoses, procedures, drug and dosing information), accuracy, continuity of variables over time, the ability to link databases, and adequate security and accessibility. In the following sections, several of these considerations are discussed as well as others as they relate to secondary use of health data.

Approvals

An important early step is to understand the process for data acquisition. Most data available from CMS are acquired through the Research Data Assistance Center (ResDAC), a CMS contractor that provides assistance to researchers using Medicare enrollment and claims records. Their web site www.resdac.umn.edu) contains much information on the process for acquisition and the associated cost but provides limited guidance on the proper use of the data. Together, CMS and ResDAC act as gatekeepers and determine who gains access to CMS data. ResDAC and CMS also make available national Medicaid data, referred to as the Medicaid Extract (MAX) files. The MAX files are a combination of the Medicaid enrollment and claims data complied by each state. Some states make Medicaid data from their state available to researchers, some do not. If you wish to use Medicaid data from a specific state, contact the Medicaid authority and determine whether the data are available and what their process for data acquisition entails.

Consult with data owners

Users should consult with data owners to understand what the data represent and ensure the proposed questions can be appropriately answered with the data. For Medicare data, this may involve discussions with ResDAC personnel and also individuals at CMS who work with the data. There are several reasons for users to seek such consultation. Administrative data are usually compiled for a specific purpose, often related to payment or program monitoring and evaluation. Users

need to understand why the administrative database was created. The reason(s) for collecting the data can have an important impact on the universe covered, data elements, variable definitions, frequency and timeliness, quality, and stability over time. A lack of understanding of what the data represent and how it may be used has lead researchers to propose research questions for which the data are poorly suited (Medi-Cal Policy Institute, 2001).

In addition, given that administrative data are often compiled for internal use by the data owners, documentation is often scant compared to survey data primarily produced for research purposes. Even with proper documentation, owners are a valuable resource for understanding technical details and should be consulted by users. The data owners have knowledge of the issues involved in working with the files, problems with specific variables, are aware of other issues not apparent from reading documentation or examining the data, and can verify that the project design is appropriate.

Such discussions also provide opportunities to clarify variable definitions. For example, Medicare enrollment files note when Medicare is a secondary payer. This occurs primarily when the beneficiary has health insurance coverage through a spouse. The person is labeled as *working aged* despite the fact that the beneficiary is not employed. Consequently, users should not assume the variable name necessarily describes the variable clearly.

Data quality

Data users should always consider the likely quality of the data for the proposed research questions. The accuracy of data is extremely important, particularly for analyses to inform public policy (Robinson & Tataryn, 1997). While the available quantity of information is often large, the accuracy and completeness is sometimes questioned. The Medi-Cal Policy Institute (2001) reported that California's Medicaid managed care data system could not be "used to make sound policy decisions" because data were inaccurate and incomplete. Most administrative data rely on reporting by individuals or firms and the information respondents provide can cause gains or losses to individuals or businesses (Wolf & Helminiak, 1998). In other cases, information can be underreported if unrelated to the gains or losses of individuals or businesses. As such, there may be biases in the information supplied.

Even if the overall database is considered complete and accurate, specific variables may differ in accuracy. Administrative files used to make payment often

have fields that are checked for completeness and reasonableness. As such, these fields are relatively accurate. Other variables may not be checked or edited, especially those that do not affect payment. Users should learn the editing rules used by the owners. Users should determine the likely extent of measurement error and decide whether it should be addressed in the research plan.

The sample

One potential benefit of administrative data is the ability to perform population based research. In theory, Medicare data may be available for all individuals age 65 and above. The analysis of population based data avoids many of the concerns with analyzing samples, whether small or large. All statistics are *actual* statistics, not *sample* statistics. Thus, conclusions can be drawn without concerns about type I or type II errors.

In practice, the Medicare program does not cover everyone age 65 and above. Individuals must qualify for Medicare based on work history (either their own or a spouse's). Some individuals never establish a sufficient work history to qualify for Medicare. For example, certain immigrant groups are less likely to qualify for Medicare because work histories were not established with the Social Security Administration. Thus, even with a database as large as the Medicare enrollment and claims data, users must be aware of who may not be adequately represented in the data and potential biases this may introduce. In addition, given the size of some administrative databases, users should consider whether they have sufficient resources (both computer and financial) to acquire, store, and analyze the data. For example, there are over a billion Medicare claims in a single year.

In almost all cases, researchers use a sample. A five or ten percent sample from a very large database is sufficient for the majority of studies. For example, many researchers use the CMS 5% Medicare Standard Analytical Files (SAF). The standard analytical files contain all enrollment and claims data for 5% of Medicare beneficiaries (approximately 800,000 people) and are created annually by CMS. Because these files are used by many researchers, the cost of acquiring the data is lower than if a researcher requests a special data pull. The SAFs are created by selecting all enrollment and claims records for individuals with 05, 20, 45, 70 or 95 in positions 8 and 9 of the health insurance claim number (i.e., the last two digits of the Medicare identification number). The sample selection criteria for the SAFs allows for individuals to be followed over time, which would not be possible with a true random sample. At the same time, this could be problematic is the last two

digits of the Medicare IDs differed across individuals in a systematic nature. However, the Medicare ID is typically the person's Social Security number (plus characters in the 10th and 11th places to denote the reason for eligibility). The last two digits of a Social Security Number are not systematically assigned based on characteristics of the individual, and thus the SAFs are generally considered to be equivalent to a random sample. If, for example, the sample was pulled based on the first three digits (where are assigned based on geographic location), then the sample would be geographically biased and not representative of the population.

While generally not a concern with large administrative databases, users must consider if the expected number of observations is sufficient to generate meaningful results. In general, power tests should be performed to determine the sample size necessary to have reasonable confidence that statistically significant results can be detected. This step is particularly important if studying a rare disease or treatment. At the same time, given the typical large sample size, users have to interpret the economic significance of their results and not simply rely on statistical significance (Drake & McHugo, 2003).

Researchers must also know the decision rules used to pull the data. For example, studies interested in the frequency of services should know if claims are "final action", or if they include denials, interim bills, or adjustments. The inclusion of interim bills and adjustments will lead to an over count of service frequency, and thus should be excluded during the analysis.

Diagnostic accuracy

Research questions often focus on specific subgroups of individuals with specific diagnoses (e.g., asthma or diabetes). Claims data contain codes that identify specific diseases using the International Classification of Diseases (ICD). ICD codes are five digit codes that can be used to identify individuals with a broad class of diseases or a very specific disease. The first three digits tend to identify a general class (e.g., 250 for diabetes), with the fourth and fifth digits being more specific (e.g., 250.41 denotes type I diabetes with renal manifestations).

Among the issues to consider is whether two years or more of data should be used to identify cases. Dombkowski, Wasilevich, and Lyon-Callo (2005) found that a diagnosis of a chronic disease (asthma) was not observed in every year. Thus, selection of cases based on diagnoses in a single year would undercount the prevalence of a disease. People still have the disease but it did not show up in the

claims data during a year for some reason. Consequently, the identification of individuals with chronic diseases may require multiple years of data.

In addition to diagnoses, prescription drug use might also identify people (e.g., insulin or perhaps metformin use for diabetes). Gilmer et al. (2001) find that the use of prescription drug records substantially increases the estimated prevalence of specific diseases. Caution must be used though since many medications are used to treat multiple conditions, and thus might not indicate a specific disease.

On the other hand, consideration might also be given to whether an individual should be included only if there are at least two records with the diagnosis of interest to rule out incorrect or miscoded chronic diagnoses. The presence of a consistent diagnosis over time provides evidence that the diagnosis is correct. Such concerns arise from studies that compare diagnoses in medical charts and claims. For example, Schwartz et al., (1980) find a relatively poor match between medical charts and claims for Medicaid enrollees; 29% of chart diagnoses of private practitioners, 37% of chart diagnoses in the free standing outpatient clinics; and 46% of diagnoses from outpatient clinics of general hospitals do not match with Medicaid claims. Interested readers should look at Virnig & McBean (2001) for a more thorough discussion of studies that assess reliability by comparing diagnostic data located in charts to claims in the database.

Security

Researchers are responsible for data security, and should have a plan for ensuring that the files cannot be accessed by unauthorized users. Some obvious steps include using automatic screen savers that can only be turned-off with a password if the data reside on an office or personal computer. If storage is on a network, only authorized users should have access, and the data should be behind a firewall if the network is connected to the internet. Email is not a safe way to transmit individually identifiable information unless adequate encryption is used. In addition, user responsibility for the data does not end when the project ends. The data use agreement (DUA) typically specifies whether the data have to be returned to the agency or destroyed.

THE PROTOCOL – DATA ANALYSIS

The protocol must also detail the analysis plan. This section provides an overview of some common analysis issues related to using administrative data.

Such analytical issues include the need to empirically assess quality, differentiate between time trends and program effects, and use medical encounters to account for the differing health status of treatment and comparison groups (Ray, 1997). Much depends on the study questions and design for the specific project. The proposed analysis should meet the standards for institutional review boards and peer reviewed publications.

Studies are discussed below that relate to the analysis issues and the solutions employed by researchers. The studies are not an exhaustive overview of questions that can be analyzed with administrative data. Readers interested in a broader discussion of health care topics that can be addressed with administrative data should see a paper by Roos, Menec, & Currie (2004), and for a broader discussion of how administrative data can be used to answer an array of social research questions see Roos et al., (2008).

Linking records

Users will often need to merge several different data files. Examples of such linkages include combining records from inpatient, outpatient, and physician claims, supplementing claims data with survey data such as the Medicare Current Beneficiary Survey, or matching individuals across years. Privacy concerns may arise when administrative records are linked to other sources and researchers should verify that the data use agreement allows such linkage (Clark, 2004).

Linking may be based on shared identifiers, deterministic matching, or probabilistic matching (Victor & Mera, 2001; Clark, 2004; Roos et al., 2008). Matching records by shared identifiers occurs when there are the same identifiers in data sets (e.g., Social Security Number or Health Insurance Claim number). Most data available from CMS can be matched using individual identifiers. However, researchers may also encounter situations when unique individual identifiers are not available. Deterministic matching examines a subset of variables and matches records that agree on this subset (e.g., name, date of birth, sex). Individuals can have the same name or date of birth or sex, but it is far less likely that different individual in two datasets will have the same name and date of birth and sex. Probabilistic linking matches based on the probability that records refer to the same person. Matching with individual identifiers or deterministic matching is typically used when attempting to draw conclusions about individuals. Probabilistic matching is used when there is limited information on which to base matches (e.g., name, date

of birth, sex). Given the difficulty in precisely matching individuals, probabilistic matching is more appropriate when drawing conclusions about populations.

The use of probabilistic matching is illustrated by Banks & Pandiani (1998). The authors derive estimates of the number of people receiving psychiatric care in state hospitals and general medical settings. Typically, the data sets would be merged based on individual identifiers or deterministic matching to avoid double counting patients who receive care in both sectors. Banks and Pandiani use probabilistic matching based only on gender and birth date to derive estimates of sample overlap, and as a result are able to estimate the number of people receiving psychiatric care. The use of probabilistic matching is likely to increase as concerns with patient privacy lead data owners to restrict the release of information that enables direct or deterministic matching to other data sources.

When records from more than one administrative source are combined it is important to be aware of potential differences in concepts, definitions, reference dates, coverage, and quality. For example, recent attention has focused on merging Medicare and Medicaid claims to study dual eligible beneficiaries (e.g., Liu, Wissoker, & Swett, 2007; Yip, Nishita, Crimmons, & Wilber, 2007). These data originate from different sources that may use different definitions, definitely have different coverage issues, and may have differences in quality. While one might expect data within the Medicare program to have consistent standards, even this may not necessarily be the case. For example, the quality of inpatient hospital claims is generally considered better than physician claims (Retchin & Ballard, 1998).

Empirically assess data quality

While data quality should be assessed for expected accuracy prior to acquisition, quality should also be assessed empirically. Once the data are linked and the sample constructed, users should examine descriptive statistics. Users should check the results for reasonableness, and if possible, compare results with alternative data sources or prior research and attempt to explain differences. Many studies have been published using Medicare and Medicaid data providing researchers with a substantial literature for comparison.

Assessing quality is particularly important when data are hand entered because data errors may be more prevalent. An example of such data entry errors occurs with beneficiary location (SSA state and county codes) in Medicare claims. Research often looks at Medicare utilization across counties in the United States. Analyzing claims, there are approximately 5,000 SSA state/county codes that appear

in the data, far greater than the 3,100 actual counties in the US. What accounts for the erroneous counties? State and county codes are often hand entered, there is no payment issue involved (payments are based on provider location, not beneficiary residence), and the field is not checked for accuracy. Such miscoding may be important for sparsely populated counties where a few miscoded observations can make a difference to the results.

Two approaches are used in the literature to address potential problems with examining geographic variation in prevalence rates across counties. For example, Cooper, Yuan, Jethva, & Rimm (2002) examine county level variation in breast cancer rates using Medicare data. The authors attempt to confirm their findings by comparing prevalence rates to the National Cancer Institute's cancer tracking database (the Surveillance, Epidemiology, and End Results, SEER, program) which tracks approximately 10-15 percent of the U.S. population. While a valid test for large counties, comparing prevalence rates in small counties could be problematic due to the (relatively) small sample size of the SEER database. Holcomb & Lin (2006) examine geographic variation of macular disease in Kansas. Because of the potential for unstable prevalence rates in small counties, the authors aggregated sparsely populated counties into larger geographic units.

Researchers should document their findings regarding quality to enable other researchers to understand why certain observations or variables were included or excluded based on data quality considerations. Documentation can also help researchers compare and reconcile studies so that others understand why decisions were made and potential implications of those choices. These are basic steps that all researchers should perform, but studies are often unable to replicate research because such steps are not taken (Dewald, Thursby, & Anderson, 1986). New users of a data set should review the literature to see how others have handled problems with the data.

Time series analysis

Over time, the Medicare and Medicaid programs have moved toward managed care, case management, and provision of prescription drugs. Consequently, it is increasingly important to track people over time to determine how participation in case management or the provision of certain prescription drugs affects health over time. For example, there has been much discussion about creating a database to track outcomes from prescription drug use after the well documented problems with Vioxx (e.g., Lohr, 2007). Multiple years of the CMS Standard

Analytical Files are often linked to examine changes over time. This is possible because, as discussed earlier, the 5% sample contains all enrollees with HIC numbers that end in specific digits. Thus, with some exceptions (some people die and new enrollees enter the data) the sample contains the same people over time.

A substantial literature using time series analysis considers the changing prevalence of specific diseases over time. For example, Lakshminarayan, Solid, Collins, Anderson, & Herzog (2006) find an increasing prevalence of atrial fibrillation diagnoses between 1992 and 2002, while Salm, Belsky, & Sloan (2006) find an increasing prevalence of eye diseases between 1991 and 2000. Lakshminarayan (2006) partly address diagnostic quality by requiring at least one inpatient claim or two outpatient claims with an atrial fibrillation diagnosis. However, both studies may be overstating the increasing prevalence of such diseases. Physicians were required to report diagnostic data on Medicare claims beginning in the early 1990s. Physician payments are typically based on procedures not diagnoses, and diagnosis is often not necessary to justify a procedure. Over time physicians have reported more thorough diagnostic data. Indeed, diagnostic reporting continues to improve more than a decade later as physicians implement electronic medical records. The point is that if one examines time trends in the prevalence of a disease, one needs to be cautious in looking at diagnostic trends in Medicare claims data. Simply looking at the increased reporting of a diagnosis is likely to overstate the increasing prevalence of a disease. While Salm (2006) at least note this possibility, neither study attempted to account for this in their analysis.

When records from different time periods are linked, they are a very rich source of information for researchers. However, users should understand whether the data will be consistent across time, and why changes may occur. The reasons for collecting the information may change over time, or variable definitions may change, or reporting may have changed. Coverage changes occur on a regular basis in Medicare based on CMS decisions and Congressional mandates. Such changes can have a substantial effect on services provided.

Accounting for individual heterogeneity

Perhaps the biggest challenges in using administrative data are to create a comparison group and decide on the appropriate analytical techniques. In clinical research, randomized control trials allow researchers to assign individuals to treatment and control groups in a random manner. Administrative data do not typically allow for this type of assignment and there are often non-random

differences between individuals that choose a treatment versus no treatment (or an alternative treatment). Pre-treatment differences may bias (typically referred to as sample selection bias) the results if such differences also correlate with the outcome.

Selection issues are common in research using administrative data, requiring researchers to account for differences between individuals. For example, administrative claims are often used to assess quality of care and examine outcomes from patient care. Hospital quality has been considered by many researchers because hospital administrative data are generally considered to be relatively high quality (e.g., Krumholz et al., 2006; Ross et al., 2007). Quality of care by physicians is also considered by Schatz et al., (2005). However, hospitals and physicians that have the most complex cases are more likely to have the highest complication and mortality rates. Consequently, accounting for case-mix is crucial to comparing the care provided across medical care settings or outcomes from alternative treatments.

There are several methods used to account for pre-treatment differences. The first two methods focus on accounting for observed differences between individuals. Many studies use risk adjusted models where control variables thought to be correlated with the outcome and the independent variable of interest (e.g., hospitals, physicians, treatment, gender, race, etc.) are included in a regression specification. Popescu, Vaughan-Sarrazinn, and Rosenthal (2007) examine racial differences in mortality after acute myocardial infarction. The authors control for sociodemographics, comorbidity, and illness severity to account for factors potentially correlated with the outcome (mortality) and variable of interest (race).

A variant on this approach is to use a diagnosis-based risk score as a measure of health (e.g., Ross et al., 2007). The score represents a measure of overall health status based on demographics and diagnoses. CMS and many States use diagnosis-based risk scores to determine compensation for managed care plans (e.g., Pope et al., 2004). While a useful measure, many researchers do not compute the scores correctly. This occurs because the models were developed using diagnoses from specific provider types (e.g., physician specialty). While this detail is contained in the technical instructions for managed care plans to submit data, it is not included in the risk adjustment publications or software published by CMS. Since most researchers do not discuss their research with people who work on risk adjustment at CMS, they often include too much diagnostic data when computing individual risk scores and overstate risk scores. As pointed out earlier, existing documentation may not provide all needed information, but such information can be learned by consulting with knowledgeable individuals. The example suggests

that users should initiate such discussions regardless of whether the researcher is aware of a lack of information.

A potential problem with risk adjusted regression models, regardless of whether specific characteristics of risk are used or an overall risk score, is that the comparison groups may not have the control variables in common. For example, if a treatment group is primarily old and a control group is primarily young, then conclusions regarding the effect of treatment may be biased given linearity assumptions in regression modeling.

Propensity score matching has become a popular alternative to regression methods in social science research for addressing selection issues when analyzing administrative files (Rosenbaum & Rubin, 1983; Imbens, 2000). Matching techniques mimic a random experiment by matching individuals in the treatment and control groups based on observed characteristics. The observed characteristics are used to estimate the probability of receiving treatment. Individuals with similar probabilities of treatment are compared, some who do and some who do not receive the treatment, to determine the effect of treatment. Using the age example, the young people in the treatment and control groups would be matched, while the older individuals in the treatment and control groups would be matched. Outcomes are then appropriately compared for similar individuals.

Numerous articles use propensity score methods to examine treatment effects when using administrative data. For example, Berg & Wadhwa (2007) examine the effects of a disease management program for elderly patients with diabetes. Propensity score methods are used to match observations in the treatment group with people in a control group who did not participate in the disease management program. Similarly, Krupski et al., (2007) examine the effects of receiving androgen deprivation therapy for individuals with prostate cancer on skeletal complications. Individuals receiving therapy are matched to individuals not receiving therapy by age, geographic region, insurance plan, and index year.

There is, however, debate about whether matching actually mimics a random experiment (Agodini & Dynarski, 2004; Smith & Todd, 2005). Research attempting to validate propensity score matching uses experimental data, and attempts to replicate the experimental results by reexamining the data using matching techniques. In other words, the data are examined under the assumption that assignment was not random and may be subject to selection biases. The majority of studies find the results from experimental data and matching methods are not similar. Thus, while matching methods may be useful, they should not be viewed as a perfect solution to problems with sample selection.

One potential problem with each of the above methods is the reliance on observed data. As such, the development of risk scores and propensity scores is challenging with administrative claims that often lack key clinical detail (Iezzoni, 1997). This issue is particularly salient for research on provider quality. Iezzoni (1997) suggests that administrative data be used as a screening tool to highlight areas for further investigation, not to draw conclusions about quality. Information on the process and appropriateness of care may not be adequate to provide accurate measures of provider quality. In general, all studies involve some degree of unobserved data. Instrumental variables methods may be appropriate if unobservable characteristics are thought to be important to the analysis. Of course, it can be extremely challenging to find suitable instruments. In conclusion, controlling for differences between treatment and control groups, or between patients seen at different hospitals, or between any two comparison groups, is crucial to drawing proper conclusions.

Research tools

There are many tools available to researchers on the internet and it may also be useful to utilize publicly available modules to develop important measures. Consistency across studies is increased if users can access such modules. Such publicly available information is typically tested by numerous users and is likely to be accurate. Much research requires manipulation of the data to create the analysis files and measures needed to answer the research questions. The internet allows researchers to utilize publicly available programs and modules that enable accurate creation of health measures such as the Charlson Index. The Manitoba Centre for Health Policy (MCHP) web site provides a web-based repository of useful tools for conducting research using administrative data (Roos, Soodeen, Bond, & Burchill, 2003). Some of the modules apply specifically to data available from the MCHP, but there are a number of statistical tools for analysis that can apply to a variety of administrative claims sources.

CONCLUSION

This paper has outlined some practice guidelines for the use of administrative data. While administrative data have great potential, there are also many pitfalls. Research using secondary data will benefit the health care of Americans only if the data are appropriately used. The growing use of such records

in research and evaluation necessitates that guidelines be developed and discussed such that the conclusions from research are valid. We hope the guidelines presented in this paper generate further discussion of the appropriate use of such data.

In summary, users of administrative data should develop a research protocol that: presents the research questions including a justification of why the research questions are important to the data owners, assesses whether the data are appropriate for the research questions (i.e., quality, sample size, available variables, and ability to link records) through reviews of the literature and discussions with the data owners, details the security plan including where the data will be stored and how access will be controlled, presents the analysis plan including an empirical assessment of the data quality and the statistical techniques that will be used to answer the research questions, discusses how potential data shortcomings will be addressed, and describes steps that will enable replication by other researchers.

Clearly, there is a need for such standards and practices in the use of administrative data given the continued increase in use. Huax (2005) outlines some of the current trends and his views on upcoming changes in health information systems. The trend continues to be towards using administrative data to inform patient care, strategic management, and clinical and epidemiological research. The future is likely to move towards the development of comprehensive electronic medical records that include information from multiple or all payers. As administrative data become more comprehensive and complex, developing and utilizing standards and practices will become even more important in the future.

REFERENCES

Agodini, R., & Dynarski, M. (2004). Are experiments the only option? A look at dropout prevention programs. *Review of Economics and Statistics, 86*, 180-194.

