ACADEMY of ECONOMICS and FINANCE JOURNAL

Volume 4 2013

Aims and Scope of the *Academy of Economics and Finance Journal*:

The *Academy of Economics and Finance Journal (AEFJ)* is the refereed publication of selected research presented at the annual meetings of the Academy of Economics and Finance. All economics and finance research presented at the meetings is automatically eligible for submission to and consideration by the journal. The journal and the AEF deal with the intersections between economics and finance, but special topics unique to each field are likewise suitable for presentation and later publication. The *AEFJ* is published annually. The submission of work to the journal will imply that it contains original work, unpublished elsewhere, that is not under review, nor a later candidate for review, at another journal. The submitted paper will become the property of the journal once accepted, and not until the paper has been rejected by the journal will it be available to the authors for submission to another journal.

Volume 4:

Volume 4 of the *AEFJ* is sponsored by the Academy of Economics and Finance and the Department of Economics and Finance at the University of North Carolina Wilmington. The papers in this volume were presented at the Fiftieth Annual Meeting of the AEF in Mobile, Alabama from February 13-16, 2013. The program was arranged by Dr. Jocelyn Evans, College of Charleston.

ACADEMY of ECONOMICS and FINANCE JOURNAL

Volume 4 | Table of Contents | 2013

Table of Contents

Modeling Mergers and Acquisitions in the Electric Utility Industry

Earl H. Davis, Nicholls State University

Abstract

The electric utility industry in the United States saw a wave of mergers and acquisitions in recent years. Beginning in the nineties, electric utilities have also faced restructuring throughout the United States. Firms face both a risk from regulation and entry as restructuring occurs. Firms mitigate losses incurred from regulatory changes in order to maintain market value and competitive viability. Facing these effects firms often attempt to merge, acquire, or be acquired to mitigate loses in market value. This paper attempts to measure the gains from merger or acquisition activities across states in the face of major industry change. The results show that expected changes in market capitalization drive mergers and acquisitions. But the results are inconclusive with respect to state determinants of market capitalization, specifically restructuring and exogenous electricity demand factors.

Introduction and Background

The electric power industry emerged throughout the country in pockets. Generation was on a small scale, usually serving cities. Being dispersed, individual utilities developed in isolation. Through time, states began to regulate the industry, attempting to fix the duplication problems and other industry irregularities. But like any industry, a certain amount of scale is necessary for the efficient production of the good or service. Through the twenties, rampant mergers and acquisitions occurred in the industry, driving down the average cost of power but leading to increased concentration. New interstate utility holding companies developed. These new holding companies were unregulated due to constitutional limitations on the individual state's ability to regulate interstate commerce. In the early thirties the federal government stepped in when the financial malfeasance of the interstate holding companies became apparent.

The Roosevelt administration halted the wave of mergers and acquisitions occurring in the industry with the passage of the Public Utility Holding Company Act of 1935 (PUHCA). PUHCA was the major legislation that ended consolidation in the utility sector. With PUHCA regulators limited the relationship that unregulated firms had with their regulated subsidiaries, empowering the states to take a more active role in the intrastate regulation of utilities.

Beginning in the late seventies several events occurred to cause legislators to reconsider the regulation of gas and electric utilities, which by this time had become intricately linked. Fossil fuel costs rose, and gas turbine technology became more efficient. The Three Mile Island incident brought renewed scrutiny to the nuclear power industry, increasing the costs of using nuclear power technology. These events changed how firms operated in the market. As a result, Congress passed the Public Utility Regulatory Policies Act (PURPA) to encourage the more efficient use of energy sources. PURPA created a class of producers called qualifying facilities that enhanced the stock of generation, leading to more efficient and diverse production of power from combined-cycle sources. (Combined-cycle generation refers to generation other than from dedicated generators, usually industrial generation from the waste heat of industrial processes.)

Beginning in the late eighties, changes in technology precipitated changes in the regulation of utilities. The Energy Policy Act of 1992 (EPAct) empowered FERC with the tools to open up energy markets without giving it explicit authority to mandate retail wheeling – the sale of energy across disconnected buyers and sellers. States responded by enacting legislation that allowed generators greater flexibility in the competitive production of electricity. With the prospects for deregulation, or at least restructuring, looming analysts believed it would be necessary for firms to form strategic combinations to remain viable in the face of restructuring (Pierobon 1995). A substantial primer on deregulation and restructuring in the electric power industry is Warwick (2002).

Throughout the nineties and into the first decade of the 21st century firms have merged and acquired other firms with the ultimate goal of increasing their competitive viability in the face of a changing regulatory environment. The present paper attempts to show that legislative changes have an effect in driving mergers and acquisitions in the electric power industry by decreasing the market value of firms in restructuring states. As firm value declines in the restructuring states, firms become attractive targets for mergers and acquisitions by potential or former competitors.

Earlier Research

A substantial body of research has examined the scale and scope of the firm. Coase (1937) established the basic theory of why firms choose the scale and scope of activities they perform. There has been a substantial body of work on the general incentives of firms to merge or acquire other firms (Andrade, Mitchell and Stafford 2001) (Gugler, et al. 2003). Many have looked at the political incentives faced by utility firms (Bergara, Henisz and Spiller 1998), and the effect on productive efficiency of electric utility firms (Delmas and Tokat 2005).

Mitchell and Mulherin have studied the conditions under which firms will merge when faced with industry shocks. Their hypothesis maintained that corporate takeovers represented the least cost means of coping with industry-specific economic shocks (Mitchell and Mulherin 1996). They studied the effect of industry-specific changes in regulation or market structure. Looking specifically at the 1980's they found that takeover activity occurred disproportionately at the industry level. The industries they observe with the greatest takeover activity also had the greatest fundamental shocks—or shocks such as deregulation, technological change, or others that caused fundamental industry change. They found that the most robust results were in industries with the greatest stock market performance. While Mitchell and Mulherin studied a variety of industries, they avoided industries that experienced significant state and federal regulation, notably the electric power and banking industries.

Many scholars have examined electricity specific mergers and acquisitions. One early study noted that utility regulation posed a major barrier to mergers and acquisitions (Ray and Thompson 1990). While potential gains were large, the study found most of the benefits would be subsequently transferred to rate payers. Other studies have contested the seemingly apparent existence of operational efficiency gains, although leaving open gains through vertical integration and coordination(Hartman 1996). One study looked at the market effects of mergers and acquisitions, finding that investors perceive few benefits to mergers in the generation sector, but stockholders favored vertical electric and gas mergers (Berry 2000). Leggio and Lien (2000) find similar results, focusing on the timing of the mergers and acquisitions with respect to the passage of the EPAct. They find significant negative returns with merger announcements after passage.

Recent research supports the view that mergers and acquisitions activities have not created significant value for shareholders. One study finds that firms that responded to deregulation or restructuring via mergers and acquisitions generally underperformed firms who did not (Becker-Blease, Kaen and Goldberg 2008). Looking at operating and total cost data in the US electricity industry, Kwoka and Pollitt (2010) find that mergers and acquisitions are not consistent with increased cost performance. But the most recent research on the potential for gains from merger and acquisition suggests that utility mergers create value (Becher, Mulherin and Walkling 2012). Furthermore they find that utility mergers are consistent with synergies and inconsistent will collusion across states. While not applicable to the United States, Keller (2010) finds a similar result in Germany.

Other authors have looked at vertical mergers in the electric utility industry. Vertical mergers have been called convergence mergers to describe mergers between upstream natural gas and downstream electric generators. While this paper does not directly examine these mergers, they are worth mentioning. Russo (1992) explores the importance of divergence and integration finding support for the transaction cost view of the firm. When faced with heightened regulation, firms responded to regulation by diversifying away from oversight as oversight intensified and disintegrating in upstream activities as regulation of those activities intensified. Others have noted that, "Convergence mergers of electric and gas companies mirror significant advances in technology, competitive dynamics, and energy economics." (Sikora 1997, pp. 47)

Brennan's analysis (2001) of vertical mergers begins by noting that vertical mergers are an "oxymoron," noting that they limit the provision of complements, rather than substitutes, as in traditional horizontal mergers. The market power that is a problem is horizontal in the input market. It notes that convergence mergers are neutral if the generator is gas fired—the gas producer already has market power. Brennan observes that convergence mergers can eliminate the double marginalization: the existence and exercise of market power in vertical markets. By eliminating the duplication of market power, the merger raises profits for the market participants and decreases price for consumers.

Hunger reaffirms Brennan's conclusions. Hunger (2003) looks at the merger review process and concluded that traditional merger analysis does not adequately capture the ability of firms to "raise rivals cost," effectively opening the door for collusion in vertical mergers and acquisitions.

Regulatory Background and Data

Mergers and acquisitions in the electric power industry were different from mergers in other industries in many ways. Regulators scrutinized mergers much more in the electric power industry due to the multi-tiered nature of regulation. Not only did the federal government through PUHCA limit the scope of mergers—into complementary industries such as gas—states often entered to limit the scale of mergers. But state regulation does not prevent the federal government from limiting scale. The Federal Energy Regulatory Commission (FERC) still maintains oversight of interstate mergers in the electric power industry.

In the late eighties the regulatory environment began to change for the industry. The first significant change occurred with the passage of the EPAct. This piece of legislation authorized FERC to begin the process of deregulating electricity markets. With Order 636, FERC mandated the unbundling of transmission and distribution from generation. The practical effect of the order was to allow for the separation of transmission and distribution costs from generation. With Order 888 FERC instituted the "open access" rule: generation was no longer limited to large co-generators or transmission and distribution owning utilities. By allowing for the separation of transmission and distribution states were now free to pass unilateral restructuring without interference from federal regulators.

With Order 888 and restructuring within the states utilities faced a regulatory environment unlike any in the last 60 years. As Mitchell and Mulherin have shown, with changing costs and regulations, industries adapt by instituting corporate restructuring: mergers, acquisitions, and takeovers. From 1996 through 2003 the electric power industry went through a wave of corporate restructuring unlike any since prior to the passage of PUHCA. Not only did private utilities begin acquiring smaller utilities and spinning off transmission and distribution assets as mandated by many states, but new classes of power producers—many of whom did not own either transmission or distribution assets—entered markets.

The data for this study was derived from COMPUSTAT financial data, Value Line investor's surveys, FERC, and Energy Information Agency (EIA) data collected from power plants across the country. The COMPUSTAT data is from 1996 through 2003. The study ignores years prior to 1996 as there were few mergers or acquisitions prior to 1996. The COMPUSTAT data was also limited to firms that had SIC numbers consistent with firms in the electricity generation, transmission, or distribution industry. Firms that specialized in the production or distribution of natural gas were excluded, as there was no EIA data on those firms. Their inclusion would be consistent with the estimation of the benefits of mergers and acquisitions hypothesized.

FERC must approve all interstate mergers and acquisitions. It has the most complete compendium of mergers and acquisitions and is therefore the source for all approved mergers and acquisitions analyzed. FERC also maintains a database of financial information in a proprietary format. Unfortunately FERC does not acquire financial data on all IOUs, only the largest, so the COMPUSTAT data was used. The market value regression uses the absolute change in sales and lagged liability to asset ratio for all of the firms who took part in a merger or acquisition from 1996 to 2003. By using absolute change, the study picks up the effect Mitchell and Mulherin observe, that changing industry sales may induce firms to merge or acquire other firms in response to industry shocks.

Results

Table 1 summarizes the principle variables used throughout this study. Each observation is an instance of a firm listed in the COMPUSTAT data that could be reconciled with the CRSP data by id number—a combination of PERMNO and CUSIP reconciled originally in a separate list. Of the total number of firms in the industry, the study uses all that took part in a merger or acquisition from 1996 through 2003. In any given year, of that sample of firms, 10% took part in a merger or acquisition. Cost per Btu is the average cost of production for electricity in the state of incorporation. Heating and cooling degree days per year are used as a measure of energy demand in each state. A heating or cooling degree day is calculated for each degree difference in temperature from 65 degrees Fahrenheit. The cumulative heating and cooling degree days in December is used as a measure of total heating and cooling degree days for the year. The cumulative deviation in December is used to determine the total cumulative deviation for the year.

Table 1: Summary Statistics

Variable	***Observations***	***Mean***	***Std. Dev.***	***Min***	***Max***
Market Capitalization	413	3705916	4660923	0	3.07e+07
Merger or Acquisition	413	.1016949	.3026133	0	1
Sales Growth	218	1.103389	.5624222	.1631399	5.619532
Lagged Liability to Asset Ratio	230	.7375282	.0928851	.3739204	1.170628
Cost per Btu	413	.1101105	.1246533	0	.7045165
Number of Utilities	413	33.32203	46.38505	1	184
Heating Days	413	1846.08	842.3508	0	3643
Heating Deviation.	413	-71.90073	231.7843	-697	479
Cooling Days	413	966.0315	886.8051	0	4641
Cooling Deviation	413	63.97821	155.9083	-299	536
Legislation Enacted	413	.4552058	.4985934	0	1

This study borrows data from a previous study (Davis forthcoming)of the effects of state restructuring on price, or specifically average revenue, based on Sass and Leigh (1991). The model estimates the probability of a merger based on the estimated value of the merger relative to not merging and the absolute percentage change in sales and liability to asset ratio:

$$M_{it}^{*} = Z_{it}\gamma + u_{0it} \qquad M_{it} = \begin{Bmatrix} 1 \text{ if } M_{it}^{*} \geq 0 \\ 0 \text{ if } M_{it}^{*} \leq 0 \end{Bmatrix} \tag{1}$$

M_{it} represents the value of the merger to an acquiring firm. If the merger has a positive net value then it proceeds with a binary value of 1. If a merger in a given year has zero or negative value, then we observe no merger and a binary value of 0.

The vector Z_{it} is a collection of firm-specific characteristics such as the absolute value of the percentage change in sales and liability to asset ratio. Absolute percentage change in sales will indicate whether the firm is facing a shock that affects sales, either positively or negatively, within the industry. When facing a shock, firms are more likely to engage in restructuring activities either to mitigate negative shocks or to take advantage of a shock by acquiring assets that were previously uncompetitive.

The lagged liability to asset ratio gives an indication of whether a firm is attempting to merge or acquire another firm in a previous period. If a firm is planning a merger or acquisition, then the firm will issue new debt in order to acquire a firm or merge in the next period.

Another determinate of the attractiveness of a merger is the potential gains from the merger. Equation 2 estimates the expected value of the natural log of market capitalization by state as a function of the standard normal probability density function (ϕ) and standard normal cumulative density function (Φ) and a vector of determinants of firm value by state (X).

$$E(\text{Market Capitalization}) = (\alpha_1 - \alpha_2)\Phi + (\beta_1 - \beta_2)X\Phi + \alpha_2 + \beta_2 X + (\sigma_{2u} - \sigma_{1u})\phi \tag{2}$$

Included in the vector of state specific effects are heating and cooling degree days for a year, the cumulative deviation in heating and cooling degree days for a year, the number of utilities in the state as of 1995, the average cost per Btu of energy within the state, and whether the state enacted restructuring legislation in a given year.

Equation 3 is the probit estimate of firm specific effects including the difference in market capitalization between firms that do not merge (C_1) and those that do (C_2).

$$M_{it} = Z_{it}\gamma + \rho(C_1 - C_2) + u_{0it} \tag{3}$$

Equation 4 is the probit estimate substituting the expected value derived in equation 2 above.

$$M_{it} = Z_{it}\gamma + \rho((\alpha_1 + \beta_1 X) - (\alpha_2 + \beta_2 X)) + \rho(u_{1it} - u_{2it}) + u_{0it} \tag{4}$$

Estimation requires running equation 4 to get estimates of Φ and ϕ. Equation 2 is then estimated using the values of Φ and ϕ to determine the coefficients α_1, α_2, β_1, and β_2. These coefficients are then used to determine the expected difference in market capitalization from merging where the subscript "1" is for non-merging firms and "2" represents merging firms.

Table 2 is the results of models (1) and (3) above:

Table 2: Regression of Probability of a Merger or Acquisition

Variable	*Expected Sign*	*Without Difference in Market Capitalization*	*Result with Difference*
Absolute Percentage change in sales	+	-0.130 (0.28)	-1.275 (1.34)
Lagged Liability to Asset Ratio	+	-3.849 (1.76)	-4.294 (1.45)
Difference in Market Cap	-		-0.142 (3.08)**
Constant		1.113 (0.69)	4.289 (1.73)
Expected Mean Difference in Market Capitalization (Mkt Cap) *E*(Non Mergers Mkt Cap - Mergers Mkt Cap)			30.377
Observations		218	218

Absolute value of t statistics in parentheses

* significant at 5%; ** significant at 1%

The predicted sign for absolute percentage change in sales is positive. To capture an effect, the absolute value is taken to determine a total shock. The negative sign implies that an increase in absolute change in sales for a year decreases the probability of a merger or acquisition in the current year. Another interpretation is that an expected change in sales will decrease the probability of a merger or acquisition in the coming year. Mitchell and Mulherin observe that changes in sales have no effect on merger or acquisition activity, but changes relative to other industries do have an effect. Because this study only looks at the electricity industry, changes with respect to other industries are unobserved. No effect from an absolute percentage change in sales is observed.

The predicted sign for lagged liability to asset ratio is positive, implying that increases in the issuance of debt in a year correspond to increases in merger and acquisition activities. This is consistent with the intuition that firms use debt to finance mergers and acquisitions as proposed by Blair (1993). Unfortunately no effect is observed on mergers and acquisitions.

The difference in market capitalization from equation (3) is included here. The coefficient of the expected difference is negative. This is consistent with the hypothesis that, as the expected difference in market capitalization falls, firms are more likely to engage in merger and acquisition activity. This could be interpreted to mean that as the expected market value of merging firms converges with that of non-merging firms, those that make the merger or acquisition decision are more likely to do so, supporting a hypothesis that IOUs engage in merger activity to increase market value. Finally, the expected market value of the non-merging firms is on average $30 million more than the merging firms, reaffirming the conclusion that merger and acquisition activity is being driven by attempts to increase the market value of the firm, relative to other firms in the industry. $30 million is a small difference considering the value of even the smallest IOUs is in the billions, but it supports a hypothesis that IOUs were partaking in convergence mergers and acquisitions in the surveyed period.

Table 3 lists the coefficients from the expected market capitalization regression. Although most of the variables are of the appropriate sign, they have a statistically insignificant effect of market capitalization in the model. This brings into question some of our assumptions about what determines the value of firms across the states. This is not inconsistent with what one could assume in a regulated environment, namely that regulation protects inefficient firms leading to perverse market outcomes in the market for electric power.

Conclusion

Given the regulatory conditions under which IOUs operate, it is difficult to explain what drives mergers and acquisitions across firms and states through time. Any explanation would have to include the incentives for the IOU to merge or acquire other firms. The model presented suggests that convergence may be one explanation. It is well understood that the manager has an incentive to increase value for the shareholder (Mulherin and Boone 2000), as well as potentially consolidating his management position (Roll 1986). It is reasonable to believe that the manager will make those decisions that increase the market value of the firm when faced with regulatory constraints imposed by the states.

Studies have shown that firms facing regulatory change and market pressure engage in merger and acquisition activity to mitigate the negative market effects and increase efficiency. Not only does evidence exist for the electric power industry, but for other industries as well (Peristiani 1997) (Cummins, Tennyson and Weiss 1999) (Brook, Hendershott and Lee 2000) (Madura, Ngo and Viale 2012) (Peyrache 2012).

Blair (1993) shows that firms use debt to finance merger and acquisition activity. These incentives could explain some of what drives merger and acquisition activities among IOUs. But this does not explain particular state-specific incentives to engage in these activities.

This paper addresses those concerns by estimating differences in market value of firms that engage in mergers and those that do not. The estimates obtained are consistent with convergence as an explanation of why firms engaged in merger and acquisition activities throughout the restructuring period. Unfortunately the results do not show absolute percentage change in sales and liability to asset ratio as having any effect on merger or acquisition activities, although the results are consistent with Mitchell and Mulherin's assertion of no effect within an industry, although not across industries. The results are consistent with the recent research showing ultimately little benefit from mergers and acquisitions in the electric power industry as measured by market returns, although inconsistent with Becher, Mulherin and Walkling (2012) who find important evidence of synergies from utility mergers and acquisitions, also controlling for geography.

Table 3: Regression of Market Capitalization on State Determinants of Firm Value

Variable	***Expected Sign***	***Coefficient***
Cost Per BTU	-	-0.516
		(0.57)
Number of Utilities	-	-0.00125
		(0.44)
Heating Days	-	-0.00032
		(1.89)
Cumulative Deviation in Heating Days	+	0.00056
		(0.50)
Cooling Days	+	0.00021
		(1.16)
Cumulative Deviation in Cooling Days	+	-0.00102
		(0.85)
Legislation Enacted	?	0.336
		(1.39)
φ	?	4.325
		(0.78)
Constant	?	14.975
		(47.88)**
Observations		218
R-squared		0.09

Absolute value of t statistics in parentheses

* significant at 5%; ** significant at 1%

References

Andrade, Gregor, Mark Mitchell, and Erik Stafford. "New Evidence and Perspectives on Mergers." *The Journal of Economic Perspectives* 15, no. 2 (2001): 103-120.

Becher, David A., J. Harold Mulherin, and Ralph A. Walkling. "Sources of Gains in Corporate Mergers: Refined Tests from a Neglected Industry." *Journal of Financial and Quantitative Analysis* 47, no. 1 (2012): 57-89.

Becker-Blease, John R., Fred R. Kaen, and Lawrence G. Goldberg. "Mergers and Acquisitions as a Response to the Deregulation of the Electric Power Industry: Value Creation or Value Destruction?" *Journal of Regulatory Economics* 33, no. 1 (2008): 21-53.

Bergara, Mario, Witold J. Henisz, and Pablo Tomas Spiller. "Political Institutions and Electric Utility Investment: A Cross-Nation Analysis." *California Management Review* 40 (1998): 18-35.

Berry, S. Keith. "Excess Returns in Electric Utility Mergers During Transition to Competition." *Journal of Regulatory Economics* 18, no. 2 (2000): 175-188.

Blair, Margaret M, ed. *The Deal Decade: What Takeovers and Leveraged Buyouts Mean for Corporate Governance.* Washington, D.C: The Brookings Institution, 1993.

Brennan, Tim. *Vertical Market Power' as Oxymoron: Getting Convergence Mergers Right.* Discussion Paper 01-39, Resources for the Future, 2001.

Brook, Yaron, Robert J. Hendershott, and Darrell Lee. "Corporate Governance and Recent Consolidation in the Banking Industry." *Journal of Corporate Finance* 6, no. 2 (2000): 141-164.

Coase, Ronald H. "The Nature of the Firm." *Economica* 16, no. 4 (1937): 386-405.

Cummins, J. David, Sharon Tennyson, and Mary A. Weiss. "Consolidation and Efficiency in the US Life Insurance Industry." *Journal of Banking & Finance* 23, no. 2 (1999): 325-357.

Davis, Earl H. "Interaction Between the Legislative and Regulatory Process: The Case of State Level Electricity Restructuring." *Journal of Business Industry and Economics* (Society of Business, Industry, and Economics), forthcoming.

Delmas, Magali, and Yesim Tokat. "Deregulation, Governance Structures, and Efficiency: The US Electric Utility Sector." *Strategic Management Journal* 26, no. 5 (2005): 441-460.

Gugler, Klaus, Dennis C. Mueller, B. Burcin Yurtoglu, and Christine Zulehner. "The Effects of Mergers: An International Comparison." *International Journal of Industrial Organization* 21, no. 5 (2003): 625-653.

Hartman, R. S. "The Efficiency Effects of Electric Utility Mergers: Lessons From Statistical Cost Analysis." *Energy Law Journal* 17, no. 2 (1996).

Hunger, David. "Analyzing Gas and Electricity Convergence Mergers: A Supply Curve is Worth a Thousand Words." *Journal of Regulatory Economics* 24, no. 2 (2003): 161-173.

Keller, Andreas. "Competition Effects of Mergers: An Event Study of the German Electricity Market." *Energy Policy* 38, no. 9 (2010): 5264-5271.

Kwoka, John, and Michael Pollitt. "Do Mergers Improve Efficiency? Evidence from Restructuring the US Electric Power Sector." *International Journal of Industrial Organization* 28, no. 6 (2010): 645-656.

Leggio, Karyl B., and Donald Lien. "Mergers in the Electric Utility Industry in a Deregulatory Environment." *Journal of Regulatory Economics* 17, no. 1 (2000): 69-85.

Madura, Jeff, Thanh Ngo, and Ariel M. Viale. "Why Do Merger Premiums Vary Across Industries and Over Time?" *The Quarterly Review of Economics and Finance* 52, no. 1 (2012): 49-62.

Mitchell, Mark L., and J. Harold Mulherin. "The Impact of Industry Shocks on Takeover and Restructuring Activity." *Journal of Financial Economics* 41, no. 2 (1996): 193-229.

Mulherin, J. Harold, and Audra L. Boone. "Comparing Acquisitions and Divestitures." *Journal of Corporate Finance* 6, no. 2 (2000): 117-139.

Peristiani, Stavros. "Do Mergers Improve the X-Efficiency and Scale Efficiency of US Banks? Evidence from the 1980s." *Journal of Money, Credit, and Banking* 28, no. 3 (1997): 326-337.

Peyrache, Antonio. *Industry Structural Inefficiency and Potential Gains From Mergers and Break-ups: An Empirical Approach.* Working Paper, School of Economics, The University of Queensland, Centre for Efficiency and Productivity Analysis (CEPA), 2012.

Pierobon, James R. "Merging for Critical Mass." *Public Utilities Fortnightly*, 1995: 16-16.

Ray, Dennis J., and Howard E. Thompson. "“Fifty in Five”: The Prospects for Merger in the Electric Utility Industry." *Journal of Regulatory Economics* 2, no. 2 (1990): 111-128.

Roll, Richard. "The Hubris Hypothesis of Corporate Takeovers." *The Journal of Business* 59, no. 2 (1986): 197-216.

Russo, Michael V. "Power Plays: Regulation, Diversification, and Backward Integration in the Electric Utility Industry." *Strategic Management Journal* 13, no. 1 (1992): 13-27.

Sass, Tim R., and Paul J. Leigh. "The Market For Safety Regulation and the Effect of Regulation on Fatalities: The Case of Motorcycle Helmet Laws." *The Review of Economics and Statistics* 73 (1991): 167-172.

Sikora, Clifford S. "Engineering Exotic Blends In Utility Mergers." *Mergers and Acquisitions - Philadelphia*, 1997: 47.

Warwick, William M. *A Primer on Electric Utilities, Deregulation, and Restructuring of US Electricity Markets.* PNNL-13906, Richland, WA: Pacific Northwest National Laboratory, 2002.

Finance-Led Growth in Developing Countries of Southern Europe: Financial Markets and Economic Development

Hugh L. Davis, III and Shawn D. Lowe, University of Southern Mississippi

Abstract

Generally, the literature holds that the Finance-Led Growth Theory, particularly that of financial deepening, promotes economic growth. Financial deepening is the increased provision of financial services geared to all sectors and levels of society. There are a number of agents promoting financial deepening. One agent, financial markets, has been determined to be an efficient allocator of funds and an important channel for the diffusion of capital and technology. This paper examines the linkage between financial markets growth and economic development in a particular group of southern European countries – Portugal, Italy, Greece, and Spain (PIGS), using panel data from 2000 to 2011.

Introduction

Bagehot (1873) emphasized the importance of the banking system in economic growth by spurring innovation and future growth through identification and funding of productive investments. Schumpeter (1911) argued that financial services are paramount in promoting economic growth as production requires credit - entrepreneurs require debt. Keynes (1930) also placed a premium on the importance of the role of the banking sector in economic growth. Gurley and Shaw (1955) demonstrated the relationship between financial markets and growth activity; the depth of the financial system, they argued, is key. Financial markets extend a borrower's financial capacity; they allow investors' to fund various capital projects.

Bencivenga *et. al.* (1996), Levine (1991), Kyle (1984), Holmstrom and Tirole, (1993) and Obstfeld (1994) demonstrate that risk-sharing and the provision of liquidity are cornerstones of the subsequent economic growth provided by financial markets. Greenwood and Smith (1996) highlight these cornerstones in funding the most productive technologies. Kim (2003) extends Greenwood and Smith's work.

Levine and Zervos (1998) identified significant causal relationships between development of stock markets and economic growth. Khan (2008) concurred. Caporale and Spagnola (2011), examined linkages between stock markets and economic growth in three developing European Economic Community countries (the Czech Republic, Hungary and Poland). Their findings suggest there is unidirectional causality running from stock markets to growth.

This paper is divided into seven sections. The first section is the Literature Review. Section Two develops the basis for the theory and the resulting hypothesis. A model is presented in Section Three. Section Four discusses the methodology. Testing the model is presented in Section Five. Section Six expands the testing of the hypothesis to include other perspectives. The final section summarizes and concludes the paper.

Theory

The *Endogenous Growth Theory* holds that economic growth is primarily the result of endogenous (internal) and not exogenous (external) forces (Romer, 1986). The theory holds that investment in human capital, innovation and knowledge are significant contributors to economic growth. Policies which create openness, competition, and innovation will promote growth (Fadare, 2010). Policies which have the effect of restricting change are likely over time to inhibit growth. Romer's (1988) *Finance Led Theory states* that channels of financial development lead a nation's economic growth by positively supporting investment in technology. Financial development primarily flows through channels like markets, insurance companies, commercial and investment banks and regulatory structure. Each channel has different proxies that can be utilized to test the Endogenous Growth Theory hypothesis (Calderón, Ce'sar and Liu, Lin, 2002).