Banks, S.M., & Pandiani, J.A. (1998). The use of state and general hospitals for inpatient psychiatric care. *American Journal of Public Health, 88*, 448-451.

Berg, G.D., & Wadhwa, S. (2007). Health services outcomes for a diabetes disease management program for the elderly. *Disease Management, 10*, 226-234.

Clark, D.E. (2004). Practical introduction to record linkage for injury research. *Injury Prevention, 10*, 186-191.

Cooper, G.S., Yuan, Z., Jethva, R.N., & Rimm, A.A. (2002). Use of Medicare claims data to measure county-level variation in breast carcinoma incidence and mammography rates. *Cancer Detection and Prevention, 26*, 197-202.

Dewald, W.G., Thursby, J.G., & Anderson, R.G. (1986). Replication in empirical economics: The *Journal of Money, Credit, and Banking* project. *American Economic Review, 76*, 587-603.

Dombkowski, K.J., Wasilevich, E.A., & Lyon-Callo, S.K. (2005). Pediatric asthma surveillance using Medicaid claims. *Public Health Reports, 120*, 515-524.

Drake, R. & McHugo, G .(2003). Large data sets can be dangerous. *Psychiatric Services, 54*, 133.

Gilmer, T., Kronick, R., Fishman, P. et al. (2001) The Medicaid Rx model: Pharmacy-based risk adjustment for public programs, *Medical Care, 39*, 1188-1202.

Haux, R. (2005). Health information systems – past, present, future. *International Journal of Medical Informatics,* September 15, 2005.

Holcomb, C.A., & Lin, M.C. (2005). Geographic variation in the prevalence of macular disease among elderly Medicare beneficiaries in Kansas. *American Journal of Public Health, 95*, 75-77.

Iezzoni, L.I. (2002). Using administrative data to study persons with disabilities. *The Milbank Quarterly, 80*, 347-378.

Iezzoni, L.I. (1997). Assessing quality using administrative data. *Annals of Internal Medicine, 127*, 666-674.

Imbens, G.W. (2000). The role of the propensity score in estimating dose-response functions. *Biometrika, 87*, 706-710.

Krumholz, H.M., Wang, Y., Mattera, J.A., Wang, Y.F., Han, L.F., Ingber, M.J., Roman, S., & Normand, S.L.T. (2006). An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction. *Circulation, 113*, 1683-1692.

Krupski, T.L., Foley, K.A., Baser, O., Long, S., Macarios, D., & Litwin, M.S. (2007). Health care cost associated with prostate cancer, androgen deprivation therapy and bone complications. *Journal of Urology, 178*, 1423-1428.

Liu, K., Wissoker, D., & Swett, A. (2007). Nursing home use by dual-eligible beneficiaries in the last year of life. *Inquiry, 44*, 88-103.

Lakshiminarayan, K., Solid, C.A., Collins, A.J., Anderson, D.C., & Herzog, C.A. (2006). Atrial fibrillation and stroke in the general Medicare population: A 10 year perspective, 1992-2002. *Stroke, 37*, 1969-1974.

Lohr, K.N. (2007). Emerging methods in comparative effectiveness and safety: Symposium overview and summary. *Medical Care, 45*, S5-S8.

Medi-Cal Policy Institute. (2001). *From Provider to Policymaker: The Rocky Path of Medi-Cal Managed Care Data*.

Pandiani, J. & Banks, S. (2003). Large data sets are powerful. *Psychiatric Services, 54*, 745.

Pope, G.C., Kautter, J., Ellis, R.P., Ash, A.S., Ayanian, J.Z., Iezzoni, L.I., Ingber, M.J., Levy, J.M., & Robst, J. (2004). Risk adjustment of Medicare capitation payments using the CMS-HCC model. *Health Care Financing Review*, 25(4), 119-141.

Popescu, I., Vaughan-Sarrazin, M.S., & Rosenthal, G.E. (2007). Differences in mortality and use of revascularization in black and white patients with acute MI admitted to hospitals with and without revascularization services. *Journal of the American Medical Association, 297*, 2489-2495.

Ray, W.A. (1997) Policy and program analysis using administrative databases. *Annals of Internal Medicine, 127*, 712-718.

Retchin, S.M., & Ballard, D.J. (1998). Establishing standards for the utility of administrative claims data. *Health Services Research, 32*, 861-866.

Robinson, J. & Tataryn, D. (1997). Reliability of the Manitoba mental health management information system for research. *Canadian Journal of Psychiatry, 42* 744-749.

Roos, L., Borwnell, M., Lix, L., Roos, N., Walld, R., & MacWilliam, L. (2008). From health research to social policy: Privacy, methods, approaches. *Social Science & Medicine, 66*, 117-129.

Roos, L.L., Menec, V., & Currie, R.J. (2004). Policy analysis in an information-rich environment. *Social Science and Medicine, 58*, 2231-2241.

Roos, L.L., Soodeen, R.A., Bond, R., & Burchill, C. (2003). Working more productively: Tools for administrative data. *Health Services Research, 38*, 1339-1357.

Ross, J. Cha, S. Epstein, A., Wang, Y., Bradley, E., Herrin, J., Lichtman, J., Normand, S., Masoudi, F., & Krumholz, H. (2007). Quality of care for acute myocardial infarction at urban safety-net hospitals. *Health Affairs, 26*, 238-248.

Rosenbaum, P.R., & Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*, 41-55.

Safran, C., Bloomrosen, M., Hammond, W.E., Labkoff, S., Markel_fox, S., Tang, P.C., & Detmer, D.E. (2007). Toward a framework for the secondary use of health data: An American Medical Informatics Association white paper. *Journal of the American Medical Informatics Association, 14*, 1-9.

Salm, M., Belsky, D., & Sloan, F.A. (2006). Trends in cost of major eye diseases to Medicare, 1991-2000. *American Journal of Ophthalmology, 142*, 976-982.

Schatz, M., Nakahiro, R., Crawford, W., Mendoza, G., Mosen, D., & Stibolt, T.B. (2005). Asthma quality-of-care markers using administrative data. *Chest, 128*, 1968-1973.

Schwartz, A.H., Perlman, B.B., Paris, M., Schmidt, K., & Thornton, J.C. (1980). Psychiatric diagnoses as reported to Medicaid and as recorded in patient charts. *American Journal of Public Health, 70*, 406-408.

Smith, J. & Todd, P.E. (2005). Does matching overcome Lalonde's critique of nonexperimental estimators? *Journal of Econometrics, 125*, 305-353.

van Eijk, M., Krist, L., Avorn, J., Porsius, A., & de Boer, A. (2001). Do the research goal and databases match? A checklist for a systematic approach. *Health Policy, 58*, 263-274.

Victor, T.W., & Mera, R.M. (2001). Record linkage of health care insurance claims. *Journal of the American Medical Informatics Association, 8*, 281-288.

Virnig B.A., & McBean A.M. (2001). Using administrative data for public health surveillance and planning. *Annual Review of Public Health, 22*, 213-230.

Wolf, N. & Helminiak, T.W. (1998). Nonsampling measurement error in administrative data: Implications for economic evaluations. *Health Economics, 5*, 501-512.

Yip, J., Nishita, C.M., Crimmons, E.M., & Wilber, K.H. (2007). High-cost users among dual eligibles in three care settings. *Journal of Health Care for the Poor and Underserved, 18*, 950-965.

SHORT-RUN EQUILIBRIUM GDP AS THE SUM OF THE ECONOMY'S MULTIPLIER EFFECTS

Thomas J. Pierce, California State University, SB

ABSTRACT

The author suggests that macro principles students' grasp of the structure and workings of the short-run economy may be enhanced by conceptualizing the demand-side equilibrium level of GDP as the sum of all individual multiplier effects at work in the economy at a given point in time. A simple numerical example, presented after the concepts of equilibrium GDP and the multiplier have been introduced and discussed initially, illustrates the point. Considering the equilibrium level of GDP from a "multipliers perspective" highlights for students the variety of short-run factors affecting the magnitude of GDP. It also helps clarify the relevance of the concept of short-run equilibrium GDP in an economic system that never actually achieves a specific equilibrium and where "other things" are rarely constant.

INTRODUCTION

While the Keynesian concepts of the marginal propensity to consume and the multiplier no longer dominate principles of macroeconomics textbooks as they once did, most instructors still spend class time discussing them. Once students are familiar with the MPC and the major types of aggregate spending, the economy's short-run equilibrium level of GDP is determined, assuming a constant price level. Discussion then typically shifts to the multiplier, with a change in autonomous spending (usually investment) establishing a new equilibrium level towards which the economy moves until the next demand shock occurs.

In the process of studying those concepts many students become proficient at determining the equilibrium level of GDP, whether by utilizing a Keynesian cross/45º line-type diagram, comparing aggregate expenditures and production in a table, or manipulating a simple algebraic model of linear equations. And they quickly learn how to use the multiplier to calculate subsequent changes in the

equilibrium level of GDP. But when asked why equilibrium is established at any specific level of output, such as $10,000 billion or $12,000 billion, or whatever the numerical example may be, about the best most students can muster is, "because that's the level of total output matched by total expenditures."

The purpose of this brief paper is to suggest that once the concepts of equilibrium GDP and the multiplier have been introduced and discussed initially per the general sequence noted above, supplementing that discussion with a slightly different take on why the equilibrium level of output is established at any particular numerical level may enhance principles students' grasp of the underlying structure and workings of the short-run economy. Specifically it is suggested that students may benefit from conceptualizing any short-run equilibrium level of GDP as the sum of all individual (demand-side) multiplier effects at work in the economy at a given point in time.

To illustrate that point here a standard principles-level example of short-run equilibrium GDP determination is presented initially. Then the resulting equilibrium GDP dollar-amount is shown to be equivalent to the sum of the economy's autonomous-spending multiplier effects, broadly construed. Next, historical precedents of that notion are cited. Finally it is argued that, while an oversimplification, conceptualizing the aggregate level of short-run economic activity as the sum of an economy's multiplier effects gives principles students, among other insights, a better sense of both the variety of short-run factors affecting the magnitude of GDP and the relevance of the concept of short-run equilibrium GDP in an economic system that never actually achieves a specific equilibrium and where "other things" are rarely constant.

DETERMINING SHORT-RUN EQUILIBRIUM GDP: A STANDARD EXAMPLE

Table 1 presents a typical textbook example of short-run (demand-side) equilibrium GDP determination. All dollar amounts are in billions and a fixed price level is assumed. The underlying consumption function is C = $300 + .8Y; there are no taxes and no induced portions of investment, net exports, or government spending.

Comparing possible levels of aggregate output with associated levels of aggregate spending in the first and last columns, respectively, yields an equilibrium GDP figure of $9000 billion. Complementing those numerical comparisons with discussion of the rationale for producers' output adjustments from non-equilibrium

output levels shows students why GDP moves naturally towards $9000 billion in the short run. Discussion then typically turns towards what happens in the event of a demand shock. A change in autonomous spending is introduced and the multiplier principle is presented to explain why the equilibrium level of GDP changes more than autonomous spending.

Table 1: Short-run Equilibrium GDP Determination

GDP (Income)	Consumption (C)	Investment (I)	Government (G)	Net Exports (X-IM)	Aggregate Expenditures
$8000	$6700	$700	$600	$200	$8200
8500	7100	700	600	200	8600
$9000	7500	700	600	200	$9000 (*equi.*)
9500	7900	700	600	200	9400
10000	8300	700	600	200	9800

Note. All dollar figures are in billions.

SHORT-RUN EQUILIBRIUM GDP AS THE SUM OF THE ECONOMY'S MULTIPLIER EFFECTS

Once students have been introduced to the concepts of equilibrium GDP and the multiplier per above, they would be well served, I'd suggest, by brief further consideration of *why* the initial equilibrium level of output in this economy happens to be $9000 billion. Having compared aggregate production and spending levels in the table, they grasp the notion that individual producers adjust their output levels to unexpectedly weak or strong demand and that, in the aggregate, those adjustments move the economy naturally towards equilibrium, other things constant, at $9000 billion.

Beyond the basic production-adjusts-to-demand explanation of short-run equilibrium, however, the $9000 billion equilibrium level may be conceived broadly as the end product of the spending-multiplier effects simultaneously at work in the economy. Comparing changes in aggregate income to changes in consumption in Table 1's columns 1 and 2, respectively, reveals an MPC of .8 and a simple multiplier of 5. Furthermore the types of autonomous spending presented in Table 1 may be viewed broadly as the basis of four separate autonomous-spending

multiplier processes generated by $300 billion of autonomous consumption,[1] $700 billion of investment, $600 billion of government spending, and $200 billion of net exports. Table 2 shows the individual effects of each of those multiplier processes on output/income as well as their total impact of $9000 billion, the short-run equilibrium level of GDP.

Table 2: Short-Run Equilibrium GDP as the Sum of the Economy's Spending-Multiplier Effects

source of autonomous spending	amount of spending	multiplier (MPC=.8)	effect on equilibrium GDP
consumption	$300	5	$1500
investment	700	5	3500
government	600	5	3000
net exports	200	5	1000
Short-run equilibrium GDP:			$9000

Relating Equilibrium GDP to its Multipliers: Historical Precedents

While, to my knowledge, no explicit references appear in current macro principles' texts to the notion that short-run equilibrium GDP might be conceptualized as the sum of the economy's multiplier effects,[2] at least two indirect references to that notion appeared in the literature during the decade or so following the publication of Keynes' *General Theory* (1936).

Colin Clark (1938) noted the relationship of the economy's multipliers to equilibrium output/income in the process of investigating a separate, but related issue: estimation of the numerical value of Great Britain's multiplier through examination of national income data. As a starting point in his inquiry, he observed that some critics of Keynes' *General Theory* had questioned the "sovereignty of investment as the sole determinant of the level of economic activity"[3] asking, "Is it not possible … there are types of originating economic impulse other than the purchase of capital goods which may affect the general level of economic activity?" (436) In pursuing that question himself, Clark stated that by "… measurement of the 'determinants' [of the level of economic activity] of which the level of private investment is one element only, and by application of the multiplier to the

determinants[,] the level of money national income can be predicted." (443) In essence he was saying that the magnitude of the short-run equilibrium level of GDP equals the sum of the individual spending-multiplier effects.

A decade later, what Clark and others had called the "determinants" of the level of economic activity and "types of originating economic impulse" were presented as "exogenous factors" in Arthur Smithies' (1948) "simple formula" for "equilibrium national income":

$$\text{National Income} = \frac{\text{Sum of the influences of exogenous factors}}{\text{1- sum of the marginal propensities}}$$

While Smithies did not explicitly state that equilibrium national income is the sum of the multiplier effects, he related the level of equilibrium GDP directly to the specific amounts of exogenous/autonomous spending in the short-run economy.[4,5]

PEDAGOGIC LICENSE, STUDENT BENEFITS, CAVEATS

Multiplier effects, almost always refer to fluctuations in autonomous spending, not to the total amount of any category of autonomous spending. Yet statements, such as Clark's and Smithies', relating the total amounts of different types of autonomous spending to equilibrium national income confirm the author's sense that, in teaching principles students, exercising a measure of pedagogic license to conceptualize the short-run equilibrium level of GDP as the sum of the economy's multiplier effects is not only reasonable, but instructive. This sum-of-multipliers perspective can play the "role of logical organizer" (Colander 1991, 232), helping students organize their thinking about how the amounts of different types of autonomous spending in combination with the multiplier concept drive the short-run demand-side economy and determine its equilibrium level of GDP.

Granted, presenting equilibrium GDP as the sum of a string of multiplier effects has a mechanical ring to it, calling to mind at first Johnson's characterization of the multiplier as "that inexhaustibly versatile mechanical toy" (1961,11) and Colander's distinction between "mechanistic" and "interpretative" Keynesian models (1999, 368). And, were equilibrium GDP presented as the sum-of-multipliers in a bare-bones manner without elaboration, students could indeed get the impression of a short-run economy operating in a lock-step, deterministic manner, devoid of dynamism. But even a modest amount of explanation dispels that notion.

Presenting short-run equilibrium GDP as the sum of multiplier effects also makes it easier for students to think realistically about the concept of equilibrium GDP. They recognize readily that producers adjust their output levels to demand but rightly doubt that such adjustments ever result in the attainment of any numerically-specific equilibrium level of GDP. By supplementing that producer-adjustment discussion of equilibrium with the "multipliers perspective" noted above, students realize that the question of whether or not short-run equilibrium is ever achieved is a non-issue as they recognize that frequent fluctuations in business investment spending, home construction, consumer confidence, export demand, etc. alter existing autonomous-spending multiplier relationships, putting the economy on a new equilibrium path that too is bound to be interrupted subsequently by yet other demand shocks. Thus this sum-of-multipliers perspective enables students to appreciate both the relevance of the concept of short-run macroeconomic equilibrium and the role of demand-side dynamism in the short-run economy.

The multipliers perspective also makes evident that multiple demand-shocks may occur simultaneously, pushing the economy in the same or opposing directions, confirming students' personal observation of a world where "other things" are not constant and macroeconomic crosscurrents are not uncommon. They also easily recognize that negative spending shocks cause recession and see the potential, however modest, for government to utilize short-run stabilization policies to prod the economy in the direction of potential GDP from a less desirable equilibrium-GDP neighborhood.

Of course the insight gained into the workings of a short-run demand-driven economy by relating equilibrium GDP directly to its multiplier effects comes at a price. As noted previously a measure of pedagogic license must be taken in discussing multipliers in terms of total amounts of different types of autonomous spending as opposed to fluctuations in their levels. Also this multipliers perspective requires placing at least a modest amount of additional emphasis on the simple 1/(1-MPC) multiplier which ignores various factors: leakages that shrink the multiplier's numerical value substantially, measurement difficulties associated with forward-looking consumption, variation in the MPC across income groups and business-cycle phases, and aggregate-supply considerations. In my estimation, however, the insight gained by students into the nuts-and-bolts workings of the short-run economy by thinking "interpretatively" (not "mechanistically") about how the equilibrium level of GDP relates to the economy's multiplier effects is worth both the pedagogic liberty taken and the modest amount of time necessary to discuss the concept.

CONCLUSION

Many principles of macroeconomics students become adept at determining the short-run equilibrium level of GDP in a simple demand-side Keynesian model of the economy and quickly learn how to use the multiplier to calculate subsequent changes in that equilibrium level. Fewer students, however, display an intuitive feel for the specific magnitude of equilibrium GDP. It is argued here that supplementing standard discussion of equilibrium GDP and the multiplier with a slightly different take on why the equilibrium level of output is established at any particular numerical level may improve students' sense of the workings of the short-run economy.

Specifically it is suggested that students may benefit from conceptualizing the short-run equilibrium level of GDP as the sum of all individual (demand-side) multiplier effects at work in the economy at a given point in time. This perspective places additional emphasis on multipliers and, unadorned, could leave students with the impression that the short-run economy operates in a lock-step, mechanical manner. A modest amount of explanation readily reveals the opposite: the dynamic nature of short-run macroeconomic relationships. In establishing that sense of dynamism within an equilibrium framework, it is argued that students may better comprehend the workings of the short-run economy observable in their daily lives.

ENDNOTES

[1] Multiplying any level of income in column 1 of Table 1 by .8 indicates that induced consumption is $300 billion less than (total) C in column 2. With that information, students deduce that $300 billion is the amount of autonomous C in this simple economy.

[2] Among the textbooks checked were recent editions of Baumol and Blinder; Case and Fair; Hall and Lieberman; Mankiw; McConnell and Brue; O'Sullivan, Sheffrin, and Perez; Stiglitz and Walsh; and Taylor and Weerapana.

[3] While Keynes's discussion of the multiplier in the *General Theory* was couched mostly in terms of investment, statements such as, "Pyramid-building, earthquakes, even wars may serve to increase wealth …" (1936, 129) make clear that he did not view investment as the "sole determinant of the level of economic activity."

[4] The 'Sum of the marginal propensities' noted in the denominator of the equation refers not to the MPC alone but to Smithies' broader assertion that, "*Each* [emphasis added] of the behavior variables --- consumption, investment,

government spending, exporting, and importing --- can be regarded as partly endogenous and partly exogenous." (300)

[5] The algebraic equivalent of Smithies' equilibrium equation (usually defined in terms of Y) appears in chapter appendices of some principles' texts and occasionally in the body of a chapter. But in no instance found by the author is that algebraic equation supplemented by a statement indicating that the short-run equilibrium level of output (Y) might be conceived broadly as the sum of the economy's multiplier effects.

REFERENCES

Clark, C. (1938). Determination of the Multiplier from National Income Statistics. *Economic Journal*, 48(190), 435-448.

Colander, D. (1991). A Consideration of the Economics Major in American Higher Education. *Journal of Economic Education,* 22(3), 227-234.

Colander, D. (1999). Teaching Keynes in the 21st Century. *Journal of Economic Education*, 30(4), 364-372.

Johnson, H. (1961). The General Theory After Twenty-Five Years. *American Economic Review, Papers and Proceedings*, 51(2), 1-17.

Keynes, J. (1936). *The General Theory of Employment, Interest, and Money*. London: Macmillan.

Smithies, A. (1948). The Multiplier. *American Economic Review*, *Papers and Proceedings*, 38(2), 299-305.

MACROECONOMIC INFLUENCES AND EQUITY MARKET RETURNS: A STUDY OF AN EMERGING EQUITY MARKET

Arshad Hasan, Muhammad Ali Jinnah University
Muhammad Tariq Javed, Quaid -i -Azam University

ABSTRACT

This study examines the short run and long run causal relationships among macroeconomic variables and equity market returns in the emerging equity market for the period of 6/1998 to 6/2008 by employing the VAR framework on monthly data. Macroeconomic variables include industrial production index, consumer price index, money supply , exchange rate, foreign portfolio investment, Treasury bill rates and oil prices. Results support the finance theory and provide evidence that long term relationship exist among equity market and macroeconomic factors. Unidirectional causality has been observed flowing from consumer price index, exchange rates, money supply and interest rate to equity market. No granger causality is observed among industrial production, foreign portfolio investment and equity market returns. This insignificant relationship with industrial production, oil indicates that market movement is not based on fundamentals and real economic activity. The cointegration analysis only captures the long-run relationship among the variables, it does not provideinformation on responsiveness of equity market returns to shocks in macroeconomic variables so impulse response function and Variance decomposition analysis based on VECM has also been performed. Variance decomposition analysis also confirms that monetary variables are a significant source of volatility in equity market.

INTRODUCTION

During last decade phenomenal growth has been observed in emerging equity markets and Pakistan is no exception. The KSE- 100 index, which is the benchmark for the Pakistani equity market, has exhibited unparalleled growth and

moved from 921 in 2002 to over 16000 points. This remarkable growth has been a subject of global interest. During said period significant changes has also been observed in macroeconomic factors. An unprecedented change has also been observed in Interest rates, inflation, exchange rates, capital flows and Oil prices in the country. So question arises whether there exists a relationship among equity markets and macroeconomic factors.