Model and Data

This study expands upon the Finance Led theory of Greenwood and Smith (1996), Levine (2005), and Caporale and Spagnola (2011). Four developing countries of southern Europe (Portugal, Italy, Greece, and Spain) were chosen to test this theory. These economies share strong similarities with their developed counterparts. The measures (proxies) this study uses testing their impact on development include: 1) Market Capitalization Ratio (***mcr***), 2) Total Value of Shares Traded Ratio (***str***), 3) Turnover Ratio (***tr***), and 4) List of Domestic Companies (***dc***). These are found in the World Bank Databank. These are measures of financial depth and are proxies that may demonstrate a unidirectional, causal effect on growth.

The following demonstrates the model:

$\boldsymbol{gdp}_t = \beta_0 + \beta_1 \boldsymbol{mcr} + \beta_2 \boldsymbol{str} + \beta_3 \boldsymbol{tr} + \beta_4 \boldsymbol{dc} + u.$

Using the model, this paper examines the causal impact of financial markets growth on economic development in Portugal, Italy, Greece, and Spain (PIGS). Using panel data from 2000 to 2011 this study investigates both the cause and strength of the linkage between financial market development and economic growth.

Variables

Gross Domestic Product (***gdp***)

The Gross Domestic Product (GDP) is the sum of gross value added by all resident producers and is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. The data used in the study is in current U.S. dollars.

Measures of stock market development used in this analysis include: 1) Market Capitalization (***mcr***), 2) Total Value of Shares Traded (***str***), 3) Turnover (***tr***), and List of Domestic Companies (***dc***).

Market Capitalization (***mcr***):

This measure equals the value of listed shares.

Total Value of Shares Traded (***str***):

This measure equals the total value of shares traded on the stock market exchange. The greater the number the greater the liquidity.

Turnover (***tr***):

This ratio equals the value of total shares traded divided by market capitalization. A small liquid market will have a high turnover ratio but a small total value traded ratio (World Bank, 2012).

Listed Domestic Companies (dc):

Listed domestic companies are the endogenous incorporated companies listed on the country's stock exchanges at the end of the year.

These measures were obtained through the World Bank's data base of indicators. The three ratios ***mcr, str*** and ***tr*** could cause complications in that the first two use ***gdp*** as a divisor and ***tr*** is a derivative of ***mcr***. To eliminate this potential error, we reversed out of the ratios to use the raw numbers in our model.

Methodology

An ordinary least squares regression results in parameter estimates from data for the years 2000 to 2011 for each of four countries available in the World Bank's databank (World Bank, 2012).
The null hypotheses can be written as: H_1: $\beta_{mcr} = 0$; H_2: $\beta_{str} = 0$; H_3: $\beta_{tr} = 0$; H_4: $\beta_{dc} = 0$;

Table One: Regression Results

	Coef	**Std.Err.**	**P**
β_0	2.64e+11	9.43e+10	
Mcr	.0178437	.0051464	0.001
Str	.0132702	.0042783	0.003
Tr	-.0001213	.0000471	0.014
Dc	-3.22e+08	7.67e+07	0.000

We write the estimate with the standard errors in parentheses:

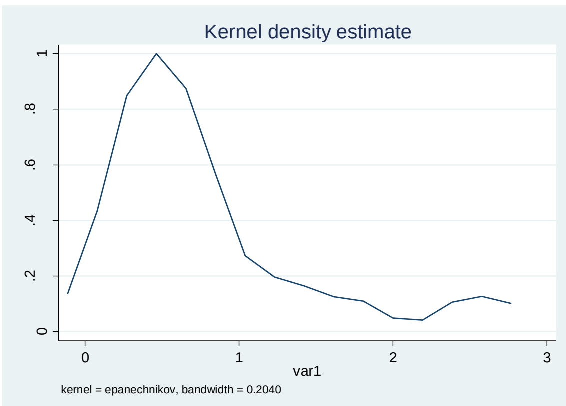
$$\underset{}{\boldsymbol{gdp}} = \underset{(9.43e+10)}{2.64e+11} + \underset{(.0051464)}{.0178437}\boldsymbol{mcr} + \underset{(.0042783)}{.0132702}\boldsymbol{str} - \underset{(.0000471)}{.0001213}\boldsymbol{tr} - \underset{(7.67e+07)}{3.22e+08}\mathbf{dc}$$

All the variables are significant. Notice that both ***tr*** and ***dc*** have a negative or inverse relationship.

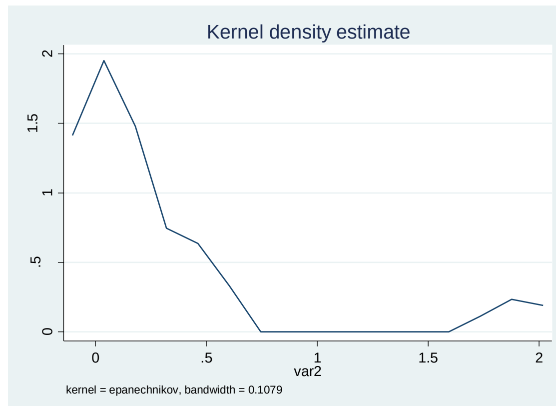
Model Testing

The goodness of fit of the regression model is high given values for F (4,43) = 21.89 with $p < .0000$ and a coefficient of determination (r^2) of 67.06% and an Adj r^2 of 0.6400. This indicates a significant amount of variation is accounted for by the explanatory variables.

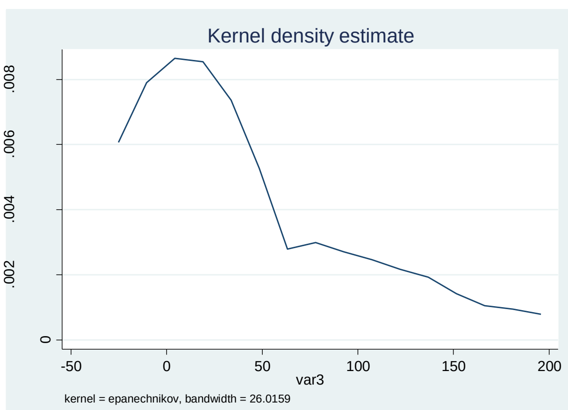
Kernel density estimations (Parzen, 1962) were performed for each variable using 2011 data and the results illustrating possible non-normality of some of these data distributions.



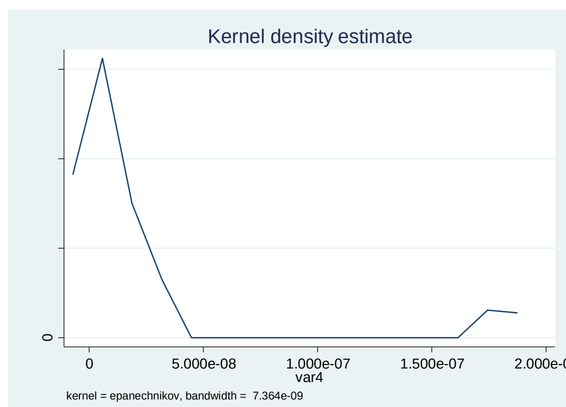
Graph 1
4 Countries reporting *mcr/gdp*



Graph 2
4 Countries reporting *str/gdp*



Graph 3
4 countries reporting *tr/gdp*



Graph 4
4 Countries reporting *dc/gdp*

The ***mcr/gdp*** and ***tr/gdp*** appear to have something approaching normal distributions, while ***str/gdp*** and ***dc/gdp*** are much less so. All four are skewed positively. The estimations warrant further assessment of data normality and possible use of nonparametric statistical methods for analysis of relationships between variables.

We test the following hypothesis: H_0: variables ***mcr, str, tr,*** and ***dc*** are normally distributed. Shapiro-Wilk tests, Shapiro-Francia, and KS Test for skewedness and kurtosis tests indicate that the data is not normally distributed. As well, a plot of the residuals against the ***gdp*** as well as the Breusch-Pagan test (Wooldridge, 2009) are used to determine the heteroskedasticity present in the model. The plot of the residuals visually demonstrates a wide variance while the Breusch-Pagan test yielded a chi square of 1.54 with a $p < .2153$. We can conclude that heteroskedasticity exists. This is a weakness in the model.

Tests for multi-collinearity using Variance Inflation Factors (VIF) method are revealing.

Table Two: test for Multi-Collinearity

Variable	VIF	1/VIF
tr	33.51	.0298
str	25.28	.0395
mcr	13.10	.0763
dc	2.44	.4099
Mean	18.58	

VIF's below .10 are suspect; the efficiency of the model is reduced due to multi-collinearity in at three factors.

We also test for model specification. The null hypothesis is H_0: the model is correctly specified. Ramsey's omitted-variable regression specification error test (RESET) yields a value of F (3,40) = 5.86 with p = .002. The high degree of significance of these results implies that we should reject the null hypothesis that the model is correctly specified. Future models should be re-specified.

Expanded Study to include Panel Data

There are 4 panels which include data that is set up with time variant variables for the years 2000 to 2011. Four tests are administered: a) Simple Regression (one variable and four variable summaries), b) Fixed Effect, c) Random Effect, and c) Hausman test. The analysis indicates that it is "strongly balanced."

Simple Regression, One Variable at a time Summary

Based upon the test run, ***mcr, str***, and ***tr*** have the best fit and are highly significant and ***dc*** is marginally significant.

Simple Regression, Four Variables Summary

In the four variable run, ***mcr, str and dc*** are highly significant with a p value less than .000 and ***tr*** is significant with a p value of .014.

Fixed-Effect Model

The fixed-effect model omits all time-invariant factors. Comparing the coefficients of the variables indicates whether ***mcr***, ***str, tr*** and ***dc*** improve growth in ***gdp.*** "The key insight is that if the unobserved variable does not change over time, then any changes in the dependent variable must be due to influences other than these fixed characteristics" (Stock and Watson, 2003, p. 289-290). Since the probability of the F is .0007 then we determine that the model is a good fit. The test indicates that **str** and **dc** are significant. The Rho is the intra-class correlation, meaning the average correlation of variables within the countries The Rho is .8888 indicating how much of the variance is due to the differences across the panels.

Table Three: FE Results

	Coef	*P*	*F*	*P*	r^2
			5.98	.0007	.3743
mcr	-.0010372	.767			
str	.0053798	.039			
tr	-.0000226	.429			
dc	2.36e+08	.037			
cons	4.66e+11	.000			
Rho			.8888		
F U=0			32.84	.0000	

We will need to utilize other tests to determine which is the most appropriate.

Random -Effect Model

Unlike the fixed effects model, the random-effect model assumes that the variation across countries is random and uncorrelated with the predictor or independent variables included in the model (Torres-Reyna, 2010). Green (2008, p. 183) states "…the crucial distinction between fixed and random effects is whether the unobserved individual effect embodies elements that are correlated with the regressors in the model, not whether these effects are stochastic or not". The random effects assumption is that the individual specific effects are uncorrelated with the independent variables.

Since the probability of the F is .0000 we determine the model is also a good fit. The test indicates that ***mcr,*** **str,** and ***dc*** are very significant and **tr** is significant.

Hausman Effect Model

To decide between fixed or random effects, a Hausman test is conducted where the null hypothesis is the preferred model is random effects and the alternative hypothesis equaling the fixed effects (Green, 2008). This model tests whether the unique errors are correlated with the regressors. The test results indicate that the Prob > chi2 (.0000); we conclude that the fixed effect model is the most useful.

Conclusion

The literature suggests various agents, including financial markets, positively affect economic development. This study utilizes four independent variables to test their predictive value. In general, the results were conflicting with the intuitive reasoning behind the hypothesis.

The Fixed Effect, confirmed by the Hausman test, indicates the two candidates for explanatory/predictor variables are ***str*** and ***dc***. ***Str*** or the Total Value of Shares Traded and ***dc***, or Listed Domestic Companies are both significant. Though **str** is not a direct measure of liquidity, it does measure the total trading value of the stock market. The higher the trading volume the greater the liquidity. ***Dc***, listed domestic companies, is the domestically incorporated companies listed on the country's stock exchanges at the end of the year. It is logical and rational that the growth in the number of domestic companies results in economic growth. Though not a complete picture of capital formation, the increase in the number of listed companies should reflect successful second stage growth in businesses. What is troubling is the negative sign indicating an inverse relationship. A closer examination of the raw data reveals that the actual number of listed domestic companies declined in three out of four panels.

The study indicates support for the Finance-Led Growth Theory in general and a semi-strong relationship between the growth of financial markets and economic development. We can conclude that the financial sector efficiently distributes funds among various economic activities that foster innovation and technology.

Future research pertaining to the Finance Led Theory might include: a) the search for more reliable proxies, b) gathering data that goes deeper than 12 years, c) increasing the number of panels, d) reviewing institutional factors that inhibit or enhance the start up of domestic companies, e) analyzing the effects of externalities in causing the creation of domestic companies, and f) re-specifying a model using ***str*** and ***dc.***

References

Bagehot, Walter. 1873. *Lombard Street: A Description of the Money Market*. London: Henry S. King. Third Edition.

Bartels, Brandon. 2008. "Beyond "Fixed Versus Random Effects: A Framework for Improving Substantive and Statistical Interpretation for TSCS, Panel, and Multilevel Data." Working Paper.

Beck, Thorsten, Ross Levine and Norman Loayza. 1999. "Finance and the sources of growth," *Policy Research Working Paper Series 2057, The World Bank.*

Bencivenga, V.R., and B. Smith. 1996. "Equity Markets, Transaction Costs, and Capital Accumulation: An Illustration*." The World Bank Economic Review,* Vol. 10: 241-65.

Calderon, Cesar & Lin Liu. 2002. The Direction of Causality Between Financial Development and Economic Growth. Working Papers Centeral Bank of Chile, 184.

Caporale, M.G., C. Rault, R. Sova, & A. Sova. 2009. Financial Depth and Economic Growth In New E.U Members.

Caporale, M.G. and Nicola Spagnola. 2011. Stock Market and Economic Growth; Evidence from Three CEECs, Center for Empirical Finance. Brunel University, London, UK.

Choong, Chee-Keong and Sok-Gee Chan. 2010. Financial Development and economic Growth: A Review. Centre for Economic Studies, Faculty of Business and Finance, Universiti Tunku Abdul Rahman (Perak Campus), Jalan

University, Bandar Barat, 31900 Kampar, Perak, Malaysia.
De Gregorio, J. and Pablo Guidotti. 1995. "Financial Development and Economic Growth," *World Development.*
Fadare, Samuel O. 2010. "Recent Banking Sector Reforms and Economic Growth in Nigeria". *Middle Eastern Finance and Economics* (Issue 8).
Forbes Financial Dictionary. 2011.
Green, W. 2008. Econometric Analysis, 3rd Ed. New York: MacMillan.
Greenwood, J. and B. Smith.1996. "Financial Markets in Development, and the Development of Financial Markets". *Journal of Economic Dynamics and Control*.21 (1997) 145-181.
Gurley, J. & E. Shaw. 1955. "Financial Aspects of Economic Development," *American Economic Review*, Vol. 45, 515-37.
Holmstrom, B., and J. Tirole.. 1993. "Market Liquidity and Performance Monitoring*." Journal of Political Economy*, Vol. 101: 678-709.
International Finance Corporation (IFC). 2012. Emerging Stock Markets (IFC Washington D.C.).
Keynes, J.M. 1930. *A Treatise on Money*. London: Macmillan.
Khan, M. S. and S.S. Abdelhak. 2000. "Financial Development and Economic Growth: An Overview." IMF Working Papers (00/209).
Khan, M. S. 2008. "Financial Growth and Economic Development in Pakistan." IMF Working Papers (00/209).
Kim, P. 2003. Three Essays on Financial Development and Economic Growth, Dissertation. Ohio State University.
Kim, P. 2007. "The effects of Financial Sector Development on Innovation as an Engine of Sustained Growth", *Seoul Journal of Economics* Vol. 20, No 1 2007.
Levine, Ross and Sara, Zervos. 1998. "Stock Markets, Banks, and Economic Growth", *American Economic Review*, 88: 537-558.
Levine, Ross. 1991. "Stock Markets, Growth, and Tax Policy," *Journal of Finance*.
Levine, Ross. 2005. "Finance and Growth: Theory and Evidence." Handbook of Economic Growth. :Philippe Aghion & Steven Durlauf (ed.), *Handbook of Economic Growth,* edition 1, volume 1, chapter 12, pages 865-934 Elsevier.
Mohtadi, H. and S. Agarwal. 2004. Stock Market Development and Economic Growth: Evidence from Developing Countries, Oxford University Press, New York.
Obstfeld, M. 1994. "Risk-Taking, Global Diversification, and Growth*." American Economic Review*, Vol. 84:1310-29.
Parzen, E. 1962. "On estimation of a probability density function and mode," *Annals of Mathematical Statistics* 33: 1065–1076.
Romer, Paul M. 1988. "Capital Accumulation In The Theory Of Long Run Growth," RCER Working Papers 123. Universityof Rochester - Center for Economic Thought.
Shaw, E.S. 1973. Financial Deepening in Economic Development. Oxford University Press, New York.
Schumpeter, J. 1911. The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest and Business Cycle, Cambridge Mass: Harvard University Press
Stock, James H. Stock, and Mark W. Watson. 2007. *Introduction to Econometrics* 2nd ed., Boston: Pearson Addison Wesley.
Torres-Reyna, Oscar. 2012. Panel Data Analysis, Fixed and Random Effects. http://dss.Priniceotn.edu/training/
Wooldridge, Jeffrey M. 2009. *Introductory Econometrics: A Modern Approach*. Mason, OH: South-Western Cengage.

Market Efficiency and Behavioral Biases in SEC Football: The Over-Under Wager

Joseph Farinella and Clay M. Moffett, University of North Carolina Wilmington

Abstract

One of the most common wagers in sports is the Over-Under wager where the bettor can wager if the total points scored by both teams is greater than (Over) or less than (Under) the total predicted. Gamblers have demonstrated a bias to bet more heavily the Over since they prefer to cheer for points being scored and receive consumption value from this wager. Several studies found betting the under is profitable in: NFL football, Arena football, NBA basketball, European soccer, and college football. These studies find that the total sample is efficient but there are sub-samples of games that are mispriced. We extend these studies by testing for market inefficiencies and profitable gambling strategies in SEC college football games. Specifically, we find that the Over-Under wager for the SEC is not efficient. We also examine results by individual team. For seven of the twelve teams betting the under would have been profitable. The most profitable bets are for Mississippi State and Vanderbilt which would have won 57.95% and 58.82%, respectively. A gambler would earn a 12.29% return on investment over the football season which translates into an annual return of more than 49%.

Introduction

Wagering on sporting events is widespread in the United States and across the world. A study performed by the National Gambling Impact Study Commission (NGISC) in June 1999 found that there is approximately $380 billion in illegal wagers annually in the United States. In 2011, there was $2.88 billion wagered in Nevada alone. This figure represents less than 1% of all sports betting nationwide.[i] The ability of gamblers to consistently profit from sports wagering has been the subject of examination by researchers for decades.

A common wager is to bet the total points scored in game will be above (over) or below a number that is set by the sportsbook. For example, if the number is 50 points then the bettor would bet the over if they believe that the total points scored by both teams will be higher than 50. Conversely, the bettor would bet the under if they believe that the total points scored by both teams will be less than 50.

There have been several studies that have found that bettors prefer the over and the total line is biased. The rationale is that is that bettors prefer to watch games and cheer for scoring thus there is a bias to bet the over. The bettor receives some consumption value and utility from betting on a game and rooting for both teams to score. The average fan does not enjoy defensive struggles when both teams are punting and the score is low. This behavioral bias causes the over to be over bet and the line to be biased.

There have been many studies that have shown that betting the under is a profitable strategy: These studies include the following leagues: the National football league (Paul and Weinbach 2002), the National Basketball Association (Paul, Weinbach and Wilson 2004), European soccer (Paul and Weinbach 2009), Arena football and College football (Paul, and Weinbach 2005). Another study found betting the under in College football is even more profitable for televised games (Paul and Weinbach 2013). In this paper, we examine if the totals bias exists for football teams in the SEC. Our study extends the existing literature on market efficiency and behavioral finance in several important ways. We use a unique time period that has not been previously analyzed. The college football totals wagering is a relatively new market. The primary study of this market was by Paul and Weinbach (2005) who use data from 1999 – 2003. As noted, the authors found a sub-sample where the under was profitable and noted that this mispricing should disappear once the market exists for a longer period. The data for our study ranges from 1985 – 2012 and will allow us to examine if the profitability of betting the under has vanished. Since 2003, the sports betting world has changed. Specifically, there is more online betting and the Southeastern conference has dominated college football.

The past seven national champions have come from the SEC. At the end of the 2012- 2013 season, there were five SEC teams ranked in the top 10. The SEC is also the most profitable conference with the strongest fan following and attendance. The revenue distributed to each school in the 2010 - 11 fiscal year was $220 million.[ii] A large portion of this revenue is generated from TV contracts. In 2008, the SEC announced a 15-year television contract with CBS worth an estimated $55 million a year. The same year, the SEC announced another contract with ESPN worth $2.25 billion or $150 million per year for the life of the contract. It is the longest and wealthiest television contract among all college conferences.[iii] The exposure from these contracts and the partnership with the SEC and ESPN and CBS may create some bettor/viewer biases. This may attract more unsophisticated bettors who bet the over for consumption value rather than investment value. This may result in the total line to continue to demonstrate upward bias.

Literature Review

Paul and Weinbach (2005) examine betting the total in college football and Arena football. The overall totals market for Arena football and college football was found to be a fair bet. However, biases were found in a subsample of games. The results show that betting the under in college football games that had a total of 52 or more, 54 or more and 58 or more points scored results in profitable betting strategies. The results for Arena football were similar. Betting the under in games that had a total number of 100, 102, 104, 106, 108 and 110 or greater all result in profitable betting strategies. The evidence is consistent with the results in the professional football and basketball markets; bettors prefer scoring thus the total lines are biased upward. The authors find that betting the under is a profitable strategy in college football and Arena football.

Paul and Weinbach (2013) examine if the profitability of betting the under in college football games is impacted by the television coverage of the game. The authors maintain that the unsophisticated bettor who prefers the over, since they enjoy cheering for points, is more likely to bet games that they are able to watch on television. Accordingly, televised games are more likely to have total lines that are biased upwards. The results from their study support their hypothesis. Specifically, they find that betting the under would be profitable for games that were televised nationally. Betting the under wins 60% of the time for nationally televised games.

The traditional model of sportsbook behavior was discussed by Pankoff (1968) and Gandar et al. (1988); in this model the sportsbook sets a line so there are equal amounts bet on each side of the wager. This allows the sportsbook to eliminate risk and lock in a 10% profit with the profit based solely on the volume of bets. If the unsophisticated bettor had a preference for the over then the line would increase till the volume of bets was equal. This would result in a total number that is biased upward and a profitable strategy of betting the under.

Levitt (2008) provided an alternative model of sportsbook behavior. He argued that sportsbooks do not necessarily need to balance the books but rather prefer to maximize profits. In this model a sportsbook would seek to profit on betting inefficiencies by the general public. The sportsbook is considered the most knowledgeable participant in the market and knows the fair line. If the public has a tendency to bet more heavily on the over then the sportsbook would not adjust the line to balance the volume on each side. This allows the volume to be higher on the overs, with the sportsbook having confidence that the betting public is consistently wrong. The sportsbook would profit from effectively having a position on the under. Both of these models are consistent with the line being biased upward if the unsophisticated bettor has a preference for the over.

In finance, mispricing or market inefficiency disappears when there is an arbitrageur who exploits the mispricing and returns the market to efficiency. However, the sportsbook can make it difficult for the arbitrageur (sophisticated bettor) to exploit this mispricing by placing limits on the size of wagers. This is a common practice for most sportsbooks. For example, Pinnacle.com had limits of $20,000 on totals bets in college football while other sites such as Sportsbook.com set limits of $2,200 on totals for college football. The limits, logistics and transaction costs in this market allow this mispricing to persist.

Methodology

The sports gambling market is considered to be efficient if there are no strategies that can consistently generate a profit. If the betting line is efficient then the bet would win 50% of the time and lose 50% of the time. The casino makes a profit by charging the losers a 10% fee. A gambler must wager $11 to win $10. Accordingly, a gambler must win more than 52.4% of their bets in order make a profit and cover the fee.

We calculate the likelihood ratio test proposed by Even and Noble (1992) to test if the market is efficient. This model allows us to test a betting strategy against the null hypothesis of a fair bet and no profitability. The unrestricted log likelihood function takes the form:

$$L^u = n [\ln (q)] + (N - n) \ln (1-q) \tag{1}$$

Where N is the total number of observations, n is the number of observations where the score is less than the posted line and q is the observed proportion of observations where the score is less than the posted line. An efficient market implies that q = .5. This creates the restricted log likelihood function L^r, which is obtained by setting q = 0.5. The likelihood ratio statistic for the null hypothesis that q = 0.5 is:

$$2(L^u - L^r) = 2 \{n [\ln (q) - \ln (0.5)] + (N - n) [\ln (1-q) - \ln (0.5)]\} \tag{2}$$

To test for a profitability given a 10% vigorish, the test changes to:

$$2(L^u - L^r) = 2 \{n [\ln (q) – \ln (0.524)] + (N – n) [\ln (1-q) – \ln (0.476)]\} \quad (3)$$

Empirical Results

The data is supplied by Covers (www.covers.com), a sports betting site that provides historical information on each team in the SEC. The site provides the betting line, score of game, and over/under for regular season and bowl games. The sample will be the twelve teams that were in the SEC since the last expansion in 1992. In 2012, the league expanded to include 14 teams (adding Texas A&M and Missouri). The following is a list of the teams examined in this study: SEC West: Alabama, Auburn, Arkansas, LSU, Mississippi, and the University of Mississippi; SEC East: Georgia, Florida, Kentucky, South Carolina, Tennessee and Vanderbilt.

Table 1 gives the results of betting on each team and the entire SEC.

Table 1: SEC Football Betting Simulations (Sub-samples are determined by Team)

Sec West	**Total Games**	**Total Bets**	**Under**	**Over**	**winpct**	**Log Likelihood: Fair Bet**		**Log Likelihood: Profitability**
Alabama	345	243	135	108	55.56	3.0062	*	0.9851
Arkansas	331	187	89	98	47.59	0.4333		1.7149
Auburn	333	219	116	103	52.97	0.7721		0.0304
LSU	337	219	114	105	52.05	0.3700		0.0093
Miss	326	172	94	78	54.65	1.4905		0.3563
Miss St	321	176	102	74	57.95	4.4735	**	2.2051
SEC East								
Florida	342	227	124	103	54.63	1.9455		0.4597
Georgia	329	224	120	104	53.57	1.1438		0.1276
Kentucky	319	168	84	84	50.00	0.0000		0.3811
South Car	205	150	77	73	51.33	0.1067		0.0658
Tenn	342	233	118	115	50.64	0.0386		0.2813
Vandy	314	136	80	56	58.82	4.2576	**	2.2795
All Teams	3844	2354	1253	1101	53.23	9.8216	***	0.6799

Notes: The log likelihood test statistic have a chi-square distribution with one degree of freedom. Critical Values are 2.706 (for an α = 0.10), 3.841 (for α = 0.05), and 6.635 (for α = 0.01). * indicates a significance at the 10 percent level of confidence, ** indicates significance at the 5 percent level and *** indicates significance at the 1 percent level.

A wager on the total points is a bet that the total score will be over or under the posted line amount. The first column shows the total number of games over the entire period for each team. The second column shows the total number of bets which means the sportsbook provided a total number and accepted these bets. It is noteworthy that over our entire sample is from 1985 – 2012 and has a total of 3,844 games. The sportsbook accepted an over/under bet for 2,079 of these games or 54% of the total games. The over/under bet in college football has grown in popularity since 2002 and the bet is available for most teams since that time. Several exceptions from the fair bet are noted among individual teams as well as the SEC as a whole. There are seven teams that have winning percentages above 52.4% meaning that wagering on the under for these teams would be profitable. Betting the under on all SEC games would be profitable and result in a winning percentage of 53.23%. Vanderbilt and MSU had winning percentages of 58.82% and 57.95%, respectively. A 58.82% winning percentages is a return of investment of 12.29% and an annual return of 49%. The likelihood statistic indicates that the lines on Alabama, MSU and Vanderbilt are not fair bets but the profitability of these bets is not statistically significant. The relatively small sample size for some of these categories contributes to their lack of statistical significance. There clearly is an upward bias on the total lines for each team in the SEC and the league as a whole. The lone exception was Arkansas with a 52.41% winning over.

Our results are consistent with Hirshleifer (2001) who notes that investors are subject to behavioral biases. He notes that the over is systematically over-bet and the contrarian strategy yields profitable bets. Paul and Weinbach (2005) find that betting the under on all college football games wins 51.7% thus was not profitable. Our results show that a betting the under for only SEC games results in a winning percentage of 53.2% and generated a profit. This may be due to some of the increased publicity the SEC receives via the TV contracts and a great sense of enthusiasm/overvaluation. Paul and Weinbach (2013) find that in nationally televised college football games, betting the under wins 60% of the time. The profitability of betting the under in the SEC may reflect some tribute to stronger SEC defenses, a trait upon which it prides itself.

Paul and Weinbach (2005) find that the magnitude of the total number has an impact on the profitability of the under wager. We replicate their selection process and divide the sample into groups based on the total number being greater than or equal to 52, 54, 56, 58, 60 and 62 points, respectively. The results are reported in Table 2.

Table 2: SEC Football Betting Simulations (Sub-samples are determined by Total Points)

Bet Under with Total Great Than or Equal to:	**Under**	**Over**	**Under Win Percentage**	**Null Hypothesis: Fair Bet**		**Null Hypothesis: No Profit**
ALL	1253	1101	0.5323	9.8216	***	0.6799
52	357	286	0.5552	7.8558	***	2.5500
54	261	209	0.5553	5.7650	**	1.8770
56	176	139	0.5587	4.3561	**	1.5455
58	112	94	0.5437	1.5748		0.3272
60	65	59	0.5242	0.2904		0.0001
62	43	35	0.5513	0.8220		0.2367

Notes: The log likelihood test statistic have a chi-square distribution with one degree of freedom. Critical Values are 2.706 (for an α = 0.10), 3.841 (for α = 0.05), and 6.635 (for α = 0.01). * indicates a significance at the 10 percent level of confidence, ** indicates significance at the 5 percent level and *** indicates significance at the 1 percent level.

The winning percentage from betting the under exceeds 52.4% and is profitable for all of the categories examined. The highest winning percentage is for games when the total number is equal to or exceeds 56. Betting the under for these games has a winning percentage of 55.87%. This implies into a return on investment of 6.67% and an annual return of 26.64%. The log-likelihood tests show that for the games with total numbers equal to or greater than 52, 54, and 56 are not a fair bet.

Conclusion

The profitability of betting the under has been documented in many sports including: European soccer, Arena football, college football, NBA, and NFL and college football. In general, the entire sample is efficient but sub-samples have been identified that are mispriced. The reason this mispricing exists is that gamblers receive consumption value from betting the over and market microstructure prevents arbitrageurs from exploiting the mispricing. We examine a sub-sample of college football games specifically SEC games. We believe these games may be more likely to be mispriced given the media exposure and nature of the conference. Our results support this notion. We find that betting the under is profitable for seven of the 12 teams in the SEC. The most profitable strategy is to be the under on all Vanderbilt games which results in a winning percentage of 58.82%. This is a return on investment of 12.29% over the football season and an annual return of 49%.

Notes

[i] See American Gaming Association's web site: www.americangaming.org/industry-resources/research/fact-sheets/sports-wagering
[ii] See www.secdigitalnetwork.com
[iii] See www.secsports.com

Acknowledgement

We would like to thank the University of North Carolina Wilmington for a summer research grant. We also want to thank Rodney Paul and Andrew Weinbach for their comments and willingness to provide us with assistance with the statistical analysis.

References

Gandar, J., Zuber, R. O'Brien, T. and Russo, B. (1988) "Testing market rationality in the point spread betting market", *Journal of Finance*, 43 995-1007.

Even, W. E., and N. R. Noble. (1992) "Testing Efficiency in Gambling Markets." Applied Economics 24: 85-88.

Hirshleifer, D. (2001). "Investor psychology and asset pricing." *Journal of Finance* 56: 1533-1597.

Levitt, S.D. (2004) "Why are gambling markets organized so differently from financial markets?" *Economic Journal*, 114, 223 – 246.

Pankoff, L. (1968) "Market efficiency in football betting", Journal of Business, 41, 203-14.

Paul, R.J., Weinbach, A.P. (2002) "Market efficiency and profitable betting rule: Evidence from totals on professional football', *Journal of Sports Economics*, 3, 256 – 263.

Paul, R.J., Weinbach, A.P. and Wilson D. M. (2004) "Efficient markets, fair bets, and profitability in NBA totals 1995 – 96 to 2001 – 02", *Quarterly Review of Economics and Finance,* 44 (4), 624 – 32.

Paul, R.J., Weinbach, A.P. (2005) "Bettor preferences and market efficiency in football totals markets", *Journal of Economics and Finance*, 29, 409 – 16.

Paul, R.J., Weinbach, A.P. (2009) "Are behavioral biases consistent across the Atlantic? The over/under market for European soccer", *The Journal of Gambling Business and Economics*, 3, 89 – 101.

Paul, R.J., Weinbach, A.P. (2013) "National television coverage and the behavioral bias of bettors: The American college football totals market," Working paper.

Elasticities of Demand for Medical Tourism

James R. Frederick, and Lydia L. Gan, University of North Carolina at Pembroke

Abstract

This paper examines the behavior of consumers from high-income countries who are thinking about traveling abroad for medical care – a phenomenon known as medical tourism. To our knowledge, this is the first paper to estimate demand elasticities for medical tourism. Using a survey of 597 consumers from North Carolina (USA), we estimate price and income elasticities of demand for five types of surgery by means of seemingly unrelated regression with feasible generalized least squares. Price elasticity estimates varied from -0.525 to -0.161. Older consumers had more elastic demand. Also, the income elasticity of demand was significantly negative for all treatments.

Introduction

Recently, as U.S. employers have sought ways to reduce costs in the face of increasing global competition, growing numbers of uninsured and under-insured Americans have been seeking medical care outside the United States. Treatments that are necessary but not urgent, such as hip replacements and some heart surgery, can be done abroad. This phenomenon of traveling to receive medical care has been called medical tourism. Countries such as Mexico, Costa Rica, Thailand, and India where labor costs are significantly lower, and where the costs of medical procedures are also correspondingly lower, have developed well staffed and well equipped modern hospitals to attract medical tourists from the United States, Europe, Japan, and other countries. The prices in these countries can easily be 60% to 90% less than they would be in the U.S. or Europe (Gan and Frederick, 2011).

Little is known about the behavior of medical tourists beyond the fact that the price differential is a major reason for the recent wave of medical tourism. How sensitive consumers are to changes in the price of medical tourism relative to U.S. health care prices is unknown. Nor do we know which types of consumers are more price sensitive than others.

The price elasticity of demand for a good or service measures consumers' sensitivity to a change in the price of the good or service. If the demand for medical tourism proves to be quite elastic, changes in exchange rates will have a significant effect on medical tourism. The depreciation of the U.S. dollar will reduce the demand for medical tourism, but if the demand is inelastic the United States will pay more in total for foreign health care.

In this study, "the demand for medical tourism" refers to consumers' desires to receive medical care outside of their home country – the demand to import medical services by traveling to the supplier's country. Furthermore, this study takes the perspective of the consumers in a developed country who seek medical care abroad.

Determinants of the Price Elasticity of Demand

According to standard demand theory, the elasticity of demand for a good or service with respect to its price will depend on various factors which include the availability of substitutes, the share of income devoted to the good, the portion of the price which is paid by third parties, and the prices of complements in consumption (Besanko and Braeutigam, 2011, pp. 49-51).

Substitutes. The availability of acceptable substitutes will increase the price elasticity of demand. The greater the cross-price elasticity, the more negative the own-price elasticity must be, *ceteris paribus*. When a consumer is considering a certain medical tourism, the consumer's choices may be summarized as

- treatment in the consumer's home country
- substitute treatments in the home country
- treatment abroad
- substitute treatments abroad
- doing without treatment

If the treatment sought is not medically necessary, such as preventive care or cosmetic procedures, then doing without treatment may be a reasonable choice. If however the treatment is medically necessary, the last option is not acceptable and the consumer's price elasticity of demand will be reduced. Manning, et al. (1987) noted different elasticities for different types of treatment. They found hospital care's elasticity of demand to be −0.14, while for less medically necessary preventative care it was −0.43. More recently, Meyerhoefer and Zuvekas (2010) using panel data estimated the price elasticity for physical medical care to be −0.12.

One of the reasons for engaging in medical tourism is to receive treatment that is not available in the consumer's home country (Brady, 2007; Demicco and Cetron, 2006). In this case, the first option on the list above would not be available, and again the consumer's price elasticity of demand would be reduced.

The degree of substitution between treatment abroad and domestic treatment may be affected by a consumer's perceptions of the quality of foreign medical care relative to that of domestic medical care. Repeat purchases of medical procedures are rare whether the procedures are done domestically or abroad, so the consumer's perceptions of quality are subject to uncertainty. Nonetheless, a consumer may have more accurate perceptions of the quality of medical care in the consumer's home country than of that in a foreign country. Risk aversion in the face of this asymmetry of information may cause consumers in developed countries to have a bias toward domestic medical care and to have less substitutability between foreign and domestic treatment. Hence, the more important this risk aversion is to a consumer, the lower the price elasticity of demand for medical tourism. Risk aversion is expected to be more important when the implications of medical complications are more severe, as with heart surgery.

Share of Budget. As the fraction of the consumer's income that is devoted to the service increases, the price elasticity of demand decreases.

Third-Party Payments. If the consumer does not bear the full cost of receiving treatment, perhaps because of health insurance benefits, the consumer will be less sensitive to changes in the price and the consumers' demand will be less elastic. Some domestic health insurance plans cover foreign medical care, but most do not. This would tilt the price ratio in favor of domestic treatment, bringing the domestic price closer to the foreign price, thus making domestic treatment a closer substitute to medical tourism. So, although consumers with adequate health insurance are less likely to use medical tourism, we expect the price elasticity of demand for medical tourism to be greater among consumers who have health insurance than among those who do not.

Complements in Consumption. When a consumer decides to get treatment, he or she not only commits to the explicit costs of medical care and travel, but to the implicit cost of forgone income while receiving treatment. If the consumer is unable to work because of his or her medical condition, then the cost of income lost while waiting for treatment must also be added. The forgone income from medical tourism may differ from the forgone income from treatment in the home country. Often, the patient is able to get treatment more quickly abroad, thereby reducing the cost of lost income. On the other hand, the costs of forgone income associated with medial tourism will be increased by the additional travel time and by the fact that doctors in a foreign country may want a patient to spend more time in recuperation due to the medical risks associated with airline travel and to the difficulty of arranging follow-up visits when the patient returns home. Thus, a 20% change in the explicit costs would cause a smaller percentage change in the overall cost of medical tourism, resulting in a less elastic demand. Thus, the larger the explicit costs are in proportion to the cost of forgone income, the greater the elasticity of demand with respect to the price of the explicit costs.

To date, there have been no attempts to estimate the price elasticity of demand for medical tourism, however there have been several studies that have estimated the price elasticity of demand for health care in general and for specific types of medical care done domestically. These studies may shed some light on the elasticity of demand for medical tourism, though the demand for medical tourism would be expected to be more elastic than the demand for health care in general, given the discussion above. Manning, et al. (1987) in the oft-cited RAND Health Insurance Experiment (HIE) estimated the overall elasticity of demand for medical care in six U.S. cities to be −0.22. Eichner (1998), analyzing patients' insurance claims data before and after their deductibles were met, found price elasticities between −0.57 and −0.79. Connolly, et al. (2009) found the elasticities of demand for two fertility procedures to be −0.41 and −0.34 in Germany.

Several consumer characteristics may have an effect on the price elasticity of demand for a good or service.

Income. It is often suggested that consumers' incomes are apt to be negatively related to consumers' responsiveness to a change in the price. The reasoning behind this is that a given change in the price of a good would represent a smaller portion of high-income consumers' incomes than of low-income consumers' income, and that they would therefore be more able to accommodate the price change. For example, Mocan, et al. (2004) found that the price elasticity of household spending in urban China was greater for poor households than for rich households. Similarly, Renwick and Green (2000), studying California, and Mieno and Braden (2011), studying metropolitan Chicago, found that the price elasticity of demand for water decreased significantly as household income increased.

In contrast, Duarte (2011) examined how the price elasticities of demand for a few types of health care in Chile depended on demographic variables, including income. He found that for doctor visits to the home and for psychologists' services, when the intensity of health care per visit is accounted for, the price elasticity increased significantly as household income increased. This effect was not found when he studied the demand for physical therapy evaluations. So, on the basis of prior empirical work, the effect of income on the price elasticity of demand is inconclusive.

In the case of medical tourism, two additional factors might cause income to have a positive effect on the price elasticity of demand: First, high-income consumers may be better educated and more confident of their ability to cope with difficulties that might arise, and thus they may be more receptive to the idea of traveling to different cultures for medical care. Second, high-income consumers may be more able to afford the higher cost of medical care in their home countries, making domestic treatment a more feasible alternative to medical tourism for them than it would be for low-income consumers. If so, then they would be more sensitive to changes in the foreign price of using medical tourism than would low-income consumers.

Education. Madden (2007) found that education had an effect on the price elasticity of demand to smoke among Irish women, however he found that the effect was not uniformly positive. Women with intermediate levels of education were found to be more sensitive to a tax-induced increase in the price of cigarettes than women with either lower or higher levels of education.

Age. In his study of Chile, Duarte (2011) found that age had significant effects on the price elasticity of the demand for some forms of medical care. Although there was no noticeable age-related trend for home visits, there was a negative relationship between age and the price elasticity of demand for psychiatrists' services and for physical therapy evaluations.

A consumer's age may have contradictory effects on the elasticity of demand to use medical tourism. Retired consumers would not have the same opportunity costs of time than an employed younger worker would have, presumably making the retired consumer more willing to use medical tourism as an alternative to domestic treatment. Among working-aged consumers, the time cost of traveling abroad for treatment would be greatest for those in their 50s or early 60s when their incomes are the highest. Because the explicit costs (medical, travel, and accommodations) of medical tourism are only part of the total cost, this effect implies that the price elasticity of demand should be the greatest among the older, retired consumers. However, when traveling for medical tourism requires a long flight, it can be stressful and can involve physical discomfort. Younger medical tourists are more able to cope with the hardships of extended travel better than older travelers. Thus, this effect implies that younger consumers would be more likely to consider medical tourism an acceptable substitute for domestic health care, and that they are therefore likely to have more elastic demand for medical tourism. The combined effect of these two effects is ambiguous, but we do expect the elasticity of demand to vary by age group.

Data

After obtaining IRB approval, a pilot questionnaire was developed early in 2010. A revised questionnaire was administered to consumers in various North Carolina communities by street intercept in June of 2011. In all, 597 responses were received. Respondents were required to be 18 years of age or older. To assess the respondent's sensitivity to prices, there were several questions similar to "Assuming that you need a knee replacement ($45,000 in the U.S.), how likely is it that you would get it done abroad if the overall cost (treatment, hospital, travel, and accommodations) is . . . " This was followed by two or three prices, and for each price there was a five-point Likert scale of likelihoods labeled "No Chance", "25%", "50%", "75%", and "Definitely would". Respondents were instructed to assume that quality of care would be the same in the U.S. and abroad. Similar questions were asked about tummy tucks, non-urgent coronary artery bypass graft (CABG), dental implants, and stem cell therapy for diabetes. The latter is not available in the U.S., so no U.S. price was given for reference. Respondents were also asked a series of questions about their gender, age, health status, ethnicity, foreign language proficiency, education, marital status, health insurance status, and household incomes.

Missing data on consumer characteristics were replaced by their respective modes, except that missing income and education data were imputed from the available consumer characteristics by multinomial regression. When a respondent answered the likelihood question for some, but not all, price levels of a treatment, the likelihood information for the nearest price level was used. In order to calculate the logarithms of the likelihoods, whenever a respondent said that likelihood of using medical tourism was zero the natural logarithm of 1% was used. The midpoint of an income bracket was used to represent all incomes in that bracket, except for the last bracket ($150,000 or more). The income level of $250,000 was arbitrarily used to represent incomes in that bracket.

Typically, a price elasticity of demand is the percentage change in the quantity demanded divided by the percentage change in the price. For the quantity demanded, this study uses the self-reported likelihood that a consumer will go abroad for treatment. The prices used are the suggested foreign prices that were given with each likelihood in the questionnaire. Respondents were instructed to assume that these prices included the costs of travel and accommodations as well as the cost of treatment. Since the questionnaire asked, for example, about the consumer's likelihood of traveling for a knee replacement at three suggested foreign prices, there are three price-quantity pairs in the data set for each consumer related to the demand for a knee replacement. The CABG and stem-cell questions also used three prices, while the tummy tuck and dental implant questions used only two prices because the latter are low-priced treatments.

Methods

To estimate the price and income elasticities, the following demand equation was estimated:

$$\ln(L_{tpi}) = \beta_{0t} + \beta_{1t}\ln(Price_{tpi}) + \beta_{2t}\ln(Income_i) + \beta_{3t}\ln(Price_{tpi})\cdot\ln(Income_i) + \varepsilon_{tpi} \quad (1)$$

L is the consumer's reported likelihood of using medical tourism for a given treatment. The first subscript, t, represents a treatment (t = knee replacement, tummy tuck, CABG, dental implants, or stem-cell therapy). The second subscript, p, represents a price level (p = 1 or 2 for tummy tuck and dental implants, and p = 1, 2, or 3 for knee replacement, CABG, or stem-cell therapy). The third subscript, i, represents a respondent (i = 1 ... 597). To interpret the interaction terms, equation (1) may be rewritten as

$$\ln(L_{tpi}) = \beta_{0t} + [\beta_{1t} + \beta_{3t}\ln(Income_i)]\cdot\ln(Price_{tpi}) + \beta_{2t}\ln(Income_i) + \varepsilon_{tpi} \quad (2)$$

Thus, if the coefficient β_{3t} is found to be statistically significant, then Income has an effect on the price elasticity of demand for treatment t. To assess the effects of other demographic variables on the price elasticity of demand, this regression model was modified as needed.

Estimation

Since the prices were posed to the respondents, rather than market-determined prices, they are exogenous. Hence, estimation of these equations by ordinary least squares would yield unbiased and consistent results. However, some unobserved consumer characteristics, such as familiarity with foreign languages, general health status, physical ability to travel, and perceptions of the quality of foreign health care, may make some consumers more favorably predisposed than others toward medical tourism in general. These individual differences were likely to have affected each of the consumers' responses for multiple treatments and multiple price points. So,

$$\begin{gathered} E(\varepsilon_{tpi}\cdot\varepsilon_{sqj}) = 0, \text{ when } i \neq j \\ E(\varepsilon_{tpi}\cdot\varepsilon_{sqi}) = \sigma_{tpsq} > 0 \end{gathered} \quad (3)$$

where s, q, and j are alternative values of the indices t, p, and i, respectively. Therefore, OLS estimates are inefficient and the OLS variance estimates are biased (Greene, 1990, pp. 382-83). To correct these problems, and to allow comparisons across equations, the five treatment equations were estimated as a seemingly unrelated regressions model by feasible generalized least squares (FGLS) (Davidson and MacKinnon, 2004, pp. 501-11) by using the model

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \quad (4)$$

where $\boldsymbol{y}$ is a 7761x1 vector of the logarithms of the respondents' reported likelihoods, and $\boldsymbol{X}$ is a 7761x20 block diagonal matrix containing columns of ones, the logarithms of the prices and the logarithms of the incomes. The twenty coefficients of $\boldsymbol{\beta}$ are the four coefficients in equation (1) repeated for each of the five treatments. The 7761 rows of $\boldsymbol{X}$ represent three prices for each of the 597 consumers for knee replacements, CABG, and stem-cell therapy and two prices for each of the 597 consumers for tummy tucks and dental implants. For all equations and all iterations, the elements of the variances and covariances in equation (3) were found to be positive, as predicted. The variances and covariances were used to construct an estimated covariance matrix for $\boldsymbol{\varepsilon}$, **S**. The feasible generalized least squares estimator of $\boldsymbol{\beta}$ and its covariance matrix are then

$$\begin{gathered} \boldsymbol{b} = (\boldsymbol{X}'\mathbf{S}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\mathbf{S}^{-1}\boldsymbol{y} \\ \boldsymbol{V} = (\boldsymbol{X}'\mathbf{S}^{-1}\boldsymbol{X})^{-1} \end{gathered} \quad (5)$$

The vector $\boldsymbol{b}$ has 20 elements. The differences between price elasticities for different treatments, $b_{1t} - b_{1s}$, and between the income elasticities, $b_{2t} - b_{2s}$, were tested by t tests. The standard errors for these differences were constructed from the elements of $\boldsymbol{V}$.

Results

Price Elasticity of Demand. On the basis of the previous literature, one might expect that the demand for medical tourism would be inelastic. The regression results of equation (1), shown in Table 1, confirm this. When interaction terms were included in the regression, none of the interaction terms was significant at the 5% level. They were therefore dropped from the model. When equation (1) was regressed without the interaction terms, the price elasticity coefficients were significantly less than zero at the 0.1% level for all five treatments. Furthermore, each of the price elasticity estimates was significantly smaller than 1.000 in absolute value, indicating that the demand is inelastic. These estimates were consistent with other health care researchers' price elasticity estimates in that the demand for each treatment was found to be inelastic. For the most part, the demand to use medical tourism was more elastic than Manning's (1987) or Meyerhoefer and Zuvekas' (2010) estimates of the demand or medical care. The exception was stem-cell therapy, which with an elasticity of –0.161 was similar to these previous estimates.

Table 1: Effects of Income on the Price Elasticity of Demand

	β_{0t} const.	β_{1t} ln(*Price*)	β_{2t} ln(*Income*)	β_{3t} ln(*Price*) · ln(*Income*)	$\beta_{1t} + \beta_{3t} \cdot$ ln(*Income*)
Knee replacement	0.562	−0.370	−0.150	−0.012	−0.493
	(0.181)	(−0.406)	(−0.515)	(−0.139)	
Tummy tuck	0.560	−0.711	−0.283	0.033	−0.362
	(0.325)	(−0.741)	(−1.753)	(0.365)	
CABG	3.990	−1.282*	−0.495*	0.089	−0.337
	(1.509)	(−2.049)	(−1.998)	(1.515)	
Dental implants	−0.246	−0.598	−0.221*	0.021	−0.379
	(−0.235)	(−0.634)	(−2.262)	(0.233)	
Stem cell therapy	2.796	−0.850*	−0.443**	0.065	−0.159
	(1.934)	(−1.988)	(−3.271)	(1.620)	
Knee replacement	1.082	−0.525***	−0.188*		
	(1.164)	(−7.362)	(−2.233)		
Tummy tuck	0.174	−0.373***	−0.244**		
	(0.180)	(−5.009)	(−2.720)		
CABG	0.466	−0.344***	−0.159*		
	(0.483)	(−7.055)	(−1.802)		
Dental implants	−0.210	−0.389***	−0.223**		
	(−0.213)	(−5.314)	(−2.418)		
Stem cell therapy	1.104	−0.161***	−0.282***		
	(1.162)	(−4.833)	(−3.180)		

The last column in the top half gives estimates of the price elasticity evaluated at the mean of ln(*Income*), which in this sample was 10.647. *t* statistics are shown in the parentheses; *** one-tailed $p < 0.001$, ** one-tailed $p < 0.01$, * one-tailed $p < 0.05$.

Table 1 also shows that the income elasticities of demand for medical tourism are negative for all five treatments and regardless of whether the interaction terms are used.

Type of treatment. The demand for elective or cosmetic surgeries, such as tummy tucks, might be expected to be more price elastic than that of medically necessary treatments. This was not the case. Table 2 shows the *t* statistics for tests of the null hypothesis $\beta_{1t} - \beta_{1s} = 0$, where t and s represent different treatments. The estimates of β_{1t} were taken from Table 1. It is evident that the difference between the price elasticities for tummy tucks (–0.373) and for CABGs (–0.344) is not significantly different from zero. Knee replacements had the most elastic demand (–0.525) and stem-cell therapy had the least elastic demand (–0.161).

The fact that the demand for stem-cell therapy abroad was significantly less price elastic than for the other treatments in the study is not surprising since stem-cell therapy for diabetes represented treatment that is not available in the United States (Table 2).

Table 2: *t* Tests of Equality of Price Elasticities across Treatments

	Knee Replacements β_{1t} = -0.525	Tummy Tucks β_{1t} = -0.373	CABGs β_{1t} = -0.344	Dental Implants β_{1t} = -0.389	Stem Cells β_{1t} = -0.161
Knee Replacements	0				
β_{1t} = -0.525					
Tummy Tucks	0.151|*	0			
β_{1t} = -0.373	(2.111)				
CABGs	0.181**	0.030	0		
β_{1t} = -0.344	(2.790)	(0.436)			
Dental Implants	0.153	-0.016	-0.046	0	
β_{1t} = -0.389	(1.565)	(-0.186)	(-0.606)		
Stem Cells	0.364***	0.212**	0.183***	0.229**	0
β_{1t} = -0.161	(5.654)	(3.059)	(4.422)	(3.214)	

The numbers are the differences ($b_{1\ row} - b_{1\ col}$). The numbers in parentheses are the *t* statistics for H_0: $\beta_{1\ row} - \beta_{1\ col} = 0$
*** one-tailed p < 0.001, ** one-tailed p < 0.01, * one-tailed p < 0.05.

Education. Education appears to make the demand for some treatments more price elastic. To assess the effect of education, the income variable in equation (1) was replaced by a dummy variable which had the value of 1 if was a master's degree or higher. A price-bachelor's degree interaction term was included in preliminary regressions, not reported, but was removed due to multicollinearity. These regressions also included the logarithm of income to control for the indirect effect education through income on the willingness to use medical tourism (Table 3). When a consumer has a master's degree, the estimate of the elasticity of demand for stem-cell therapy increases by –0.225 from −0.130 to −0.355 (p = 0.0242); having a master's degree made a consumer's demand for stem-cell therapy more elastic. This effect appears to be true for the most of other treatments, but the effects were not large enough to be significant. Since this effect was detected in only one of the five treatments, and since the p value was 2.42%, this finding should be interpreted with caution, due to the multiple comparisons problem.

Table 3: Effects of education and income on the price elasticity of demand

	β_{0t} const	β_{1t} ln(*Price*)	β_{2t} ln(*Income*)	β_{3t} *BA*	β_{4t} *MA*	β_{5t} ln(*Price*) · *MA*	$\beta_{1t} + \beta_{5t} \cdot \overline{MA}$
Knee replacement	1.398	−0.464***	−0.249**	0.377*	1.178	−0.262	−0.479
	(1.455)	(−6.145)	(−2.820)	(2.117)	(1.618)	(−1.230)	
Tummy tuck	0.603	−0.332***	−0.302**	0.419*	0.587	−0.239	−0.362
	(0.605)	(−4.177)	(−3.210)	(2.209)	(1.432)	(−1.066)	
CABG	0.278	−0.302***	−0.161	0.164	0.977	−0.269	−0.336
	(0.278)	(−5.825)	(−1.732)	(0.874)	(1.567)	(−1.841)	
Dental implants	−0.233	−0.394***	−0.222*	0.048	−0.093	0.121	−0.379
	(−0.227)	(−5.033)	(−2.284)	(0.245)	(−0.360)	(0.547)	
Stem cell therapy	1.678	−0.130***	−0.360***	0.539**	0.961**	−0.225*	−0.159
	(1.718)	(−3.685)	(−3.889)	(2.891)	(2.790)	(−2.255)	

The last column gives the price elasticity evaluated at the mean of the dummy variable *MA*, which was 0.1256 in this sample. *t* statistics are shown in the parentheses; *** one-tailed p < 0.001, ** one-tailed p < 0.01, * one-tailed p < 0.05.

There was some evidence that the more education a consumer has, the more likely the consumer is to use medical tourism. For all five of the treatments, the coefficients on the bachelor's degree variable, β_{3t}, were positive. With one exception, the coefficients on the master's degree variable, β_{4t}, were also positive and larger than β_{3t}, and this difference was large enough to be significant in the case of stem-cell therapy. As a joint test the effect of having a degree against having only a high school diploma or less, the average of the ten coefficients, β_{3t} and β_{4t}, was tested and found to be significantly positive (t = 2.70, p = 0.0071). So, having a university degree increases the likelihood of using medical tourism, but it does not affect the elasticity of demand for medical tourism.

Insurance Coverage. The extent of consumers' health insurance coverage was divided into three categories: uninsured, under-insured (relying on Medicaid, mini-med policies, or individually purchased insurance), or well insured (relying on employer-based insurance, military insurance or TriCare, or Medicare). Regressions with 0-1 dummy variables for being under-insured and well insured suggest, as would be expected, that the better a consumer's health insurance, the less likely the consumer was to consider using medical tourism (Table 4). This effect was statistically significant at the 5% level in the case of dental implants. It is to be expected that the demand for dental implants would be the exception, since most U.S. health plans would not cover dental implants. None of the price-insurance interaction terms was statistically significant in the top half of Table 4.