The link among macroeconomic variables and the equity market has always attracted the curiosity of academicians and practitioners as it has an innate appeal. Finance theory suggests that prices of financial instruments are based on expected cash flows and discount factor. Macroeconomic variables affect both expected cash flows as well as discount rates. Therefore macroeconomic changes should be priced by market. The traditional dividend discount model is also based on above theoretical framework.

Therefore it is a well established fact that equity prices are influenced by economic information but theory is silent about specific variables which may influence equity prices. The empirical work has attempted to establish the relationship but results are yet inconclusive

Chen, Roll, and Ross (1986) explore this new avenue by examining the link among equity prices and macroeconomic variables by employing a multifactor model which provides evidence that macroeconomic factors are priced. Pearce and Roley (1985), Hardouvelis (1987), McElroy and Burmeister(1988), Hamao (1988) and Cutler, Potterba and Summers (1989) also confirm that equity prices react to arrival of macroeconomic information. At the same time, Poon and Taylor(1991), Shanken(1992) contradict the results. Some studies are in partial agreement. Flannery and Protopapadakis (2002) are of opinion that macroeconomic variables can predict future equity market returns to some extent and exact relationship among is difficult to establish. Therefore empirical evidence on relationship among macroeconomic variables and equity market is mixed

Under this cloud of uncertainty, number of studies has been conducted in various parts of globe by using various methods of exploring long term relationship among time series data. Mukherjee and Naka (1995), Cheung and Ng (1998), Nasseh and Strauss (2000), McMillan (2001) and Chaudhuri and Smiles (2004) employs cointegration analysis and granger causality test to explore long run relationship among equity prices and macroeconomic variables.

According to Humpe and Macmillan(2007) significant research has been done to investigate the relationship between equity market returns and a broad range of macroeconomic factors , across a number of equity markets and over a

range of different time horizon. But this research is generally focused on developed markets or emerging markets of Asia Pacific Rim. Only few studies are available with reference to Pakistan which is one of the major countries of south Asia and lies on cross roads of Central Asia, Middle East. And these studies only explore few variables.

The objective of this paper is to analyze the long-term relationship between the KSE and a broad set of macroeconomic factors for a longer time period by employing conitegration approach proposed by Johnson and Jusilius. Direction of causal flow has been captured by using Granger causality test. Other dynamic of time series data have also been explored by using impulse response analysis and variance decomposition analysis. The broad set of macroeconomic variable include industrial production index , consumer price index, money supply , exchange rate, foreign portfolio investment, Treasury bill rates and oil prices. This set of data has been used first time in Pakistan. Karachi stock exchange index return has been used as proxy for equity market returns. The study's main contribution is to examine the short run and long run relationships between Karachi stock market and macroeconomic variables , which have been relatively neglected by previous researchers

The rest of the paper is organized as follows: Section II incorporates a brief over view of recent empirical work. Section III describes the macro economic variables and Methodology used in the study. Empirical results are reported in Section IV and finally Section V concludes the results.

LITERATURE REVIEW

The relationship between equity market returns and economic fundamentals has been extensively researched in developed markets e.g.Chen et al. (1986), Fama(1990), Chen(1991), Cheung and Ng (1998) , Choi et al.(1999), Dickinson (2000), Nasseh and Strauss(2000). However the literature with reference to transition economies is limited and that too is focused on Asia pacific rim.

Chen, Roll and Ross (1986) investigate the existence of long run relationship among equity prices and industrial production, inflation, risk premium, market return, oil prices, term structure and consumption for US. Study assumes that the variables are uncorrelated and changes in variables are unexpected. . Results provide evidence about the existence of long run relationship between the macroeconomic variables and the expected equity returns. It has been observed that industrial production, risk premium, yield curve, and unanticipated inflation can

explain expected returns during periods of high volatility. However, oil prices, market index, and consumption are not priced in the market. CRR also investigate the sensitivity of US stock returns to the unanticipated news and conclude that equity returns responds to arrival of macroeconomic news and this responsiveness is priced by the market.

Beenstock and Chan (1988) investigate the presence of long term relationship among export volume, fuel and material cost, relative export prices, money supply, inflation, and interests rates and equity markets by employing IN UK equity market and find that unanticipated increase in fuel and material costs and interest rate leads to reduction in equity returns. Study also provides evidence about existence of positive relationship among equity returns and money supply and inflation. However export prices and export volume are not priced by equity market.

Hamao (1988) uses the methodology proposed by Chen, Roll and Ross (1986) for Japanese economy and reveals that variations in expected inflation and unexpected variations in risk premium and term structure of interest rates influence equity returns significantly. However, variations in macroeconomic activities are found weakly priced in Japanese economy in comparison to variations priced in U.S.A.

Mukherjee and Naka(1995) examine the relationship between exchange rate, inflation , long term government bond rate, money supply, real economic activity and call money rate in the Japanese stock market and find that cointegration is present among macroeconomic variables and positive relationship exist between the industrial production and equity market return.

Habbibullah *et al* (1996) explores the long run relationship among Malaysian equity market and money supply(M1 and M2) and output(GDP) by using monthly data and finds equity market of Malaysia is informationally efficient with respect to money supply as well as output

Cheung and Ng (1998) provides evidence about long term interlinkages among equity market indices and real oil price, real consumption, real money, and real output by employing Johansen cointegration framework. Equity market returns are found related to transitory deviations from the long run relationship and to changes in the macroeconomic variables. Cointegration analysis under constrained environment provide insight about equity market return variation that is not already captured through dividend yields, interest rate spreads, and GNP growth rates.

Fazal and Mahmood (2001) explore causal relationship between equity prices and economic activity, investment spending, and consumption expenditure

for the period 7/1959 to 6/99 by employing cointegration analysis and VECM and provide evidence about existence of long run relationship among above stated variables. Unidirectional causality has also been found flowing from macro variables to equity prices. However it is observed Pakistani equity in unable to influence aggregate demand. Fazal(2006) again examines relationship to investigate the stochastic properties of the variables by considering the shifts as a result of economic liberalization and finds unidirectional causality between the real sector and equity prices. No significant change in patterns is observed.

Ibrahim and Yusoff (2001) examine dynamic relationship among macroeconomic variables and equity prices for Malaysian capital market for the period 1/1977 to 7/1998 by employing VAR framework. Macro economic variable include industrial production, consumer price index , money supply, exchange rate, and equity prices. Results indicate that equity prices are being influenced by money supply. Money supply is found positively associated with equity prices in short run and negatively associated with equity prices in the long run. A negative impact of depreciation shocks has also been observed on equity prices. Maysami *et al* (2004) examines the long run relationship among macroeconomic variables and STI and sectoral indices like the property index, finance index and the hotel index and finds STI and the property index have long term relationship with industrial production, inflation , exchange rate , changes in the short and long-term interest rates and money supply.

Al-Sharkas(2004) investigates the relationship among equity market and real economic activity, money supply, inflation, and interest rate for Jordanian equity market by using Johansen Approach and provides evidence about presence of long run relationship among equity market and macroeconomic variables. Gay(2008)investigates the relationship among Indian equity market and exchange rate and oil price for Brazil, Russia, India, and China (BRIC) by employing ARIMA model and finds no evidence about existence of significant relationship among variables. It is further observed that equity markets of Brazil, Russia, India, and China are weak form efficient

Shahid (2008) explores causal relationships among equity prices and industrial production, money supply , exports, exchange rate , foreign direct investment and interest rates for the period 3/95 to 3/2007 by employing cointegration analysis and Toda and Yamamoto Granger causality test on quarterly data. Short run relationships among variables have also been investigated by using Bivariate Vector Autoregressive Model for variance decomposition and impulse

response functions. The study concludes that equity prices in India lead economic activity in general. However, Interest rate is found to lead the equity prices.

DATA DESCRIPTION AND METHODOLOGY

This study explores the long term causal relationship among macro economic variables and Pakistani capital market for the period 6/1998 to 6/2008 by using monthly data. The macroeconomic variables include Industrial Production Index, Broad Money, Oil Prices, Foreign Exchange Rate, Inflation and Interest Rate. Monthly time series has been chosen as it is consistent with earlier work done by Chan and Faff (1998) to explore the long run relation ship between macroeconomic variables and equity markets. Variables have been constructed and measured by using following proxies

Data Description

Equity Market Returns

Equity market returns has been calculated by using following equation

$$R_t = \ln (P_t / P_{t-1})$$

Where: R_t is Return for month 't';and P_t and P_{t-1} are closing values of KSE- 100 Index for month 't' and 't-1' respectively.

Industrial Growth rate

Industrial production index has been used as proxy to measure the growth rate in real sector and it has been calculated by using log difference of industrial production index.

$$\text{Growth Rate} = \ln (IIP_t / IIP_{t-1})$$

Studies that explore the relationship among industrial production and equity market returns include Chan, Chen and Hsieh (1985), Chen, Roll and Ross (1986), Burnmeister and Wall (1986), Beenstock and Chan (1988), Chang and Pinegar

(1990), Kryzanowski and Zhang (1992), Chen and Jordan (1993), Sauer (1994), Rahman, Coggin and Lee (1998).

It is hypothesized that an increase in growth rate is positively related to equity market returns.

Money Supply

Broad Money (M_1) is used as a proxy of money supply. Money growth rate has been calculated by using log difference of broad money (M_2)

$$\text{Money growth rate} = \ln (M_t / M_{t-1})$$

Studies that explore the relationship among money supply and equity market returns include Beenstock and Chan (1988), Sauer (1994)

It is hypothesized that an increase in money supply is positively related to equity market returns

Inflation Rate

Consumer Price Index is used as a proxy of inflation rate. CPI is chosen as it is a broad base measure to calculate average change in prices of goods and services during a specific period.

$$\text{Inflation Rate} = \ln (CPI_t / CPI_{t-1})$$

Studies that explore the relationship among inflation and equity market returns include Chan, Chen and Hsieh (1985), Chen, Roll and Ross (1986), Burnmeister and Wall (1986), Burmeister and MacElroy (1988), Chang and Pinegar (1990), Defina (1991) Kryzanowski and Zhang (1992), Chen and Jordan (1993), Sauer(1994), Rahman, Coggin and Lee (1998).

It is hypothesized that an increase in inflation is negatively related to equity market returns.

Change in oil prices

Brent oil prices has been used as proxy for oil prices and change in oil prices has been measured by using log difference i.e

$$\text{Change in oil prices} = \ln (\text{Brent}_{t} / \text{Brent}_{t-1})$$

Chan, Chen and Hsieh (1985), Chen and Jordan (1993) investigate the relationship among oil prices and equity markets for US market.

It is hypothesized that an increase in oil rates is negatively related to equity market returns

Change in Foreign Exchange Rate

Change in Foreign exchange rate is measured by employing end of month US$/Rs exchange rate and change in value is worked out through log difference i.e

$$\text{Change in foreign Exchange Rate} = \ln (FER_{t} / FER_{t-1})$$

Where FER is foreign exchange rate US $/Rs

Kryzanowski and Zhang (1992), Sauer (1994) also explore the relationship between foreign exchange rate and equity market returns.

It is hypothesized that depreciation in home currency is negatively related to equity market returns

Change in Interest Rate

Treasury bill rates have been used as proxy of Interest rate. Change in interest rate has been measured by using log difference to T bill rates.

$$\text{Change in Interest Rate} = \ln (TB_{t} / TB_{t-1})$$

Burmeister and MacElroy (1988) study the relationship between short term interest rates and equity market return.

It is hypothesized that an increase in interest rate is negatively related to equity market returns

Change in Foreign Portfolio Investment

Foreign portfolio Investment has been used as proxy of Investor confidence. Change in Foreign portfolio Investment has been measured by using log difference to Foreign portfolio Investment.

$$\text{Change in Interest Rate} = \ln (FPI_t / FPI_{t-1})$$

It is hypothesized that an increase in foreign portfolio investment is positively related to equity market returns

Methodology

There are several techniques for testing the long term causal and dynamic relationship among equity prices and macro economic variables. In this study the emphasis is given to test the relationship among macro economic variables and Karachi stock exchange by employing via;(i) Descriptive Statistics ,(ii) Correlation Matrix,(iii) JJ cointegration Tests,(iv) Granger Causality Test,(v) Impulse Response Analysis and (vi) Variance Decomposition Analysis

Stationarity of data is tested by using unit root tests. Null hypothesis of a unit root is tested by using Augmented Dickey-Fuller Test and Phillips-Perron Test. The ADF test examines the presence of unit root in an autoregressive model. A basic autoregressive model is $Zt = áZ_{t-1} + u_t$, where Z_t is the variable studied, t is the time period, á is a coefficient, and u_t is the disturbance term. The regression model can be written as $ÄZ_t = (á - 1)Z_{t-1} + u_t = äZ_{t-1} + u_t$, where Ä is the first difference operator. Here testing for a unit root is equivalent to testing. ä = 0.

The Dickey-Fuller tests assume that the error terms are statistically independent and have a constant variance. This assumption may not be true in some of the data used so Phillip Perron test is also used that relaxes above assumptions and permits the error disturbances to be heterogeneously distributed and it can be represented mathematically by

$$Z_t = á_o + á_1 Z_{t-1} + á_t \{t - T/2\} + u_t$$

Test statistics for the regression coefficients under the null hypothesis that the data are generated by $Z_t = Z_{t-1} + u_t$, where $E(u_t) = 0$.

If a time series is non stationary but it becomes stationary after differencing then said time series is said to be integrated of order one i.e. I (1). If two series are integrated of order one, there may exist a linear combination that is stationary without differencing. If such linear combination exists then such streams of variables are called cointegrated.

Cointegration tests are divided into two broader categories ;(i). Residual based test ;(ii). Maximum likelihood based tests. Residual based test include the

Engle-Granger (1987) test whereas Maximum likelihood based tests include Johansen (1988; 1991) and Johansen-Juselius (1990) tests. During this study we apply Johansen and Juselius test to determine the presence of cointegrating vectors in a set of non stationary time series. The null hypothesis is that there is no cointegration among the series. Vector Autoregressive (VAR) approach is employed to test multivariate cointegration. This assumes all the variables in the model are endogenous. The Johansen and Juselius procedure is employed to test for a long run relationship between the variables. Johansen and Juselius suggest two likelihood ratio tests for the determination of the number of cointegrated vectors. Maximal eigenvalue test evaluates the null hypothesis that there are at most r cointegrating vectors against the alternative of r + 1 cointegrating vectors. The maximum eigen value statistic is given by,

$$\lambda max = - T \ln (1 - \lambda r+1)$$

where λ r+1,…,λn are the n-r smallest squared canonical correlations and T = the number of observations.

Trace statistic tests the null hypothesis of r cointegrating vectors against the alternative of r or more cointegrating vectors. This statistic is given by

$$\lambda trace = -T \Sigma \ln (1 - \lambda i)$$

In order to apply the Johansen procedure, Lag length is selected on the basis of the Akaike Information Criterion (AIC).

If co-integration in the long run is present then the system of equations is restructured by inserting an Error Correction Term to capture the short-run deviation of variables from their relevant equilibrium values. This investigation is necessary as impact of financial development is generally more apparent in the short-run and disappears in the long run as economy expands and matures. According to Granger (1988) presence of cointegrating vectors indicates that granger causality must exist in at least one direction. A variable granger causes the other variable if it helps forecast its future values. In cointegrated series, as variables may possibly share common stochastic trends so dependent variables in the VECM must be Granger-caused by lagged values of the error-correction terms. This is possible because error-correction terms are functions of the lagged values of the level variables. Thus an evidence of cointegration between variables itself provides the basis for construction of error correction model. ECM permits the introduction of past disequilibrium as

explanatory variables in the dynamic behavior of existing variables thus facilitates in capturing both the short-run dynamics and long-run relationships between the variable. The chronological granger causality between the variables can be explored by using a joint F-test to the coefficients of each explanatory variable in the VECM. The variance decomposition of the equity returns is based on the analysis of responses of the variables to shocks. When there is a shock through the error term we study the influence of this shock to the other variables of the system and thus get information about the time horizon and percentage of the error variance F test is in fact a within-sample causality tests and does not allow us to gauge the relative strength of the of causality among variables beyond the sample period.

In order to examine the out of sample causality we use variance decomposition analysis which partitions the variance of the forecast error of a certain variable into proportions attributable to shocks in each variable in the system. Variance decomposition analysis present a factual breakup of the change in the value of the variable in a particular period resulting from changes in the same variable in addition to other variables in preceding periods. The impulse response analysis investigates the influence of random shock in a variable on other variables of interest. Impulse responses of returns in various markets to a shock in oil innovations are also examined. Impulse responses show the effect of shocks for different days separately whereas variance decomposition analysis exhibits the cumulative effect of shocks.

EMPIRICAL RESULTS

Table 1 displays the descriptive statistics regarding changes in macroeconomic variables and equity market returns. The average monthly returns earned at Karachi stock exchange during last ten years is 2.2 % which is equivalent to an annualized return of 29.28%. This is one of the highest returns offered by emerging equity markets. The highest returns achieved during one month are 24.11% and maximum loss incurred in one month is 27.8%.

Average monthly industrial growth rate is 0.22% which is not appreciating at all. Oil prices increased at an average monthly rate of 2.09%. Narrow money growth rate is 1.67% per month which is significantly high. Average change in consumer price index is 0.56% per month whereas T bill rates appear to change at a rate of 0.25% per month. Average decrease in value of Pakistani currency is 0.35%. Percentage changes in exchange rates ranges from a minimum of -7.62% to a maximum value of 3.07% percent. Foreign portfolio investment is on average

increased by 0.55% per month. Average change in Treasury bill is 1.81%. However, significantly high volatility is observed in equity returns, industrial production, oil prices and t bill rates. Unstable macroeconomic variables lead to high risk and affect over all quality of decisions.

Table 1: Descriptive Statistics

	ÄKse100	Ä IPI	Ä Oil	ÄX Rate	ÄT Bill	ÄCPI	ÄFPI	ÄM1
Mean	0.0220	0.0022	0.0209	-0.0035	-0.0025	0.0056	0.0055	0.0167
Median	0.0219	0.0016	0.0310	-0.0006	0.0000	0.0047	0.0018	0.0091
Std Dev	0.0912	0.1121	0.0788	0.0121	0.0985	0.0070	0.0238	0.0422
Skewness	-0.3055	-0.4653	-0.6324	-2.4291	-0.6279	0.9219	3.5235	4.2966
Min	-0.2780	-0.4857	-0.2161	-0.0762	-0.4242	-0.0088	-0.0605	-0.0646
Max	0.2411	0.3533	0.2241	0.0307	0.3200	0.0303	0.1651	0.3481

Table 2 shows the correlation among equity returns and macroeconomic variables. Weak correlation is generally observed between the equity return and monetary variables.

Table 2: Correlation Matrix

	ÄKse100	Ä IPI	Ä Oil	ÄX Rate	ÄT Bill	ÄCPI	ÄFPI	ÄM1
ÄKse100	1							
Ä IPI	-0.0257	1						
Ä Oil	-0.0391	-0.1321	1					
ÄX Rate	0.1219	0.0579	-0.0943	1				
ÄT Bill	-0.1429	-0.1637	0.0325	-0.1974	1			
ÄCPI	-0.1698	-0.0169	0.1892	-0.2029	0.2557	1		
ÄFPI	0.1490	-0.0146	-0.0655	0.0956	0.0221	-0.0172	1	
ÄM1	0.0241	0.1560	-0.0183	0.1455	-0.0198	-0.0145	0.0498	1

Interest rates are negatively correlated with equity returns which are logical as increase in interest rates leads to increase in discount rate and it ultimately results

in decrease in present value of future cash flows which represent fair intrinsic value of shares. However this relationship is found insignificant. The relationship between inflation and equity returns can also be viewed on the basis of above analogy. This relationship is also found insignificant. Foreign portfolio investment increases liquidity in market and higher demand leads to increase in market prices of shares so relationship should be positive. But this relation ship is found insignificant. Increase in oil prices increase the cost of production and decrease the earning of the corporate sector due to decrease in profit margins or decrease in demand of product. So negative relation ship is in line with economic ration but it is again insignificant. Money growth rate is positively correlated with returns that are in line with results drawn by Maysami and Koh (2000). The possible reason is that increase in money supply leads to increase in liquidity that ultimately results in upward movement of nominal equity prices. However relationship is insignificant and weak. Similarly interest rate parity theory is also confirmed from results as interest rate is negatively correlated with exchange rates.

Table 3: Unit Root Analysis

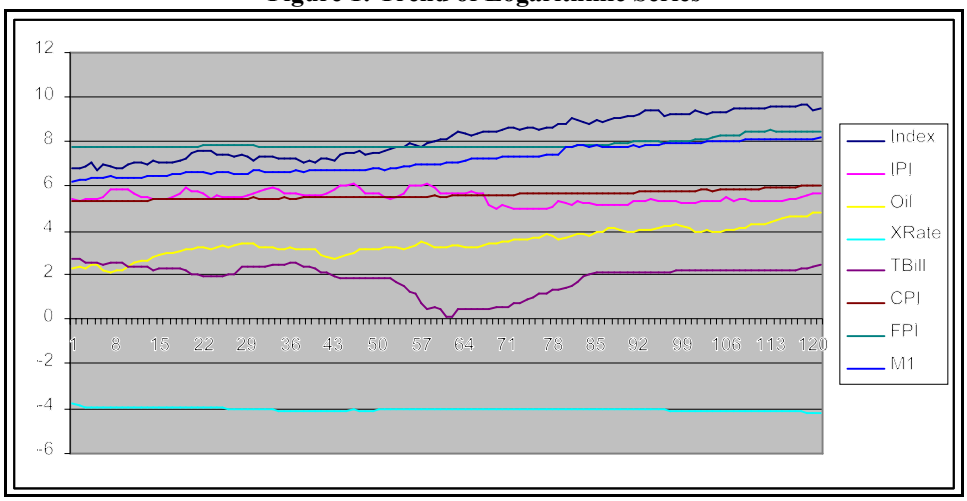
	ADF- Level	ADF- Ist Diff	PP- Level	PP- Ist Diff
Ln Kse100	-2.1686	-12.015	-2.0872	-12.2821
Ln IPI	-3.1322	-8.9420	-2.8182	-8.7609
Ln Oil	-2.3550	-8.3208	-2.0543	-8.2033
Ln X Rate	-2.3659	-6.6074	-3.1003	-6.4168
Ln T Bill	-1.6981	-3.6063	-1.3595	-7.8162
Ln CPI	2.9023	-8.6160	2.6215	-8.6190
Ln FPI	0.4762	-3.6651	-0.4640	-10.8700
Ln M1	-1.8832	-10.245	-1.9545	-10.2284
1% Critic. Value	-4.0363	-4.0370	-4.0363	-4.0370
5% Critic. Value	-3.4477	-3.4480	-3.4477	-3.4480
10%Critic Value	-3.1489	-3.1491	-3.1489	-3.1491

Correlation analysis is relatively weaker technique. Therefore causal nexus among the monetary variables has been investigated by employing multivariate cointegration analysis. Cointegration analysis tells us about the long term relationship among equity returns and set of monetary variables. Cointegration tests

involve two steps. In first step, each time series is scrutinized to determine its order of integration. For this purpose ADF test and Phillips-Perron test for unit has been used at level and first difference. Results of unit root test under assumption of constant and trend have been summarized in Tables 3.