Table 4: Effects of insurance coverage on the price elasticity of demand

	β_{0t} const	β_{1t} ln(Price)	β_{2t} Under Ins.	β_{3t} Well Ins.	β_{4t} ln(Price) · Under	β_{5t} ln(Price) · Well
Knee replacement	−0.720	−0.501**	0.159	−0.540	−0.032	0.019
	(−1.143)	(−2.710)	(0.196)	(−0.773)	(−0.133)	(0.094)
Tummy tuck	−2.070***	−0.541**	−0.048	−0.593	0.196	0.213
	(−5.907)	(−2.775)	(−0.106)	(−1.525)	(0.783)	(0.986)
CABG	−0.620	−0.372**	0.113	−1.073	−0.058	0.077
	(−1.158)	(−2.927)	(0.164)	(−1.805)	(−0.352)	(0.544)
Dental implants	−2.228***	−0.208	−0.210	−0.519*	−0.093	−0.234
	(−10.497)	(−1.088)	(−0.769)	(−2.204)	(−0.376)	(−1.103)
Stem cell therapy	−1.868***	−0.166	0.194	−0.155	0.060	−0.009
	(−6.312)	(−1.901)	(0.509)	(−0.472)	(0.538)	(−0.092)
Knee replacement	−1.034***	−0.467***			0.043	−0.058
	(−4.290)	(−5.773)			(0.789)	(−1.234)
Tummy tuck	−2.461***	−0.423***			0.233	0.017
	(−18.335)	(−3.711)			(1.944)	(0.166)
CABG	−1.286***	−0.255***			−0.012	−0.121**
	(−6.248)	(−4.257)			(−0.241)	(−2.899)
Dental implants	−2.606***	−0.068			−0.125	−0.442*
	(−32.009)	(−0.386)			(−0.560)	(−2.304)
Stem cell therapy	−1.926***	−0.175**			0.128	−0.016
	(−17.027)	(−2.979)			(1.894)	(−0.282)

The upper half of the table shows the full model, whereas the lower half omits the main effect of the two insurance dummy variables on the likelihood of using medical tourism.
t statistics are shown in the parentheses; *** one-tailed p < 0.001, ** one-tailed p < 0.01, * one-tailed p < 0.05.

Age. To test the effect of age on consumers' price elasticity of demand for medical tourism, dummy variables were created to classify consumers into four age groups: 18 to 30, 31 to 50, 51 to 64, and 65 or older. The 18 to 30 year olds comprised the baseline age group in the regressions (Table 5). Older consumers were less inclined to use medical tourism for dental implants and tummy tucks. Being in the oldest age group significantly increased the price elasticity for all treatments but tummy tucks, as did being in the middle age group for CABGs and stem-cell therapies. Since the price elasticities, β_{1t}, were all negative and the interaction effects were all negative, the implication is that older consumers have more elastic demands than the 18 to 30 year olds. Although the youngest age group was the most inclined to use medical tourism, they had the least elastic demand.

Table 5: Effects of age on the price elasticity of demand

	β_{0t} const	β_{1t} ln(Price)	β_{2t} Age 31-50	β_{3t} Age51-64	β_{4t} Age65+	β_{5t} ln(Price) · Age 31-50	β_{6t} ln(Price) · Age51-64
Knee replacement	−1.446***	−0.288*	0.903	0.334	0.254	−0.300	−0.284
	(−3.624)	(−2.461)	(1.335)	(0.496)	(0.415)	(−1.516)	(−1.397)
Tummy tuck	−2.254***	−0.209	−0.0003	−0.082	−1.729***	−0.242	−0.357
	(−10.186)	(−1.697)	(−0.0007)	(−0.214)	(−5.088)	(−1.157)	(−1.667)
CABG	−1.758***	−0.142	1.013	0.619	0.038	−0.268*	−0.346*
	(−5.158)	(−1.765)	(1.752)	(1.045)	(0.072)	(−1.972)	(−2.480)
Dental implants	−2.339***	−0.242*	−0.188	−0.474*	−1.097***	−0.069	−0.389
	(−17.487)	(−1.998)	(−0.827)	(−2.042)	(−5.342)	(−0.338)	(−1.850)
Stem cell therapy	−2.112***	0.018	0.618	−0.144	−0.138	−0.250**	−0.312***
	(−11.320)	(0.335)	(1.952)	(−0.444)	(−0.480)	(−2.701)	(−3.298)

t statistics are shown in the parentheses; *** one-tailed p < 0.001, ** one-tailed p < 0.01, * one-tailed p < 0

Discussion

Limitations

This study used a short list of treatments for medical tourists. It included only one treatment which is not available in the United States and only one cosmetic procedure. Caution must be used when trying to generalize from stem-cell therapy to all treatments which are not available in the U.S. or from tummy tucks to all cosmetic surgeries.

The respondents for this study were chosen by street intercept. Probability sampling was not employed, so the price and income elasticities might not represent those of the population of North Carolina consumers, nor of all American consumers. Because of multicollinearity in the data set, it was impossible to use large, multifactor equations that might have been able to separate the effects of education, income, health insurance coverage, and age from each other.

This study made no attempt to determine consumers' perceptions of the quality or diversity of medical care in hospitals in destination countries. The instructions on the questionnaire told respondents to think of foreign care as having quality equal to that in the United States. Some respondents might not have heeded that instruction. Alternatively, the respondents may have been considering the difficulties of traveling to a foreign country to receive medical care.

When consumers have difficulty determining quality for themselves, they often assume that the price is a proxy for quality, which would affect the elasticity estimates.

Summary

As far as the authors have been able to determine, no previous study has examined consumers' price elasticity of demand for medical tourism. This study found strong evidence that it is inelastic. Stem-cell therapy had the least price elasticity among the five treatments studied. This was the only treatment in the study that was not available in the United States. Consumers who were over thirty years of age were found to have more elastic demand than those between the ages of 18 and 30 for CABG and stem-cell therapy, and to a lesser degree for knee replacements and dental implants. Consumers who had graduate or professional degrees had more price elastic demand for stem-cell therapy.

This study also found that medical tourism is an inferior service in the technical sense that its income elasticity of demand is negative. Despite the efforts that some hospitals and clinics in destination countries have made to obtain the most recent medical technology, to hire well qualified (often Western trained) medical staff, and to obtain international accreditation, the consumers in this study would prefer to use U.S. hospitals when they can afford to do so. This could be either because their image of medical care in destination countries is of uniformly low quality or because they are uncertain of the quality of such care.

Our results showed that the demand for knee replacements was the most elastic and that the demand for stem-cell therapy was the least elastic. It may be that the elasticity of demand for CABGs was less than that for knee replacements because the consumers in the survey considered the medical risks associated with heart surgery to be much greater than those of knee surgery. It was noted above that consumers treat medical tourism as an inferior service. Thus, it might be that they tended to prefer heart surgery in the United States, regardless of the price of such treatment abroad. This medical risk aversion may have also been an additional factor in explaining why the elasticity of demand for stem-cell therapy was less than that for knee replacements. The fact that stem-cell therapy is not available in the United States may have carried a suggestion that stem-cell therapy is experimental and therefore risky. Tummy tucks and most dental implants are relatively low-cost procedures in comparison to the other three treatments considered in this study. So, the elasticity of demand for tummy tucks and dental implants may have been less than that of knee replacements because they would constitute a smaller portion of the consumer's budget than the other procedures would.

Finding that the demand for medical tourism is inelastic implies that a depreciation of the value of the U.S. dollar will increase the amount of money Americans spend on medical tourism. There is some evidence (Turner, 2007; Lee, 2010) that several of the foreign health care providers charge medical tourists higher prices than they would charge their local patients. This suggests that prices are determined by what medical tourists are willing to pay, rather than by competitive market conditions in the host countries. It this is the case, then the extent of a U.S. dollar depreciation that is passed through in the form of higher dollar prices would be lessened, as would the increase in the number of dollars spent on medical tourism.

Another implication of inelastic demand for medical tourism is that, as the influx of medical tourists adds to the demand for health care in the host countries, the increase in the price of health care in the host countries will be greater than it would be otherwise. If the demand were more elastic, rising prices in the host country would reduce the influx of medical tourists. This increase in host-country prices is exactly what advocates for the poor in the host countries fear. (Vijaya, 2010; Cattaneo, 2009; Hazarika, 2009; Chinai & Goswami, 2007) They argue that when medical resources in the host country are diverted to medical tourism, the poor will be unable to afford basic health care in their own countries.

Acknowledgments

The authors are grateful for funding from the School of Business of the University of North Carolina at Pembroke.

References

Besanko David, and Ronald Braeutigam. 2011. *Microeconomics*, 4th edition. Wiley: New York.

Brady, Christopher J. 2007. "Offshore Gambling: Medical Outsourcing Versus ERISA's Fiduciary Duty Requirement." *Washington Lee Law Review* 64(3): 1073-1114.

Cattaneo, Olivier. 2009. "Trade in Health Services: What's in it for Developing Countries?" *World Bank Policy Research Working Paper*, 5115.

Chinai, Rupa and Rahul Goswami. 2007. "Medical Visas Mark Growth of Indian Medical Tourism." *Bulletin of the World Health Organization* 85(3): 164-165.

Connolly, Mark P., Georg Griesinger, William Ledger, and Maarten J. Postma. 2009. "The Impact of Introducing Patient Co-Payments in Germany on the Use of IVF And ICSI: A Price-Elasticity of Demand Assessment." *Human Reproduction* 24(11): 2796-2800.

Davidson, Russell and James G. MacKinnon. 2004. *Econometric Theory and Methods*. Oxford University Press: New York.

DeMicco, Frederick J. and Marvin Cetron. 2006. "Club Medic." *Asia Pacific Biotech News* 10(10): 527-531.

Duarte, Fabian. 2011. "Price Elasticity of Expenditure Across Health Care Services." *The Selected Works of Fabian Duarte, Rand Corporation* [Online]. Available at: http://works.bepress.com/fabian_duarte/1 [Accessed: 19 December 2012].

Eichner, Matthew. 1998. "The Demand for Medical Care: What People Pay Does Matter." *American Economic Review* 88(2): 117-121

Gan, Lydia L. and James R. Frederick. 2011. "Medical Tourism Facilitators: Patterns of Service Differentiation." *Journal of Vacation Marketing* 17(3): 165-183.

Greene, William H. 1990. *Econometric Analysis*. Macmillan: New York.

Hazarika, Indrajit. 2009. "Medical Tourism: Its Potential Impact on the Health Workforce and Health Systems in India." *Health Policy and Planning* 25(3): 248-251.

Lee, Tae-hoon. 2010. "Foreigners Victims of Inflated Medical Fees." *Korea Times* 19 Feb 2010. Available at: http://www.koreatimes.co.kr/www/news/nation/2010/02/116_61022.html [Accessed: 19 December 2012].

Madden, David. 2007. "Tobacco Taxes and Starting and Quitting Smoking: Does the Effect Differ by Education?" *Applied Economics* 39(5): 613-627. DOI : 10.1080/00036840500447898

Manning, Willard G., Joseph P. Newhouse, Naihua Duan, Emmett B. Keeler, and Arleen Leibowitz. 1987. "Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment." *The American Economic Review* 77(3): 251-277.

Meyerhoefer Chad D. and Samuel H. Zuvekas. 2010. "New Estimates of the Demand for Physical and Mental Health Treatment." *Health Economics* 19(3): 297-315.

Mieno Taro and John B. Braden. 2011. "Residential Demand for Water in the Chicago Metropolitan Area." *Journal of the American Water Resources Association* 47(4): 713-723.

Mocan H. Naci, Erdal Tekin, and Jeffrey S. Zax. 2004. "The Demand for Medical Care in Urban China." *World Development* 32(2): 289-304.

Renwick Mary E. and Richard D. Green. 2000. "Do Residential Water Demand Side Management Policies Measure Up? An Analysis of Eight California Water Agencies." *Journal of Environmental Economics and Management* 40(1): 37-55.

Turner, Leigh. 2007. "First World Health Care at Third World Prices: Globalization, Bioethics, and Medical Tourism." *BioSocieties* 2: 303-325.

Vijaya, Ramya M. 2010. "Medical Tourism: Revenue Generation or International Transfer of Healthcare Problems?" *Journal of Economic Issues* 44(1): 53-69.

White, Halbert. 1980 "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* 48: 817-838.

When IPOs Migrate

Marlin R.H. Jensen, Beverly B. Marshall, and John S. Jahera, Jr., Auburn University

Abstract

We show that IPOs in countries with well-developed capital markets exhibit a home bias when going public. There appears to be an increasing trend of migrating IPOs to destinations other than the U.S. We find that migrating IPOs are more likely to choose the U.S. as their IPO market destination when 1) the U.S. economic freedom index is higher and 2) the IPO is larger. IPOs are more likely to migrate to other markets when 1) the IPO is smaller, 2) in the period following Sarbanes-Oxley (SOX), and 3) if the IPO home and destination country are part of the same regional trading bloc.

Introduction

Over the past 15 years, the number of companies going public has plummeted relative to historical norms. The peak of publicly traded companies occurred in 1997 when there were 7,888 listed firms. By 2011, the number of public companies had decreased by 38 percent in America and 48 percent in Britain's main markets (2012 The Economist). The U.S. Department of Treasury reports the average number of companies going public in the U.S. from 1991 through 1999 was 547 and the average number of companies going public from 2000 through June of 2011 was 192 per year. In addition, they point out the decline in small company IPOs is even greater. The report goes on to examine just venture-backed emerging growth companies to illustrate the extreme drop in smaller IPOs. From 1991 through 2000, approximately 2000 smaller companies went public and from 2001 through 2010, only 477 companies did IPOs. Even the President's Council on Jobs and Competitiveness report in October of 2011 indicated the importance of new public companies by noting that firms under five years old have accounted for about 40 million new jobs in the past three decades. Weild and Kim (2009) estimate the decline in U.S. IPO market has cost as many as 22 million jobs through 2009.

As the world has become more global, access to markets has increased and as was mentioned, the U.S. has seen a decline in the number of IPOs. The objective of this research is to examine country specific factors that may play a role in a firm's decision to move their IPO from their home country to another country. Using data from several sources, we consider the level of economic freedom in a country as well as the relative exchange rate among countries. We find that IPOs that deviate from their domestic market are more likely to choose the U.S. as their IPO market destination when 1) the U.S. economic freedom index is higher and 2) the IPO is larger. There is some evidence, although marginal, that a strong dollar helps. IPOs are more likely to migrate to other markets if 1) they are smaller, 2) the IPO is after the passage of SOX, and 3) if the IPO home country is part of a regional trading bloc with the market chosen (such as being a member of the EU or ASEAN, PIF, etc.).

Literature Review

Luchetti (2011) shows the decline in the number of stocks listed in the U.S. and compares U.S. stock listings to the rest of the world. Part of the decline may be attributed to the fact that U.S. exchanges aren't as competitive due to relatively high listing costs. The main reason however why there is a decline in number of stocks listed is the prolonged slump in U.S. IPOs. He says that at least some U.S. IPOs, that in the past might have turned to a U.S. exchange, list abroad where fees are lower and they don't face such high compliance costs such under Sarbanes-Oxley. Luchetti points to a Dealogic report that indicates that five large corporations gobbled up 134 private U.S. companies in 2010, which is nearly equal to the entire crop of IPOs on the nation's two big exchanges. Jay Ritter, Professor of Finance at the University of Florida says, "Companies going public have generated a lot of jobs and economic growth. The prolonged drought in IPOs in the U.S. raises concerns about whether an important engine of growth in the U. S. economy has come to an end."

When private companies do proceed with IPO plans, at least some look abroad. Besides the compliance costs of being a public corporation, listing fees are also more expensive in the U.S. HaloSource Inc., a water-purification company in Seattle decided to go public on the London Exchange AIM in 2010. Its annual listing fee is about \$8600 whereas Nasdaq fees range from \$27,500 to \$99,500 and NYSE fees range from \$38,000 to \$500,000. Besides the lower fees, the AIM market does not require a minimum number of shares publicly held, bid price, or market capital in order to be listed. Although the AIM market has surpassed the number of IPOs in U.S. markets for several years, more money was raised from IPOs on AIM than NASDAQ for the first time in 2006. Over the past five years, 45 U.S. companies, worth about \$5.2 billion, have gone public on the AIM market. Luchetti (2011) reports that 74 U.S. companies have gone public in foreign countries since 2005 which is a small fraction of the 650 U.S. companies that have gone public on U.S. exchanges. Mary Schapiro, Chair of the SEC, informed the House Committee on Oversight and Government Reform that the SEC is reviewing how the SEC regulates security offerings in both the private and public markets. In private offerings, in which securities are sold to qualified investors or to large institutions, Rule 144A applies. Firms use Rule 144A to avoid the mandatory disclosure rules applicable to public offerings such as Section 404 of the Sarbanes-Oxley Act (SOX). The SEC's survey in 2009 indicated that costs to comply with Section 404 of SOX had declined but that in 2009, the average direct costs of compliance for an average firm was still two million dollars. Scott (2011) points out that these costs may help explain why U.S. companies prefer to remain private. He also noted that in 2010, 79.3 percent of the funds raised in the U.S. by foreign companies in global IPOs were raised through the private Rule 144A market. In the public markets, Stephens (2011) points out that the lack of rebound of our economy can be attributed in part to the lack of small firm IPOs. He indicates the primary reason for the lack of IPOs is the excessive regulation and compliance and subsequent costs that are required of public companies.

IPO Task Force

In response to the declining IPOs and the perceived loss of jobs, the U.S. Treasury Department in March of 2011 formed the Access to Capital Conference to gather information and make recommendations of how to make capital markets more accessible to emerging companies. Emerging companies were defined as any non-reporting issuer with total annual gross revenue of less than one billion dollars. From this conference, the IPO Task Force (2011) was created to examine what caused the IPO decline and to make recommendations to bring the emerging growth companies back to the public markets through IPOs. "The mandate of the IPO Task Force was to examine the root causes of the current U.S. IPO crisis and quickly develop reasonable and actionable steps that can restore access for emerging growth companies to the capital they need to create jobs and expand their businesses globally," (Kate Mitchell, chairman of the IPO Task Force). William Sahlman, professor at the Harvard Business School and IPO Task force member, stated, "As a country, we need to stop debating how to slice a shrinking pie and start working to grow the pie." If the capital market system is to expand, some type of reform needs to be made to grow the economy. The IPO Task Force noted there were a number of regulatory actions rather than one event that caused the decline in IPOs.

The IPO Task Force presented their findings to the U.S. Department of the Treasury on October 20, 2011. They state the regulatory changes have driven up the costs of going public for smaller companies. Young companies on average are spending $2.5 million in legal and compliance costs to go public and then $1.5 million per year to meet ongoing compliance costs. The IPO Task Force went on to say they found that there is little information known about smaller companies when they try to go public and there is little incentive to invest in small companies for the long term.

The IPO Task Force stated that their recommendations would preserve investor protection while also bringing existing regulatory structure in line with current market realities. It was noted that government policies targeted at just small business or at just big business affect all firms, not just firms of a particular size. The one-size-fits-all regulatory regime should be replaced with a scaled approach whereby emerging growth companies would transition in the SEC's disclosure and compliance regime over a period of years following an IPO. The first recommendation by the IPO Task Force was to reduce costs for companies trying to go public. If the firm's total annual gross revenue is less than one billion dollars at IPO registration and the firm is not recognized as a well-known seasoned issuer, the firm will be given up to five years from the date of the IPO to move up to meet full SEC compliance. The scaled regulations focus on the high cost compliance areas and will not compromise investor protection or disclosure.

The second recommendation was to increase visibility for emerging growth companies while maintaining transparency for investors. During the capital formation process, they suggest enacting modifications to existing restrictions on banking research, and expanding permissible pre-filing communications. The final recommendation was to lower the capital gains tax rate for investors who purchase shares of an IPO and hold the shares for a minimum of two years. This will help emerging growth companies to attract long-term investors at the initial allocation so the firm will raise sufficient capital through the IPO. The goal of the three recommendations is to efficiently connect the investors to the sellers of emerging company stocks. The IPO Task Force argued the need for meaningful small-cap IPO reform is urgent as the U.S. continues to lose public offerings and jobs to foreign markets. Cowan (2011) in an interview with Henri Leveque, head of PricewaterhouseCoopers's, stated that more clients are interested in going public in capital markets outside of the U.S. According to Leveque, 278 U.S. companies did IPOs in London in the five years ending in 2010, to 177 in the U.S. and 25 in Hong Kong.

Weild and Kim (2010) examined a number of these regulatory actions mentioned by the IPO Task Force and also argue that they caused the decline in IPOs. Some of the events behind what they called the perfect storm were End of the Four Horsemen by 1999, Regulation Fair Disclosure in 2000, decimalization in 2001, Sarbanes-Oxley in 2002, the Global Settlement in 2003, and Regulation NMS in 2005. The IPO Task Force (2011) and Weild and Kim (2010) point out that all the events were meant to champion a pro-consumer agenda to help the individual investor.

The End of the Four Horsemen refers to merger of four investment banks into commercial banks: Alex Brown, Montgomery Securities, Robertson Stephens, and Hambrecht & Quist. All four when independent played a critical role in connecting investors with innovators by supporting smaller venture-funded companies to market. The merger of these investment banks with commercial banks can be attributed to the Gramm-Leach-Bliley Act of 1999, which led to increased concentration of the financial services industry. The argument is these four investment banks were merged out of existence because it was viewed their economic model of equity research, equity sales, and equity trading did not work anymore.

Regulation Fair Disclosure in 2000 leveled the information playing field for all investors by requiring all public companies to disclose material information at the same time. The unintended consequence was the deterioration of the depth and breadth of company research coverage available to investors. What is the value of the research if all investors have the same information? The problem is that Fair Disclosure limited the ability of small stock issuers to freely and fully communicate one-on-one with investors and analysts. Weild and Kim (2010) argue that Regulation Fair Disclosure enabled an increase in speculative trading and a decrease in long-term investing.

Decimalization lowered trading costs but as spreads disappeared so did economic incentives for firms to provide research and liquidity support for smaller stocks. The decimalization started with the SEC approving Regulation ATS (Alternative Trading System) in 1998. Conversion to a decimal system from a fraction system made sense for some but it's another area where we find a set of unintended consequences. Diminished spreads increased the risk to market makers displaying limit orders which decreased the liquidity provided by the orders. Consequently, the buy side moved to algorithmic trading breaking up block orders that could no longer be handled efficiently. The bottom line was that the process of moving from pricing stocks in cents rather than fractions took profits out of trading that had previously given brokers the incentive to list small companies. This lack of liquidity has pushed institutional investors away from small-cap growth markets.

The passage of Sarbanes-Oxley (SOX) in 2002 further reduced the U.S.'s international competitive position by creating a regulatory burden for public companies that has discouraged foreign and domestic issuers from going public in the U.S. In addition to the added regulatory burden, SOX has increased the cost of outside legal and accounting experts through higher insurance. This increased imposition of personal liability on officers by SOX reporting is another factor that deters companies from listing in the U.S.

The Global Settlement was an enforced agreement between the ten largest U.S. securities firms to address conflicts between research and investment banking in their businesses. The settlement separated equity research from investment banking. For IPOs, investment banking was prevented from having any input into research compensation or coverage decisions. Further, research analysts were prohibited from going with investment bankers on road shows to market new issues. The settlement disconnected the economic incentive for brokers to generate investment research on companies with which they had banking relationships. This led to a further decline in the equity research coverage and support for small-cap stocks because it eliminated the ability to pay analysts directly with investment banking fees. While the IPO market has become more honest and transparent as a result of regulatory change, smaller IPOs have become harder to market. Weild and Kim (2010) argue there is no one event that has led to the demise of the IPO in the U.S. but instead it is a culmination and interaction of many factors.

In 2005, the rules promoting the National Market System were consolidated into regulation NMS. The main rules were order protection, access rule, sub-penny rule and market data rules. The order protection (or trade through) rule provides intermarket price priority for quotations that are immediately and automatically accessible. The order protection rule has been controversial because it requires traders to transact at the lowest price rather than trades offering the quickest execution or the most reliability. The rule is viewed by many as an improper government intervention into private business affairs. The access rule tries to improve access to quotations from trading centers in the National Market System by requiring greater linking and lower access fees. Sub-penny rule tries to help small investors by not allowing high-frequency traders to jump to the front of the line in the National Best Bid and Offer (NBBO). The NBBO is the highest posted bid and the lower posted offer for a trading instrument, by making a price improvement in increments of 1/100th of a penny. Thus a high-frequency trader transaction would be executed first and provide the trader with the best opportunity to capture the spread. The rule specifies that the minimum pricing increment for stocks over $1.00 must be $0.01; stocks under $1.00 have a minimum pricing increment of $0.0001. The market data rules are designed to promote the wide availability of market data and to allocate revenues to self-regulatory organizations that produce the most useful data for investors. The goal was to strengthen the existing market data system, which provides investors in the U.S. equity markets with real-time access to the best quotations and most recent trades. For each stock, quotations and trades are continuously collected from many different trading centers and then disseminated to the public in a consolidated stream of data. Therefore, investors of all types have access to a reliable source of information for the best prices.

JOBS Act 2012

Based on the recommendations of the IPO Task Force and the perceived need to reduce barriers to growth and investment, the Jump-Start Our Business Start-Ups Act (JOBS) was made law on April 5, 2012. As the IPO Task Force pointed out, this JOBS act was needed to create more potential jobs. The task force showed that when emerging growth firms go public, research shows that more than 90 percent of their job creation happens after the IPO. The new law will clearly reduce the regulatory burden on upstart businesses trying to raise capital and allow them more freedom to communicate with investors.

The JOBS act will allow many new public companies to avoid some of the accounting rules and restrictions that larger, more established firms face. What the act calls “emerging growth companies” are those with annual revenues under $1 billion and publicly traded for less than five years. The JOBS act intends to create an on-ramp that eases the regulatory burden and costs for these emerging growth companies. First, these emerging growth companies will be given five years before being subjected to the full weight of federal regulation such as section 404 of SOX regarding the redundant external audit of internal audit of their business processes. However, after the five year period, all companies will have to comply with all regulations in place and regulators will still have antifraud enforcement authority when the firms go public.

Another provision in the JOBS act allows aspiring public companies to share information confidentially with the SEC, accept feedback and disclose SEC-mandated changes to their disclosures later. This provision will allow companies to avoid disclosing potentially sensitive information to competitors. The companies will still have to release the financial documents 21 days before they try to persuade institutional investors to buy into the IPO during the road show.

Nonetheless, firms are moving IPOs to countries around the world as market conditions change in the home country. They are no longer bound to the home country for such IPO offerings. Clearly, greater globalization has led to increased mobility of capital.

Data

Table 1 provides a distribution of the number of total global IPO offerings and U.S. IPOs by year from 2000 through 2009. The IPO data is from the Securities Data Corporation (SDC) database. We exclude offers with share prices below $1, firms in the 6000 SIC code, offerings less than one million dollars, and IPOs where the market exchange is missing. As Table 1 indicates, the percentage of U.S. IPOs to total global IPOs has declined over the time period. Table 1 also shows the number of IPOs that go outside their home country market. Of the IPOs that go abroad, the percentage of those coming to the U.S. appears to be declining. This raises the concern that the U.S. may have lost its competitive edge in attracting foreign IPOs.

Table 1: U.S. versus World IPOs that go abroad

Year	Total # of IPOs		Percentage of Total U.S. to World IPOs	IPOs that go abroad		U.S. IPOs going abroad (3)	% of World IPOs that go abroad	% of non-U.S. IPOs that come to U.S. 2/(1-3)
	World	U.S.		World (1)	To U.S. (2)			
2000	1037	348	33.6%	103	66	6	9.9%	68.0%
2001	457	135	29.5%	20	11	0	4.4%	55.0%
2002	519	167	32.2%	18	13	2	3.5%	81.2%
2003	478	130	27.2%	8	8	0	1.7%	100.0%
2004	843	298	35.3%	58	28	5	6.9%	52.8%
2005	873	276	31.6%	93	36	13	10.7%	45.0%
2006	1109	245	22.1%	134	45	16	12.1%	38.1%
2007	1456	331	22.7%	177	68	17	12.2%	42.5%
2008	428	50	11.7%	43	16	4	10.0%	41.0%
2009	444	78	17.6%	38	17	2	8.6%	47.2%

Table 2 shows the export and import of IPOs by country for the 2000 through 2009 time period. The table shows the home country bias of listing but also shows whether each country is a net importer or exporter of IPOs. As shown in the table, most countries are net exporters of IPOs. The largest net exporter of IPOs is China. Of the few countries that are net importers, the U.S. and U.K. stand out as the main net importers of IPOs. Over the time period examined, 85.5% of companies listing in the U.S. were domestic companies and 96.6% of U.S. firms going public chose the U.S. as their listing destination.

Table 2: The Importing and Exporting of IPOs by Country

	Domestic as a % of listing companies	Percentage Listing domestically	IPOs exported	IPOs imported	Net Import (Export) of IPOs
Argentina	100.0%	36.4%	7	0	(7)
Australia	98.4	93.1	9	2	(7)
Austria	96.4	75.0	9	1	(8)
Belgium	90.9	92.6	4	5	1
Brazil	97.3	90.7	11	3	(8)
Bulgaria	100.0	91.7	1	0	(1)
Canada	96.9	94.8	31	18	(13)
China	99.3	84.1	140	5	(135)
Cyprus	100.0	16.7	10	0	(10)
Denmark	97.4	97.4	1	1	0
Finland	100.0	88.9	3	0	(3)
France	93.0	96.8	8	18	10
Germany	90.7	96.0	9	22	13
Greece	100.0	81.6	21	0	(21)
India	100.0	95.7	13	0	(13)
Ireland	72.7	26.7	30	11	(19)
Israel	90.9	17.9	46	1	(45)
Italy	99.2	91.9	11	1	(10)
Japan	99.5	99.8	2	4	2
Jersey	0.0	0.0	27	0	(27)
Luxembourg	50.0	7.7	12	1	(11)
Mexico	100.0	61.1	7	0	(7)
Netherlands	44.8	28.3	33	16	(17)
New Zealand	91.7	84.6	2	1	(1)
Norway	86.1	94.9	4	12	8
Poland	95.3	100.0	0	7	7
Portugal	90.0	90.0	1	1	0
Russia	100.0	61.4	17	0	(17)
Singapore	63.6	58.3	5	4	(1)
South Africa	100.0	46.7	8	0	(8)
South Korea	98.8	99.0	7	8	1
Spain	100.0	93.7	2	0	(2)
Sweden	94.9	93.3	4	3	(1)
Switzerland	80.4	77.6	13	11	(2)
Thailand	100.0	80.0	2	0	(2)
Turkey	100.0	88.9	2	0	(2)
United Kingdom	55.2	80.8	66	225	159
United States	85.5	96.6	64	308	244
Vietnam	100.0	85.7	6	0	(6)
All others			44	3	(41)

There may be a number of reasons why a firm would migrate from their home country market to offer an IPO. Two important reasons may be economic freedom and the ability to capture a larger value for their IPO. The Economic Freedom Index (EFI) is published annually the Fraser Institute (Gwartney and Lawson, various) and is based on 42 data items designed to capture, in a broad sense, the role of government, the rule of law, money, barriers to trade internationally, and overall business regulation. Research suggests that countries with more economic freedom have greater economic output, higher income, and better overall amenities in their society. Therefore, a company may migrate out of their home country in order to try and enjoy of benefits associated with a country with a higher EFI. Table 3 shows the EFI for the U.S., the average EFI for non-U.S. countries importing IPOs, and the average EFI of countries exporting IPOs. The U.S. has historically been rated very highly and continues to enjoy a high ranking although recent years have seen a decline from as high as number two in the world down to number eighteen in the most recent update of the EFI. Table 3 also shows the U.S. Dollar Index which is used to track the dollar against the currencies of six major U.S. trading partners. The higher dollar index reflects the fact that the U.S. has a dominant currency and is an indicator of the economic health of the U.S. Foreign firms that migrate here for their public offering may want to issue in the U.S. to take advantage of having dollars for business transactions. Higher values of the dollar would also be associated with higher returns in the foreign firm's home currency.

Table 3: Economic Freedom Index (EFI) of the World and U.S. Dollar Index

	EFI for U.S.	Average EFI of non-U.S. Import Markets for IPOs	Average EFI of Nations Exporting IPOs	U.S. Dollar Index
2000	8.45	7.83	7.53	118.387
2001	8.23	7.60	7.50	125.347
2002	8.23	7.76	7.26	130.183
2003	8.18	---*	7.24	127.021
2004	8.15	7.88	7.28	115.066
2005	8.09	8.01	7.19	108.816
2006	8.03	7.91	7.33	111.753
2007	8.10	7.80	7.06	107.545
2008	7.91	7.87	7.15	100.196
2009	7.60	7.13	7.10	110.025

* All migrating IPOs in 2003 came to the U.S.

The Model and Empirical Results

We expect the decision for a foreign firm to go public in the U.S. is a function of various firm and offer characteristics and U.S. market conditions. We estimate a logit model where the dependent variable, U.S. IPO, takes the value of one if the foreign firm went public in the U.S., and zero otherwise.

$$U.S.\ IPO\ (1/0) = \beta_0 + \beta_1 EFI + \beta_2 SOX + \beta_3 Size + \beta_4 Dollar\ Index + \beta_5 Economic\ Union \qquad (1)$$

Table 4 presents the results of the logistic regression analysis. Model 1 presents the choice of a U.S. IPO as a function of the relative economic freedom of the target market, a dummy indicating whether the IPO was pre or post-Sarbanes Oxley (SOX), the size of the IPO, the U.S. dollar index and a dummy variable indicating membership in a trading bloc such as the EU or ASEAN. To capture the relative economic freedom of the U.S. market we use two measures. *Higher U.S. EFI* takes the value of one if the U.S. EFI is higher than the domestic market and zero otherwise and is used in Model 1. *EFI difference* is the U.S. EFI less the EFI of the domestic market and is used in the alternative Model 2.

Table 4: Logistic Regression of U.S. Market Choice for Non-U.S. Migrating IPOs

	(1)	Wald Chi-Square	(2)	Wald Chi-Square
Intercept	-5.06	2.47	-3.95	1.55
Higher U.S. EFI	0.62**	3.95		
EFI Difference			0.37***	9.85
Post Sox	-0.91**	5.51	-0.87**	5.08
Log of IPO proceeds	0.30***	20.28		
Inverse of proceeds			-3.29**	6.19
U.S. Dollar Index	0.40	2.28	0.44*	2.84
Union bias	-2.07***	69.39	-1.82***	53.02
Number of observations	582		582	
Likelihood ratio Chi-Square	143.60***		135.53***	

* significant at the 1% level , ** significant at the 5% level, *** significant at the 10% level

The results suggest that the choice to remain in the U.S. for an IPO is positively related to the level of economic freedom as well as the overall size of the IPO. The probability that a migrating IPO firm will choose the U.S. market as its destination is higher when the U.S. economic freedom index is greater than their home market and when the IPO is larger and is reduced post-SOX and when the IPO firm's home country is part of a regional trading bloc.

Model 2 in Table 4 presents a slight modification of Model 1 using *EFI difference* and measuring IPO size using the inverse of IPO proceeds. These results further support the view that firms do choose the U.S. for IPO offerings based largely on the level of economic freedom as well as size of the offering. Further, there is some slight evidence in this model that a strong dollar does help. Also, the size proxy here suggests that smaller IPOs tend to not offer in the U.S. Likewise, IPOs subsequent to the passage of Sarbanes-Oxley tend to migrate to other markets rather than the U.S.

Summary and Conclusions

This study of 692 IPOs, covering the period from 2000 through 2009, is intended to examine migration of IPOs from the home country. While most countries are net exporters of IPOs, the United States and the United Kingdom are the primary net importers of IPOs. However, the data suggest that there is an increasing trend for migration to countries other than the U.S. Overall, we find that IPOs that deviate from their domestic market are still more likely to choose the U.S. when the level of U.S. economic freedom is higher and the IPO is larger. There remains some slight evidence that a strong dollar may play some role in this choice. IPOs are more likely to migrate to other markets when they are smaller IPOs, when the IPO is done after the passage of Sarbanes-Oxley and when the home country is part of a larger trading bloc such as the EU or ASEAN. Clearly, the migration of IPOs is important and given the drop in economic freedom in the U.S. as evidenced by the Economic Freedom Index, one may find that IPOs continue to decline in the U.S.

Acknowledgements

We thank the Academy of Economics and Finance for the opportunity to present this paper in 2013 at the 50th annual meeting and the participants for the feedback and comments received.

References

Bargeron, Leonce L., Kenneth M. Lehn, and Chad J. Zutter. 2010. "Sarbanes-Oxley and Corporate Risk-taking." *Journal of Accounting and Economics* 49: 34-52.

Content First, LLC. 2011. "Venture Impact: The Economic Importance of Venture Capital-Backed Companies to the U.S. Economy." *Global Insight* 4.

Cowan, Lynn. 2011. "More U.S. Firms Asking About IPOs Abroad." *Wall Street Journal*. 11 November, Online.

Crovitz, Gordon. 2011. "Information Age: Exporting Wall Street." *Wall Street Journal*. 28 February, A17.

Doidge, Craig, G. Andrew Karolyi, and Rene M. Stulz. 2011. "The U.S. Left Behind: The Rise of IPO Activity Around the World." Working paper, University of Toronto.

Freeman, James. 2009. "The Supreme Case Against Sarbanes-Oxley." *Wall Street Journal*. 7 December, A23.

Gao, Xiaohui, Jay R. Ritter, and Zhongyan Zhu. 2011. "Where Have All the IPOs Gone?" Working paper, University of Hong Kong.

Greifeld, Bob. 2011. "Sarbox and Immigration Reform for Jobs." *Wall Street Journal*. 4 October, A21.

Gwatney, James and Robert Lawson. Various. "*Economic Freedom of the World Index*." Fraser Institute, Vancouver. www.freetheworld.com.

IPO Task Force. 2011. "Rebuilding the IPO On-Ramp, Putting Emerging Companies and the Job Market Back on the Road to Growth." 20 October. Presented to The U.S. Department of the Treasury.

Kaserer, Christoph and Dirk Schiereck. 2007. "Going Public and Being Public – A Global Comparison of the Impact of the Listing Decision on the Cost of Capital." Presented at the German Equity Forum in Frankfurt.

Lucchetti, Arron. 2011. "U.S. Falls Behind in Stock Listings." *Wall Street Journal*. 26 May, A1.

Scott, Hal S. 2011. "Capital Market Regulation Needs an Overhaul." *Wall Street Journal*. 20 April, A13.

Stephens, Warren. 2011. "Business Regulation vs. Growth: The View from Middle America." *Wall Street Journal*. 25 August, A15.

The Economist Intelligence Unit. 2012. "The Endangered Public Company: The Big Engine that Couldn't." *The Economist*. 19 May.

Weild, David and Edward Kim. 2009. "Why are IPOs in the ICU?" Grant Thornton white paper. Presented to the NYSE and National Venture Capital Association's Blue Ribbon Task Force. 19 November.

Weild, David and Edward Kim. 2010. "Market Structure is Causing the IPO Crisis – and More." Grant Thornton Captial Market Series, June.

Stock Returns, Oil Price Movements, and Real Options

Steve Johnson and Robert Stretcher, Sam Houston State University

Abstract

In this paper, we introduce a new method for detecting the presence of real options. Using option theory, we develop a "real option term" that is a function of both crude oil price movement and volatility. After controlling for risk factors in panel regressions, the coefficient on the real option term is positive and statistically significant, providing support for the hypothesis that real options contribute significantly to the value of firms in the oil and gas industry. During the period1998-2002, this impact is only detectable for small firms, while during the periods 1993-1997 and 2003-2008, movements in oil-price volatility impact stock prices across the oil industry.

Introduction

The oil and gas industry provides a natural laboratory for testing theories of real options. One of the biggest challenges in empirical testing of real option theories at the firm level is the identification and measurement of the volatility that is relevant to the pricing of the firm's real options. In the oil and gas industry, firms' revenues (holding quantity constant) are an increasing function of the price of oil. Oil and gas companies' reserves become economically feasible to develop when the price of oil is sufficiently high. Thus, it is reasonable to use oil-price volatility as a measure of real option volatility of the underlying asset in the oil and gas industry.

Our goal in this paper is to detect the presence of real options on crude oil by examining stock returns in the oil and gas industry. We develop a "real option term" that is a function of crude oil price movement and volatility. The value of an option holder's position increases when the volatility of the underlying asset increases. However, if both the price and the volatility of underlying asset price movements change, the resulting effect on the value of the option is complex. After controlling for other effects shown in prior studies to impact stock prices, our results show that the stock returns of publicly-traded oil and gas firms in the US have a positive exposure to fluctuations in oil-price volatility during periods when price movements are relatively small and volatility fluctuations are relatively large. This methodology is fundamentally different from that of previous studies that focus on the effects of movements in the price of oil.

There are at least three streams of literature that are relevant to this paper. The first stream is the literature on real options. Another stream is the literature on asset pricing models that include the effects of firms' strategic decision-making. A third stream concerns the examination of the effects of oil price movements on stock prices.

Real options arise when managers can make strategic investment decisions in response to movements in prices. The theory of real options predicts that, ceteris paribus, real option value increases when the underlying asset return volatility increases. Empirically, this effect has been difficult to identify, due to the confounding effects of movements in price. The impact of real options on investment include delaying investment to capture real option value at the individual project level (Quigg 1993) and an increase in real option value when underlying asset return volatility increases (Ingersoll and Ross, 1992; McDonald and Siegel, 1986; Venezia and Brenner, 1979; Grullon, Lyandres, and Zhdanov 2010). Henriques and Sadorsky (2010) note that, consistent with real option theory, oil price volatility has a nonlinear effect on firm investment. Grullon, Lyandres, and Zhdanov (2010) find that changes in firm-specific volatility have an effect on firm value.

An active area of research is the design of models of firm investment that explain the size and book-to-market effects. Berk, Green, and Naik (1999) link firms' real investment decisions and asset return dynamics. Gomes, Kogan, and Zhang (2003), Kogan (2004), Carlson, Fisher and Giammarino (2004), Cooper (2006), Zhang (2005), and Aguerrevere (2009) provide models that relate risk and return dynamics to firm-specific characteristics such as size and book-to-market. Carlson, Fisher and Giammarino (2006) note that size and/or book-to-market are frequently correlated with growth options and/or real options.

Stock prices are commonly believed to react to economic news, such as movements in the price of oil. Unanticipated movements in the price of oil are believed to impact the entire market. Two typical Wall Street Journal articles written from this viewpoint are "Oil Spike Raises Portfolio Questions" (Kansas 2009) and "Oil Shock: Did High Oil Prices Cause the Recession?" (Johnson 2009). Some researchers have found that movements in oil prices have an effect on stock prices. For oil, gas, and/or mining firms, Al-Mudhaf and Goodwin (1993), Sadorsky (2001), and Boyer and Filion (2007), find evidence that oil price shocks have a positive impact. For non-extraction industries, empirical frequently notes that oil price shocks have a negative impact. Driesprong, Jacobsen, and Maat (2008) find demonstrate that oil price shocks have a negative impact on stock prices overall. Sadorsky (1999) reports that oil price increases have significantly negative impacts on US stocks. Also, the magnitude of the impact may have increased since the 1980s. Nandha and Faff (2008) find that oil price rises have a negative impact on equity returns for all sectors except for the mining and the oil and gas industries. Ciner (2001) finds evidence of a nonlinear relationship between oil price movements and stock returns. Jones and Kaul (1996) suggest, but do not pursue the conjecture, that oil price information might be useful for forecasting stock prices.

Other researchers have found that movements in oil prices do not have a significant net effect on stock prices. Chen, Roll, and Ross (1990) investigate a number of possible factors and find evidence that oil price risk is not rewarded in the stock market. Huang, Masulis, and Stoll (1996) also find no evidence that oil price risk is rewarded in the stock market.

There is also evidence that oil price movements have a lagged effect on stock prices in addition to a contemporaneous effect. Park and Ratti (2009) find that oil price shocks have a significant impact on real stock returns contemporaneously and/or within the following month. Driesprong, Jacobsen, and Maat (2008) also find that oil price shocks have a one-month lagged effect on stock returns.

These studies have generally focused on movements in the level of oil prices, rather than oil-price return volatility. One major exception is Grullon, Lyandres, and Zhdanov (2010), who find that movements in oil-price volatility do indeed affect stock prices in a manner consistent with option pricing theory. Our analysis differs from theirs in that we use option pricing theory to disentangle the effects of movements in oil price returns from those of oil price volatility.

Option Pricing Theory

Movements in the underlying asset price and movements in the underlying asset return volatility will both have an effect on option value. Over any given period of time, there will be movements in both underlying asset price and volatility. For example, it may be the case that volatility drops, lowering the value of the option, but the stock price increases, increasing the value. The dominance of one effect or the other depends on the magnitudes of the changes in volatility and price. The entanglement of the effects of changes in price and changes in volatility makes estimation of real option value challenging. We address this problem by focusing on periods when changes in oil return volatility are relatively large and changes in oil prices are relatively small. Our choice of the value of the threshold is driven by the underlying theory of option pricing.

Based on option theory, an option has the greatest time value when the underlying asset price is around-the-money (close to at-the-money). Real options that are deep in-the-money or deep out-of-the-money have relatively small time values and these are not very sensitive to changes in volatility. Unfortunately, data concerning the strike price of specific projects is generally not publicly available. To overcome this complication, we use a three-part strategy.

First, movements in volatility of the underlying asset generate the most significant effect on the value of the firm's real options relative to the firm's overall value when the underlying asset price is near the strike price. Deep in-the-money and deep out-of-the-money options have little time value overall and thus the time value of such options probably contributes little to overall firm value. Around-the-money options have a relatively large time value and are more likely to contribute significantly to firm value. Also, changes in volatility have a significant impact on around-the-money options. We implicitly assume that any options that we will be able to detect will be around-the-money.

Second, small movements in volatility may be dominated by large movements in the underlying asset price. While a change in oil price may have an impact on the value of an oil and gas firm due to the direct effect on revenues, even if the firm holds no real options, changes in oil-price volatility would not have an immediate direct effect on revenues. Option-pricing theory predicts that an increase in the volatility of oil prices will increase the value of options on oil. Empirically, however, the effects of small movements in oil-price return volatility may be "drowned out" by noise and the effects of movements in other variables. Therefore, when we create our real options term, we focus on periods when changes in volatility are relatively large.

Finally, large movements in price may dominate the effects of relatively small movements in volatility. In particular, a large downward price movement may be associated with an increase in volatility. These two changes have complex, opposite effects on the value of the option, making disentanglement challenging. Also, large upward movements could have a positive effect on the value of an oil and gas firm even if the firm holds no real options, due to the fact that revenues of firms in the oil and gas industry depend, at least in part, on the price of oil. Therefore, when we create our real options term, we focus on the times when changes in oil prices are relatively small.

By focusing on periods where changes in volatility are relatively large and where changes in price are relatively small, we are more likely to be able to detect the effects of changes in volatility on real options held by the firm. Note that this focus will not eliminate all periods where changes in real option values are small. For example, if a firm's real options are all currently deep out-of-the-money or deep-in-the-money, real option-based volatility effects will be small even when movements in volatility are large. In this case, it is an empirical question whether this method will be successful or not in detecting real option effects.

In order to ameliorate the problem of entanglement of price movement and volatility movement effects, we attempt to select a threshold value that is both high enough to serve as a "floor" for volatility movements and low enough to serve as a "ceiling" for price movements. Since real options are probably only a part of any given firm's total assets, we look for a threshold value that will limit the results to those months with changes of at least 10% of option value. Based on the results of table I, we select a value of 3% for the threshold. OPEC notes that the time from exploration, to discovery of a new well, through the development of the new well, can be three to eight years (OPEC 2010). Since the motivation of our model involves oil reserves that have already been discovered but not developed, we choose a period of three years. Our results are not qualitatively sensitive to the choice of three years and remain robust to choices of three through eight years, while a change to two years results in only the value in the last row of panel B dropping below an absolute value of 10% (9.8%).

In panel A of table I, we examine the results of a three percent increase in volatility and a negative three percent return. The volatility increase will drive the price of the option up, while the negative return represents a drop in price that will drive the price of the option down. We use the historical distribution of monthly return oil price volatility over the period 1993-2007. We use daily closing oil prices to estimate monthly oil price return volatility. Using base values of volatility drawn from the 25%, 50%, and 75% percentiles (second column of table 1, "%ile of σ") of the distribution of monthly volatility over the sample period and base values of the underlying asset equal to 90%, 100%, and 110% of the spot price, we examine the effect of changes of a three percent increase in volatility and a negative three percent return on the value of the option. The price of oil is in units of X, the strike price. Our results indicate that, in each case, the value of the option increases by at least ten percent. Since larger increases in volatility and smaller decreases in the price will only result in larger increases in option value, the threshold value of three percent is appropriate.

In panel B of table I, we examine the results of a three percent decrease in volatility and a positive three percent return. The volatility decrease will drive the price of the option down, while the positive return represents an increase in price that will drive the price of the option up. Using the same three base values of volatility and the same spot-strike price pairs as in panel A, we find that in each case, the value of the option decreases by at least ten percent. Since larger decreases in volatility and smaller increases in price will result in larger decreases in option value, the threshold value of 3 percent is appropriate.

Table 1: Black-Scholes Option Price Movements

Panel A: $\Delta\sigma = 0.03$, $\Delta r = -0.03$

Spot/Strike	%ile of σ	σ	C(S,σ)	C(S+ΔS,σ+Δσ)	ΔC	ΔC/C
0.90	25%	0.0764	0.1575	0.2038	0.0463	0.2939
0.90	50%	0.0940	0.1945	0.2392	0.0447	0.2298
0.90	75%	0.1156	0.2392	0.2817	0.0426	0.1780
1.00	25%	0.0764	0.2198	0.2649	0.0451	0.2053
1.00	50%	0.0940	0.2581	0.3021	0.0440	0.1704
1.00	75%	0.1156	0.3048	0.3469	0.0422	0.1383
1.10	25%	0.0764	0.2900	0.3315	0.0416	0.1434
1.10	50%	0.0940	0.3282	0.3698	0.0415	0.1266
1.10	75%	0.1156	0.3755	0.4160	0.0405	0.1079

Panel B: $\Delta\sigma = -0.03$, $\Delta r = 0.03$

Spot/Strike	%ile of σ	σ	C(S,σ)	C(S+ΔS,σ+Δσ)	ΔC	ΔC/C
0.90	25%	0.0764	0.1575	0.1085	-0.0490	-0.3109
0.90	50%	0.0940	0.1945	0.1468	-0.0477	-0.2454
0.90	75%	0.1156	0.2392	0.1932	-0.0460	-0.1922
1.00	25%	0.0764	0.2198	0.1753	-0.0445	-0.2023
1.00	50%	0.0940	0.2581	0.2130	-0.0451	-0.1748
1.00	75%	0.1156	0.3048	0.2603	-0.0445	-0.1460
1.10	25%	0.0764	0.2900	0.2546	-0.0353	-0.1219
1.10	50%	0.0940	0.3282	0.2887	-0.0395	-0.1204
1.10	75%	0.1156	0.3755	0.3344	-0.0411	-0.1094

NOTES: Table I reports the results of the valuations of changes in option value for near-the-money call options on the spot price of WTI crude oil using the Black-Scholes formula. The first column, "S = % of X," in each panel describes the moneyness of the option, with values of S equal to 90%, 100%, and 110% of the strike price. The second column, "%ile of σ," lists the percentile of the base value of volatility for each row. The third column, "σ," lists the value of oil-price volatility at the corresponding percentile from the second column. The column labeled "C(S,σ)" lists the Black-Scholes option price for the option with the characteristics listed in the first three columns, and with T = 36 months and the spot price in units of X. The column with header C(S+ΔS,σ+Δσ) lists the Black-Scholes option price of the option with strike price equal to S+ΔS and volatility equal to σ+Δσ. ΔC and ΔC/C are the change in value and the percent change in value moving from C(S,σ) to C(S+ΔS,σ+Δσ). Panel A reports results for the changes Δσ= 0.03 and Δr= -0.03, while Panel B reports results for the changes Δσ= -0.03 and -Δr= -0.03.

There are two more possibilities that should be considered. Both the volatility and the price could increase, or both the volatility and price could decrease. There is no significant entanglement problem in this case, because both changes have the same effect on option value. Both a volatility increase and a price increase result in an increase in the value of a Black-Scholes call option, while both a volatility decrease and a price decrease result in a decrease in the value of a Black-Scholes call option. Based on the results of table I, we choose a threshold value of three percent for the creation of the real options term used in the empirical tests in this paper.

Note that the use of this method does not guarantee success. We make the assumption that the price of oil is sufficiently close to the strike price frequently enough to be significantly affected by movements in oil-price volatility. As oil prices have fluctuated greatly, it is likely that at least some of the time, many of the real options held by oil and gas firms have been deep out-of-the-money or deep in-the-money. Therefore, failure to detect real options with our model can not be taken as evidence that oil and gas firms do not hold real options. However, if evidence of real options is detected in spite of the existence of time periods where options are deep in-the-money or deep out-of-the-money, it would not be unreasonable to use this result to support the hypothesis that oil and gas firms hold significant real options on the price of oil.

Firm Size and Real Options

An additional consideration is that firms with different characteristics may be more or less likely to hold real options. Determining the type of firm that is most likely to hold real options would be a significant aid in detecting the presence of real options. Many investigators, beginning with Myers (1977) and Myers and Majluf (1984), describe the value of a firm as the value of assets in place and growth opportunities. Growth opportunities that involve strategic decision-making can frequently be modeled as real options. However, at least some growth opportunities, that have no significant strategic interactions, could also be modeled using traditional NPV methods rather than real options. Finance theory notes several variables that may be correlated with the presence of real options. One seminal paper that reports a model of the ties between firms' investment decisions and asset return dynamics is Berk, Green, and Naik (1999). Gomes, Kogan, and Zhang (2003), Kogan (2004), Carlson, Fisher and Giammarino (2004), Cooper (2006), Zhang (2005), and Aguerrevere (2009) provide models that relate risk and return dynamics to firm-specific characteristics such as size and book-to-market. If assets in place and real options have different sensitivities to changing economic conditions, their systematic risks will be different.

Berk, Green, and Naik (1999) demonstrate that the expected return equation can be written in terms of size and book-to-market. In their model, book-to-market is a proxy for systematic risk, which changes over time as assets turn over, while size is a proxy for the relative value of growth options. Carlson, Fisher and Giammarino (2006) note that size and/or book-to-market are frequently correlated with growth options and/or real options. If at least some growth options are also real options, it follows that there may be a relationship between size and real option value. Also, Portillo (2000) notes that smaller firms in the oil and gas industry (SIC 1311) tend to focus more on exploration, while the larger firms hold higher reserves. This qualitative difference in focus may drive a difference in the quantity of real options in a firms' portfolio that is correlated with the difference in firm size.

Estimation

The method used in this study is a panel regression, controlling for hetereoskedasticity and autocorrelation using the White-Newey-West estimator (White, 1980; Newey and West 1987). Asset-pricing research has demonstrated that a number of factors have an effect on stock price movements. These factors include the excess return on the market, the Fama-French value and size factors (Fama and French 1992; 1993; 1996), and the Carhart momentum factor (Carhart 1997). Since this is a study of the oil and gas industry, both lagged and contemporaneous oil-price returns may influence stock returns (Park and Ratti 2009; Driesprong, Jacobsen, and Maat 2008). Changes in oil-price volatility may also be a factor that prices oil and gas stocks (Grullon, Lyandres, and Zhdanov 2010). The excess return on the market, the size factor, the value factor, and the momentum factor are obtained from Ken French's website. Spot oil-price data is taken from the Department of Energy's Energy Information Administration website (www.eia.doe.gov). RC Research, Inc. provided oil futures price data. Monthly oil and gas company stock returns are taken from CRSP. Financial statement data is taken from Research Insight/COMPUSTAT.

We run panel regressions over three 5-year segments: 1993-1997, 1998-2002, and 2003-2007. In order to reduce irregularities from newly-listed firms, each firm must have at least twenty-four months of stock-price return data before the beginning of each period. We match COMPUSTAT data from the fiscal year immediately preceding the beginning of each period. In order to be included in our sample, a firm must not be missing any stock-price return data during either the five-year estimation period or the two-year period preceding the estimation period, and must also have the necessary COMPUSTAT data from the fiscal year preceding the estimation period.

The empirical focus of this study is the test of significance of the change in the volatility of oil-price returns during periods when the change in the price of oil is small. According to option-pricing theory, an increase in the volatility of the underlying asset returns will result in an increase in a call option on the underlying asset. However, empirical estimation is complicated by at least two effects. First, the impact of the change in the price of the underlying asset may be greater than the impact of the change in the volatility, depending on the relative sizes of the changes in price and volatilty and the moneyness of the option. Second, movements in volatility result in the largest changes in option value when the option is around-the-money, with much smaller effects when deeper in-the-money or deeper out-of-the-money. We control for the first effect by creating the variable FlatVar, which is equal to the change in oil-price volatility when the oil-price return is less than 3% and the change in volatility is greater than 3%, and zero otherwise. The justification for the choice of the 3% threshold is derived in the previous section of this article. We do not directly control for the second effect. However, we note that if oil companies' real options are a significant contribution to the value of the firm, these real options must be near-the-money. If oil companies' real options were deep out-of-the-money, the value would be negligible, and if the real options were deep in-the-money, most of the value would not come from the real option aspect of the projects.

An important consideration is that different types of oil companies may hold different proportions of real options. Investigators have determined that, based on the theories of asset pricing and real options, firm size may be correlated with higher levels of real options (Berk, Green, and Naik, 1999; Carlson, Fisher and Giammarino, 2006). In order to determine whether some types of firms hold more real options than others, we create two variables. The first variable is FlatVar, which is equal to the change in the standard deviation of the price returns on crude oil. The second variable is RealOption, which is equal to the product of a dummy variable equal to one if the firm is in the upper half of the distribution of firm size, as measured by market capitalization.

The regressions used in this study are:

$$R_i = b_0 + \text{CONTROLS} + b_7\text{*diff} + b_8\text{*FlatVar} + b_9\text{*RealOption} \qquad (1)$$
$$R_i = b_0 \qquad + b_7\text{*diff} + b_8\text{*FlatVar} + b_9\text{*RealOption}$$

where CONTROLS = $b_0 + b_1$*RMRF + b_2*SMB + b_3*HML + b_4*UMD + b_5*ChOil + b_6*lagChOil

The definitions of the variables used in the regressions and in variable definition are:

RMRF	Excess return on the market
SMB	Size factor
HML	Value factor
UMD	Momentum factor
ChOil	Oil-price return
diff	Change in oil-price return volatility
d1	dummy variable equal to one if ChOil < 3%
d2	dummy variable equal to one if diff > 3%
FlatVar	diff * d1 * d2
d3	dummy variable equal to one if the firm is in the lower 50% of size
RealOption	FlatVar * d3

After running the panel regressions described previously, we find evidence of real option value in the oil and gas industry. For robustness, two measures of oil prices are used. The results for the West Texas Intermediate Spot price are reported in table 2. The results for crude oil futures prices are reported in Table 3. In table 2, where oil prices are measured by using the spot price of oil, over the years 1993-1997 and 2003-2008, the coefficient on FlatVar is positive and significant at the 5% level for market capitalization when control variables are included in the regression. In table 3, where oil prices are measured by using the futures price of oil, the coefficient on FlatVar is positive and significant at the 1% level over the periods 1993-1997 and 2003-2008 when control variables are included in the regression. The coefficient remains marginally (10% significance level) significant even when the control variables are omitted. Also, during the period 1998-2002, while the coefficient on FlatVar is not statistically significant, the coefficient on RealOption is positive and signficant at the 1% level in both regressions, including and omitting the control variables. These results support the conjecture that firms in the oil industry hold real options on the price of oil, and that these options contribute to the value of the firm.