Results clearly indicate that the index series are not stationary at level but the first differences of the logarithmic transformations of the series are stationary. Therefore, it can safely said that series are integrated of order one I (1).It is worth mentioning that results are robust under assumption of constant trend as well as no trend.

Figure 1: Trend of Logarithmic Series



In second step, time series is analyzed for Cointegration by using likelihood ratio test which include (i) trace statistics and (ii) maximum Eigen value statistics.

Table 4 exhibits the results of trace statistics at a lag length of three months. On the basis of above results null hypothesis of no cointegration between the equity indices and macroeconomic variables for the period 6/1998 to 3/2008 can not be rejected in Pakistani equity market. Trace test indicates the presence of 4 cointegrating vectors among variables at the $\alpha = 0.05$. In order to confirm the results Maximum Eigen value test has also been employed and Max Eigen value test also confirms the presence of cointegration at the $\alpha = 0.05$. Therefore, study provides evidence about existence of long term relationship among macroeconomic variables and equity returns.

Table 4: **Multivariate Cointegration Analysis Trace Statistic**

Hypothesized No. of CE(s)	Eigen value	Trace Statistic	Critical Value0.05	Prob.
None *	0.3923	193.3427	159.5297	0.0002
At most 1 *	0.2630	135.0690	125.6154	0.0117
At most 2 *	0.2087	99.3636	95.7537	0.0276
At most 3 *	0.1958	71.9817	69.8189	0.0333
At most 4	0.1507	46.4931	47.8561	0.0668
At most 5	0.1259	27.3791	29.7971	0.0927
At most 6	0.0667	11.6342	15.4947	0.1753
At most 7	0.0300	3.5632	3.8415	0.0591

It is worth mentioning that Johansen and Jusilius cointegration tests do not account for structural breaks in the data.

As variables are cointegrated so Granger Causality must exist among the variables. This requirement of granger representation theorem is helps us to identify the direction of causality flow. Table 5 reports the results Granger causality.

Above table provides evidence about existence of unidirectional causality from X Rate , T Bill , Money Supply and CPI to equity market returns at $\alpha = 0.05$. However no granger causality is observed in industrial production and equity market returns. Results can be summarized as that unidirectional causality flowing from monetary variables to equity market and this lead- lag relationship makes it imperative for financial and economic mangers of country to be more careful and vigilant in decision making as these decisions are priced in equity market and sets the trends in capital market which is considered as barometer of economy. However insignificant relationship with industrial production, oil indicates that market movement is not based on fundamentals and real economic activity.

Impulse response analysis provides information about the response of equity market returns to one standard deviation change in industrial production, oil, money growth rate, foreign portfolio investment, inflation, T bill and exchange rate. Fig 2 is graphical presentation of relationship between innovations in macroeconomic variables and equity market returns in the VAR system. Statistical significance of the impulse response functions has been examined at 95% confidence bounds.

Table 5: Granger Causality Test			
Null Hypothesis:	Obs	F-Statistic	Probability
IPI does not Granger Cause INDEX	117	0.5518	0.648
INDEX does not Granger Cause IPI		0.6710	0.5716
OIL does not Granger Cause INDEX	117	0.6649	0.5753
INDEX does not Granger Cause OIL		3.3713	0.0211
XRATE does not Granger Cause INDEX	117	6.1909	0.0006
INDEX does not Granger Cause XRATE		0.0989	0.9604
TBILL does not Granger Cause INDEX	117	3.5113	0.0177
INDEX does not Granger Cause TBILL		0.9056	0.4409
CPI does not Granger Cause INDEX	117	2.9798	0.0345
INDEX does not Granger Cause CPI		0.3946	0.7571
FPI does not Granger Cause INDEX	117	0.3015	0.8242
INDEX does not Granger Cause FPI		0.3832	0.7653
M1 does not Granger Cause INDEX	117	2.8654	0.0399
INDEX does not Granger Cause M1		0.5660	0.6385

Results confirm that one standard deviation change in money supply leads to increase in equity prices due to increase in liquidity and this result is consistent with results of Maysami and Koh(2000). Similarly one standard deviation change in Treasury bill rate leads to reduction in prices of equity due to increased discount rates. No statistically significant impact has been observed with reference to variation in exchange rates. It is acceptable because in Pakistan a managed floating rate system has been observed and during last five years exchange rates has been managed within a small range by state bank of Pakistan through open market operation. These results are in conformity with earlier work.

Impulse response function captures the response of an endogenous variable over time to a given innovation whereas variance decomposition analysis expresses the contributions of each source of innovation to the forecast error variance for each variable. Moreover, it helps to identify the pattern of responses transmission over time. Therefore variance decomposition analysis is natural choice to examine the reaction of equity markets to system vide shocks arising from changes in industrial

production, inflation, oil, money supply, Treasury bill rates, foreign portfolio investment and exchange rates. Table 7 exhibits the results of VDC Analysis..

Figure 2: Impulse Response Analysis

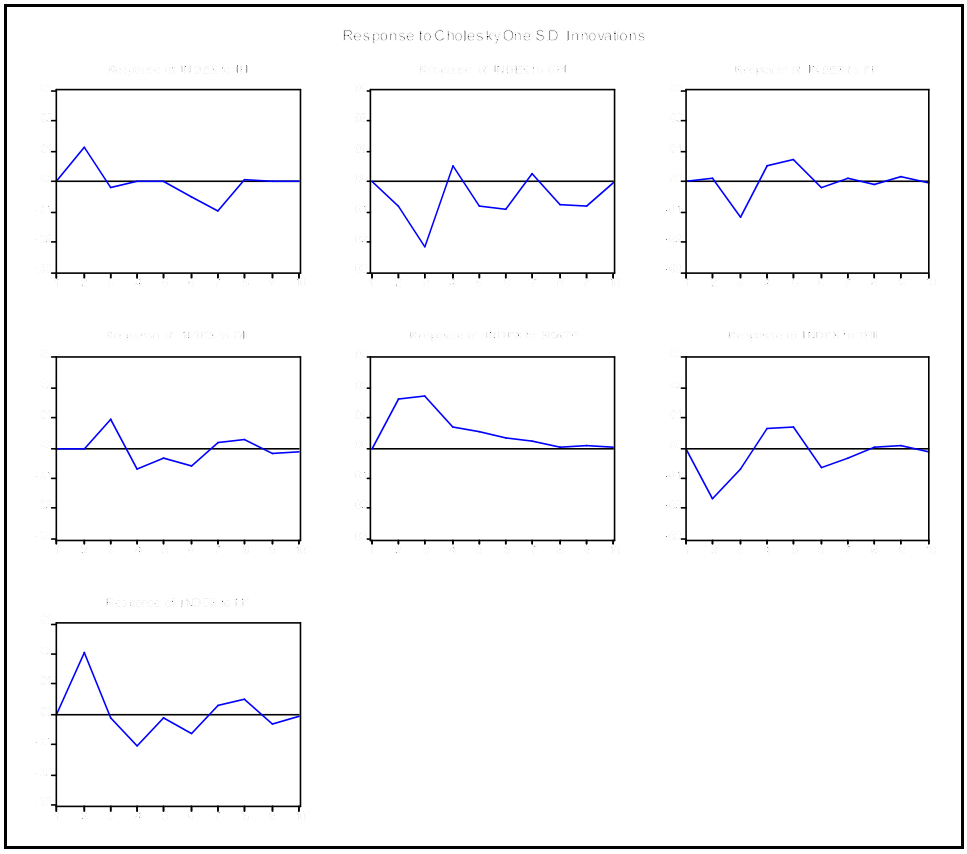


Table 7: Variance Decomposition Analysis									
Period	S.E.	INDEX	IPI	CPI	FPI	OIL	XRATE	TBILL	M1
1	0.08	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.09	86.18	1.56	0.77	0.01	0.00	3.17	3.29	5.02
3	0.10	76.68	1.44	5.58	1.45	0.98	5.97	3.43	4.46
4	0.10	74.47	1.39	5.68	1.67	1.40	6.25	3.70	5.44
5	0.10	72.98	1.36	6.18	2.16	1.47	6.42	4.09	5.33
6	0.10	71.32	1.59	6.82	2.14	1.75	6.36	4.41	5.60
7	0.10	70.50	2.48	6.78	2.12	1.76	6.31	4.44	5.60
8	0.10	69.88	2.46	7.27	2.11	1.83	6.26	4.41	5.80
9	0.10	69.37	2.44	7.80	2.12	1.84	6.22	4.38	5.84
10	0.10	69.36	2.44	7.80	2.12	1.84	6.21	4.39	5.84

Results confirm that monetary variables are a significant source of the volatility of equity market The contribution of an inflation shock to the equity returns ranges from 0.77 % to 7.8%. Similarly the contribution of T bill rates ranges from 3.29% to 4.39% and contribution of X rate ranges from 3.17% to 6.42% which is also significant. Money supply is also one of major contributor of volatility. Role of IPI and Oil in equity market volatility also increase gradually. The pattern of transmission of shocks is also apparent and indicates an increasing trend. This may be helpful to stake holders in their decision making process

CONCLUSION

This paper examines the long run relationship among equity market returns and seven important macroeconomic variables which include industrial production, Money Supply, , foreign portfolio investment, Treasury Bill Rates, oil prices, foreign Exchange Rates and consumer price index for the period 6/1998 to 6 /2008 by using Multivariate Cointegration Analysis and Granger Causality Test. Result provide evidence about existence of long run relationship among equity market and macroeconomic variables and explains the impact of changes at macroeconomic front on the stock market. Multivariate regression analysis provides evidence about the presence of four cointegrating vectors among variables at the $\alpha = 0.05$. Maximum Eigen value test also confirms the results.

Granger causality test indicates that T bill rates, exchange rates, inflation and money growth rate granger causes returns. This relationship has economic rational as increase interest rates , inflation leads to increase in discount rates and it ultimately results in reduction of prices. Impulse response analysis exhibits that one standard deviation change in money supply leads to increase in equity prices due to increase in liquidity and this result is consistent with results of Maysami and Koh(2000). No statistically significant impact has been observed among equity market and industrial production, oil prices and portfolio investment. Results can be summarized as that unidirectional causality flowing from monetary variables to equity market and this lead- lag relationship makes it imperative for financial and economic mangers of country to be more careful and vigilant in decision making as these decisions are priced in equity market and sets the trends in capital market which is considered as barometer of economy. However insignificant relationship with industrial production, oil indicates that market movement is not based on fundamentals and real economic activity.

Variance decomposition analysis is also performed that reveals that confirm that monetary variables are a significant source of the volatility of equity market The contribution of an inflation shock to the equity returns ranges from 0.77 % to 7.8%. Similarly the contribution of T bill rates ranges from 3.29% to 4.39% and contribution of X rate ranges from 3.17% to 6.42% which is also significant. Money supply is also one of major contributor of volatility.

These results reveal that identification of direction of relationship between the macroeconomic variables and capital market behavior facilitates the investors in taking effective investment decisions as by estimating the expected trends in exchange rates and interest they can estimate the future direction of equity prices and can allocate their resources more efficiently. Similarly, architects of monetary policy should be careful in revision of interest rates as capital market responds to such decisions in the form of reduction of prices. Similarly, Central bank should also consider the impact of money supply on capital markets as has significant relationship with dynamic of equity returns. As under efficient market hypothesis capital markets respond to arrival of new information so macroeconomic policies should be designed to provide stability to the capital market.

REFERENCES

Al-Sharkas (2004), The dynamic relationship between Macroeconomic factors and the Jordanian stock market, *International Journal of Applied Econometrics and Quantitative Studies* , 1(1),

Beenstock, M. and Chan, K.F. (1998), Economic Forces in the London Stock Market, *Oxford Bulletin of Economics and Statistics*, 50 (1): 27-39.

Burmeister, E. and K.D.Wall (1986), The Arbitrage Pricing Theory and Macroeconomic Factor Measures, *The Financial Review*, 21(1), 1-20.

Chan, K.C., N.F.Chen and D.A.Hsieh (1985),"An Exploratory Investigation of the Firm Size effect, *Journal of Financial Economics*, 14,.451-471

Chang, E.C. and J.M.Pinegar (1990),"Stock Market Seasonals and Prespecified Multifactor Pricing Relations, *Journal of Financial and Quantitative Analysis*, .25(.4), 517-533.

Chaudhuri, K. and S. Smile, (2004), Stock market and aggregate economic activity: evidence from Australia, *Applied Financial Economics*, 14, 121-129.

Chen, N.(1983), Some Empirical Tests of the Theory of Arbitrage Pricing, *The Journal of Finance*, 38,.1393-1414.

Chen, S.J. and B. D. Jordan (1993), Some Empirical Tests of the Arbitrage Pricing Theory: Macro variables vs. Derived Factors, *Journal of Banking and Finance*,17, .65-89.

Chen, Nai-fu, (1991), Financial Investment Opportunities and the macroeconomy, *Journal of Finance,* June, 529-554.

Chen, N. F., Roll, R., & Ross, S. (1986), Economic forces and the stock market, *Journal of Business,* 59, 383-403.

Cheung, Y. Wong and Ng, L. K (1998), International evidence on the stock market and aggregate economic activity, *Journal of Empirical Finance*, 5, 281–296.

Cutler, D. M., J. M. Poterba and L. H. Summers (1989), What moves stock prices?, *Journal of Portfolio Management,* 15, 4-12.

Dickey, D. & Fuller, W. (1979), Distribution for the Estimators for Autoregressive Time Series with Unit Root, *Journal of American Statistical Association*, 74, 427-431.

DeFina, R.H., (1991), Does Inflation Depress the Stock Market?, *Business Revie*w, Federal Reserve Bank of Philadelphia, 3-12.

Engle, R., and C. Granger (1987), Cointegration and Error Correction: Representation, Estimation, and Testing, *Econometrica,* 55(2), 251–276.

Fazal H and Mahmood T.(2001), The Stock Market and the Economy in Pakistan, *The Pakistan Development Review,* 40(2), 107–114.

Fazal H. (2006), Stock Prices, Real Sector and the Causal Analysis: The Case of Pakistan, *Journal of Management and Social Sciences*, 2(2), 179-185.

Flannery, M. J. & Protopapadakis, A. A. (2002), Macroeconomic factors do influence aggregate stock returns, *The Review of Financial Studies*, 15(3), 751-782.

Habbibullah M.S and A. Z. Baharumshah (1996), Money, output and stock prices in Malaysia: an Application of the cointegration tests, *International Economic Journal*, 10(2)

Hamao, Y. (1988), An empirical examination of the arbitrage pricing theory: Using Japanese data, *Japan and the World Economy,* 1, 45-61.

Hardouvelis, G. (1987), Macroeconomic information and stock prices, *Journal of Economic and Business,* 131-139.

Humpe, A. & Macmillan, P.(2007), Can Macroeconomic Variables Explain Long Term Stock Market Movements? A Comparison of the US and Japan, CDMA Working Paper No. 07/20

Johansen, S (1991), Estimation and hypothesis testing of cointegrating vectors in Gaussian vector autoregressive models, *Econometrica*,59, 1551–1580.

Johansen, S.and Juselius, K.(1990), Maximum likelihood estimation and inference on cointegration – With applications to the demand for money, *Oxford Bulletin of Economics and Statistics* 51, 169–210.

Kryzanowski, L. and H.Zang (1992), Economic Forces and Seasonality in Security Returns, *Review of Quantitative Finance and Accounting*, .2, 227-244.

Maysami, R. C. & Koh, T.S. (2000), A vector error correction model of the Singapore stock market, *International Review of Economics and Finance*, 9, 79-96.

Maysami R Cooper, Lee Chuin Howe L. C. and Hamzah A. (2004), Relationship between Macroeconomic Variables and Stock Market Indices: Cointegration Evidence from Stock Exchange of Singapore's All-S Sector Indices, *Jurnal Pengurusan* 24 47-77.

McMillan, D.G. (2001), Cointegration Relationships between Stock Market Indices and Economic Activity: Evidence from US Data, University of St Andrews, Discussion Paper 0104.

Mukherjee, T. K., and A. Naka (1995), Dynamic Relations Between Macroeconomic Variables and the Japanese Stock Market: An Application of a Vector Error-Correction Model, *The Journal of Financial Research* , 18, 223-37

Nasseh A. & Strauss J (2000), Stock prices and domestic and international macroeconomic activity: A cointegration approach, *Quarterly Review of Economics and Finance,* 40(2): 229-245

Pearce, D. K. & Roley, V.V. (1988), Firm characteristics, unanticipated inflation, and stock returns, *Journal of Finance,* 43, 965-981.

Phillips, P. & P. Perron (1988), Testing for a Unit Root in Time Series Regression, *Biometrika*, 75, 335-46.

Poon, S. and S.J. Taylor, (1991), Macroeconomic Factors and the U.K. Stock Market, *Journal of Business, Finance and Accounting*, 18, 619–636.

Robert D. Gay, Jr.(2008), Effect of macroeconomic variables On stock market returns For four emerging economies: Brazil, Russia, India, and China, *International Business & Economics Research Journal*, 7(3), 1

Ross, S. (1976), The arbitrage theory of capital assets ,*Journal of Economic Theory,* 341-60.

Shahid A (2008), Aggregate Economic Variables and Stock Markets in India, *International Research Journal of Finance and Economics*, Issue 14

THE EFFECTS OF ALCOHOL USE ON SCHOOL ENROLLMENT

Wesley A. Austin, University of Louisiana at Lafayette

ABSTRACT

Considerable controversy surrounds the effects youth alcohol use has on educational outcomes. This article addresses the question of whether youth drinking leads, in causal ways, to lower school enrollment, or is the widely reported negative correlation between drinking and this educational outcome caused by common unobservable factors? An instrumental variable model is estimated to study the effects of several drinking measures on the probability school enrollment for a sample of high school and college age individuals. Extensive testing is conducted to verify instrument strength and exogeneity. Results indicate that alcohol use reduces school enrollment among those of high school and college age and results are consistent across instrument specifications.

INTRODUCTION

In many health-related and social science fields, there has been considerable concern about the various harmful effects of alcohol use. Recent evidence indicates drinking, coupled with smoking, reduces income (Auld 2005). Another related consequence of alcohol use is the potential reduction in human capital accumulation by drinkers. This issue is particularly acute during adolescence and early adulthood, in which decisions regarding high school completion and college attendance are first considered, and academic performance realizations that affect longer-term educational and economic outcomes are initially observed. Excessive drinking has been associated with this age group despite its illegality until the age of 21. For instance, data from the 2006 and 2007 National Survey on Drug Use and Health (NSDUH) found approximately 18 percent of youths ages 15 – 18 (high school age) and approximately 43 percent of young adults ages 18 – 25 (college age) engaged in binge drinking, i.e. the consumption of at least five alcoholic beverages in one sitting, in the past month.

Several reasons might lead heavy drinking to impair human capital formation. Intoxication potentially interferes with class attendance and learning, and

time spent in activities where drinking occurs could substitute away from time allocated to studying. This hurts academic performance in the short term, which might diminish the ability or incentive to continue schooling over the longer term. Risks stemming from intoxication, such as injury from accidents or fights, pregnancy and disease from unsafe sex, conflicts with parents or law enforcement, and a tarnished reputation with school authorities can also limit the capability of a student to remain in school (Cook and Moore 1993). Alternatively, social interactions associated with drinking might improve academic achievement by providing a means of relieving stress (Williams et al. 2003).

Much evidence has established a negative relationship between the regularity and intensity of drinking and human capital measures such as school completion. But distinguishing whether these relationships are causal, such that increased alcohol consumption directly reduces, for example, probable school enrollment, or merely correlational, with changes in other confounding variables simultaneously leading to drinking and lower enrollment rates, is critical.

Thus, for economists and policy makers, obtaining an accurate estimate of the magnitude of the causal effect that alcohol use has on educational outcomes should be a top priority. This task is a natural one to tackle by using econometric techniques such as instrumental variables (IV) regression – a method specifically designed to estimate the causal impact of a variable that does not otherwise vary independently with other unobserved determinants of the outcome being examined.

Why is the potential impact of alcohol use on school enrollment relevant for the discipline of economics? Human capital accumulation bears directly and heavily on earning potential and it is widely accepted that strong and statistically significant relationships link individual health and human capital formation. Moreover, variables such as school completion and enrollment are commonly examined education outcomes among broader literatures on human capital accumulation, given that they are easily measured and have a clear marginal impact on future wages that economists have long focused on estimating.

LITERATURE OVERVIEW

Only recently has the relationship between alcohol use and human capital accumulation been addressed by economists, and research on the topic had been fairly limited, with measures of drinking and schooling as well as conclusions varying across studies. Comparatively early research produces evidence of a negative relationship, but either makes no attempt to econometrically deal with the

potential endogeneity of drinking in education equations, or does so in a way that has since been criticized as unsatisfactory, so it is unclear whether this negative correlation indeed represents declines in educational outcomes that are caused by drinking.

Cook and Moore (1993), estimate IV models in which the effect of current alcohol use on post-secondary schooling was identified by the state excise tax on beer and an indicator for whether the student could legally drink based on the state's MLDA. Results from three separate specifications show that heavy drinking in 12th grade decreased subsequent schooling. Dee and Evans (2003) call into question the causal effect interpretation of these results. They argue that the use of cross-state alcohol policy variation to identify the effects of drinking on other outcomes is potentially problematic because such variation might be correlated with unobservable attributes that affect both alcohol use and educational attainment.

Mullahy and Sindelar (1994), use ordinary least squares (OLS) regressions, and find that the onset of alcoholism symptoms by age 22 is associated with a five percent reduction in completed schooling. Yamada et al. (1996) use single equation probit models that do not account for the possibility that alcohol use is endogenous. Results show that the probability of high school graduation is 6.5 percent lower for students who consumed alcohol on at least two occasions in the previous week. In addition, drinking is inversely related to beer taxes, liquor prices, MLDAs and marijuana decriminalization, meaning that each is positively associated with high school graduation rates through its covariance with alcohol use.

Koch and Ribar (2001) examine the relationship between age of drinking onset and educational attainment by age. Estimates from IV models that specify sibling onset age as the instrument for respondent onset age imply that delaying alcohol initiation by a year increases subsequent schooling by 0.22 years. However, they argue that this represents an upper bound for the effect size based on the sign of the bias if the assumptions needed for consistency are not met, and indeed OLS and family fixed effects models produce estimates that are three to four times smaller for males, and still smaller and sometimes insignificant for females.

More recent evidence comes from Chatterji and DeSimone (2005), who estimate the effect of binge and frequent drinking by adolescents on subsequent high school dropout using an IV model with an indicator of any past month alcohol use as the identifying instrument. In contrast to the last two studies cited above, the authors find that OLS yields conservative estimates of the causal impact of heavy drinking on dropping out, such that binge or frequent drinking among 15–16 year old students lowers the probability of having graduated or being enrolled in high

school four years later by at least 11 percent. The results of overidentification tests using two measures of maternal youthful alcohol use as additional instruments provide support for their empirical strategy. Also, Oreopoulos (2006) finds that the gains from policies requiring compulsory schooling up to a certain age are quite large, regardless of whether "these laws impact on a majority or minority of those exposed."