Table 2. WTI spot oil price results

Panel A: 1993-1997

	diff	*FlatVar*	*RealOption*
Intercept only	-0.26228	1.684563	0.017424
	(0.388)	(1.219)	(0.824)
Intercept + controls	0.005693	1.596836	0.017424
	(0.202)	(0.748) **	(0.824)

Panel B: 1998-2002

	diff	*FlatVar*	*RealOption*
Intercept only	-0.183	-0.539	1.172
	(0.959)	(1.290)	(0.539)
	diff	*FlatVar*	*RealOption*
Intercept + controls	-0.305	-0.426	1.172
	(0.975)	(0.732)	(0.538)

Panel C: 2003-2008

	diff	*FlatVar*	*RealOption*
Intercept only	-0.26228	1.684563	0.017424
	(0.388)	(1.219)	(0.824)
	diff	*FlatVar*	*RealOption*
Intercept + controls	0.005693	1.596836	0.017424
	(0.202)	(0.748) **	(0.824)

Table 3. Oil futures price results

Panel A: 1993-1997

	diff	*FlatVar*	*RealOption*
Intercept only	-0.45634	3.102152	-0.89824
	(0.442)	(1.743) *	(1.257)
Intercept + controls	0.014695	3.785853	-0.89824
	(0.258)	(1.367) ***	(1.257)

Panel B: 1998-2002

	diff	*FlatVar*	*RealOption*
Intercept only	-0.605	0.057	0.702 ***
	(1.122)	(1.062)	(0.109)
	chSTDevFut	*FlatFutVar*	*RealOptionFut*
Intercept + controls	-0.324	-0.366	0.706 ***
	(1.083)	(1.034)	(0.116)

Panel C: 2003-2008

	diff	*FlatVar*	*RealOption*
Intercept only	-0.45634	3.102152	-0.89824
	(0.442)	(1.743) *	(1.257)
	chSTDevFut	*FlatFutVar*	*RealOptionFut*
Intercept + controls	0.014695	3.785853	-0.89824
	(0.258)	(1.367) ***	(1.257)

Conclusion

One major challenge for empirical research in real options is identifying and measuring the volatility of the underlying asset. By using basic concepts from option pricing theory, we conjecture that the effect of movements in volatility on real option value should be greatest when the price level remains constant. We develop a methodology that focuses on periods when oil prices remain relatively constant but oil-price volatilty fluctuates significantly. Using this methdology, we find empirical evidence that movements in oil-price volatility have a measurable impact on firms in the oil industry. During the period1998-2002, this impact is only detectable for small firms, while during the periods 1993-1997 and 2003-2008, movements in oil-price volatility impact stock prices across the oil industry.

Acknowledgment

The authors would like to express their gratitude to RC Research, Inc. for providing oil futures price data for this research project.

References

Aguerrevere, FL. 2009. "Real Options, Product Market Competition, and Asset Returns." *Journal of Finance*, 64: 957–983.

Al-Mudhaf, A, Goodwin, TH. 1993. "Oil Shocks and Oil Stocks: evidence from the 1970s." *Applied Economics* 25, 181-190.

Berk, J, Green, RC, and Naik, V. 1999. "Optimal Investment, Growth Options and Security Returns." *Journal of Finance* 54, 1153-1607.

Boyer, MM, Filion, D. 2007. "Common and Fundamental Factors in Stock Returns of Canadian Oil and Gas Companies." *Energy Economics* 29 (3), 428-453.

Carhart, MM. 1997. "On Persistence in Mutual Fund Performance." *Journal of Finance* 52, 57-82.

Carlson, M, Fisher, A, and Giammarino, R. 2004. "Corporate Investment and Asset Price Dynamics: Implications for the Cross-Section of Returns." *Journal of Finance* 59, 2577-2603.

Carlson, M, Fisher, A, and Giammarino, R. 2006. "Corporate Investment and Asset Price Dynamics: Implications for SEO Event Studies and Long-run Performance." *Journal of Finance* 61, 1009-1034.

Chen, NF, Roll, R, Ross, SA. 1986. "Economic Forces and the Stock Market." *The Journal of Business* 59, 383-403.

Ciner, C. 2001. "Energy Shocks and Financial Markets: Nonlinear Linkages." *Studies in Non-Linear Dynamics and Econometrics* 5, 203-212.

Cooper, I. 2006. "Asset Pricing Implications of Nonconvex Adjustment Costs and Irreversibility of Investment." *Journal of Finance* 61, 139-170.

Driesprong, G, Jacobsen, B, Maat, B. 2008. "Striking Oil: Another Puzzle?" *Journal of Financial Economics* 89, 307-327.

Fama, EF, and French, KR. 1992. "The Cross-Section of Expected Stock Returns." *Journal of Finance* 47, 427-465.

Fama, EF, and French, KR. 1993. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics* 33, 3-56.

Fama, EF, and French, KR. 1996. "Multifactor Explanations of Asset Pricing Anomalies." *Journal of Finance* 51, 55-84.

Gomes, J, Kogan, L, and Zhang, L. 2003. "Equilibrium Cross Section of Returns." *Journal of Political Economy* 111, 693-732.

Grullon, G, Lyandres, E, and Zhdanov, A. 2010. "Real Options, Volatility, and Stock Returns." Working Paper, Rice University and University of Lausanne

Guidi, M, Russell, A, Tarbert, H. 2006. "The Effect of OPEC Policy Decisions on Oil and Stock Prices." *OPEC Review* 30, 1-18.

Henriques, I, and Sadorsky, P. 2010. "The Effect of Oil Price Volatility on Strategic Investment." *Energy Economics*, forthcoming, doi 10.1016/j.eneco.2010.09.001

Huang, RD, Masulis, RW, Stoll, HR. 1996. "Energy Shocks and Financial Markets." *Journal of Futures Markets* 16, 1-27.

Ingersoll JE Jr, Ross, SA. 1992. "Waiting to Invest: Investment and Uncertainty." *Journal of Business* 65:1-29.

Johnson, K. 2009. "Oil Shock: Did High Oil Prices Cause the Recession?" *Wall Street Journal*, April 22, 2009, available at http://blogs.wsj.com/environmentalcapital/2009/04/22/oil-shock-did-high-oil-prices-cause-the-recession/

Jones, CM, Kaul, G. 1996. "Oil and the Stock Markets." *Journal of Finance* 51, 463-491.

Kansas, D. 2009. "Oil Spike Raises Portfolio Questions." *Wall Street Journal*, May 17, 2009, available at http://online.wsj.com/article/SB124251684528027290.html

Kogan, L. 2004. "Asset Prices and Real Investment." *Journal of Financial Economics* 73, 411-431.

McDonald R, Siegel D. 1986. "The Value of Waiting to Invest." *Quarterly Journal of Economics* 101: 707-727.

Myers, SC. 1977. "Determinants of Corporate Borrowing." *Journal of Financial Economics* 5, 147-175.

Myers, SC, and Majluf, NS. 1984. "Corporate Financing and Investment Decisions when Firms have Information that Investors do not Have." *Journal of Financial Economics* 13, 187-221.

Nandha, M, Faff, R. 2008. "Does Oil Move Equity Prices? A global view." *Energy Economics* 30, 986-997.

Newey, W, and West, KD. 1987. "A Simple, Positive Semi-definite, Hetereoskedasticity and Autocorrelation Consistent Convariance Matrix." *Econometrica* 55, 703-708.

OPEC, 2010. "The Oil Industry: How Long does it take to Discover Oil and Bring it to the Market?" http://www.opec.org/opec_web/en/press_room/179.htm

Park, JW, Ratti, RA. 2008 "Oil Price Shocks and Stock Markets in the U.S. and 13 European Countries." *Energy Economics* 30, 2587-2608.

Portillo, J. 2000. "Dynamic Behavior of the U.S. Oil Industry." *Resource and Energy Economics* 22, 125-145.

Quigg, L. 1993. "Empirical Testing of Real Option-Pricing Models." *Journal of Finance* 48, 621-640.

Sadorsky, P. 1999. "Oil Price Shocks and Stock Market Activity." *Energy Economics* 21, 449-469.

Sadorsky, P. 2001. "Risk Factors in Stock Returns of Canadian Oil and Gas Companies." *Energy Economics* 23, 17-28.

Venezia I, Brenner M. 1979. "The Optimal Duration of Growth Investments and Search." *Journal of Business* 52: 393-407.

White, H. 1980. "A Heteroskedasticity-consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* 48, 817-838.

Zhang, L. 2005. "The Value Premium." *Journal of Finance* 60, 67-103.

The U.S. Personal Income Tax Reform: Linear and Gradual Tax Simplifications

Robert Kao, Park University
John Lee, Rigel Technology Corporation

Abstract

The complexity of the U.S. personal taxations has long been an imminent subject discussed by many legislators and policymakers. Many recommendations and proposals on the possible impacts on the economy, revenue, equity, and efficiency are emphasized. A great increase of citizens and government offices would like to reduce tax processing time and operating costs with a simplified tax system. In this paper, a new simple linear and gradual (LG) tax system has been developed and analyzed to compare with the current progressive tax system. The implied LG tax system could not only simplify our current complicated federal tax systems but also reduce substantial processing time and managing costs for individuals, businesses, and governments. The research findings include the estimations and projections of tax revenues from business and individual income taxes for federal government.

Introduction

The re-election of President Obama has set in motion over the expiration of the Bush-era tax cuts. Nearly $100 billion in automatic spending cuts and more than 50 expiring tax extenders, which include the alternative minimum tax patch for tens of millions of taxpayers. The elected President has consistently called for higher tax with upper bracket individual income tax rates and also advocated maximum capital gains rate increases. On the whole, we will anticipate some significant changes with the dynamics for the tax reform in the long run.

The existing complex federal tax systems have 6 tax brackets for individual taxes. More individuals and government officers would like to simplify our tax systems to reduce tax processing time and operating cost. This research paper has been developed a new tax system with fair tax rates by linear and gradual (LG) formulas, which can be used to simplify and replace the existing complex 12-page Tax Table, multi-tax brackets, schedules, and computations.

LG tax rates can be modified easily and reasonably during a special situation, such as a recession or a booming economy. Tax rate differences from the proposed LG tax system and the existing tax systems can be miniscule (0 to 1%). The LG tax system could be applied for combining filing status, taxable income, tax rate formula, and tax rate range check together in a simplified format for tax rate classifications and tax calculations, which can be proceeded automatically or manually. The range check is used as a tool for checking the accuracy of the calculations. The LG tax system can also be used for both federal and state individuals to simplify our existing tax systems and save tax processing time and costs.

Tax Law Enactment and Literature Reviews

Congress and the president collaborated to enact the Taxpayer Relief Act of 1997 to reduce the burden of middle-income taxpayers with children, homes, and other assets earning long-term capital gains. The result of income the tax changes benefited tax credits to those families with children, tax credits for college tuition expenses, and new incentives to save for retirement. Most of these changes have affected the distribution of some tax burdens while increasing both the complexity of the U.S. tax code and the excess encumbrance of the tax system in the U.S.

In 2001, Congress enacted another tax relief policy, called the Economic Growth and Tax Relief Reconciliation Act (EGTRRA 2001.) The main content of this new tax legislation embraced a reduction in marginal tax rates (MTRs) for all brackets. The top rate, which was set at 39.6 percent in 2000, was reduced to 35 percent. In addition, many other changes were made in the tax law to reduce the “marriage penalty” controversy that was phased in beginning of 2005 and taken into a full effect in 2009. The EGTRRA has increased in tuition credits, child credits, and in allowable contributions to tax-deferred retirement accounts.

Later, in response to the president’s effort to stimulate a sluggish economy, Congress enacted the Jobs and Growth Tax Relief Reconciliation Act of 2003 (JGTRRA 2003). This legislation accelerated many of the provisions of EGTRRA to become more effective in 2003 and 2004. The Act lowered MTRs and also accelerated the marriage penalty relief in these two years. The legislation not only reduced tax rates for capital gains and dividends, but also enacted special provisions to encourage business investment activities.

All the tax changes enacted would expire at the end of 2010 unless renewed by the Congress. The reason for this restriction was because Congress was apprehensive about the budget balance projections. As a precaution, they have automatically programmed in sharp tax increases for 2011, which obligates future lawmakers to raise taxes unless funds are available to finance the tax rate reductions. Tax relief was to be phased in for 10 years unless Congress renewed it. The reduction in tax rates on capital gains and dividends enacted in 2003 was set to expire by 2009. Several changes in the income tax code were made in the following two years.

President Obama campaigned with a pledge to cut income taxes for households with less than $250,000 annual income. The major tax cut for middle-income families would be accompanied by increases in MTRs on the upper 2 percent of income earners to levels that prevailed prior to enactment of EGTRRA. Also, he promised tax cuts for senior citizens, for those who without health insurance, first-time homebuyers, and families with college age children. Another promise included consolidation of tax credits and measures to simplify the income tax. Any changes made to the tax code and redistribution of the tax burden toward upper income groups would take effect by 2011. However, some special tax cuts were incorporated into the American Recovery and Reinvestment Act of 2009 (ARRA 2009). These special emergency tax benefits were designed to encourage spending for homes and automobiles so as to stimulate sectors of the depressed economy.

Several linear income tax proposals have been discussed and analyzed in the last four or five of decades. Based upon the sum of individual utilities, the income taxation theoretically should improve income distribution and social welfare. Eytan Sheshinski (1972) supported the theorem that if the supply of labor in the economy is a non-decreasing function of the net wage rate then the optimal tax schedule of linear tax function would provide a positive lump-sum at zero income. Also, the optimal marginal tax schedule is bounded by a fraction of minimum of elasticity of the labor supply.

Slemrod, Yitzhaki, and Mayshar (1994) have investigated two-bracket piecewise linear income tax structures. They analyzed the optimal structure of taxation for the social welfare function, utility function, and distribution with the standard optimal linear income tax system. They concluded that a linear income tax offers considerable administrative advantages over more complex graduated income tax systems. They criticized that "historically, countries have accepted the higher administrative costs in order to achieve a more progressive distribution of the tax burden than that offered by a linear income tax system." They challenged the pervasive policy and showed the benefits of allowing two brackets rather than one to parameterize the problem.

In recent study, Diamond and Saez (2011) has analyzed the optimal progressivity of earnings taxation and considered if capital income should be taxed. They argued that "a result from basic research is relevant for policy only if it is based on economic mechanisms that are empirically relevant and first order to the problem." Hence, they placed high implicit marginal tax rates on low earners in the models with only an intensive margin for low earners. Moreover, they used a decreasing marginal tax rate on very high earners by using the lognormal distribution for the Pareto efficiency. They opposed zero taxation of capital income from the aggregate efficiency result, and debated the Atkinson-Stiglitz theorem because savings rates are not uniform in the civilian.

Federal Personal Income Tax Rate Models

The existing federal individual tax rates are presented as below. *TI* is taxable income and *Yi* represent different levels of taxable income. The base tax rates are *f, h* and *j*. The tax rate (TR), marginal tax rate (m), tax amount (T), and Tax Table (TT, Tax Guide for Individuals 2011) can be expressed as the formula below.

1. The existing federal personal income tax rate models

- *If $TI \leq Y_1$ ($100,000), T = TT and average tax rate is T/TI (See Tax Table 2011)*
- *If $Y_1 < TI \leq Y_2$, then T = 0.25 TI – a with marginal tax rate of 0.25 and tax rate is 0.25 - a / TI (Tax rate increases fast at first and then slow)*
- *If $Y_2 < TI \leq Y_3$, then T = 0.28 TI - b with marginal tax rate of 0.28 and tax rate is 0.28 – b / TI*
- *If $Y_3 < TI \leq Y_4$, then T = 0.33 TI - c with marginal tax rate of 0.33 and tax rate is 0.33 – c / TI*
- *If $TI > Y_4$, then T = 0.35 TI – d with marginal tax rate of 0.35 and tax rate is 0.35 – d / TI*

The total levied tax amount can then be summed up the above equations. However, it is not a function of *TI*. A, B, C or D is the constant of a, b, c or d times tax return number during its tax income range.

$$\text{Total levied tax amount} = \sum T(\text{Tax Table}) + 0.25\sum TI - A + 0.28\sum TI - B + 0.33\sum TI - C + 0.35\sum TI - D \quad \ldots (1)$$

2. The proposed federal individual tax rate models

- *If* $TI \leq Y_1$*, then* $TR = f + TI/g$ — *The levied tax is* $TR \times TI$
- *If* $Y_1 < TI \leq Y_{2'}$*, then* $TR = h + TI/i$ — *The levied tax is* $TR \times TI$ *(Tax rate increases linearly)*
- *If* $Y_{2'} < TI \leq Y_4$*, then* $TR = j + TI/k$ — *The levied tax is* $TR \times TI$
- $TI > Y_4$*, then* $TR = m - d/TI$ — *The levied tax is* $TR \times TI$ *(Tax rate increases gradually)*

The total levied tax amount can then be summed up the above equations and shown as a function of *TI*. The formula can be expressed as below.

Total levied tax amount$=f(TI)=f\sum TI + \sum TI^2/g + h\sum TI + \sum TI^2/i + j\sum TI + \sum TI^2/k + m\sum TI - D$..... (2)

By the equation (2), total tax can be calculated when *TI* values are known or as a function of *TI*. These constants such as *f*, *g*, *m*, and *d* can be adjusted quickly according to actual situations and needs. By the equation (1), the first part is not a function of *TI*, which needs to be obtained from the complex Tax Table. Also, most of the personal taxable incomes are less than $100,000. The subtotal tax in the right parts in the equation (2) can be calculated when *TI* values are known.

Implications

1. Existing federal tax system for individuals

In our existing federal tax system for individuals, there are 6 tax brackets, which are 10%, 15%, 25%, 28%, 33% and 35% with tax rates from 10% to 35%. There are four filing statuses: (1) Married filing jointly or qualifying widow(er); (2) Head of household; (3) Single and (4) Married filing separately. For married filing jointly, the tax rate schedule shows the tax rate is at 10% to tax incomes from 0 to $17,000, which is shown in Table 3. When taxable incomes are from 0 to $100,000, the 12-page Tax Table (2011) is used for individuals to search and find their tax payments. These tax numbers in the 12-page Tax Table (2011) have no direct connection or relationship. The tax data in the Tax Table can be stored into a tax software product with more data space and search program, which is used for automatic search. Table 2 shows the differences of the Tax Table, taxable income ranges and tax computations in 2011 and 2010. These differences make our existing tax system even more complex and increase related processing time and operating costs to the government and individuals.

Table 1: 2011 Federal Tax Table for Married Filing Jointly or Qualifying Widow(er) (12 pages)

Taxable income (TI)	*Tax is*	*Taxable income (TI)*	*Tax is*	*Taxable income (TI)*	*Tax is*
0 – 5	0	10,000 – 10,050	1,003	………………………..	
…………………………		10,050 – 10,100	1,008	85,900-85,950	13,731
5,000-5,050	503	……………………………		85,950-86,000	13,744
5,050-5,100	508	50,000 –50,050	6,654	……………………………	
………………………		50,050 –50,100	6,661	99,950-100,000	17,244

Table 2: Federal Tax Systems (2010 and 2011) for Married Filing Jointly or Qualifying Widow(er)

Taxable income (TI)		*2010 Tax*	*Taxable income (TI)*		*2011 Tax*
Over	*Not over*		*Over*	*Not over*	
0	100,000	Tax Table (12 pages)	0	100,000	Tax Table (12 pages)
100,000	137,300	0.25×TI - 7,637.5	100,000	139,350	0.25×TI-7,750
137,300	209,250	0.28×TI - 11,756.5	139,350	212,300	0.28×TI-11,930.5
209,250	373,650	0.33×TI - 22,219	212,300	379,150	0.33×TI-22,545.5
373,650		0.35×TI - 29,692	379,150		0.35×TI-30,128.5

Table 3: Federal Tax Rate Schedule for Married Filing Jointly or Qualifying Widow(er) (2011)

Taxable income (TI) *Over* *Not over*	*Tax is*	*The Amount is over*	*Tax Computation*
0 - 17,000	10%		
17,000 - 69,000	$1,700 + 15%	$17,000	$1,700 + 0.15 × (TI - 17000)
69,000 - 139,350	$9,500 + 25%	69,000	9,500 + 0.25 × (TI - 69000)
139,350 - 212,300	$27,087.5 + 28%	139,350	27,087.5 + 0.28 × (TI - 139350)
212,300 - 379,150	$47,513.5 + 33%	212,300	47,513.5 + 0.33 × (TI - 212300)
379,150	$102,574 + 35%	379,150	10,257.4 + 0.35 × (TI - 379150)

Table 4 below shows the federal tax system for Head of household with 12-page of Tax Table, different taxable income ranges, and tax computations. Also, there is no self-check tool for tax rates. For Single and Married filing separately, their situations are similar, in which their Tax Table, taxable income ranges and tax computations are also different in 2010 and 2011. They have been checked and are different every year since 2004. Our existing federal tax system for individual income is very complex, which could involve long processing time and high operating cost.

Table 4: Federal Tax Systems for Head of Household

Taxable income (TI) *Over*	*Not over*	*2010 Tax*	*Taxable income (TI)* *Over*	*Not over*	*2011 Tax*
0	100,000	Tax Table (12 pages)	0	100,000	Tax Table (12 pages)
100,000	117,650	0.25×TI - 5,152.5	100,000	119,400	0.25×TI - 5,232.5
117,650	190,550	0.28×TI - 8,682	119,400	193,350	0.28×TI - 8,814.5
190,550	373,650	0.33×TI - 18,209.5	193,350	379,150	0.33×TI - 18,482
373,650		0.35×TI - 25,682.5	379,150		0.35×TI - 26,065

The complexity of the existing federal tax systems on Tax Tables, tax schedules, taxable income ranges, and tax computations could be simplified and improved. The processing time and operating cost could then be reduced or saved significantly. The Tax Table provides unfair tax rates to individuals with low taxable incomes (less than $1,000), which are such as 19.6%, 15.9% or 11%, which are shown in Table 7.

2. The proposed linear and gradual tax rate formula

For the existing tax system for Married Filing Jointly or Qualifying Widow(er), there are two parts. One part is the 12-page Tax Table and the second part includes the four tax computations, which are shown in Tables 2 and 3. When individuals have their taxable incomes from 0 to $100,000, they need to locate their tax numbers from the 12-page Tax Table (2011), which is rather lengthy and time consuming. Also, there is no self-check tool for checking tax rates.

The significant challenge is how to build relatively simple and fair formula(s) to replace the tax rates and tax data from the 12-page Tax Table. An individual with more taxable income (*TI*) should pay more tax at a reasonable high tax rate. For individuals with taxable incomes from 0 to $100,000, a linear formula of $y = a + bx$ is found to match the tax rates from the 12-page Tax Table reasonably.

Tax rate = 0.1 + TI/1,380,453 (tax rate range check: 0.1- 0.173 at 0 - $100,000) ………..…. (1/1)

Here 1/1,380,453 (b) is a constant, which is the slope of $y = a + bx$ from 1/100,000/(0.17244-0.1)=1/1,380,453. The tax number of 0.17244 (17,244/100,000) is from the Tax Table (2011). Tax rates change linearly over taxable incomes from 0 to $100,000. The bottom tax rate is 0.1 or 10% (a).

Example: When a Married filing jointly has a taxable income of $39,960, the tax rate formula is 0.1+TI/1,380,453 with the range check (10%-17.3%). Then 0.1+39,960/1,380,453=12.89% is the tax rate, which is within the range. The tax is $5,152.72 (= Tax rate × TI). Tax rate and tax calculations can be done automatically or manually. The tax from $39,950 to $40,000 in the 2011 Tax Table is $5,146 at 12.88%. The difference $50 (40,000-39,950) from the Tax Table causes 0.13% (50/39,975). Their tax rate difference from the LG formula (1/1) and 2011 Tax Table is only 0.01% (12.89%-12.88%).

When the simple LG tax rate formula (1/1) is used to replace the 12-page Tax Table (2011), the situations have been simplified and improved significantly. Their results are compatible in the following discussions. The LG tax system has the narrow tax rate range check from 10% to 17.3% as a self-check tool to avoid or reduce calculation mistakes.

The other 3 linear and gradual (LG) tax rate formulas of linear $y = a + bx$ and gradual $y = c - d/x$ are found to match the tax rate schedules and tax computations (2011), which are shown in Table 5. The four easy taxable income ranges are designed into 0-\$100,000, \$100,000-\$250,000, \$250,000-\$400,000 and over \$400,000. Their filing status, taxable income, tax rate formula and range check are related together, which may be used for tax rate and tax calculations, tax projection and tax data analysis.

Table 5: LG Tax System for Married Jointly or Qualifying Widow(er)

Filing Status	*Taxable Income (TI)* *Over*	*Not over*	*Tax Rate Formula*	*(Range Check)*
(1/1)	0	100,000	0.1 + TI/1,380,453	(0.1-0.173)
(1/2)	100,000	250,000	0.1275 + TI/2,226,312	(0.172-0.240)
(1/3)	250,000	400,000	0.1817 + TI/4,299,791	(0.239-0.275)
(1/4)	400,000		0.35 - 30,128.5/TI	(0.274-0.35)

For Head of household, there are also two parts. One part is the 12-page Tax Table and the second part includes the four tax computations with different taxable income ranges, which are shown in Tables 6. For Head of household individuals with taxable incomes from 0 to \$100,000, a linear formula of $y = a + bx$ is found to match the tax rates and tax data from the 12-page Tax Table reasonably.

Tax rate = 0.1 + TI/1,024,485 (tax rate range check: 0.1-0.198 at 0 - \$100,000) …………..… (2/1)

Here 1/1,024,485 (b) is a constant, which is the slope for $y = a + bx$ from 1/100,000/(0.19761-0.1)=1/1,380,453. The tax number of 0.19761 (19,761/100,000) is from the Tax Table (2011). Tax rates change linearly. The bottom tax rate is 0.1 or 10% (a).

When the taxable income range is from over \$100,000 to \$250,000, the linear tax rate formula is:

Tax rate = 0.1586 + TI/2,565,769……………………………………………………………….. (2/2)

Here b=1/(250000-100000)/(0.256072-0.19761)=1/2,565,769 is the slope. The number of 0.256072 is from 0.33-18482/250000 at TI=250000. The number a is from 0.19761-100,000/2,565,769=0.1586. The tax rate range check is within 0.197-0.257 at \$100,000 to \$250,000. These numbers and formula are found to match the existing tax schedules and tax computations (2011).

When the taxable income range is over \$400,000, a gradual tax rate formula of $y = c - d/x$ is:

Tax rate = 0.35 – 26,065 / TI …..…………………………………………………………….. (2/4)

Here the formula 2/4 is converted from (0.35×TI-26,065)/TI (2011). When taxable incomes are over \$400,000, their tax rate range is within 0.284-0.35.

Besides (1) Married filing joint or Qualifying Widow(er) and (2) Head of household, other LG formulas and tax rate range checks for other filing statuses of (3) Single and (4) Married filing separately are also found, which are shown in Table 6.

Table 6: LG Tax System for Federal Individuals

Filing Status: (1) Married filing jointly (2) Head of household (3) Single (4) Married filing separately
(The formulas match 2011 Tax Table, computations and schedules)

Filing Status	*Taxable income (TI) Over*	*Not over*	*Tax rate formula*	*(Range check)*
(1/1)	0	100,000	0.1 + TI/1,380,453	(0.1-0.173)
(1/2)	100,000	250,000	0.1275 + TI/2,226,312	(0.172-0.240)
(1/3)	250,000	400,000	0.1817 + TI/4,299,791	(0.239-0.275)
(1/4)	400,000		0.35 - 30,128.5/TI	(0.274-0.35)
(2/1)	0	100,000	0.1 + TI/1,024,485	(0.1-0.198)
(2/2)	100,000	250,000	0.1586 + TI/2,565,769	(0.197-0.257)
(2/3)	250,000	400,000	0.2080 + TI/5,208,184	(0.256-0.285)
(2/4)	400,000		0.35 - 26,065/TI	(0.284-0.35)
(3/1)	0	60,000	0.1 + TI/701,643	(0.1-0.186)
(3/2)	60,000	100,000	0.1396 + TI/1,307,770	(0.185-0.217)
(3/3)	100,000	250,000	0.1804 + TI/2,804,367	(0.216-0.270)
(3/4)	250,000	400,000	0.2300 + TI/6,318,896	(0.269-0.291)
(3/5)	400,000		0.35 - 22,686/TI	(0.29-0.35)
(4/1)	0	70,000	0.1 + TI/737,730	(0.1-0.195)
(4/2)	70,000	100,000	0.1357 + TI/1,183,236	(0.194-0.221)
(4/3)	100,000	150,000	0.1511 + TI/1,446,480	(0.22-0.255)
(4/4)	150,000	200,000	0.1953 + TI/2,521,008	(0.254-0.275)
(4/5)	200,000		0.35 - 15,064/TI	(0.274-0.35)

There are about 79 million individual tax returns in the U.S. each year. These simple linear and gradual tax rate formulas in the LG tax system provide a good tool for the government and individuals to calculate tax, estimate tax projection and analyze tax data.

3. Tax rate difference from the existing tax system and LG tax system

The tax rate differences between the 2011 Tax Table (for Head of household) and the LG formula (2/1) are compared in Table 8 with compatible results (0-1% differences). Also in the Tax Table (2011), the tax rates with 19.6%, 15.9%, 12.0%, and 10.1% (*) for the low taxable income ranges from $5 to $1,000 are unreasonable because the tax rates are from 10.0% to 19.8% for the taxable incomes from $10,000 to $100,000. These tax rates for the taxable incomes from 0 to $100,000 should increase gradually from 10% to 19.8%. The tax rates calculated from the formula (2/1) are from 10.0% to 19.8% linearly and gradually, which are simple, reasonable and practical.

When taxable incomes are from $100,000 to $10,000,000, their tax rate differences between the 2011 tax computations and the LG formulas are compared and shown in Tables 7 with very compatible results (0-0.4% differences). The tax rate data in these tables may also be converted into figures to show their differences visually. These minor differences mean the results from LG tax rate formulas and our existing Tax Table, tax schedules and computations are matched each other. The LG tax system is much more simple and practical. Similar data and situations are found in other filing statuses of Married filing jointly, Single and Married filing separately.

Table 7: Comparison of Tax Rates between 2011 Tax Table and LG Formula for Head of Household (Taxable Income: 0 - $10,000,000)

Taxable Income ($)	*2011 Tax Table Tax ($)*	*2011 Tax Table Tax Rate*	*from Formula (2/1) (0.1+TI/1,024,485)*	*Difference (Absolute Value)*
5.1	1	19.6% *	10.0%	9.6%
25.1	4	15.9% *	10.0%	5.9%
50.1	6	12.0% *	10.0%	2.0%
100.1	11	11.0% *	10.0%	1.0%
1,000	101	10.1% *	10.1%	0.0%
10,000	1,003	10.0%	11.0 %	1.0%
20,000	2,396	12.0%	12.0%	0.0%
30,000	3,896	13.0%	12.9%	0.1%
40,000	5,396	13.5%	13.9%	0.4%
50,000	7,274	14.6%	14.9%	0.3%
60,000	9,774	16.3%	15.9%	0.4%
70,000	12,274	17.5%	16.8%	0.7%
80,000	14,774	18.5%	17.8%	0.7%
90,000	17,274	19.2%	18.8%	0.4%
100,000	19,761	19.8%	19.8%	0.0%

Taxable Income ($)	*2011 Tax Rates (Tax/Taxable Income)*	*LG Formulas (Table 6)*	*Difference (Absolute Value)*
100,000	19.8%	19.8%	0.0%
120,000	20.7%	20.5%	0.2%
140,000	21.7%	21.3%	0.4%
160,000	22.5%	22.1%	0.4%
180,000	23.1%	23.0%	0.1%
200,000	23.8%	23.7%	0.1%
230,000	25.0%	24.8%	0.2%
250,000	25.6%	25.6%	0.0%
290,000	26.6%	26.4%	0.2%
320,000	27.2%	26.9%	0.3%
350,000	27.7%	27.6%	0.1%
380,000	28.1%	28.1%	0.0%
400,000	28.5%	28.5%	0.0%
1,000,000	32.4%	32.4%	0.0%
10,000,000	34.7%	34.7%	0.0%

4. Tax rate modification and tax data analysis

If a special situation is taking place such as a recession, tax rates may be reduced from 10%-35% to such as 5%-30% for people to pay less tax and have more money for purchasing more goods, which may promote more sale and production and reduce unemployment rate, then LG tax rate formulas are subtracted by 3%, 5% or any reduction easily. For example, for Married Filing Jointly or Qualifying Widow(er) individuals with their taxable incomes from 0 to $100,000, their LG tax rate formula of 0.1+TI/1,380,453 may be modified into of 0.05+TI/1,380,453 easily. Their tax rate range is changed from 0.1-0.173 into 0.05-0.123. A reasonable reduction may be applied according actual situations.

Tax data analysis such as total tax or tax difference from the group may be analyzed with the LG formulas easily. When the existing Tax Table (12 pages) is changed to meet the reduction such as 5% and analyze tax data, there is much more work involved to meet these goals. For Married jointly or Qualifying Widow(er) (2011), total tax is:

$$Total\ Tax = 0.1\sum TIo + \sum(TI^2)o/1,380,453 + 0.1275\sum TIp + \sum(TI^2)p/2,226,312 + 0.1817\sum TIq + \sum(TI^2)q/4,299,791 + 0.35\sum TIr - 30,128.5\times r \quad \ldots\ldots\ldots\ldots\ldots\ldots (7)$$

The equation(s) can be used for tax analysis and projection. Here o, p, q and r are individual numbers. The two simple math equations of $y = a + bx$ (linear) and $y = c - d/x$ (gradual) are used in the LG Tax System, in which the constants of a, b, c and d may be modified easily according to actual situations.

Another example is for a taxable income range of $100,000 to $250,000, we may design tax rates to 20% at taxable income $100,000 and 26% at $250,000. Then their linear tax rate formula is 0.16+TI/2,500,000. Here 2,500,000 is from 1/(250,000-100,000)/(0.26-0.2)=1/2,500,000. Their tax rates change linearly within the taxable income range and tax rate range from 20% to 26%.

Conclusion

The complication of the U.S. personal taxations has long been recognized as an imminent subject discussed by many legislators and policymakers. As a result of the expired Bush tax cuts, there is an anticipation of tax increase in 2013. In this paper, a proposed new simple linear and gradual (LG) tax system has been developed and analyzed through a comparison of the current progressive tax system with this simplified LG system. Besides 2011, the LG tax system can also be modified for 2012 and 2013.

The proposed tax simplification process suggests that all filing statuses, taxable incomes, incomes, tax rate formulas, tax rate range checks, taxpayer information, tax rate, and tax calculations be combined into short tax simplification phases. Tax rate formulas could then connect to the related filing statuses, taxable income, and income brackets. Subsequently, the total tax amounts will be calculated automatically or manually with simple procedures.

Numerous benefits can be demonstrated by applying this LG system. The taxable income ranges would reduce to three simple fixed levels instead of the current Tax Table plus several variable levels. The proposed method would simplify the taxpayers' reporting process and reduce the preparation time. Consequently, the new LG formulas could provide more fine-tuned tax rates for the tax amount calculations. Furthermore, the proposed tax rates could be easily tested and monitored by the tax office and taxpayers.

In addition, the long list of tax tables can be eliminated and replaced by the modest formulas. The proposed tax rates represent the minor tax rate differences comparing to the current system, which are resulting only within an insignificant 1% of margins. For the what-if tax rate analysis, the tax office could change the rates and estimate the revised tax revenue easily. The streamline of tax analysis and revenue projection would enhance the efficiency of the Revenue Office. Overall, we could anticipate the increase of total tax revenue, which is resulted from the reduction of tax preparation time, tax rates analysis, and tax amount projections.

References

Diamond, Peter and Emmanuel Saez (August, 2011). "The Case for A Progressive Tax: From Basic Research to Policy Recommendations." *CESIFO* Working Paper, NO. 3548, Category 1: Public Finance.

David N. Hyman (2011). *Public Finance, A Contemporary Application of Theory to Policy*, 10th Ed., South-Western, Cengage Learning.

Economic Growth and Tax Relief Reconciliation Act of 2001 (EGTRRA June 7, 2001). Public Law 107–16, 115 Stat. 38.

Sheshinski, Eytan (July, 1972). The Optimal Linear Income-Tax, *The review of Economic Studies*, Volume 39, Issue 3:297-302.

Slemrod, Joel, Shlomo Yitzhak,i and Joram Mayshar (August, 1994). "The Optimal Two-Bracket Linear Income Tax." *Journal of Public Economics* 53: 269-290. North-Holland, The Prime Minister's Office, Stockholm S-103 33, Sweden.

Tax Brackets and Tax Rates 2012 (January 23, 2012). U.S. Tax Center, Elizabeth Rosen, http://www.irs.com/articles/tax-brackets-and-tax-rates.

Tax Guide 2011 for Individuals, Internal Revenue Service, Department of Treasury, Federal Catalog Number 10311G (http://www.irs.gov/pub/irs-pdf/p17.pdf).

Taxpayer Relief Act of 1997 (August 5, 1997). Pub.L, 105–34, H.R. 2014, 111 Stat. 787.

The American Recovery and Reinvestment Act of 2009 (ARRA February 17, 2009). Public Law 111–5.

The Jobs and Growth Tax Relief Reconciliation Act of 2003 (JGTRRA May 23, 2003). Public Law 108–27, 117 Stat. 752.

The Taxpayer Relief Act of 1997 (TRA August 5, 1997). Public Law 105–34, H.R. 2014, 111 Stat. 787.

Appendix

Figure 1.1: Filing Status – Head of Household with Taxable Income below $100,000

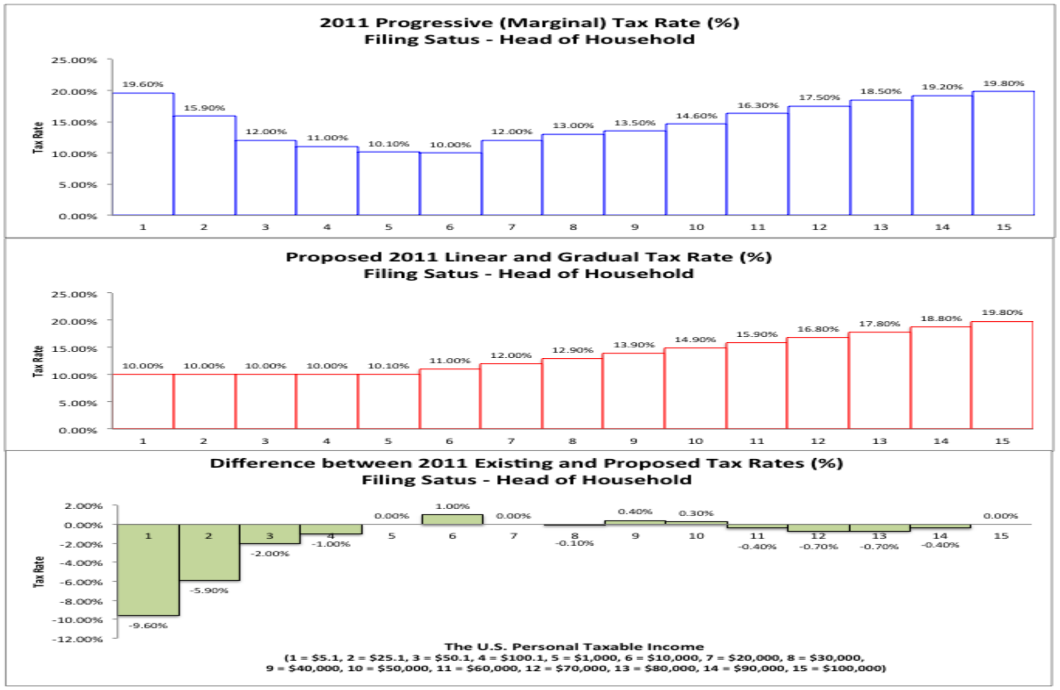
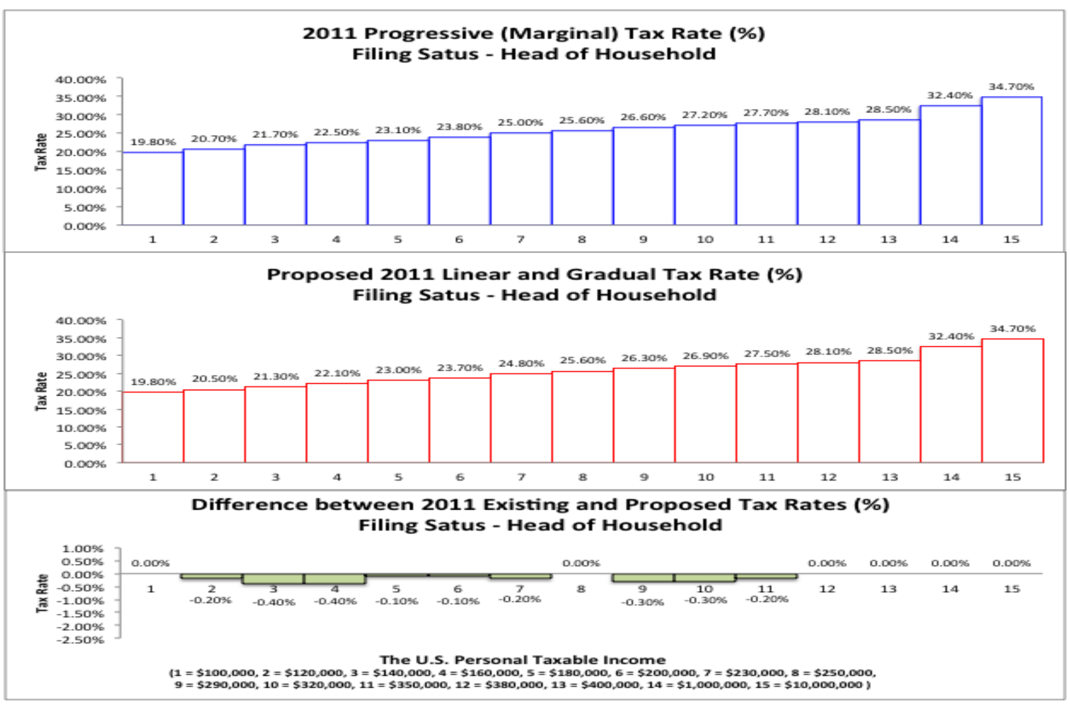


Figure 1.2: Filing Status – Head of Household with Taxable Income between $100,000 and $10,000,000



The Rate of Return Convergence and the Value Anomaly

Gary Moore, Doina Chichernea, and Mei Zhang, University of Toledo

Abstract

Convergence in the market to book ratios is thought to play an important role in explaining the value premium. Trends in the convergences of market-to-book ratios and rates of return between value and growth portfolios are examined over the 1970-2011 period. While the convergence of market-to-book ratios does not vary substantially over time, the speed of convergence in returns increased significantly in recent periods. This is consistent with both the presence of more intensive competition and the increasingly widespread use of style investing based on value and growth portfolio designations.

Introduction

It has been long documented in the finance literature that high book-to-market (mature, past growth or "value") firms exhibit higher average returns than low book-to-market ("growth") firms. This value premium is probably one of the most studied and robust financial anomalies, as it has been shown to persist across different time periods and different markets around the world (see Fama and French, 1998). Building on academic developments, value and growth are now widely recognized investment styles followed by money managers, and many of the style-specific benchmarks used in their analysis are based on the ratio of book-to-market value of equity (see Chan and Lakonishok, 2004, for a good review of the academic research literature on value and growth investing). While the existence of the value premium remains largely uncontested, its interpretation does not. The literature is rich with alternative proposed explanations for the mechanism behind this observed phenomenon, ranging from various sources of risk (rational pricing) to investors' biased estimations (irrational pricing).[1]

In the current study, we take an agnostic view on what are the underlying causes generating the cross-sectional return difference between value and growth stocks. Rather, we focus on the notion of persistence of the difference in value/growth characteristics and their respective return premiums through time. Specifically, we are interested in answering the following questions: (1) does the cross-sectional difference in value/growth characteristics (i.e. market-to-book ratios) converge to zero in the periods after portfolio formation and if so, how fast?; (2) does the cross-sectional difference in returns between value and growth stock (i.e. the value premium) converge to zero after portfolio formation and if so, how fast?; and finally (3) does the convergence pattern in returns follow the convergence pattern in market-to-book ratios? We are particularly interested whether the answers to these questions are time invariant, or are sensitive to general market conditions.

We find a strong tendency for the market-to-book (M/B) ratios of value and growth portfolios to converge through time. Our estimates suggest that there should be little significant difference in the market-to-book ratios between these groups in 5-7 years after portfolio formation. Furthermore the convergence of M/B ratios does not seem to be very sensitive to the sample period taken into account (i.e. there is no evidence that the convergence pattern is dependent on market conditions). However, the convergence of returns paints a different picture. Overall, returns of the extreme portfolios (value and growth) do seem to converge towards the mean, with an estimated horizon of about 6 years. However, in the early period of our sample (before style investing was popular), we see a much weaker convergence pattern than in the more recent period. Specifically, after the 1990s, we see a strong convergence pattern in returns, which appears to be much faster than before the 1990s.

The notion of convergence of value and growth characteristics of stocks and their respective returns has received some attention in previous literature. For example, Fama and French (2007a, b) investigate how the migration of stocks across style groups (i.e. value and size) contributes to the size and value premium in average stock returns. The authors define convergence as the expected decline in the high price-to-book (P/B) of growth stocks and the expected increase in the low P/Bs of value stocks. They propose a simple rational explanation supporting the view that convergence in profitability and growth is anticipated by investors and thus incorporated into prices. Value firms present a combination of weak growth in book equity (generated by low investment) and strong capital gains after they are categorized as value stocks, which together produce upward convergence in the P/Bs of value portfolios after formation. On the other hand, growth firms have high investment generating high book equity growth rates in the year after they are allocated to growth portfolios, which in turn creates the negative convergence of P/Bs as these companies exercise their growth options.

Alternatively, the behavioral view of the value anomaly also supports the notion of convergence, but for entirely different reasons. For example, Lakonishok, Shleifer, and Vishny (1994) argue that investors constantly overestimate (underestimate) the improvement (deterioration) in profitability and growth after stocks are allocated to value (growth) portfolios. This story would also manifest in the observed convergence, albeit in this case convergence is largely unexpected and consistently surprises investors.

Our study differs from the aforementioned literature in several aspects. First, most of the previous studies are motivated by the attempt to uncover the underlying mechanism generating the value premium. Consequently, these studies focus on the anatomy of the value/growth characteristics during year t+1, i.e. the year after portfolio formation (which is the year when the return premium is observed). Since our focus is on how long the process continues after portfolio formation, we are investigating a much longer horizon. We focus on a buy and hold strategy with formation at time t, and then follow the characteristics of our portfolios through event time for up to 4 years after portfolio formation.[2] Thus our definition of convergence is technically different than that employed by Fama and French (2007b), in the sense that ours basically refers to the speed of reversal of characteristics/returns through time.

Second, our methodology is entirely different, given that we are interested in how the market-to-book characteristics and the returns evolve month by month after portfolio formation. We employ a variation of the event-time regression methodology employed by Liu and Zhang (2008) and Ball and Kothari (1989). Specifically, each month, we form quintile portfolios based on stocks' market-to-book characteristics and then we calculate equal average returns and average market-to-book ratios for each of the quintiles for t+m, where m=1,2,….48. This methodology has the distinct advantage of allowing us not only to calculate the rate of convergence, but also to investigate whether the convergence process is similar for growth and value firms or whether some asymmetry exists.

Finally, although we do not take a stance on what is the underlying mechanism generating convergence, investigating the connection between the pattern of convergence for book-to-market-ratios and returns can help disentangle between various explanations proposed for the value premium. We discuss implications of our results, but we leave rigorous empirical testing of these hypotheses for future research.

Our results are of interest to several groups. From a practitioner's perspective, the obvious question is how often does a value investing strategy needs portfolio rebalancing? While we offer estimates over our entire sample period, our results imply that the answer to this question varies with economic conditions. From an academic perspective, investigating the connection between the two rates of convergence can help disentangle between various explanations proposed for the value premium. We further elaborate on these implications in the next section.

Background and Hypothesis Development

As briefly mentioned in the introduction, previous literature has paid some attention to the idea of the evolution (convergence) of the book-to-market characteristic, as well as the evolution of the returns of portfolios sorted on this characteristic. We briefly review the relevant papers in order to provide the framework and some structure for the current study.

Several studies that focus on the variation of firms' book-to-market ratios over time exist in the current literature. Some of these studies investigate the explanations behind the time variation in B/M, while others focus on how value-glamour style evolution helps explain future stock returns in the cross-section (i.e. the value premium). For example, Fama and French (2008) and Daniel and Titman (2006) investigate to what extent particular information reflected in the B/M evolution process drives the value effect in stock returns. Similarly, Watanabe (2008) examines the relationship between the B/M evolution and the value premium in international markets.

Specifically, the framework employed in these papers originates with the one proposed by Fama and French (2008), according to which B/M evolves through time through a process that can be illustrated by the following equation:

$$bm_{it} = log\frac{B_{it}}{M_{it}} = log\frac{B_{it-k}}{M_{it-k}} + log\frac{B_{it}}{B_{it-k}} - log\frac{M_{it}}{M_{it-k}} \tag{1}$$

where B_{it} is the book equity value of firm *i* at time *t*, M_{it} is the market value of firm *i* at time *t*, $\frac{B_{it}}{B_{it-k}}$ is the book return (rate of growth for book value) for firm *i* between the month *t-k* and *t*, and $\frac{M_{it}}{M_{it-k}}$ is the market return for firm *i* between the month *t-k* and *t*. The process in equation (1) basically describes the current *B/M* of firm *i* as a function of its past *B/M* (the first component), changes in its book value or book returns (the second component), and changes in its market values or market returns (the third component). Based on equation (1), B/M increases when book return is high/positive and exceeds market returns and it decreases when book return is low/negative and is dominated by market returns. Empirical evidence shows that growth stocks exhibit the first type of behavior and value stocks exhibit the second,[3] thus resulting in some form of conversion to the mean of the B/M characteristics of stocks in these extreme portfolios.

Fama and French (2008) show that the interplay between book returns and market returns is important in explaining how stocks migrate between value and growth. Based on the arguments put forth in Fama and French (1995), the authors argue that stocks move to high expected return value portfolios (high B/M) as a result of poor past profitability and low (often negative) growth in book equity accompanied by even sharper declines in stock prices. Conversely, stocks that move to lower expected return growth portfolios (high B/M) typically have high past profitability and growth in book equity along with even sharper past increases in stock prices.

The question of interest is whether the evolution of the returns process follows in any way the evolution of the book-to-market characteristic described above. Fama and French (2007 a, b) describe the return evolution as follows:

$$1 + R_{t,t+1} = \frac{D_{t+1}}{P_t} + \frac{P_{t+1}}{P_t} = \frac{D_{t+1}}{P_t} + \frac{B_{t+1}}{B_t} * \frac{P/B_{t,t+1}}{P/B_t} \tag{2}$$

Equation (2) above decomposes the return over the period *t* to *t+1* as a function dividend yield $\left(\frac{D_{t+1}}{P_t}\right)$, rate of growth in book equity (or book return) $\left(\frac{B_{t+1}}{B_t}\right)$ and rate of growth in P/B (growth in M/B ratio) $\left(\frac{P/B_{t,t+1}}{P/B_t}\right)$. The empirical results presented by Fama and French (2007b) focus on a period of one year after portfolio formation. Their findings confirm that in the year after stocks are allocated to value portfolios (high B/M), growth in book equity is minor or even negative, and the large capital gains of value stocks show up as increases in P/B (i.e. decreases in the B/M ratio). In contrast, companies invest heavily after they are allocated to growth portfolios (low B/M), resulting in growth rate of book equity much larger that the growth rate of stocks price (i.e. B/M increases on average). Consequently, the positive average rates of capital gain of growth portfolios trace to increases in book equity that more than offset the increases in B/M.

The story that develops based on the above summarized literature with respect to the evolution of B/M is relatively straight forward. High expected profitability and growth combine with low expected stock returns (i.e. low cost of equity capital) to produce low B/M for growth stocks, while in the same time low profitability, slow growth and high expected returns produce high B/M for value stocks. After stocks are categorized, evidence seems to suggest that there exists reversion towards the mean (convergence) in their B/M characteristics. From a rational perspective, the explanation advanced by Fama and French (2007 a, b) is based on the fact that competition erodes the high profitability of growth firms (additionally, their profitability also declines as they exercise some of their growth options). Therefore, each year, some of them cease to be profitable, and their B/M increases, making them eventually move from the growth category. Conversely, some value companies restructure, improve profitability and are rewarded by the market with lower costs of equity and higher stock prices, which results in lower B/M ratios, making them eventually move down from the high B/M category. From a behavioral perspective, the same type of convergence is expected, albeit for entirely different reasons. Regardless of whether one takes a rational or behavioral view of the world, evidence seems to imply that through time we would expect that B/M ratios converge towards the mean (i.e. the difference between the B/M of extreme portfolios converges towards zero). The questions that are of interest for our study are (1) how fast does this convergence occur and (2) is it the case that the speed of convergence varies through time. To the extent that competition has increased during recent times, we would expect a higher rate of convergence in B/M ratios in the later periods compared to that observed in earlier periods.

In terms of returns, the empirical evidence presented by Fama and French (2007b) seems to indicate that convergence in P/B and book equity growth are the most important components driving the return dynamics of value and growth stocks.[4] However, their study is limited to a one year horizon. To the extent that the result holds over longer horizons, this motivates us to believe that we should observe some similarities in terms of the convergence rates of the return process for extreme groups relative to the convergence rates of the B/M process.

Data and Methodology

Firm financial statement data are obtained from the annual COMPUSTAT file, and stock return data are obtained from the monthly CRSP files for the period from 1970 to 2011. Our initial sample includes all ordinary common shares listed on the New York Stock Exchange, the American Stock Exchange, and the NASDAQ.[5] We then eliminate firms with a negative book value of equity. We calculate book equity (BE) following the definitions in Davis, Fama and French (2000).[6] When constructing the B/M ratio, in order to ensure that accounting information is available to investors prior to portfolio formation and return accumulation periods, we match CRSP stock return and market values data from t+3 through t+14 (12 months) with accounting information from COMPUSTAT as of time t (i.e. we allow for one quarter for the accounting information to be revealed to the market).

To provide a better characterization of the stocks in our sample, we calculate simple measures of risk and risk adjusted returns such as beta, total volatility, and alpha. We estimate alphas and betas based on market model regressions over a window including the previous 36 months. We require a window of minimum 12 months for the stock to be included in the estimation.

We perform our analysis at a monthly level. Our final sample contains on average approximately 4000 stocks per monthly cross-section, over the 504 months covering our entire sample period. We sort stocks cross-sectionally into quintiles based on the value of their M/B ratio at the beginning of each month. We present descriptive statics for the overall final sample as well as for each of the M/B quintiles in Table 1 below.

Table 1: Descriptive Statistics

Panel A: Overall Sample Descriptive Statistics (monthly)

	Mean	***Median***	***p25***	***P75***	***StDev***
RET	0.012	0.001	-0.064	0.071	0.155
M/B	6.194	1.537	0.951	2.837	107.459
Beta	1.152	1.047	0.601	1.597	0.865
ME	1273.381	120.979	32.596	506.019	6389.854
Total Volatility	0.144	0.127	0.089	0.178	0.082
Alpha	0.003	0.003	-0.010	0.016	0.026

Panel B: Selected Descriptive Statistics Within Market-to-Book Quintiles

		Q1 (Value)	***Q2***	***Q3***	***Q4***	***Q5(Growth)***
RET	***Mean***	0.021	0.014	0.012	0.009	0.007
	Median	0.003	0.004	0.002	0.000	-0.004
	StDev	0.182	0.133	0.1325	0.141	0.165
M/B	***Mean***	0.647	1.083	1.597	2.550	23.322
	Median	0.627	1.051	1.529	2.419	5.767
	StDev	0.927	0.476	1.156	1.446	215.466
Beta	***Mean***	1.041	1.033	1.107	1.229	1.383
	Median	0.946	0.930	1.009	1.135	1.295
	StDev	0.817	0.765	0.777	0.847	1.021
ME	***Mean***	392.946	816.659	1121.414	1760.874	2312.908
	Median	35.748	94.449	156.855	237.338	237.545
	StDev	1883.755	3488.225	4813.021	7600.199	9225.374
Total Vol	***Mean***	0.144	0.126	0.130	0.144	0.175
	Median	0.131	0.112	0.114	0.129	0.157
	StDev	0.073	0.068	0.070	0.076	0.101
Alpha	***Mean***	-0.010	-0.001	0.003	0.008	0.019
	Median	-0.008	0.001	0.004	0.008	0.017
	StDev	0.023	0.020	0.020	0.023	0.031

The results presented in Panel A show that the characteristics of our sample are comparable to those reported in previous literature (see, for example, Fama and French (1992, 1995)). The mean monthly return of our stocks during the 1970-2011 sample period is about 1.2% per month. The market- to-book ratio is extremely skewed with a mean value of 6.19, but a median value of 1.54. As observed in Panel B, most of the skewness is driven by the Q5 growth portfolio. The difference in means and medians for the other quintiles is not that extreme. The other simple stock characteristics are also comparable to previous literature. The median beta for our sample period is 1.04 while the mean market capitalization is 1.27 billion. The mean monthly total volatility is approximately 14%. Panel B shows that the value premium is present over the entire sample period – the value portfolio (Q1) has the highest mean return at 2.1% per month while the growth portfolio (Q5) had the lowest mean return at 0.7% per month. The growth portfolio also has the highest average total volatility at 17.53%, while the value portfolio is the second most volatile. The pattern of U-shaped volatility across M/B quintiles is consistent with previous studies such as Fama and French (1992, 1995). Extreme skewness in the market-to-book ratio is seen in the growth portfolio. The mean M/B of the growth portfolio is 23.3 while the median M/B of the same group is 5.77. Because of the high degree of skewness of the M/B variable, we conduct our analysis of this characteristic using medians rather than means, to insure that our results are not driven by influential outliers.