DATA

The National Survey on Drug Use and Health (NSDUH), sponsored by the Substance Abuse and Mental Health Services Administration (SAMHSA), is administered to approximately 55,000 civilian, non-institutionalized individuals age 12 and over, chosen so that the application of sample weights produces a nationally representative sample, with approximately equal numbers of respondents from the 12–17, 18–25 and 26 and over age groups. Data from the NSDUH allow for both breadth and depth of coverage on the topic. Breadth comes from the ability to study aspects of educational outcomes using data from an elaborate questionnaire covering a wide array of youth experiences. Depth is provided by numerous variables on demographics, family income, family composition and relocation.

An equally important facet of the NSDUH data is that they are conducive for the use of the IV regression methodology to estimate the causal effect of alcohol use on human capital. Abundant information is collected on experiences related to alcohol consumption, including measures of religiosity and the perceived risks involved in alcohol/ drug use. An assortment of variables are observed, therefore, that have the potential to serve as instruments for the proposed model, in the sense that they are very likely to be highly correlated with alcohol use but would not have any obvious reason to be otherwise associated with educational outcomes.

A potentially problematic attribute of the data is non-random measurement error emanating from the self-reported nature of responses. Although IV will eliminate bias from random measurement error, it cannot salvage data plagued by systematic measurement error. However, studies on the quality of self-reported academic variables and drinking data suggest that such reporting bias should be minimal. Cassady (2001) finds that self-reported GPA values are "remarkably similar to official records" and therefore are "highly reliable" and "sufficiently adequate for research use." Grant et al. (1988), Midanik (1988) and Reinisch et al. (1991) conclude that youth drinking self-reports are reliable, based on the consistency of responses to alcohol use questions from repeated interviews. Harrison

and Hughes (1997) find that survey methods not requiring subjects to verbally answer questions, as in the NSDUH, increase the accuracy of substance use self-reports.

RESEARCH METHOD AND EMPIRICAL SPECIFICATION

In determining causation, the primary methodological question is whether drinking is properly specified as an exogenous variable with respect to educational outcomes or should instead be treated as endogenous. Consider the following equations, in which drinking (D) is a function of exogenous factors and an educational variable such as school enrollment (E) is a function of some (but not all) of the same exogenous determinants as well as D,

$$(1)\ D = \alpha_0 + Z\alpha_1 + X\alpha_2 + \omega,$$

$$(2)\ E = \beta_0 + \beta_1 D + X\beta_2 + \epsilon.$$

In the above equations, which apply to individual NSDUH respondents (with the corresponding observation-level subscript suppressed), vectors X and Z represent sets of exogenous variables that affect both drinking and enrollment (X), and drinking but not enrollment (Z), ω and ϵ are error terms that encompass all factors influencing the corresponding dependent variable that are not explicitly controlled for on the right hand side of the equations, and the α's and β's are parameters to be estimated. Econometrically, alcohol use is exogenous in equation 2 if it is uncorrelated with the error term ϵ. This condition holds, by definition, if none of the unobserved schooling determinants are related to drinking. If so, there is no need to estimate equation 1; a single equation regression method such as OLS will produce consistent estimates of the causal effect of drinking, β_1.

However, two sources of endogeneity could possibly lead to a nonzero correlation between alcohol use (D) and the error term in (2). One is unobserved heterogeneity, which would occur if any unmeasured educational outcome (e.g. enrollment) determinants that are subsumed in the error term ϵ are correlated with alcohol use; the resulting estimate of β_1 in (2) would suffer from omitted variable bias, which cannot be eliminated directly because the omitted variables are not recorded in the data. Disruptive events such as parental separation or divorce might simultaneously be responsible for greater alcohol consumption and lower school enrollment rates.

Such events are not observed and thus are not held constant in the regression. The negative correlation between drinking and school enrollment that they induce becomes embedded into the alcohol use coefficient, which is thus biased negatively as an estimate of the causal drinking effect. Conversely, unmeasured ability or socioeconomic background could create a positive bias in the estimated drinking effect if higher ability individuals are better able to function normally after alcohol consumption, or students who have more money to spend on alcohol also enjoy greater academic success and are more likely to be enrolled in school.

The other potential source of endogeneity is reverse causation. If alcohol use and educational outcomes like enrollment are simultaneously determined, the outcome will not only be a function of drinking, as specified in equation 2, but also will be a contributing factor to the decision regarding whether and how much alcohol to consume. In terms of equation 2, shocks to the error term ϵ that, by definition, influence educational outcomes will ultimately extend to drinking through the feedback effect of educational outcomes on alcohol consumption, thus creating a correlation between alcohol use and ϵ that renders the estimate of the causal drinking effect β_1 inconsistent. To investigate the possibility that alcohol use is endogenous as an explanatory factor for school enrollment, this analysis utilizes the method of instrumental variables (IV).

To use IV, there must be at least one, preferably two or more, variables (i.e. instruments or IVs) that affect alcohol use but have no direct impact on enrollment. In the case of exactly one instrument Z, the IV method works by estimating the causal drinking effect β_1 as the ratio of the sample correlation between the instrument and school enrollment to the sample correlation between the instrument and alcohol use, i.e.

$$(3)\ \beta_1 = \text{corr}[Z, E]/\ \text{corr}[Z, D],$$

where the quantity is estimated from the data and the correlations are estimated while holding constant the vector X of explanatory factors. Because the instrument is exogenous and related to enrollment only through drinking, the sample correlation between the instrument and enrollment is purely a product of that between drinking and enrollment. Thus, the sample correlation between the instrument and enrollment merely needs to be standardized by that between the instrument and drinking in order to be used as an estimate for the causal effect of drinking on school enrollment. In the case of two or more instruments, $\hat{D}$, the linear projection of Z onto D, takes the place of Z in equation 3.

Equation 3 makes transparent the two important conditions that the instrument vector Z must satisfy in order for IV to produce consistent estimates of the causal drinking effect β_1: First, the instruments must be highly correlated with alcohol use but not correlated with school enrollment through any other mechanism besides drinking. If the correlation between the instruments and drinking is not statistically significant, the denominator in (3) is statistically equal to zero, thus rendering the expression for β_1 indeterminate. The strength of this correlation is judged from the F-statistic for the joint significance of α_1 in equation 1. Minimally, α_1 should be significant at the 1 percent level; beyond this, Staiger and Stock (1997) advise a more stringent requirement that the associated F-statistic be at least 10.

Second, if a direct correlation between the instruments and school enrollment exists outside of the pathway from the instruments to drinking to enrollment, the numerator in (3) includes variation that is not part of the relationship between drinking and enrollment, and consequently the expression is no longer a consistent estimate of the causal effect of drinking. The reason multiple instruments are preferred is this overidentifies equation 2, which allows for specification tests to determine the empirical validity of excluding the instrument set Z from (2).

Under the null hypothesis that the instruments are not separately correlated with school enrollment, the sample size multiplied by the R-squared from a regression of the residual in (2), $\hat{\varepsilon}$, on all the exogenous variables (i.e. a constant, X and Z) is distributed as chi-square with degrees of freedom equal to one less than the number of instruments. Typically, the estimator represented by equation 3 is generated by a two-stage least squares (2SLS) procedure. The first stage estimates equation 1 above using OLS. From the estimated parameters, predicted values of alcohol use, $\hat{D}$, are constructed for each respondent using their corresponding values of the explanatory variables X and instruments Z. The second stage estimates equation 2 using the fitted values $\hat{D}$ in place of observed drinking D.

2SLS yields consistent estimates even when alcohol use and/or education variables are represented by a binary indicator. However, for binary drinking measures, e.g. an indicator of any past month binge drinking, an approach suggested by Wooldridge (2003) to improve efficiency is utilized. It is similar to 2SLS with two modifications. First, before running 2SLS, a preliminary probit regression for equation 1 is estimated. Second, the ensuing 2SLS procedure uses the predicted probabilities of drinking from the probit regression as instruments in place of Z. The resulting estimates are likely to be similar in magnitude to those that would be generated by the analogous 2SLS regression, but standard errors will be slightly smaller.

One other methodological point merits attention. Although IV estimates are consistent if the instrument strength and exogeneity conditions outlined above are satisfied, they are inefficient relative to OLS if it turns out that alcohol use is truly exogenous with respect to school enrollment, in which case the OLS estimates can be interpreted as causal effects. Thus, it is desirable to econometrically test the null hypothesis that drinking is exogenous in the enrollment equation. This is done using a Hausman (1978) test, which proffers that, if drinking and the error term are uncorrelated, IV and OLS estimates should differ only by sampling error. If the null hypothesis of exogeneity is rejected, OLS estimates are inconsistent and hence conclusions should be based on IV estimates; failure to reject the null means that OLS estimates are preferable because of their smaller standard errors.

SCHOOL ENROLLMENT

Current school enrollment is a binary variable indicating whether the respondent is currently enrolled in middle or high school (including those who are home schooled) or a college/ university. Approximately 99 percent of youth ages 15 and under report attending school, and individuals ages 26 and above who have not graduated from college are particularly likely to have experienced previous gaps in school enrollment, not currently be enrolled and not return to school in the future. The enrollment analysis is conducted utilizing a sample of high-school age students (15-18 years old) and college age students (19-25 years old). For the high school age sample, age 15 is the omitted category in the regressions thus mitigating the effects of compulsory attendance laws which typically require school attendance up to age 16.

DRINKING VARIABLES

Among the varied measures utilized are: the number of days the respondent drank in the past year (which is coded as '0' for nondrinkers and those that consumed no drinks in the previous year) and the number of drinks consumed in the previous month (which is coded as '0' for nondrinkers and those that consumed no drinks in the previous month). Binge drinking is defined as consuming five or more drinks on the same occasion on at least one day in the past thirty days. Although the timing of the number of drinks and binge drinking variables is not an ideal match for the enrolment measure, in the sense that past month consumption cannot literally affect behavior that preceded the past month, this work will follow that of previous

studies in assuming that previous month drinking patterns proxy those occurring in the recent period prior to the previous month.

The impact on enrollment from alcohol abuse or dependence in the past year is also examined. This is accomplished by an indicator in the NSDUH of whether respondents exhibited symptoms of alcohol abuse or dependence in the past year. This is retrospectively coded by SAMHSA based on responses to questions corresponding to criteria outlined in the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders*, the clinical standard for establishing drug abuse and dependence.

EXOGENOUS VARIABLES

Several variables from the NSDUH data are considered exogenous (i.e. explanatory) in the model: family income is measured in seven categories: $10,000 or less; $10,000-$19,999; $20,000-$29,999; $30,000-$39,999; $40,000-$49,999; $50,000-$74,999; and $75,000 or greater, with $10,000 or less as the omitted category. Population density is represented by indicators for two categories: an MSA with one million persons or greater and an MSA of less than one million persons, with non-MSA areas as the omitted category. A binary measure is included for whether the respondent has ever been arrested. For race, indicators are specified for African Americans, Native Americans, Asians, non-white Hispanics and multiracial, with Caucasians as the omitted category in the regressions. Family size is measured using two variables: the number of members if the household has one to five members and an indicator for those with over five members. A binary measure of gender is included as well.

Age indicators for the high school age sample are 16, 17, or 18 years old and 19, 20, 21, 22 or 23, 24 or 25 years old for the college age sample. Indicators for the last grade completed is 9^{th}, 10^{th} or 11^{th} grade (with 12^{th} as the omitted grade) for the high school age sample and freshman or sophomore/ junior (with senior as the omitted category) for the college age sample.

INSTRUMENTAL VARIABLES

Several NSDUH variables conceivably influence drinking without having direct effects on school enrollment and are thus candidates to serve as instrumental variables. The specific variables utilized for the high school age sample are: perceived risk of bodily harm from alcohol use; whether religious beliefs are

important and whether religious beliefs influence decisions. The specific variables utilized for the college age sample are: perceived risk of bodily harm from alcohol use; perceived risk of bodily harm from marijuana use and whether religious beliefs influence decisions.

For alcohol risk, a binary measure indicates if the respondent feels there are great/ moderate risks or slight/ no risks of harm, physically or otherwise, from consuming four to five drinks once or twice a week. For marijuana risk, a binary measure indicates if the respondent feels there are great/ moderate risks or slight/ no risks of harm, physically or otherwise, from using marijuana once or twice a week. Given that these variables only pertain to consuming illegal substances, it is presumed that there is no direct influence on school enrollment.

For both religion variables, a binary variable is created and coded as '0' if religion is not important or does not influence decisions and '1' otherwise. Religiosity has been linked to drinking behaviors (Kenkel and Ribar, 1994) but some evidence has established exogeneity with respect to educational outcomes (Wolaver, 2002). All instrumental variables undergo extensive testing in the following section.

EMPIRICAL FINDINGS

The causal effect drinking has on the probability of school enrollment is estimated using the three instrumental variables listed above. The main results of the IV analysis are also compared with parameter estimates obtained using OLS methodology. While discussion that follows concentrates on the effects of alcohol consumption and specification tests, appendix 1, for the binge drinking measure, shows the IV coefficients and marginal effect standard errors of all exogenous variables on the probability of enrollment for the high school age sample. Appendix 2 does the same for the college age sample.

Table 1. Descriptive Statistics (high school age sample) (n=19,022)

Variable	Mean	Std. Deviation
Number of days drank-past year	17.823	45.594
Number of drinks in previous month	5.703	32.916
Binge drinking in the past 30 days	0.119	0.324
Abuse/ Dependence on alcohol classification	0.080	0.272
Respondent perceives risk of harm from	0.762	0.426

Table 1. Descriptive Statistics (high school age sample) (n=19,022)		
Variable	**Mean**	**Std. Deviation**
drinking		
Religious beliefs are important in life	0.720	0.449
Religion influences your decisions	0.633	0.482
Probability of school enrollment	0.931	0.253
Family income ($10,000-$19,999)	0.108	0.310
Family income ($20,000-$29,999)	0.116	0.320
Family income ($30,000-$39,999)	0.105	0.307
Family income ($40,000-$49,999)	0.106	0.308
Family income ($50,000-$74,999)	0.190	0.392
Family income ($75,000 or more)	0.287	0.452
MSA segment with 1+ million persons	0.417	0.493
MSA segment of less than 1 million	0.486	0.500
Age of student (15 years old)	0.282	0.450
Age of student (16 years old)	0.278	0.448
Age of student (17 years old)	0.272	0.445
Age of student (18 years old)	0.255	0.436
Last grade in (9th grade)	0.015	0.123
Last grade in (10th grade)	0.135	0.342
Last grade in (11th grade)	0.306	0.461
Last grade in (12th grade)	0.300	0.458
Ever been arrested	0.096	0.498
Race (African American)	0.146	0.354
Race (Native American)	0.016	0.124
Race (Asian)	0.033	0.179
Race (non-white Hispanic)	0.165	0.371
Number in family	3.191	1.543
Number in family (>5)	0.139	0.346

Table 2. Descriptive Statistics (college age sample) (n=20,666)		
Variable	**Mean**	**Std. Deviation**
Number of days drank-past year	49.773	76.094
Number of drinks in previous month	15.536	50.292
Binge drinking in the past 30 days	0.300	0.458
Abuse/ Dependence on alcohol classification	0.148	0.355
Respondent perceives risk of harm from drinking	0.891	0.310
Religion influences your decisions	0.627	0.483
Respondent perceives risk of harm from marijuana	0.790	3.506
Probability of school enrollment	0.441	0.496
Family income ($10,000-$19,999)	0.156	0.362
Family income ($20,000-$29,999)	0.139	0.346
Family income ($30,000-$39,999)	0.116	0.321
Family income ($40,000-$49,999)	0.111	0.314
Family income ($50,000-$74,999)	0.140	0.347
Family income ($75,000 or more)	0.161	0.367
MSA segment with 1+ million persons	0.399	0.489
MSA segment of less than 1 million	0.516	0.499
Age of student (19 years old)	0.157	0.364
Age of student (20 years old)	0.140	0.347
Age of student (21 years old)	0.126	0.332
Age of student (22 or 23 years old)	0.205	0.403
Age of student (24 or 25 years old)	0.189	0.392
Freshman	0.148	0.355
Sophomore/ Junior	0.191	0.393
Ever been arrested	0.193	0.395
Race (African American)	0.142	0.349
Race (Native American)	0.017	0.129
Race (Asian)	0.031	0.174
Race (non-white Hispanic)	0.192	0.394

Table 2. Descriptive Statistics (college age sample) (n=20,666)		
Variable	**Mean**	**Std. Deviation**
Number in family	2.950	1.388
Number in family (>5)	0.104	0.305

Tables 1 and 2 present select summary statistics. The mean number of days drinks were consumed in the past year is about 18 (high school age) and 50 (college age) while the mean number of drinks consumed in the past month is 5.7 (high school age) and 15.5 (college age). Mean alcohol abuse/ dependence is 0.08 (high school age) and 0.14 (college age). Mean school enrollment is 0.44 for those of college age, and as expected, very high (0.93) for the high school age sample. Mean reported family income for college age sample is lower across the board as individuals of this age have moved out of the parental household. About 90 percent of respondents in both samples live in an MSA, roughly equally split between MSAs with populations greater than and less than one million. African Americans comprise about 14 percent of both samples while non-white Hispanics account for about 16 percent of the high school sample and 19 percent of the college sample.

FIRST STAGE REGRESSION RESULTS

Table 3 presents the probit results for the drinking measures on the instruments for the high school age sample. Of those who perceive that there is moderate to great risk of harm from consuming alcohol, the number of days drinking occurred in the past year is lowered by about 23 days. The number of drinks consumed in the past month is reduced by 11, while the likelihood of binge drinking in the last 30 days falls by 0.13 percentage points. The likelihood of being categorized as abusive/ dependent on alcohol falls by 0.09 points.

Importance of religious beliefs reduces all alcohol use measures. For those that report that religion is important in life, the number of days drinking occurred in the past year is lowered by approximately one day. The number of drinks consumed in the past month is reduced by 0.30, while the probability of binge drinking in the last 30 days falls by 0.02 percentage points. The likelihood of being categorized as abusive/ dependent on alcohol falls by 0.007 points.

Table 3. First stage regression estimates for the probability of enrollment (high school age) (n=19,022)				
Exogeneous Variables	number of days drank in past year	number of drinks in past month	binge drinking	abuse/ dependence on alcohol
Risk of bodily harm from drinking	-22.895	-10.946	-0.130	-0.089
	(1.012)	(0.766)	(0.007)	(0.006)
Religious beliefs are important in life	-0.891	-0.030	-0.016	-0.007
	(0.912)	(0.691)	(0.006)	(0.006)
Religion influences your decisions	-8.676	-2.830	-0.045	-0.036
	(0.854)	(0.646)	(0.006)	(0.005)
F stat/ chi2-coefficient of joint significance	249.05	82.12	418.29	272.28
P-value of significance level	(0.0000)	(0.0000)	(0.0000)	(0.0000)

When religiosity impacts decisions, the effects on the drinking measures are more pronounced. The number of days drinking occurred in the past year is lowered by nine days. The number of drinks consumed in the past month is reduced by about two, while the probability of binge drinking in the last 30 days falls by 0.45 points. The likelihood of being categorized as abusive/ dependent on alcohol falls by 0.04 points. The χ^2 coefficients and associated p-values indicate that the instruments are jointly significant for all the drinking measures.

Table 4 presents the probit results for the instruments for the college age group. For this age group, if moderate to great risk of harm from consuming alcohol is perceived, the number of days in which drinking occurred in the past year is lowered by 42 days. The number of drinks consumed in the past month is reduced by roughly18, while the probability of binge drinking in the last 30 days falls by 0.20 percentage points. The likelihood of being categorized as abusive/ dependent on alcohol decreases by 0.11 points.

If moderate to great risk of harm from using marijuana is perceived, the number of days in which drinking occurred in the past year is lowered by one day. The number of drinks consumed in the past month is reduced by 0.28, while the probability of binge drinking in the last 30 days falls by 0.003 percentage points. The likelihood of being categorized as abusive/ dependent on alcohol falls by 0.002 points. When religiosity impacts decisions, the number of days in which drinking occurred in the past year is reduced by 15 and the number of drinks consumed in the past month is reduced by four. The probability of binge drinking in the last 30 days

falls by 0.09 percentage points while the likelihood of being categorized as abusive/ dependent on alcohol falls by 0.04 points. The F statistics and χ^2 p-values signify support for the hypothesis of joint instrument significance for all the drinking measures.

Table 4. First stage regression estimates for the probability of enrollment (college age) (n=20,666)

Exogeneous Variables	number of days drank in past year	number of drinks in past month	binge drinking	abuse/ dependence on alcohol
Risk of bodily harm from drinking	-42.628	-18.468	-0.201	-0.105
	(1.579)	(1.067)	(0.009)	(0.007)
Risk of bodily harm from using marijuana	-0.816	-0.280	-0.003	-0.002
	(0.138)	(0.093)	(0.008)	(0.001)
Religion influences your decisions	-15.077	-4.690	-0.086	-0.039
	(1.018)	(0.688)	(0.006)	(0.005)
F stat/ chi2-coefficient of joint significance	352.67	125.76	665.92	241.11
P-value of significance level	(0.0000)	(0.0000)	(0.0000)	(0.0000)

THE EFFECTS OF DRINKING ON THE PROBABILITY OF SCHOOL ENROLLMENT (HIGH SCHOOL AGE)

As shown in table 5, drinking has significant, negative effects on the probability of being enrolled. For each daily increase in past year drinking, the probability of being enrolled is subsequently lowered by 0.001. For each additional drink increase in the past month, the probability of enrollment is also lowered by 0.003. If, for instance, the respondent reports drinking 52 days in the previous year, the likelihood of enrollment is diminished by approximately 0.052 points compared to not drinking at all. If the student reports consuming 30 drinks in the previous month, the probability of enrollment decreases by 0.09 points.

Table 5. IV estimates of drinking on the probability of enrollment (high school age) All three instruments (n=19,022)		
Alcohol variables	**IV**	**OLS**
number of days drank-past year	-0.001*	-0.0002*
Marginal Effect Standard Error	(0.0002)	(0.0000)
P-value of overidentification test	0.828	
Hausman statistic (p-value)	-5.243 (0.000)	
number of drinks in past month	-0.003*	-0.0003*
Marginal Effect Standard Error	(0.0006)	(0.0001)
P-value of overidentification test	0.303	
Hausman statistic (p-value)	-4.483 (0.000)	
binge drinking	-0.230*	-0.0042*
Marginal Effect Standard Error	(0.040)	(0.0054)
P-value of overidentification test	0.649	
Hausman statistic (p-value)	-5.772 (0.000)	
abuse/ dependence on alcohol	-0.329*	0.0017*
Marginal Effect Standard Error	(0.060)	(0.0060)
P-value of overidentification test	0.825	
Hausman statistic (p-value)	-5.624 (0.000)	
*Statistically significant at 1%		

Binge drinking further reduces the probability of enrollment by 0.23 points. For students who have engaged in binge drinking, the probability of school enrollment declines by approximately 24 percent compared to not binging. For those classified as abusive/ dependent with respect to alcohol, the probability of enrollment decreases by 0.32 points and this categorization reduces the probability of school enrollment by 35 percent. For all drinking indicators, the overidentification tests have associated p-values that offer strong evidence in support of the assumption of instrument exogeneity at the 10 percent level. The p-values associated with the Hausman coefficient signify that there are statistically significant differences between the OLS and IV parameter estimates for all the drinking measures.

Overall, in the high school sample, there is a strong indication that drinking, possibly by raising the opportunity cost of high school education, impairing cognitive functioning, etc., reduces enrollment in high school. And, considering the additional resources the student devotes toward drinking if the student binge drinks or is abusive/ dependent on alcohol, there is compelling evidence that the probability of high school enrollment is largely and negatively impacted.

INSTRUMENT ROBUSTNESS AND THE PROBABILITY OF ENROLLMENT (HIGH SCHOOL AGE)

To determine if there is any sensitivity in the main results attributable to changes in the instrument set, regressions are performed with varying pairs of instruments with results presented in table 6. The instrument that is omitted from the IV combination is utilized as an explanatory variable and its coefficient and standard error is reported.

Table 6. IV estimates of drinking on the probability of enrollment using IV pairs (high school age) (n=19,022)

Alcohol variables	religion important and alcohol risk	religious decisions and alcohol risk	religion important and religious decisions
number of days drank-past year	-0.001*	-0.001*	-0.002*
Marginal Effect Standard Error	(0.0003)	(0.0003)	(0.0004)
P-value of overidentification test	0.942	0.828	0.931
Hausman statistic (p-value)	-3.958 (0.000)	-4.759 (0.000)	-3.360 (0.000)
Coefficient (Standard Error) of omitted IV	0.002 (0.005)	-0.0002 (0.004)	-0.005 (0.012)
number of drinks in past month	-0.003*	-0.003*	-0.005*
Marginal Effect Standard Error	(0.0007)	(0.0006)	(0.0016)
P-value of overidentification test	0.992	0.429	0.995
Hausman statistic (p-value)	-3.627 (0.000)	-4.128 (0.000)	-3.024 (0.000)
Coefficient (Standard Error) of omitted IV	0.006 (0.004)	0.004 (0.004)	-0.025 (0.020)
binge drinking	-0.220*	-0.239*	-0.240*
Marginal Effect Standard Error	(0.051)	(0.047)	(0.067)
P-value of overidentification test	0.702	0.739	0.662

Table 6. IV estimates of drinking on the probability of enrollment using IV pairs (high school age) (n=19,022)			
Alcohol variables	**religion important and alcohol risk**	**religious decisions and alcohol risk**	**religion important and religious decisions**
Hausman statistic (p-value)	-4.354 (0.000)	-5.197 (0.000)	-3.577 (0.000)
Coefficient (Standard Error) of omitted IV	0.002 (0.005)	-0.002 (0.005)	-0.002 (0.011)
abuse/ dependence on alcohol	-0.323*	-0.341*	-0.333*
Marginal Effect Standard Error	(0.078)	(0.069)	(0.095)
P-value of overidentification test	0.834	0.906	0.826
Hausman statistic (p-value)	-4.238 (0.000)	-5.092 (0.000)	-3.602 (0.000)
Coefficient (Standard Error) of omitted IV	0.001 (0.005)	-0.002 (0.005)	-0.001 (0.011)
*Statistically significant at 1%			

For all drinking variables, the effect on enrollment using IV pairs is remarkably similar to those in the main regression where all three instruments are employed. For all drinking variables the overidentification test results support exogeneity for all IV pairs. Hausman tests indicate there are statistically significant differences between IV and OLS estimates in all specifications and the additional instrument not used to identify drinking is never significant in the enrollment equation.

THE EFFECTS OF DRINKING ON THE PROBABILITY OF SCHOOL ENROLLMENT (COLLEGE AGE)

As shown in table 7, drinking has significant, negative effects on the probability of being enrolled for the college age group. For each daily increase in past year drinking, the probability of being enrolled is subsequently lowered by 0.001. For each additional drink increase in the past month, the probability of enrollment is also lowered by 0.002. If, for instance, the respondent reports drinking 52 days in the previous year, the likelihood of enrollment is diminished by approximately 0.052 points compared to not drinking at all. If the student reports consuming 30 drinks in the previous month, the probability of enrollment decreases by 0.06 points.

Table 7. IV estimates of drinking on the probability of enrollment (college age) All three instruments (n=20,666)		
Alcohol variables	**IV**	**OLS**
number of days drank-past year	-0.001*	-0.0001*
Marginal Effect Standard Error	(0.0002)	(0.0000)
P-value of overidentification test	0.162	
Hausman statistic (p-value)	-5.043 (0.000)	
number of drinks in past month	-0.002*	-0.0002*
Marginal Effect Standard Error	(0.0004)	(0.0001)
P-value of overidentification test	0.082	
Hausman statistic (p-value)	-4.528 (0.000)	
binge drinking	-0.191*	-0.0112*
Marginal Effect Standard Error	(0.0359)	(0.0070)
P-value of overidentification test	0.263	
Hausman statistic (p-value)	-5.963 (0.000)	
abuse/ dependence on alcohol	-0.376*	0.0127*
Marginal Effect Standard Error	(0.0756)	(0.0080)
P-value of overidentification test	0.225	
Hausman statistic (p-value)	-5.258 (0.000)	
*Statistically significant at 1%		

Binge drinking and abuse/ dependence on alcohol further reduce the probability of enrollment by 0.19 points. For students who have engaged in binge drinking, the probability of school enrollment declines by approximately 43 percent compared to not binging. For those classified as abusive/ dependent with respect to alcohol, the probability of enrollment decreases by 0.37 points. Categorization as abusive/ dependent reduces the probability of school enrollment by 83 percent.

For number of days drinking occurred in the past year, binging and abuse/ dependence on alcohol, the overidentification tests have associated p-values that afford strong evidence in support of the assumption of instrument exogeneity at the 10 percent level. Even for the past month drinking variable, instrument exogeneity is not rejected at the 5 percent level. The p-values associated with the Hausman

coefficient signify that OLS and IV estimates statistically differ for all the drinking measures.

The estimated effects for binge drinking and abuse/ dependence are quite large, possibly indicating that for college age individuals, resources (monetary and otherwise) spent on drinking undercut the probability of post high school education, especially considering that there are greater costs (especially monetary) associated with obtaining education at that age. In addition, if the college age person has a history of drinking, especially at abuse and dependence levels, pre-college academic achievement might have been much lower thus precluding post high school enrollment in colleges, universities and other institutions.

INSTRUMENT ROBUSTNESS AND THE PROBABILITY OF SCHOOL ENROLLMENT (COLLEGE AGE)

To determine if there is any sensitivity in the main results attributable to changes in the instrument set, regressions are performed with varying pairs of instruments with results presented in table 8. Again, the instrument that is omitted from the IV combination is utilized as an explanatory variable and its coefficient and standard error is reported.

Table 8. IV estimates of drinking on the probability of enrollment using IV pairs (college age) (n=20,666)

Alcohol variables	**religious decisions and alcohol risk**	**religious decisions and marijuana risk**	**alcohol risk and marijuana risk**
number of days drank-past year	-0.001*	-0.001*	-0.001*
Marginal Effect Standard Error	(0.0002)	(0.0003)	(0.0002)
P-value of overidentification test	0.456	0.215	0.353
Hausman statistic (p-value)	-5.211 (0.000)	-3.081 (0.000)	-3.574 (0.000)
Coefficient (Standard Error) of omitted IV	0.001 (0.001)	-0.013 (0.018)	-0.001 (0.007)
number of drinks in past month	-0.002*	-0.004*	-0.002*
Marginal Effect Standard Error	(0.0004)	(0.0010)	(0.0005)
P-value of overidentification test	0.177	0.213	0.447
Hausman statistic (p-value)	-4.627 (0.000)	-2.865 (0.000)	-3.448 (0.000)
Coefficient (Standard Error) of	0.001 (0.001)	0.030 (0.025)	-0.003 (0.007)

Table 8. IV estimates of drinking on the probability of enrollment using IV pairs (college age) (n=20,666)			
Alcohol variables	**religious decisions and alcohol risk**	**religious decisions and marijuana risk**	**alcohol risk and marijuana risk**
omitted IV			
binge drinking	-0.202*	-0.213*	-0.165*
Marginal Effect Standard Error	(0.036)	(0.064)	(0.043)
P-value of overidentification test	0.718	0.289	0.350
Hausman statistic (p-value)	-6.102 (0.000)	-3.605 (0.000)	-4.287 (0.000)
Coefficient (Srd Error) of omitted IV	0.001 (0.001)	-0.006 (0.016)	-0.002 (0.007)
abuse/ dependence on alcohol	-0.396*	-0.458*	-0.320*
Marginal Effect Standard Error	(0.078)	(0.148)	(0.086)
P-value of overidentification test	0.550	0.295	0.401
Hausman statistic (p-value)	-5.357 (0.000)	-3.216 (0.000)	-3.911 (0.000)
Coefficient (Std Error) of omitted IV	0.001 (0.001)	-0.012 (0.020)	-0.002 (0.007)
*Statistically significant at 1%			

For all drinking variables, the effect on enrollment is remarkably similar to those in the main regression. For all drinking variables the overidentification test results support the exogeneity hypothesis for all IV pairs. Hausman tests indicate there are statistically significant differences between IV and OLS estimates in all specifications and the additional instrument not used to identify drinking is never significant in the enrollment equation.

Overall, the robustness evaluation for both samples offers strong evidence to support the hypothesis that instruments are exogeneous. Throughout the analyses, OLS parameter estimates consistently underestimate the magnitude of the negative effects in the main specification for enrollment. This could be ascribed to the prospect that higher ability (i.e. higher achieving) students perform better academically even when they drink. And these higher achievers are more likely to be enrolled in school. In addition, higher income students (who spend more on alcohol and therefore drink more) also command more resources that can be channeled toward education, such as test preparation for the SAT, and simply have more money to pay for college, and, once in college, funds to pay for tutoring services, etc. This in turn could serve to keep enrollment elevated.

CONCLUDING REMARKS

This paper contributes to the literature by examining the effects of youth drinking on the probability of school enrollment while accounting for unobserved endogeneity. The literature has established a negative link between drinking and educational variables, but many of these studies do not account for the possibility that the negative correlation between these factors may be the result of unobserved variables that cause simultaneous increases in drinking and reductions in educational variables. And, for studies that have incorporated unobserved endogeneity, instrumental variable procedures have been subject to criticism.

This study finds strong evidence that the probability of school enrollment is lowered when students use alcohol more frequently and intensely. Binge drinking and abuse of alcohol have the most detrimental impact on enrollment. Throughout the analysis, overidentification tests generally confirm instrument exogeneity and thus show that adolescent alcohol consumption should be treated as endogenous. OLS regressions consistently underestimate the effects of alcohol use on enrollment. Although there is no direct analysis of the effectiveness of laws and other programs designed to curtail youth drinking, the conclusions in this paper support the premise that reducing adolescent alcohol use enhances human capital accumulation. Minimum legal drinking ages, high school anti-drug programs and other policies aimed at lowering youth drinking may well be justified on human capital grounds. Although the instrumental variables prove to be very effective and useful, further research should include continued exploration for reliable instruments to ensure that the relationship between drinking and academic outcomes is properly identified. A further examination of the effectiveness of public policies that purport to reduce youth drinking would also prove valuable.

REFERENCES

Auld, M. (2005). Smoking, Drinking, and Income. *Journal of Human Resources, 15*(2), 505-518.

Bryant, A., Schulenberg, J., O'Malley, P., Bachman, J., & L. Johnston (2003). How academic achievement, attitudes, and behaviors relate to the course of substance use during adolescence: A 6-year, multiwave longitudinal study. *Journal of Research on Adolescence*, 13(3), 361–397.

Chatterji, P. and J. DeSimone (2005). Adolescent drinking and high school dropout. Working paper 11337, National Bureau of Economic Research.

Cook, P. J., & M. J. Moore (1993). Drinking and schooling. *Journal of Health Economics*, 12(4), 411–429.

Dee, T. S., & W. N. Evans (2003). Teen drinking and educational attainment: Evidence from two-sample instrumental variables estimates. *Journal of Labor Economics*, 21(1), 178–209.

Evans, W., Oates, W., & R. Schwab (1992). Measuring peer group effects: A study of teenage behavior. *Journal of Political Economy*, 100(51), 966–991.

Grant, B. F., Harford, T. C., & M. B. Grigson (1998). Stability of alcohol consumption among youth: a national longitudinal survey. *Journal of Studies on Alcohol*, 49(3), 253–260.

Harrison, L., & A. Hughes (1997). The validity of self-reported drug use: Improving the accuracy of survey estimates. *NIDA Research Monograph*, 167, 1–16.

Hausman, J. (1978). Specification tests in econometrics. *Econometrica*, 46(6), 1251-1271.

Hoyt, G., & F. Chaloupka (1994). Effect of survey conditions on self-reported substance use. *Contemporary Economic Policy*, 12(3), 109–121.

Koch, S. F., & D. C. Ribar (2001). A siblings analysis of the effects of alcohol consumption onset on educational attainment. *Contemporary Economic Policy*, 19(2), 162–174.

Midanik, L. (1998). Validity of self-reported alcohol use: a literature review and assessment. *British Journal of Addiction*, 83, 1019–1029.

Mullahy, J., & J. L. Sindelar (1994). Alcoholism and income: The role of indirect effects. *The Milbank Quarterly*, 72(2), 359–375.

Oreopoulos, P. (2006). Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws Really Matter. *American Economic Review*, 96(1), 152–175.

Reinisch, E. J., Bell, R. M., & P. Ellickson (1991). *How accurate are adolescent reports of drug use?* Santa Monica, CA: Rand Corporation.

Staiger, D., & J. H. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65(3), 557–586.

Williams, J., Powell, L. M., & H. Wechsler (2003). Does alcohol consumption reduce human capital accumulation? Evidence from the college alcohol study. *Applied Economics*, 35(10) 1227–1239.

Wolaver, A M. (2002). Effects of heavy drinking in college on study effort, grade point average, and major choice. *Contemporary Economic Policy*, 20(4), 415–428.

Wooldridge, J. M. (2003). *Introductory econometrics: a modern approach* (second edition). South-Western College Publishing.

Yamada, T., Kendix, M., & T. Yamada (1996). The impact of alcohol consumption and marijuana use on high school graduation. *Health Economics*, 5(1), 77–92.

APPENDIX

Appendix 1. All IV estimates on the probability of enrollment for binge drinking (high school age) (n=19,022)	
Explanatory variables	**IV coefficient (Marginal Effect SE)**
Binge drinking	-0.229 (0.040)
Female	-0.005 (0.003)
Race (African American)	-0.003 (0.006)
Race (Native American)	-0.026 (0.017)
Race (Asian)	0.028 (0.007)
Race (non-white Hispanic)	-0.034 (0.005)
Age of student (16 years old)	-0.034 (0.005)
Age of student (17 years old)	-0.124 (0.007)
Age of student (18 years old)	-0.255 (0.009)
Last grade completed (9th grade)	0.001 (0.005)
Last grade completed (10th grade)	0.044 (0.007)
Last grade completed (11th grade)	0.141 (0.008)
Ever been arrested	-0.031 (0.010)
Number in family	-0.007 (0.002)
Number in family (>5)	-0.058 (0.015)
Family income ($10,000-$19,999)	-0.045 (0.011)
Family income ($20,000-$29,999)	-0.017 (0.109)
Family income ($30,000-$39,999)	-0.005 (0.010)
Family income ($40,000-$49,999)	0.011 (0.010)
Family income ($50,000-$74,999)	0.024 (0.009)
Family income ($75,000 or more)	0.032 (0.009)
MSA segment with 1+ million persons	-0.003 (0.006)
MSA segment of less than 1 million	-0.007 (0.006)
Year 2006 indicator	-0.027 (0.006)

Appendix 2. All IV estimates on the probability of enrollment for binge drinking (college sample) (n=20,666)	
Explanatory variables	**IV coefficient (Marginal Effect SE)**
Binge drinking	-0.191 (0.035)
Female	-0.027 (0.007)
Race (African American)	-0.009 (0.011)
Race (Native American)	-0.026 (0.022)
Race (Asian)	0.111 (0.016)
Race (non-white Hispanic)	-0.068 (0.008)
Age of student (19 years old)	-0.271 (0.007)
Age of student (20 years old)	-0.434 (0.010)
Age of student (21 years old)	-0.503 (0.011)
Age of student (22-23 years old)	-0.599 (0.010)
Age of student (24-25 years old)	-0.690 (0.009)
Last grade completed (Freshman)	0.350 (0.008)
Last grade completed (Sophomore/ Junior)	0.512 (0.008)
Ever been arrested	-0.030 (0.010)
Number in family	-0.012 (0.003)
Number in family (>5)	-0.103 (0.014)
Family income ($10,000-$19,999)	-0.115 (0.010)
Family income ($20,000-$29,999)	-0.133 (0.010)
Family income ($30,000-$39,999)	-0.122 (0.010)
Family income ($40,000-$49,999)	0.125 (0.011)
Family income ($50,000-$74,999)	0.086 (0.010)
Family income ($75,000 or more)	0.027 (0.010)
MSA segment with 1+ million persons	0.082 (0.011)
MSA segment of less than 1 million	0.060 (0.010)
Year 2006 indicator	-0.056 (0.010)

EVALUATING SCIENTIFIC JOURNALS: PROPERTIES, LIMITS AND CONDITIONS OF EFFECTIVENESS OF CLASSIFICATION METHODOLOGIES

Alessandro Carretta, University of Rome "Tor Vergata"
Gianluca Mattarocci, University of Rome "Tor Vergata"

ABSTRACT

During last years a large debate was developed among academics about research evaluation procedure. Approaches proposed in literature considers quantitative and/or qualitative aspects of each journal but each methodology has limits and characteristics high heterogeneous. The meaning and implication of results achieved with a ranking procedure is strictly influenced by the approach selected for the journal evaluation.

The paper presents a literature review of the main qualitative and quantitative approaches proposed for journal ranking focusing the attention on the main differences of approaches. The study is completed with an empirical analysis on the database Thompson Scientific, one of the main provider of quantitative rankings. The analysis considers the characteristics of quantitative rankings proposed, look at the qualitative characteristics of best and worst ranked journals and compares results obtained with those achieved by an international qualitative survey (Harzing database). Results obtained demonstrate the low degree of coherence of ranking based on different approaches and point out some risks related to the use of these approaches to evaluate research.

INTRODUCTION

The evaluation of the scientific production of an individual or an institution has always been an issue subject to a great deal of debate (Liner, 2002), and, in recent times, the lack of resources for research gives new impulse on the studies of the usefulness of these approaches (Addis et al., 2002).

First approaches used to evaluate the quality of the research make a survey among academics and/or readers in order to evaluate the reputation of the journal but all rankings based on these approaches are highly influenced by the criteria adopted in the selection of interviewers (Kostoff et al. 2001). On the second half of the Nineties, new approaches based on more objective data are proposed to substitute these subjective measures (Garfield, 1952) and during the next years a growing number of criteria available are defined for the journal ranking (OECD 1997).

This paper is meant to contribute to the debate on the reliability of the indicators used in classifying scientific journals, proposing an empirical control meant to gauge the stability over time and the consistency of the results obtained with the different methodologies, in addition to establishing a "classification capacity" of the various indicators. Results obtained demonstrate the lack of coherence among the ranking based on different bibliometric indicators, the high variability of results over time and the failure to define ranking that rewards some qualitative journal characteristics.

The paper is organized as follows: section 2 presents a literature review of qualitative criteria (section 2.1) and bibliometric approaches (section 2.2), section 3 attains the empirical analysis presenting the sample analysed (section 3.1), looking at the consistency and persistency of quantitative rankings (section 3.2), at the characteristics of best and worst journals (section 3.3) and at the coherence with qualitative rankings (sections 3.4). The last section (section 4) presents some brief conclusions and implications of results achieved.

LITERATURE REVIEW

Qualitative Evaluation of Journals

Qualitative evaluations assess the general opinion of the academic community with regard to the quality and the scientific value of the articles published in a given journal (Benjamin and Brenner, 1974).

The qualitative approach normally entails the distribution of questionnaires, in order to obtain a ranking for a group of journals held to be similar in terms of the topics addressed and/or the readership target (Webb and Albert, 1995).

In order to be able to use the data collected during the individual surveys not only for one-off, non-recurring, evaluations but also for system-wide analyses of research, databases holding the results of the surveys carried out by different

interviewers have been created[1]. Though the use of different sources of qualitative research does not guarantee that the results will be objective, the higher the number of surveys, the lower the probability that the results will be distorted by the specific interests of interviewed (Barman et al., 2001).

The first result that emerges from these surveys is the fact that the perceived value of English-language journals is generally higher respect to other publications for the high diffusion of this language in world countries (Garfield and Welljams-Dorof, 1990).

Furthermore, authors appear to prefer, with all other conditions being equal, journals that are not particularly specialised. The leaning towards publishing one's article in journals that address a wide variety of topics and that are not concerned only with certain specific themes, makes possible greater visibility with the scientific community because the number of potential purchasers of the journal is higher (Brauninger and Haucap, 2002).

The quality of the editorial staff, and especially the reputation of the editor, constitute another qualifying aspect considered in the journal selection process. The presence of a qualified editorial staff is viewed as a guarantee of the quality of the journal and of the scientific significance of the articles published therein (Smith, 2004).

Qualitative analyses of periodicals demonstrate the importance given by the academic community on the mechanism used in selecting the articles published. The factors taken into consideration are the number of articles sent to the journal, the acceptance rate, the average time from the date of acceptance and actual publication and the type of the referee process (if it is scheduled).

The presence and the characteristics of a referee process influence the evaluation of the journal, because articles published in refereed journals are selected through a procedure, structured to varying degrees, which involves outside parties, and not merely the author and the editor, meaning the so-called referees (Surinach et al., 2002). If the identity of the author of the article is not communicated to the referee (blind referee), then the process, as a rule, should reinforced the objectivity of the selection of the articles. Regardless of the specific features of the process for selecting the articles, the use of a refereeing system can be considered an indicator of the reliability of a journal, and empirical analysis proposed in the literature have demonstrated noteworthy differences in the quality of the service offered, depending on the experience, the age and the academic backgrounds of the referees (Nylenna et al., 1994).

Journals can choose between two blind-referee options: the single blind referee and the double blind referee. Under the single blind referee procedure the referee does not know the identity of the author, though the author, when results of the refereeing process are notified, will know the identity of the those who have evaluated his or her article. With a double blind referee, not only every possible reference to the identity of the author is removed from the article submitted to the referee, but the author is informed of the final judgment of the refereeing process without receiving any data about the referee's identity. Empirical analysis have shown that, on the average, the presence of a double blind referee is related to an higher frequency of citations for a given article (Laband and Piette, 1994a).

Bibliometric Indicators

The quantitative evaluation of a journal is released constructing measures that regards its circulation, measured from various perspectives (Beattie and Goodacre, 2004). The majority of the quantitative approaches are based on the assumption that the bibliographies cited in scientific articles constitute a key instrument for assessing the quality of journals, and that an analysis of the articles cited makes it possible to identify the highest quality articles in a given discipline (Wang and White, 1999). The idea behind these approaches is that, in terms of the citations included in their articles, authors tend to favour articles that analyse topics of particular relevance, propose innovative approaches, present useful points of inspiration for research or lay the groundwork for a certain discipline (Small, 2004). Seen in this perspective, attention is focussed on studying the references of published articles, with the aim of evaluating the impact on the scientific community of the publication of a particular article (Kostoff, 2002).

One of the first measures proposed was *total citations*, an indicator calculated as the sum total of the citations of the articles published in a given journal within a sample group of journals during a established time horizon ($\sum_j\sum_i$citations(i,j)). The formula is:

$$TC_j = (\sum_j\sum_i \text{citations}(i,j)) \qquad (1)$$

Its limit is the influence of the time framework, plus the greater the difference in the number of articles published by the individual journals during the reference period, the lower the reliability of the ranking (Garfield, 1971). Those who support this approach have proposed updates of the measure in order to consider the

natural and progressive growth over time of the number of citations of a given journal and the practice of *self citation* (Linton and Narongsak, 2004).

Article effectiveness is calculated as the ratio between the number of citations of articles published in a given journal ($\sum_j\sum_i$citations(i,j)) and the average number of articles published during the period under consideration by the k^{th} journal ($MeanPub_k$) (Arnold, et al., 2003). The formula is:

$$AE_k = (\sum_j\sum_i citations(i,j))/MeanPub_k \qquad (2)$$

The strength of this indicator is that it defines a ranking which takes into account the rate of production of scientific articles during the period observed, in addition to providing a useful instrument when a noteworthy change in the volume of scientific articles has occurred during the time period considered.

Impact efficiency represents the number of citations of articles published in a given journal ($\sum_j\sum_i$citations(i,j)) for each 10,000 words published in the journals taken into consideration ($\sum_i words_i$). The formula is:

$$IE_k = (\sum_j\sum_i citations(i,j))/(\sum_i words_i) \qquad (3)$$

The decision to normalise the number of citations on the basis of the number of words found in the journal makes it possible to obtain measurements that can be used to establish comparisons between journals of different sizes, though the approach is highly influenced by the style adopted by the authors in writing the articles, as well as by the topic addressed, which can call for a greater or lesser frequency of citations (Chan et al., 2004).

The most widely used indicator is the *impact factor*, an index that represents the ratio between the number of citations in a journal ($\sum_j\sum_i$citations(i,j)) and the total number of articles published over the last two years by different journals ($\sum TotArt_i$) (Garfield, 1955). The formula is:

$$IF_k = (\sum_j\sum_i citations(i,j))/ (\sum_i TotArt_i) \qquad (4)$$

The indicators presented do not make it possible to illustrate factors of noteworthy relevance, such as the immediate impact of the article published and the persistence of the citations over time. The supporting instruments proposed for analysing these factors are the *immediacy index* and the *cited half life*.

The *immediacy index* is calculated by taking the number of citations of a journal ($\sum_j\sum_i$citations(i,j)) and considers the ratio to the total number of articles published during the year of publication by the journal ($TotArt_k$), with the resulting index representing an estimate of the immediate reaction of the scientific community following the appearance of the article (Harter, 1998). Its formula is:

$$II_k = (\sum_j\sum_i \text{citations}(i,j))/(TotArt_k) \qquad (5)$$

A high index points to the most valuable articles, meaning those which are immediately considered to be relevant by authors who write on a given topic (Ahmed et al., 2004).

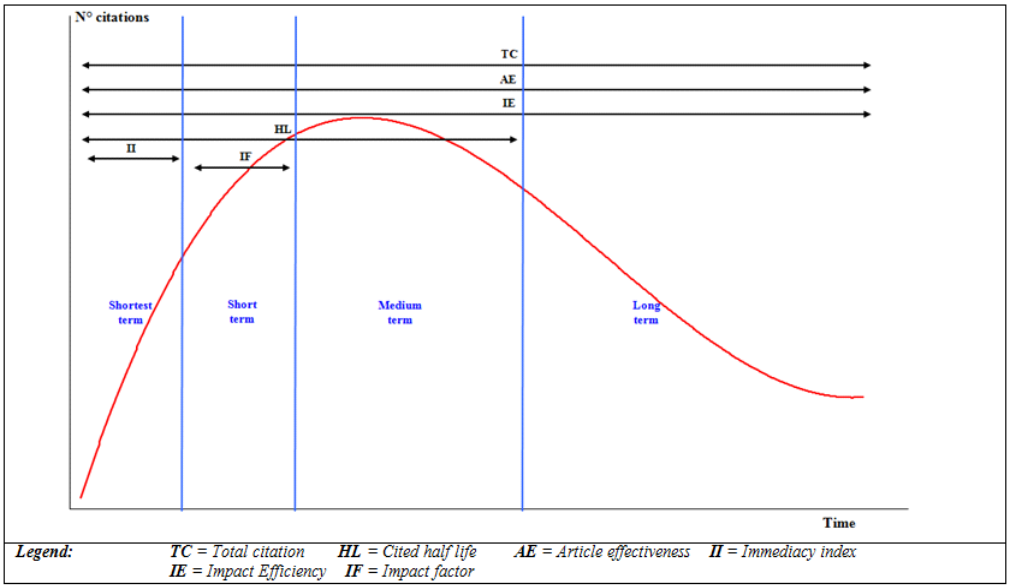
The *cited half life* represents the number of years needed for the number of citations of an article in a journal to decrease to a level that is half that of the maximum number of citations registered (Kademani and Kumar, 2002). Assessments of journals that fail to take into consideration the persistence of citations could lead to results that are not reliable, though the final result should always be examined with care: in certain cases, a high index value can simply mean that there has been a gradual deterioration in the average level of the publications on the subject, so that the citations always return to the consolidated sources in the literature (Diamond, 1989).

Based on the time frameworks normally used for application of the individual indexes, a relationship can be established between the type of index used and the curve estimating the general performance, in terms of citations, of a journal and/or article (Figure 1).

All indicators constructed through an analysis of bibliographies are affected by problems of self-citation and citations identity, which lead researchers to cite certain articles less because of their contents than on account of their authors (Fang and Rosseau, 2001).

The term self-citation refers to the natural tendency of authors to cite their own articles and/or the articles of colleagues who are part of the same sector of disciplines (Glanzel et al., 2004). The fact that such a practice is impossible to eliminate does not justify ignoring the impact it can have on estimates; in recent years there has also been a noteworthy increase in the number of articles by more than one author (Hudson, 1996), adding to the potential effect of self-citation on the evaluation of journals (Axarloglou and Theoharakis, 2003).

Figure 1: Citation results and the time framework of reference for the main quantitative indicators



Source: the authors' reformulation of the "generalized citation curve" proposed by Amin and Mabe (2000)

In writing articles, the tendency of authors, over time, is to consolidate their reference bibliography for a given topic, with the result that the citations of articles by a given author on a given subject shall not show noteworthy changes over time (citations identity). Analyses of journals that include authors who have published a number of articles on a given subject will thus be influenced by this circumstance, with the result that the greater the number of articles published by an author on a given argument, the greater the importance placed by the methodology of quantitative analysis on the reference articles used by that author (White, 2001).

The styles followed in drawing up articles differ from country to country, and there are even noteworthy differences in the approaches taken by authors to the existing literature. The differences between countries can be reflected in the use of bibliographies that are longer or shorter, or that go into greater or lesser historical depth. It follows that analyses which examine journals published in different countries can be influenced by the editorial styles followed by the authors of a given country (Bordons et al., 2002).

The number of citations registered for a journal is also tied to the characteristics of the authors whose articles it has published, meaning that, if the quality of the articles by those authors has been especially high, then the articles published in that journal will have greater visibility. It follows that the quantitative result is distorted by the importance given to results achieved in the past by subjects who shall not necessarily be presenting new contributions in the journal. Empirical analyses have shown that classifications drawn up with these indicators are relatively stable over time, making it difficult for new journals to enter the leading positions in a relatively short period of time (Laband and Piette, 1994b).

The results obtained using the quantitative approach are influenced by the time horizon considered, and an erroneous choice of valuator can render the estimate pointless for determining the current importance of the journal. Data collected at brief intervals exclude from the analysis citations in journals characterised by lengthy processes of refereeing, though such publications, thanks precisely to such procedures, could offer articles of greater scientific worth (Garfield, 2000). With the time framework defined in terms of the average article publication times, the approaches penalises the more innovative articles: an article that differs too greatly from the existing literature may not be accepted in the short run by the academic community, meaning that journals which choose to publish such articles would be penalised by the limits of the approach used (Hogson and Rothman, 1999).

Finally, in evaluating the results obtained, consideration must be given to a limitation that cannot be eliminated from the methodology, being tied to its inability to analyse the nature of the citation (Posner, 1999). In fact, an author can decide to cite another article as support of his or her own thesis (positive citation) or in order to criticise a particular approach or the results obtained (negative citation): using quantitative approaches, there is no way of distinguishing between the two types of references (MacRoberts and MacRoberts, 1989). The assessments obtained through these approaches must, therefore, take into consideration the possibility that negative citations were included in the calculation of the indicators, with the result that, the greater the number of negative citations, the less reliable the rankings obtained.

EMPIRICAL ANALYSIS

The Sample

The sample is constructed starting from the database of the journals constructed by one of the main data provider specialised in quantitative bibliometric evaluation, Thomson Scientific. The database consists of more than 1800 journals regarding 54 subjects, all falling under the category of the social sciences (Table 1)[2].

Table 1: The sample

Subject	Relevance in the sample	Subject	Relevance in the sample
Anthropology	1.96%	History of social sciences	0.21%
Applied linguistic	1.27%	Industrial relations & labor	0.48%
Applied Psychology	1.54%	Information science & library science	2.60%
Area studies	2.07%	Interdisciplinary social sciences	1.33%
Biological Psychology	0.37%	International relations	1.12%
Biomedical social sciences	0.05%	Law	3.45%
Business	1.38%	Management	1.49%
Business Finance	1.22%	Mathematical methods social sciences	0.21%
Clinical Psychology	2.66%	Mathematical psychology	0.05%
Communication	1.54%	Multi-disciplinary	30.38%
Criminology and penology	0.85%	Multi-disciplinary psychology	3.35%
Demography	0.42%	Nursing	1.81%
Developmental psychology	1.43%	Planning & development	0.69%
Economics	4.83%	Political science	2.50%
Education/educational research	3.66%	Psychiatry	2.71%
Education special	0.58%	Psychoanalysis psychology	0.53%
Educational psychology	1.12%	Public administration	0.74%
Enviromental studies	0.64%	Public environmental & education health	1.49%
Ergonomics	0.32%	Rehabilitation	0.85%
Ethics	0.42%	Social issues	0.64%

Table 1: The sample

Subject	Relevance in the sample	Subject	Relevance in the sample
Ethnic studies	0.27%	Social Psychology	1.75%
Experimental Psychology	2.02%	Social work	0.90%
Family studies	0.11%	Sociology	2.87%
Geography	1.06%	Substance abuse	0.74%
Gerontology	0.58%	Transportation	0.42%
Health policy & services	1.22%	Urban studies	0.42%
History	0.69%	Women's studies	0.64%
History & philosophy of science	1.33%		

* The category multi-disciplinary considers all journals classified by Thompson Scientific in two or more subjects

Source: the authors' processing of Thompson Scientific data

Data collected from Thompson Financial regard all the main quantitative indicators calculated for each journal during the time period 2000-2006[3]. These data are integrated with qualitative data collected from journal websites that, on the basis on a review of the literature previously presented, are relevant to distinguish among different academic journals and with results of qualitative surveys collected in the Harzing database[4].

Persistence and Consistency of the Indicators in Classifying Journals: an Empirical Control

The degree of consistency between the classifications is released by calculating the main quantitative indicators for the period considered and constructing the ranking for each one.

On the constructed ranking, journals are grouped in ten subclasses on the basis of percentiles distribution. The analysis studies the frequency with which the subclasses attributed on the basis of each quantitative indicator coincide with the subclasses determined on the basis of the other indicators (Table 2).

Table 2: Consistency between the classifications based on different approaches, year by year for the period 2000-2006 (frequency as a %)

		TC	IF	II	HL	AE
TC	Mean	100.00%	-	-	-	-
	Max					
	Min					
IF	Mean	39.20%	100.00%	-	-	-
	Max	44.62%				
	Min	28.59%				
II	Mean	28.28%	32.93%	100.00%	-	-
	Max	32.53%	37.48%			
	Min	23.04%	28.79%			
HL	Mean	30.92%	26.29%	28.23%	100.00%	-
	Max	38.26%	29.63%	39.60%		
	Min	24.26%	20.81%	23.01%		
AE	Mean	36.23%	59.46%	28.88%	25.05%	100.00%
	Max	38.26%	29.63%	39.60%	38.26%	
	Min	10.69%	10.69%	10.69%	11.65%	

Legend: ***TC*** *= Total Citations* ***IF*** *= Impact Factor* ***HL*** *= Half Life*
AE = Article effectiveness II = Immediacy Index

Source: the authors' processing of Thompson Scientific data

The empirical evidence points out a significantly low degree of coherence of the results obtained by using different indicators, seeing that the average ratio of correspondence for the classifications falls below 35% for the entire time interval considered.

The lack of correspondence between the classifications is extensive in all the disciplines considered[5], even though marked differences are more observable for certain disciplines because for certain discipline, on certain years, almost all the journals belonging to the category are affected[6].

For effective use of the results obtained from classifying journals, it is obviously best that the rankings proposed following application of the indicator

present a noteworthy level of persistence over time (Garfield, 1972), especially if the time intervals for collecting data are not too long.

The persistence over time of the different classifications was analysed by considering, for each journal, the ranking assigned for time period *t* by the various quantitative indicators and controlling whether or not there were differences, in terms of the positioning of the journals within the deciles, compared to the ranking assigned for time period *t+n*. The comparison was established between the situation for the previous year and that for the subsequent year, as well the situation at the start and the end of the overall period considered (Graph 1).

Graph 1: Persistence over time of the classifications established with the different approaches per year

Source: the authors' processing of Thompson Scientific data

The analysis of persistence for the entire period points to a scarce correspondence of the rankings in the different time intervals, especially when measures expressing the short-term impact of the publication of the articles in the journals are considered (the *impact factor* presents average levels of correspondence of less than 30%, while the *immediacy index* is equal to approximately 15%). Furthermore, a year-by-year comparison of the classifications points to noteworthy variability within the period, and especially in the case of certain sub-periods.

Quantitative Rankings and Qualitative Standards of Excellence

The formulation of quantitative rankings should make possible identification, above all else, of publications that guarantee published articles at high

standards of quality, meaning that they should represent the best sources for citations.

Looking at literature available one of the desired characteristics of a journal is represented by the international language and the Thompson scientific database is so characterized by an higher preponderance of English language journal (Table 3)

Table 3: Statistics of quantitative indicators by language for the period 2000-2006
(n° journal, mean, maximum and minimum ranking position)

Language	N° journals in the database	Percentile ranking on the basis of bibliometric indicator														
		TC			IF			II			AE			HL		
		Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min
Cezch	3	VIII	VIII	X	VIII	VIII	X	VIII	VIII	X	VIII	VIII	X	VIII	VIII	X
Dutch	1	X	X	X	IX	IX	IX	VII	VII	VII	IX	IX	IX	VII	VII	VII
English	1776	IV	I	X	IV	I	X	IV	I	X	IV	I	X	IV	I	X
French	8	VII	I	X	VII	I	X	VII	I	X	VI	I	X	VI	I	X
German	28	VI	I	X	VI	I	X	VI	I	X	VI	I	X	VI	I	X
Greek	1	VIII	VIII	VIII	VIII	VIII	VIII	VII	VII	VII	IV	IV	IV	VI	VI	VI
Japanese	3	VIII	VIII	X	VIII	VIII	X	VII	VII	X	VII	VII	X	VII	VII	X
Multi lang-uage	76	V	I	X	V	I	X	V	I	X	V	I	X	V	I	X
Portu-guese	1	VII	VII	VII	VIII	VIII	VIII	VIII	VIII	VIII	VIII	VIII	VIII	VIII	VIII	VIII
Spanish	6	VI	I	X	VII	II	X	VI	II	X	VII	II	X	VII	III	X
Turkish	1	VIII	VIII	VIII	VIII	VIII	VIII	VIII	VIII	VIII	VIII	VIII	VIII	VII	VII	VII

Source: the authors' processing of Thompson Scientific data

Looking at the relationship between ranking position and the language it is possible to assume that normally journals written in English are, in mean, ranked in the better position of the ranking (IV percentile) even there is an high variability of ranking position among these journal (the range of variation considers all the possible position on the ranking). Considering journals written in other languages the analysis point out that in mean they are ranked in the second half of the

percentile distribution but the variability among different journals written in the same language do not allow to exclude the possibility to find some outliers classified in the best position of the ranking.

Many studies have highlighted, in terms of profiles of quality, the significant role of the reputation of the editorial staff in determining the success of a journal, though no established set of criteria has yet been identified to evaluate this profile.

Based on the information available, the present assessment considers, from among the qualitative profiles subject to analysis, only the characteristics of the refereeing process.

An analysis of the refereeing process makes possible the identification of three categories: journals with a system of refereeing that is not declared, or that is characterised by a noteworthy degree of arbitrary discretion on the part of the editor with respect to the procedures of the refereeing (approximately 58% of the sample); journals that have a blind referee (less than 39% of the sample); journals that utilise a double blind referee (roughly 3% of the sample).

Using the same sample presented earlier, the study of the relation between the refereeing process and the ranking shows that the quantitative indicators are capable only in part of valorising in decisive fashion the type of refereeing utilised by the journal (Table 4).

The journals that utilise a *blind* or *double blind* refereeing system obtain, on the average, higher rankings than the journals without refereeing, or than those whose refereeing is not declared, in more than 80% of the cases examined. Furthermore, the indicators for such journals register a variability that tends to be greater than that recorded for refereed journals. As a result, it is possible that the latter may obtain better positions than journals which utilise an explicit process for the selection of contributions.

Comparing Quantitative and Qualitative Rankings

The scarce relevance of certain qualitative factors in rankings based on bibliometric indicators can result in noteworthy discrepancies between qualitative and quantitative rankings.

Table 4: Descriptive statistics of quantitative indicators by type of refereeing
(mean and St. Dev.)

	Type of referee	Statistics	Type of bibliometric indicator				
			TC	IF	II	HL	AE
Time period	Not declared	Mean	746.6042	0.8568	0.1843	6.0542	0.0683
		St. Dev.	2078.754	1.0353	0.3254	3.5908	0.1912
		Mean worst year	612.5	0.7854	0.1645	36.1385	3.2803
		Mean best year	2005.783	1.008	0.3087	45.0254	6.0197
	Blind referee	Mean	649.2492	0.7954	0.1682	5.619	0.1539
		St. Dev.	1854.8876	0.9946	0.3037	3.7488	0.4401
		Mean worst year	683.1382	0.8607	0.1755	30.8666	2.7223
		Mean best year	1426.189	0.946	0.2739	38.9327	6.4956
	Double blind referee	Mean	819.6205	0.9029	0.1964	6.3805	0.0042
		St. Dev.	2246.6539	1.0659	0.3417	3.4723	0.0046
		Mean worst year	1264.69	0.7616	0.1983	58.2914	2.1091
		Mean best year	2296.601	1.099	0.2449	64.541	7.2689

Legend
TC = Total Citations HL = Half Life,,,,,
IF = Impact Factor AE = Article effectiveness,,,,,
II = Immediacy Index,,,,,,

Source: the authors' processing of Thompson Scientific data

In the interests of selecting an extensive database of qualitative surveys on international journals, consideration was given to the Journal Quality List (JQL), a database that is updated quarterly and holds the results of scientific surveys published in recent years[7]. The journals placed in the ranking and utilised in the survey regard the following disciplines: Economics (Eco), distributed among two subject headings; Finance & Accounting (F&A); Management (Mgmt), distributed among 7 subject headings; Marketing (Mkt), distributed among 5 subject headings.

Studies presented in the literature have shown that the rankings assigned to individual journals by the different surveys are, on the average, consistent with one another, meaning that the indications provided by this dataset can be considered reliable (Mingers and Harzig, 2007).

To verify the consistency with the data sample available for the bibliometric indicators, it was decided to use the data collected in the twenty-fifth edition of the JQL, published on 1 February 2007, which holds data for the surveys carried out through 2006[8].

The analysis took into consideration the relevance of the journals examined in Thompson Financial with respect both to the all the journals considered in the survey and to the journals belonging to the individual classes or disciplines (See Table 5).

The results obtained demonstrate that, on the average, only slightly less than half of the journals considered in the qualitative surveys constitute journals included in the Thompson Financial database, even though the percentage weight of the journals in the database varies significantly from survey to survey.

A detailed analysis of the role of journals for which bibliometric indicators are available within the qualitative rankings considered shows that their percentage weight in the higher quality rankings is greater. It would appear, therefore, that the presence of citation indexes ensures greater notoriety for the journal, thus increasing the probability that those interviewed will place it in the top positions of the classification (Clark, 1957).

After considering the disciplines of the journals found on both the JQL and the Thompson Financial database, a higher frequency can be noted for journals belonging to the disciplines of Economics and Management, while the bibliometric indicators prove less relevant for the disciplines of Marketing and Finance, and Accounting. Underlying the difference in relevance for certain disciplines there would appear to be a discrepancy in the extent to which the qualitative/quantitative rankings can be interchanged.

CONCLUSIONS

Classifications of journals can be drawn up using a variety of procedures designed either to make direct registration of the opinion of the scientific community or to measure the level of readership and distribution of the journal, primarily by recording the citations found in the literature of reference. The use of quantitative bibliometric indicators is the most widely used approach, given that the procedure can be replicated and verified by third parties other than the promoter of the analysis, though it presents a number of limitations that have given rise to proposals for revision of the formulation of the indicators (Hartes and Nisonger, 1997) and do not allow to achieve uniform results for the different choices taken by the different

subjects involved (authors, editors, evaluators) (Judge et al. 2007; Cacciafesta, 2007) and for the specific characteristics of the subject.

Table 5 : The relation between the rankings included on the Journal Quality List and the Thompson Financial Database

Survey data			Percentage of journals registered also in Thompson Financial database							
	Total		Break-down by class				Break-down by subject			
Survey code JQL database*	Journals considered		Class A	Class B	Class C	Class D	Eco	F&A	Mgmt	Mkt
Not95	208	53.37%	18.27%	26.92%	8.17%	0.00%	12.98%	27.20%	6.62%	14.52%
US98	125	72.00%	16.80%	28.80%	23.20%	3.20%	6.25%	12.00%	16.18%	1.61%
NL99	136	58.82%	15.44%	20.59%	22.79%	0.00%	25.48%	23.20%	26.47%	51.61%
SMJ99	62	91.94%	22.58%	22.58%	19.35%	27.42%	8.65%	9.60%	10.29%	24.19%
Wiei01	483	57.97%	44.72%	12.42%	0.62%	0.21%	12.98%	27.20%	6.62%	14.52%
UQ03	345	46.67%	8.12%	17.68%	16.52%	4.35%	6.25%	12.00%	16.18%	1.61%
VHB03	393	52.42%	28.75%	15.27%	6.11%	2.29%	25.48%	23.20%	26.47%	51.61%
BJM04	401	58.35%	27.68%	21.95%	8.48%	0.25%	8.65%	9.60%	10.29%	24.19%
CNRS04	290	71.03%	23.45%	20.00%	17.59%	10.00%	12.98%	27.20%	6.62%	14.52%
ESS05	257	60.31%	48.25%	11.67%	0.39%	0.00%	6.25%	12.00%	16.18%	1.61%
HKB05	322	55.28%	13.04%	15.22%	20.19%	6.83%	25.48%	23.20%	26.47%	51.61%
Theo05	257	64.59%	5.06%	3.50%	10.51%	45.53%	8.65%	9.60%	10.29%	24.19%
Ast06	598	56.02%	37.12%	12.37%	6.19%	0.33%	12.98%	27.20%	6.62%	14.52%
Cra06	361	62.60%	17.73%	28.53%	13.02%	3.32%	6.25%	12.00%	16.18%	1.61%
EJL06	268	57.09%	26.12%	0.37%	30.60%	0.00%	25.48%	23.20%	26.47%	51.61%
HMB06	474	54.22%	25.32%	22.15%	5.70%	1.05%	8.65%	9.60%	10.29%	24.19%
FT06	34	94.12%	94.12%	0.00%	0.00%	0.00%	12.98%	27.20%	6.62%	14.52%

* For further details on the meaning of the symbols proposed, see table A1 in the appendix, which presents the main data from the different surveys

Source: the authors' processing of data from Thompson Scientific and the Journal Quality List

The empirical assessment performed in the paper highlights certain instances of discontinuity in the classifications, as well as a low level of consistency in terms of the results. This last observation is not entirely justifiable, given that the different indicators all use the same source of information (the bibliographic

citations of the articles) (Block and Gary, 2001), in addition to which there exists a general relation between the indicators and the life cycle of the journal and its citations, which should produce essentially consistent results, apart from a few individual differences tied to the specific characteristics of each indicator.

The combined analysis of the qualitative and quantitative profiles shows that the quantitative indicators lack an adequate capacity to reward journals that present objectively better qualitative features, such as a rigorous process of refereeing. As a consequence, there are noteworthy differences in the results produced by the qualitative and quantitative evaluations that do not allow to consider these approaches as substitutes and in order to understand the journal raking is necessary to consider limits and characteristics of the approach used for the classification.

Furthermore, an evaluation of academic production carried out by analysing the articles published in journals provides only a partial vision of research activities (Moore et al., 2002), one judged a priori to be representative of the sum total of the most innovative contributions for a given discipline (Ding et al, 2000; Carretta, 2006), even though it should be supplemented with the study of other profiles. Along these lines, it is worth noting the profile of academic works produced in different formats, pointing to the advisability of valorising published monographs as well (Gray et al., 1997), in order to avoid unjustifiably penalising (apart from an arbitrary ex ante selection) those subjects who prefer to publish monographs or volumes (Seglen, 1997) and excluding all the citations of journals made by such authors in their books (Johnes and Johnes, 1993). The next step of the analysis will be to define criteria, methodologies and databases that allow to evaluate also these type of contribution in order to define more complete journal ranking.

ENDNOTES

* This article is taken from a wider-ranging work published in Italian as a supplement in Banking & Finance Lab, no. 1, 2008, and also available on the site of the Association of Professors in Economics of Financial Intermediaries and Markets (www.adeimf.it). Reference should be made to the larger work for an all-encompassing overview of the issue in both theoretical and empirical terms. The work is a joint effort by the two authors. Alessandro Carretta wrote sections 1 and 4, while Gianluca Mattarocci wrote sections 2 and 3. Authors are grateful to Fabrizio Cacciafesta, Roberto Cafferata, Giacomo de Laurentis, Franco Fiordelisi, Mario Masini, Paolo Mottura, Luciano Munari, Claudio Porzio, Daniele

Previati, Francesco Ranalli, Paola Schwizer and Giulio Tagliavini for all suggestions given.

[1] See, for example, Harzing, A.W. (various years), *Journal Quality List,* www.harzing.com.

[2] For further details about journals included in each category, see the internet website of Thompson Scientific (http://scientific.thomson.com/).

[3] In the Thomson Scientific database are available data about all bibliometric indicators previously presented except for the impact efficiency. In the analysis this measure had to be excluded.

[4] For further details about the Harzing database see www.harzing.com.

[5] The sample group of journals was segmented in accordance with the 54 sub-classes proposed by Thompson Scientific.

[6] For further details see Table A1 in the appendix.

[7] Cfr. Harzing (various years), *Journal quality list*, www.harzing.com.

[8] For more detailed information on the surveys considered as part of the analysis, as well as on the characteristics of the individual surveys and the criteria followed to ensure that the number of classes of journals considered in the different rankings is uniform, see Table A2 in the appendix.

REFERENCES

Addis M., M. Costabile & K. Matzler (2002). La valutazione delle pubblicazioni italiane per gli studiosi di marketing. *Finanza, Marketing e Produzione,* 20 (3) 70-95

Ahmed T., B. Johnson, C. Oppenheim & C. Peck (2004). Highly cited old papers and the reasons why they continue to be cited, Part II. *Scientometrics,* 61 (2)147-156

Amin M. & M. Mabe (2000), Impact Factors: use and abuse. *Perspectives in Publishing,* 1(10) 1-5

Arnold T., A.W. Butler, T.F. Crack & A. Altintig (2003). Impact: what influences finance research?. *Journal of Business,* 72 (3) 343-361

Axarloglou K. & V. Theoharakis (2003). Diversity in economics: an analysis of journal quality perceptions. *Journal of the European Economic Association,* 1(6), 1402-1423

Barman S., M.D. Hanna & R.L. Laforge (2001). Perceived relevance and quality of POM journals: a decade later. *Journal of Operations Management,* 19(3), 367-385

Beattie V. & A. Goodacre (2004). Publishing patterns within the UK accounting and finance academic community. *British Accounting Review,* 36 (1) 7-44

Benjamin J.J. & V. Brenner (1974). Perceptions of journal quality. *Accounting Review*, 49(2) 360-362

Block S. & W. Gary (2001). The Impact Factor: time for change. Australian and New Zealand *Journal of Psychiatry*, 35 (5) 563-568

Bordons M., M.T. Fernandez & I. Gomez (2002). Advantages and limitations in the use of the impact factor measures for the assessment of research performance in peripheral countries. *Scientometrics,* 53 (2) 195-206

Brauninger M. & J. Haucap (2002), What economists think of their journals and how they use them: reputation and relevance of economic journals, University of the Federal Armed Forces discussion papers in economic policy n° 111, Hamburg

Cacciafesta F., (2007). Sulla valutazione (dei ricercatori e degli articoli di ricerca). *Notiziario della unione matematica italiana,* 7 (3) 31-33

Carretta A. (2006). Gli studi di economia degli intermediari finanziari: una prospettiva di analisi basata sulla letteratura citata. in L. MUNARI (eds), *Attualità e prospettive negli studi di economia dei mercati e degli intermediari finanziari* (pp. 1103-1109). Parma: Monte Università Parma Editore

Chan K.C., C.R. Chen & T.L. Steiner (2004). Who is publishing? An analysis of finance research productivity in the European region. *Journal of Business, Finance and Accounting,* 31(3-4), 401-437

Clark K.E. (1957), *America's Psychologists: a survey of growing profession,* Washington: American Psychological Association

Diamond A.M. (1989). The core journals of economics, *Current Contents,* 1 (1) 4-11

Ding Y., G.G. Chowdhury & S. Foo (2000). Journals as makers of intellectual space: journal co-citation analysis of information retrieval area, 1987-1997, *Scientometrics*, 47(1), 53-73

Fang Y. & R. Rosseau (2001), Lattices in citations networks: an investigation into the structure of citation graphs. *Scientometrics,* 50(2), 273-287

Garfield E. (1955). Citation indexes for science: a new dimension in documentation through association of ideas *Science,* 122(3159), 108-111

Garfield E. (1971). Citation indexing, historio-bibliography and the sociology of science. *Essays of an Information Scientist,* 1(1), 158-174

Garfield E. (1972). Citation analysis as a tool in journal evaluation. *Science*, 178(11), 471-479

Garfield E. (2000). Use of journal citation reports and journal performance indicators in measuring short and long term journal impact. *Croatian Medical Journal,* 41(4), 368-374

Garfield E. & A. Welljams-Dorof (1990), Language use in international research: a citation analysis. *Annals of the American Academy of Political and Social Science,* 511(1), 10-24

Glanzel W., B. Thijs & B. Schlemmer (2004). A bibliometric approach to the role of author self-citations in scientific communication. *Scientometrics,* 59(1), 63-77

Gray R.H., J. Guthrie & L.D. Parker (1997), Publish or perish: the social construction of quality in accounting and management research, *Fifth Interdisciplinary Perspectives on Accounting Conference* 4, 1111 - 1123

Hartes S.P. & T.E. Nisonger (1997). ISI's impact factor as misnomer: a proposed new measure to assess journal impact. *Journal of American Society for Information Science*, 48 (12) 1146-1148

Harzing A.W. (different years), *Journal Quality List,* Retrieved December, 2008, from http://www.harzig.com

Hogson G.M. & H. Rothman (1999). The editors and authors of economics journals: a case of institutional oligopoly. *Economic Journal,* 109 (453) 165-186

Hudson J. (1996). Trends in multi-authored papers in economics. *Journal of Economic Perspectives,* 10 (3) 153-158

Johnes G. & J. Johnes (1993). Measuring the research performance of UK economics departments: an application of data envelopment analysis. *Oxford Economic Papers*, 45 (2) 332-347

Judge T.A., D.M. Cable, A.E. Colbert & S.L. Rynes (2007). What causes a management article to be cited - article, author or journal?. *Academy of Management Journal,* 50(3) 491-506

Kademani B.S. & V. Kumar (2002), Citation index: an indispensable retrieval tool for research and evaluation, presented to the ICSSR Sponsored Training Workshop on Exploring Social Science Information in a Digital Environment, Mumbai, India

Kostoff R.N. (2002). Citation analysis of research performer quality. *Scientometrics*, 53(1) 49-71

Kostoff R.N., J.A. De Rio, J.A. Hurnenik, E.O. Garzia & A.M. Ramirez (2001). Citation mining: integrating text mining and bibliometrics for research user profiling. *Journal of the American Society for Information Science and Technology*, 52(13), 1148-1156.

Laband D.N. & M.J. Piette (1994a). A citation analysis of the impact of the blinded peer review. *Journal of the American Medical Association,* 272 (2)147-149

Laband D.N. & M.J. Piette (1994b). The relative impacts of economics journals: 1970-1990. *Journal of Economic Literature*, 32 (2) 640-666

Liner G.H. (2002). Core journals in economics. *Economic Inquiry*, 40 (1) 138-145

Linton J.D. & T.T. Narongsak (2004). Perspective: ranking the technology innovation management journals. *Journal of Product Innovation Management*, 21 (2) 123-139

Macroberts M.H. & B. Macroberts (1989). Problems of citation analysis: a critical review. *Journal of American Society for Information Science,* 40 (5) 342-349

Mingers J. & A.W. Harzing (2007). Ranking journals in business and management: a statistical analysis of the Harzing dataset, *European Journal of Information Science,* 16 (4) 303-316

Moore W.J., R.J. Newman, P.J. Sloane & J.D. Steely (2002), Productivity Effects of Research Assessment Exercises, Louisiana State University Departmental Working Papers n˚ 2002-15, Baton Rouge

Nylenna M., P. Riis & Y. Karlsson (1994). Multiple blinded reviews of the same two manuscripts. *Journal of the American Medical Association,* 272 (2) 149-151

OECD (1997), The evaluation of scientific research: selected experience, presented to the OECD Workshop on the Evaluation of Basic Research, Paris, France

Posner R.A. (1999), The theory and practice of citations analysis with special reference to law and economics, University of Chicago Law School, John M. Olin Law & Economics Working Paper No. 83 Chicago

Seglen P.O. (1997). Why the impact factor of journals should not be used for evaluating research. *British Medical Journal*, 314 (7079), 498-502

Small H. (2004). Why authors think their papers are highly cited. *Scientometrics,* 60 (3), 305-316

Smith S.D. (2004). Is an article in a top journal a top article. *Financial Management*, 33 (4)133-149

Surinach J., J.C. Duque, R. Ramos & V Royuela (2002), Authors', institutions' and countries' ranking in regional and urban science: an analysis for nine top international journals from 1991 to 2000, University of Barcelona working paper n˚ 92, Barcelona

Wang P. & M.D. White (1999). A cognitive model of document use during a research project, study ii. decision at the reading and citing stage. *Journal of American Society for Information Science*, 50(2) 98-114

Webb J.R. & J.D. Albert (1995). Evaluating the real estate journals: the mainstream finance perspective. *Journal of Real Estate Research,* 10 (2) 217-226

White H.D. (2001). Authors as citers over time. *Journal of the American Society for Information Science and Technology,* 52 (2) 87-108

APPENDIX

Table A1: Consistency between the classifications based on different approaches for subject (frequency as a %)

	2000	2001	2002	2003	2004	2005	2006		2000	2001	2002	2003	2004	2005	2006
Anthropology	24.59%	29.73%	27.30%	24.59%	31.08%	29.46%	12.43%	History of social sciences	10.00%	17.50%	20.00%	25.00%	37.50%	67.50%	60.00%
Applied linguistic	38.33%	56.67%	48.75%	47.08%	47.50%	38.75%	29.58%	Industrial relations & labor	22.22%	21.11%	18.89%	22.22%	24.44%	18.89%	10.00%
Applied psychology	35.33%	34.00%	29.00%	28.67%	29.67%	24.67%	12.67%	Information science & library science	22.08%	30.42%	30.42%	29.79%	26.88%	35.42%	22.71%
Area studies	26.41%	37.95%	39.74%	44.36%	40.00%	41.79%	33.08%	Interdisciplinary social sciences	25.20%	26.00%	26.80%	22.40%	20.00%	20.80%	12.80%
Biological psychology	42.86%	40.00%	37.14%	31.43%	41.43%	41.43%	25.71%	Intern-ational relations	14.29%	20.00%	26.19%	26.67%	25.71%	32.38%	23.33%
Biomedical social sciences	100.00%	100.00%	20.00%	30.00%	30.00%	40.00%	100.00%	Law	29.54%	32.31%	28.77%	29.85%	30.77%	28.77%	17.38%
Business	39.63%	39.26%	45.19%	42.59%	46.67%	32.59%	26.67%	Manage-ment	35.00%	40.71%	36.07%	33.21%	25.36%	22.86%	13.93%
Business finance	45.22%	54.35%	53.04%	47.39%	40.00%	29.13%	17.39%	Mathematical methods social sciences	35.00%	52.50%	42.50%	35.00%	7.50%	10.00%	30.00%
Clinical psychology	26.47%	33.53%	36.08%	34.90%	27.45%	28.04%	12.75%	Mathematical psychology	0.00%	10.00%	10.00%	10.00%	0.00%	30.00%	10.00%
Communication	26.21%	43.10%	42.41%	39.31%	45.86%	36.55%	31.03%	Multidisciplinary	24.34%	27.41%	25.71%	25.50%	24.97%	25.42%	14.67%
Criminology and penology	48.82%	47.06%	44.12%	44.71%	34.71%	36.47%	16.47%	Multidiscip-linary psychology	21.75%	26.03%	24.92%	29.21%	28.73%	30.32%	17.30%
Demography	16.25%	15.00%	22.50%	18.75%	33.75%	18.75%	8.75%	Nursing	28.33%	34.72%	39.17%	35.83%	37.22%	36.11%	22.78%
Developmental psychology	30.00%	29.64%	30.00%	23.93%	29.64%	24.29%	16.43%	Planning & development	54.62%	53.08%	36.15%	21.54%	20.77%	30.77%	11.54%
Economics	26.81%	32.75%	29.34%	28.68%	28.90%	29.45%	16.92%	Political science	25.96%	33.40%	32.34%	34.04%	31.28%	29.57%	18.51%
Education and educational research	35.14%	36.11%	35.00%	35.14%	34.58%	30.97%	19.58%	Psychiatry	32.55%	29.22%	32.75%	33.14%	35.29%	37.25%	25.10%
Education special	26.36%	42.73%	31.82%	20.00%	30.00%	49.09%	44.55%	Psychoanalysis psychology	28.00%	51.00%	42.00%	38.00%	45.00%	44.00%	28.00%

Table A1: Consistency between the classifications based on different approaches for subject (frequency as a %)

	2000	2001	2002	2003	2004	2005	2006		2000	2001	2002	2003	2004	2005	2006
Educational psychology	16.67%	16.19%	29.05%	31.90%	17.14%	18.57%	11.43%	Public administration	21.43%	37.86%	41.43%	25.00%	26.43%	20.00%	19.29%
Environmental studies	53.85%	48.46%	49.23%	50.00%	27.69%	27.69%	8.46%	Public environmental & education health	46.77%	46.45%	37.10%	40.65%	33.55%	35.48%	16.13%
Ergonomics	26.67%	40.00%	18.33%	20.00%	15.00%	40.00%	13.33%	Rehabilitation	14.38%	26.88%	22.50%	20.00%	29.38%	21.88%	10.63%
Ethics	45.00%	42.50%	35.00%	30.00%	25.00%	26.25%	12.50%	Social issues	30.83%	49.17%	51.67%	50.83%	49.17%	35.00%	29.17%
Ethnic studies	34.00%	56.00%	48.00%	40.00%	48.00%	44.00%	34.00%	Social psychology	40.30%	39.70%	34.24%	29.70%	26.97%	30.91%	11.21%
Experimental psychology	39.74%	43.08%	40.51%	41.03%	39.49%	37.95%	17.44%	Social work	15.29%	24.71%	25.29%	27.65%	31.18%	32.94%	22.35%
Family studies	10.00%	20.00%	10.00%	15.00%	5.00%	5.00%	5.00%	Sociology	23.70%	30.74%	30.93%	33.70%	37.78%	28.89%	20.74%
Geography	20.50%	20.50%	23.50%	23.50%	22.00%	21.00%	10.50%	Substance abuse	17.69%	20.77%	19.23%	19.23%	26.15%	25.38%	16.92%
Gerontology	21.82%	22.73%	25.45%	35.45%	29.09%	24.55%	20.91%	Transportation	22.22%	31.11%	36.67%	25.56%	42.22%	38.89%	27.78%
Health policy & services	29.13%	34.35%	40.00%	38.26%	34.78%	28.26%	23.48%	Urban studies	36.25%	42.50%	30.00%	31.25%	40.00%	51.25%	30.00%
History	16.15%	34.62%	30.00%	30.00%	32.31%	40.00%	24.62%	Women's studies	20.83%	27.50%	34.17%	34.17%	23.33%	27.50%	21.67%
History & philosophy of science	16.80%	26.80%	25.60%	27.60%	18.80%	21.20%	18.00%								

Source: the authors' processing of data on Thompson Scientific

Table A2: The Qualitative Rankings considered on the Journal Quality List

JQL symbol	Promoting organisation	Nation and Year	Distinctive features of the survey	Criteria for ensuring uniformity of the number of classes			
				Class A	Class B	Class C	Class D
Not95	Nottingham University	UK -1995	Scholars from 27 different institutions	4<Rank≤5	3<Rank≤4	3≤Rank<4	1≤Rank≤2
US98	Virginia Commonwealth University	USA-1998	Scholars from American universities	0.75<Rank≤1	0.5<Rank≤0.75	0.25<Rank≤0.5	0≤Rank≤0.25
NL99	Netherlands Academics in Business Administration	Holland-1999	Dutch scholars of Business Administration	Rank=A or A(P)	Rank=B or BP	Rank=C or CP	-
SMJ99	Strategic Management Journal	1999	Frequencies of citations in a sample group of 17 journals	0<Rank≤15	15<Rank≤30	30<Rank≤45	45<Rank≤65
Wiei01	Wirtschaftsuniv erisitat Wien	Austria-2001	In-house survey	Rank A or A+	Rank B	Rank C	Rank D
UQ03	University of Queensland	USA-2003	In-house survey	Rank 1	Rank 2	Rank 3	Rank 4 or 5
VHB03	Association of Professors of Management in German speaking countries	Germany-2003	Scholars ad researchers in Germany	Rank=A or A+	Rank=B	Rank=C	Rank=D or E
BJM04	British Journal of Management	UK-2004	Ranking of scientific and academic institutions in the UK	5.5<Rank≤7	4<Rank≤5.5	2.5<Rank≤4	1<Rank≤2.5
CNRS04	Centre de la Recherche Scientifique	France-2004	Opinion of select experts	Rank=5 or 4	Rank=3	Rank=2	Rank=1
ESS05	ESSEC Business School Paris	France-2005	Opinion of 7 scholars of the ESSEC	Rank=0 or 1	Rank=2	Rank=3	Rank=4
HKB05	Hong Kong Baptist University School of Business	Hong Kong-2005	List approved by the HBKU Executive Committee	Rank=A	Rank=B+	Rank=B	Rank=B-
Theo05	Survey performed by Theoharakis et al.	2005	Opinions of scholars and doctoral candidates in 7 different disciplines	71<Rank≤95	47<Rank≤71	23<Rank≤47	0≤Rank≤23
Ast06	Aston Business School	UK-2006	Opinions of scholars of the University of the Midlands	Rank=3	Rank=2	Rank=1	Rank=0
Cra06	Cranfield University School of Management	UK-2006	In-house survey	Rank=4	Rank=3	Rank=2	Rank=1
EJL06	Erasmus Research Institute of Management	Holland -2006	In-house survey	Rank=STAR or P	Rank=PA	Rank=S	Rank=SD

Table A2: The Qualitative Rankings considered on the Journal Quality List

JQL symbol	Promoting organisation	Nation and Year	Distinctive features of the survey	Criteria for ensuring uniformity of the number of classes			
				Class A	Class B	Class C	Class D
HMB06	Harvey Morris Business Journal Listing	UK-2006	Opinions of university rectors and directors of research centres in the UK	Rank=4 or 4*	Rank=3	Rank=2	Rank=1
FT06	Financial Times Survey	2006	Research objective is a ranking of top business schools	Classified	-	-	-

Source: the authors' processing of data on the Journal Quality List

www.alliedacademies.org