After we sort stocks into quintile portfolios based on their M/B characteristic, we focus on a buy and hold strategy with formation at time t, and then follow the characteristics of our portfolios (i.e. M/B and raw returns) through event time for months t+m, where m=1,2…48 (that is, up to 4 years after portfolio formation). Thus, portfolios formed each month are followed through time without refreshing. This contrasts to most market-to-book studies (Fama and French 1995) where the portfolios are "refreshed" (reassigned based upon their new market to book ratios) once a year. We average median M/B ratios and mean returns across event time for each "un-refreshed" quintile portfolio and observe the convergence patterns of these processes. In order to investigate whether results are sensitive to the chosen time period, we perform this analysis on the overall sample, as well as subsamples of approximately 10 years at a time. Specifically, we isolate four time periods, from 1970-1981, 1982-1991, 1992-2000, and 2000-2011. We present and discuss our results in the following section.

Results

We first look at the long term convergence pattern of the M/B characteristic. Given the highly skewed distribution of the M/B variable (see descriptive statistics presented in Table 1), we try to mitigate the influence of abnormally high outliers by reporting the average medians of each quintile portfolio through event time after portfolio formation (rather than the average means). Based on the arguments and empirical evidence put forth by Fama and French (2007 a,b), we would expect that, at least during the first year after portfolio formation, M/B values slowly reverse towards the mean (i.e. the median M/B of stocks in the value portfolio Q1 slowly increases and the median M/B of stocks in the growth portfolio Q5 slowly decreases). Table 2 presents the evolution of median M/B ratios for selected event times after portfolio formation.

Table 2: Convergence in Market-to-Book Ratios for Value and Growth Portfolios

Event Time	*Q1(Value)*	*Q2*	*Q3*	*Q4*	*Q5(Growth)*	*Q5-Q1*
1	0.628	1.052	1.529	2.411	5.738	5.110
12	0.686	1.062	1.468	2.176	4.164	3.478
24	0.765	1.083	1.444	2.026	3.433	2.669
36	0.823	1.101	1.426	1.912	3.023	2.201
48	0.875	1.114	1.412	1.843	2.758	1.884

The median M/B in the growth portfolio converges from 5.73 in month 1 to 4.16 in month 12. The M/B of the value portfolio converges from .628 in month 1 to .686 in month 12. This trend continues into the second, third and fourth year following formation of the portfolios. Table 2 shows that there is clear pattern of convergence of M/B ratios among our quintile portfolios during the four years after portfolio formation. Over 48 months, the M/B in the growth group, Q_5, decreases from 5.73 to 2.75, while the market to book of the value group, Q_1, increased from .628 to .875.

Table 2 reports the average medians for the M/B characteristic for each quintile portfolio, as well as the average difference in medians between the value and growth portfolio, which can be used as a simple measure of convergence. Technically, if these extreme portfolios eventually converge towards the mean, the difference between Q5 and Q1 should converge towards 0 – the last column of table 2 shows that after 48 months, the difference between Q5 and Q1 has indeed decreased from 5.11 to 1.88.

Although the last column of Table 2 does show that Q5-Q1 seems to trend towards zero, we cannot make any statement about the speed of convergence in this characteristic and we cannot use the magnitude of the difference to compare this speed of convergence with the convergence in returns. Using Q_5-Q_1 as a market-to-book convergence measure, we run a simple trend regression to test the existence of a statistically significant trend (i.e. we regress the differences in median M/B ratios on event time). In order to determine whether one of the extremes (value or growth) converges towards the mean faster or whether any other asymmetries in the convergence behavior of value/growth stocks exist, we apply the same methodology to the median M/B of value and growth portfolios separately. We consider the coefficient on event time as an estimate of the speed/rate of convergence towards the mean. We present these results in Table 3.

As suspected, the trends are statistically significant. Our convergence portfolio measure (Q5-Q1), converges at -.061 units per month. The intercept term is 4.416. Dividing the intercept term by the absolute value of the trend coefficient, we can estimate that the convergence term will be zero in about 72 months or six years if the predicted trend continued. Thus, we expect to see little or no difference in market-to-book ratios of the two portfolios in about 6 years. Our regression on the Q_1 (value) portfolio indicates that its market-to-book ratio increased by .006 per month over the 1970-2011 period. This is consistent with the one year estimates of market-to-book convergence documented by Fama and French (2007a,b). Our results suggest that this continues for much longer than a year. Alternatively, the market-to-book ratio of the growth portfolio (Q_5) decreases by -.056 per month. There is an asymmetry in the convergence rate between the extreme growth and value portfolios. In terms of the magnitude, the trend coefficient of the growth portfolio is approximate 10.2 larger than that of the value portfolio. Normalizing by the intercept gives the percentage movement in the first month of the portfolios existence. In this case the percentage movement of the value portfolio (coefficient divided by intercept) is a positive .0088 while the percentage movement in the growth portfolio is a -.0111. This seems to indicate that the growth portfolio's market-to-book ratio regresses to the mean at a slightly greater rate than that of the value portfolio.

Table 3: Trend Regressions for M/B Ratios: 1970-2011 (48 months)

	$Q_5 - Q_1 = b_1 + b_2\ (time) + e_1$			
	Intercept	Time	Adjusted R Square	F
Convergence Portfolios	4.416	-0.061	0.899	417.5
Q_5-Q_1 (1970-2011)	(52.18)	(-20.43)		
	$Q_1 = b_1 + b_2\ (time) + e_1$			
	Intercept	Time	Adjusted R Square	F
Value Portfolios	0.624	0.006	0.993	6169.9
Q_1 (1970-2011)	(317.07)	(78.54)		
	$Q_5 = b_1 + b_2\ (time) + e_1$			
	Intercept	Time	Adjusted R Square	F
Growth Portfolios	5.040	-0.056	0.884	358.7
Q_5 (1970-2011)	(60.63)	(-18.94)		

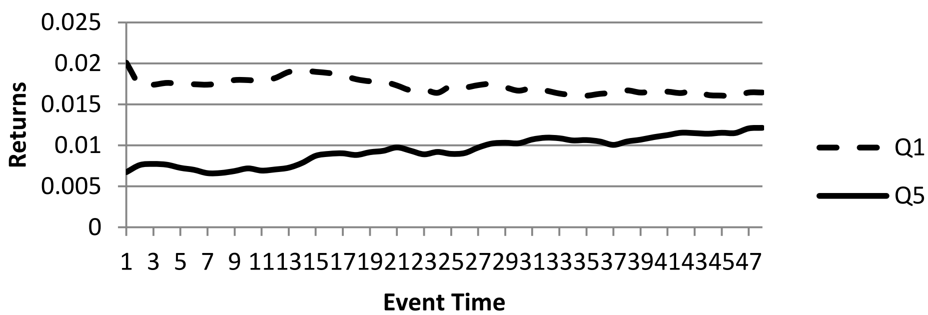
After documenting the pattern in M/B convergence, our question of interest is whether the pattern in returns for these portfolios exhibits similarities. As discussed in our Section 2, the growth rate in M/B ratio is expected to be an important component of capital gains returns. As such, we would expect that returns of value portfolios decrease through time as their M/B increases, while the opposite happens for growth portfolio. To investigate the validity of this claim, we apply the same event time analysis to the returns of our quintile portfolios. We start by following their returns up to 48 months after portfolio formation. Table 4 presents mean returns for selected event times after portfolio formation for our entire sample period.

Table 4: Convergence in Returns for Value and Growth Portfolios

Event Time	*Q1(Value)*	*Q2*	*Q3*	*Q4*	*Q5(Growth)*	*Q5-Q1*
1	0.020	0.013	0.011	0.009	0.007	-0.013
12	0.018	0.015	0.013	0.011	0.007	-0.011
24	0.016	0.015	0.014	0.012	0.009	-0.007
36	0.016	0.013	0.014	0.012	0.010	-0.006
48	0.016	0.015	0.014	0.014	0.012	-0.004

Looking at Table 4 it is relatively clear that over the whole period there appears to be a trend toward convergence in returns. For a more complete picture, we also graph the average returns of value and growth portfolios through event time in Figure 1. For the growth portfolios the market-to-book ratios are falling, but the returns (which start lower than the value group) are rising through time. Conversely, the value stocks returns start out to be the highest but they are declining through time, converging toward the mean of the entire sample.

Figure 1: 1970 – 2011 Returns Ranked by M/B



Do the returns of growth and value groups converge enough so that there is little difference between the two groups at some time? For true convergence in the returns between the respective groups the difference between the growth portfolio, Q5 and the value portfolio Q1 should approach 0. However, it is clear that the process for true convergence in returns is longer than the 48 months event time. The difference between the returns in the growth portfolio and the value in 48 months is a little over a third of the magnitude at event time 0. To further characterize the convergence phenomenon observed in Table 4, we run trend regressions on the returns of the convergence portfolio, as well as the returns of value and growth portfolios separately. Results are presented in Table 5.

Table 5: Trend Regressions for Returns: 1970-2011 (48 months)

	$Q_5 - Q_1 = b_1 + b_2\,(time) + e_1$			
	Intercept	Time	Adjusted R Square	F
Convergence Portfolios	-0.0119	0.0001648	.892	390.0
Q_5- Q_1 (1970-2001)	(-50.72)	(19.74)		
	$Q_1 = b_1 + b_2\,(time) + e_1$			
	Intercept	Time	Adjusted R Square	F
Value Portfolios	0.0184	-0.0000504	.559	60.6
Q_1 (1970-2001)	(101.52)	(7.42)		
	$Q_5 = b_1 + b_2\,(time) + e_1$			
	Intercept	Time	Adjusted R Square	F
Growth Portfolios	0.0065	0.0001144	0.919	487.92
Q_5 (1970-2001)	(44.96)	(22.08)		

The trend regression on returns indicates a statistically significant convergence in the returns over the 1970-2011 period. The t-statistic on the trend variable is 19.74, meaning that the convergence trend is strongly significant. Dividing the absolute value of the intercept, .0119, by the trend estimate coefficient, 0.0001648, gives us a theoretical estimate of 72 months for convergence to 0. Interestingly, the convergence rate is asymmetric between the value group and the growth group. According to the regression estimates there is a.0000504 decrease in the return of the value portfolio each month. In the first month this represents a 0.274 percent decrease in the expected return on the value portfolio. On the other hand the regression estimates suggest a .000115 increase in the expected return on the growth portfolio each month. This represents a 1.76 percent increase in the expected return in the first month. For some reason the returns for the growth portfolio converge toward the mean at a faster rate than the value portfolio during this period. This difference in rates can also be seen by examining Figure 1.

Previous literature does seem to indicate that the value premium may be related to macroeconomic conditions (see for example, Berk, Green and Naik, 1999, for a proposed explanation). Additionally, Asness, Friedman, Krail, and Liew (2000) and Cohen, Polk, and Vuolteenaho (2003) provide evidence of predictable time variation in the return premium earned by value stocks over growth stocks. Specifically, the dispersion in book-to-market ratios, or earnings-to-price ratios, between value stocks and growth stocks predicts high value premium. To see whether the convergence patterns are also time-varying, we repeat our analysis in four separate subsamples, which are approximately equal in length and exhibit relatively similar market conditions. Specifically, we break our sample into the following four sub-periods: 1970 to 1981 (a bear market), 1982 to 1991 (a bull market), 1992 to 2000 (a super bull market) and 2001 to 2011(a bearish market), respectively. Within each one of these sub-samples we repeat our trend regression analyses for the B/M and return convergence processes. We ran trend analysis for both groups over 24 and 48 months. The results indicated that the market-to-book convergence was relatively stable and occurred over a period longer than 48 months. However, the return convergence is highly variable over different subsamples, with returns converging in about two years in the latest period. The variability of the return convergence can be seen in Figures 2 through 5. Because of the much shorter period of return convergence in the latter periods we present the results of our trend regressions for 48 months for market-to-book ratios and 24 months for the rate of return convergence. Results are presented in Tables 6 and 7, respectively.

As suggested by the results in Table 6, the M/B ratios converge with a relatively uniform rate over the 4 subsample periods. We can divide the intercept by the absolute value of the trend coefficient to obtain estimates for convergence to 0 – these estimates amount to 74 months for the 1970-1981 period, 64 months for the 1982-1991 period, 67 months for the 1992-2000 period, and 83 months for the 2001-2011 period. All of these estimates are consistent with the estimate of M/B convergence of 72 months for the entire 1970-2011 period.

Table 6: Trend Regressions for M/B Ratios by Subsamples (48 months)

$Q_5 - Q_1 = b_1 + b_2 (time) + e_1$

	Intercept	Time	Adjusted R Square	F
1970-1981				
Convergence Portfolios	2.950	-0.03976	.972	1644.8
(Q_5-Q_1)	(106.92)	(-40.55)		
1982-1991				
Convergence Portfolios	4.583	-0.07164	.904	444.8
(Q_5-Q_1)	(47.94)	(-21.09)		
1992-2000				
Convergence Portfolios	5.854	-0.08732	.814	206.9
(Q_5-Q_1)	(19.71)	(-14.38)		
2001-2011				
Convergence Portfolios	4.689	-0.05634	.931	636.0
(Q_5-Q_1)	(74.6)	(25.22)		

When we examine the rate of convergence in returns over the same four periods, we observe that there is no correspondence between convergence in M/B ratios and rate of return convergence. Our measure of return convergence is the difference in the returns between the growth portfolio and the value portfolio. Trend regression results for returns in each corresponding sub-sample are presented in Table 7.

Table 7: Trend Regressions for Returns by Subsamples (24 months)

Q_5-$Q_1 = b_1 + b_2$ (time) + e_1

	Intercept	Time	Adjusted R Square	F
1970-1981				
Convergence Portfolios	-0.00704	-.0000994	.171	5.7
(Q_5-Q_1)	(-11.90)	(-2.40)		
1982-1991				
Convergence Portfolios	-0.00981	0.000149	.241	8.3
(Q_5-Q_1)	(-13.28)	(2.88)		
1992-2000				
Convergence Portfolios	-0.01208	0.0001824	.147	4.97
(Q_5-Q_1)	(-10.34)	(2.22)		
2001-2011				
Convergence Portfolios	-0.01463	0.0004935	.925	287.4
(Q_5-Q_1)	(-35.19)	(16.95)		

There is a large disparity in the rates of return convergence over the different sample periods. Although we calculated the trend coefficient statistics for both 24 and 48 months we report only the 24 month trend coefficient because of the relatively fast convergence of the rate of return in the later periods. This can be clearly seen in Figures 2, 3, 4, and 5. In terms of the 24 month estimates it can be seen that in the first 24 months there is no rate of return convergence in the 1970-1981 period. The trend coefficient even has the wrong sign. The estimate for convergence to zero found by dividing the intercept term by the absolute value of the trend coefficient the 1982-1991 period is 60 months, and for the 1992-2000 period it is 66 months. But the estimate for 2001-2011 is 29 months. Examining the visual evidence in figure 3, 4, and 5 we can see that the actual rate of return convergence might actually be slightly faster than the regression estimates for the later periods. A clear pattern develops when examining the rate of return convergence through time, i.e. the rate of return convergence seems to be getting more rapid through time. This can be easily observed looking at the graphs of the convergence of returns through event time after portfolio formation for each one of our subsamples.

Figure 2: 1970 – 1981 Returns Ranked by M/B

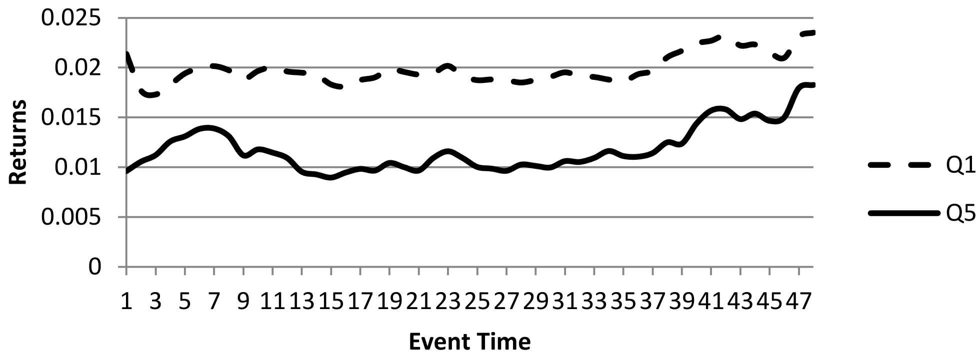


Figure 3: 1982 – 1991 Returns Ranked by M/B

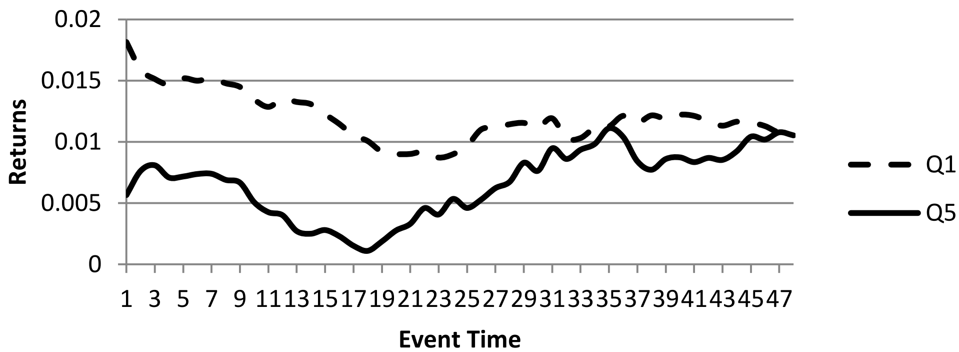


Figure 4: 1992 – 2000 Returns Ranked by M/B

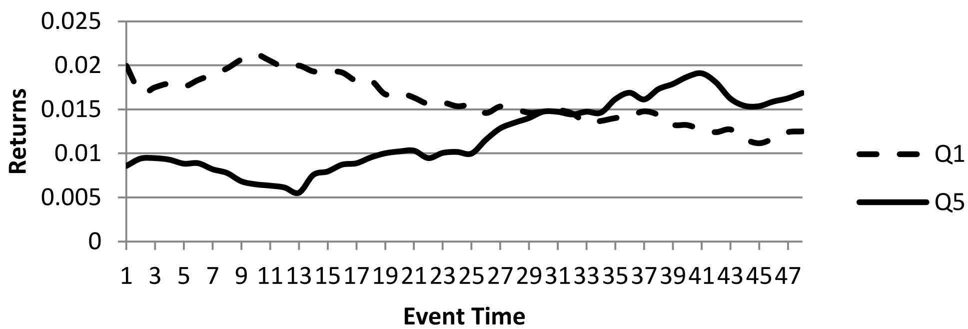
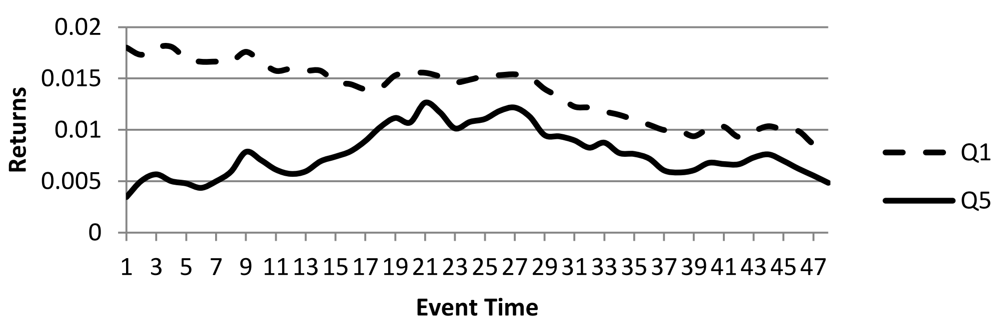


Figure 5: 2000 – 2011 Returns Ranked by M/B



Conclusion

We examine the convergence patterns in M/B ratios and returns for value and growth portfolios over up to four years after portfolio formation, during a sample covering 1970-2011. We document a clear pattern of convergence towards the mean for the extreme value and growth portfolios in both M/B ratios and returns. However, the speed of the convergence for returns does not follow the same pattern as the one for the M/B ratios. Furthermore, we observe that the returns convergence patterns are time-varying. Specifically, later periods exhibit much faster convergence in returns than earlier periods of our sample, while the same is not true for convergence in M/B ratios.

The change of investors' behavior and the competitive economic environment can both explain our documented results. The style strategy literature dates to at least the 1980s (Rosenberg, Reid and Lanstein, 1985) but became very mainstream in early 1990, after prominent research done by Fama and French. After 1991, the rate of return convergence almost triples, which would be consistent with a wider audience taking the phenomenon into account. With wide recognition, investing in value stocks became a popular investing strategy and then resulted in the faster convergence trend in 1990s. The more investors obey the rules of value investing, the more obvious the convergence trend of average returns becomes, which would explain our documented results. Conversely, the lack of obvious return convergence in the data in the earlier period could be explained by the scarcity of research/knowledge of the pattern by investors participating in the market before the 1980s.

Another possible explanation for the faster return convergence is that the economic environment became more and more competitive. Since Fama and French (2007) indicate that it is competition that causes the declining market-to-book phenomenon in growth companies, it would seem that more intense competition would cause the phenomenon to occur faster. Compared to the 1970s and 1980s, the economy in the United States today is much more competitive, with companies facing not only domestic but also overseas competitors, which makes it harder for them to extend profit margins. Strong competition leads directly to profit cutoffs. For example, a value company investing in projects which require large financial input, such as purchasing fixed property, will definitely increase its book value. However, the resulting profits on these projects might be lower than expected because of the competitive market, leading to a decrease in the M/B ratio. More intense competition would increase the speed of this process, potentially explaining our results.

Notes

1. Risk based explanations originate with Fama and French (1993, 1995), who argue that value firms are fundamentally riskier than growth firms, and hence commend higher expected returns. On the other hand, behavioral explanations such as those advanced by Lakonishok et al (1994) argue that investors overreact to previous fundamentals, depressing the average price of value stocks and inflating the price of growth stocks.
2. In this sense, our methodology is closer to that employed by Fama and French (1995), who follow portfolio characteristics up to five years after formation period.
3. Results show that book return (growth in book equity) is high for growth stocks because of high earnings and reinvestment, and it is low to negative for value stocks as these companies exhibit low earnings and reinvestment. In the same time, market returns are high for growth stocks and low for value stocks.
4. According to Fama and French (2007) the growth in unrefreshed P/B ratios for each portfolio (which is our method of studying convergence) technically contains both a drift and a convergence component. However, the study shows that for a one year horizon, the convergence is the most important part influencing returns. It remains an open question whether this is the case for longer horizons also.
5. As a robustness check, we repeat our analysis on a sample where we eliminate financial firms (6000<=SIC<=6999). Results remain qualitatively unchanged.
6. Specifically, the total book equity value is given by stockholders' book equity plus balance sheet deferred taxes and investment tax credit (if available) minus the book value of preferred stock, which is given by the redemption, liquidation, or par value in the order of availability. If the stockholders' equity is missing, it is measured by the book value of common equity plus the par value of preferred stock, or else by the book value of assets minus total liabilities.

References

Asness, Clifford, Jacques Friedman, Robert Krail, and John M. Liew. 2000. "Style timing: Value Versus Growth." *Journal of Portfolio Management* 26(3): 50-60.

Ball, Ray and S.P. Kothari. 1989. "Nonstationary Expected Returns: Implications for Tests of Market Efficiency and Serial Correlation in Returns." *Journal of Financial Economics* 25: 51-74

Berk, Jonathan B., Richard C. Green, and Vasant Naik. 1999. "Optimal Investment, Growth Options, and Security Returns." *The Journal of Finance* 54(5): 1553-1607.
Chan, Louis, K. C. and Josef Lakonishok. 2004. "Value and Growth Investing: Review and Update." *Financial Analysts Journal* 60(1): 71-86
Cohen, Randolph B., Christopher Polk, and Tuomo Vuolteenaho. 2003. "The Value Spread." *Journal of Finance* 58(2): 609-641.
Daniel, Kent, and Sheridan Titman. 2006. "Market Reaction to Tangible and Intangible Information." *Journal of Finance* 61(4): 1605-1643.
Fama, Eugene F., and Kenneth R. French. 1992. "The Cross-Section of Expected Stock Returns." *Journal of Finance* 47(2): 427-465.
Fama, Eugene F., and Kenneth R. French. 1993. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economic* 33(1): 3-56.
Fama, Eugene F., and Kenneth R. French. 1995. "Size and Book-to-Market Factors in Earnings and Returns." *Journal of Finance* 50(1): 131-155.
Fama, Eugene F., and Kenneth R. French. 1996. "Multifactor Explanations of Asset Pricing Anomalies." *Journal of Finance* 51(1): 55-84.
Fama, Eugene F., and Kenneth R. French. 1998. "Value versus Growth: The International Evidence." *Journal of Finance* 53(6): 1975-1999.
Fama, Eugene F. and Kenneth R. French. 2007. "Migration." *Financial Analyst Journal* 63(3): 48-58.
Fama, Eugene F. and Kenneth R. French. 2007. "The Anatomy of Value and Growth Stocks." *Financial Analysts Journal* 63(6): 44-54.
Fama, Eugene F., and Kenneth R. French. 2008. "Average Returns, B/M, and Share Issues." *Journal of Finance* 63(6): 2971-2995.
Lakonishok, Josef, Andrei Shleifer, and Robert W. Vishny. 1994. "Contrarian Investment, Extrapolation, and Risk." *Journal of Finance* 49(5): 1541-1578.
Liu, Naiping and Zhang, Lu. 2008. "Is the value spread a useful predictor of returns?" *Journal of Financial Markets* 11(3): 199-227.
Watanabe, Akiko. 2008. "Style Evolution, Equity Issuance, and International Value Premium." Working Paper, University of Alberta.

Sportsbook Pricing and Informed Bettors in the Early and Late Season in the NBA

Rodney Paul, Syracuse University
Andrew Weinbach, Coastal Carolina University
Brad Humphreys, University of Alberta

Abstract

Detailed data from the NBA wagering market over six seasons is analyzed to identify the presence of informed bettors in betting patterns. About 2.5% more bets are made on underdogs early and late in the season, relative to other times. These bets are likely placed by informed bettors, as the betting public prefers betting on favorites. A simple strategy of wagering on underdogs early in the season wins more than implied by efficiency. The same strategy at the end of the season, however, does not reveal statistically significant profits. The NBA betting market appears to contain heterogeneous bettors.

Introduction

Common assumptions about the composition of participants in betting markets focus on the presence of a group of recreational (or uninformed) bettors and a separate group of professional (or informed) bettors (Shin 1991, 1992). Both of these groups of bettors participate in the market, and under the assumption of efficient markets, informed bettors counter any of biases generated by bets placed by uninformed bettors. Bets made by informed bettors ultimately reflect full information in the betting market and make closing prices optimal and unbiased forecasts of game outcomes and combined scoring in betting markets.

Although betting markets have been shown, in general, to be efficient, the NBA betting market appears to contain subsets of games where simple betting strategies reject the efficient markets hypothesis and generate excess returns. In a sample of games from 1995-96 to 2001-02, Paul and Weinbach (2005) showed that betting big underdogs won more than implied by efficiency and betting big home underdogs won enough to reject the null of no profitability. In the same study, the "hot-hand" hypothesis, another form of market inefficiency, was tested, based on the original hot hand studies by Camerer (1989) and Brown and Sauer (1993). Betting against teams on winning streaks was shown to also be profitable in this sample. The "hot-hand" hypothesis has also been analyzed more detailed NBA data, and the public was found to wager more on teams on streaks, while point spreads were still shown to be optimal and unbiased forecasts of game outcomes (Paul, Humphreys, and Weinbach, 2011).

Much past research on betting markets based the analysis on the balanced book hypothesis, the idea that sports books will set point spreads to balance the betting on the favorite and the underdog. If there are substantial betting imbalances, the balanced book hypothesis predicts that the point spread will move to address this imbalance in an attempt to equalize betting. If successful in balancing the book, a sports book maker is not an active participant in the wager as they simply earn the commission on losing bets for the service of offering wagers in this market (at a rule of bet $11 to win $10).

Recently, the balanced book hypothesis has been challenged in the literature. With more detailed betting market data on betting percentages now available, direct tests of the balanced book hypothesis are now possible. Using this data, the balanced book hypothesis was soundly rejected in the NBA as bettors were shown to heavily prefer to place bets on the favorite, with road favorites being even more popular (Paul and Weinbach, 2008). The balanced book hypothesis has also been rejected in other settings. Across sports, including the NBA, although bettor biases are strong and easily predictable, the market is still found to be efficient as the point spread represents an optimal and unbiased forecast of the game. Therefore, even though the bettors are quite biased, the sportsbook appears to price as a forecast of game outcomes, allowing the betting biases to persist in the market.

With this new evidence on sports book behavior and more detailed data to work with, it is now possible to address questions that were impossible to study previously. One outstanding question about bettor behavior is the existence of participant heterogeneity, in the form of groups of informed and uninformed bettors, in betting markets. Anecdotal evidence about informed bettors, such as Alan Boston in the book *The Odds* (Millman, 2001), exists but little formal empirical evidence exists and informed bettors were difficult to identify in previous betting market data; the presence of informed bettors is more assumed than confirmed. One possible way to develop evidence of the presence of informed bettors is to identifying situations that an informed bettor might exploit and test for difference in betting patterns during these situations. One setting where patterns consistent with informed betting might take place is betting early and late in sports seasons. If some bettors are better informed early in the season, and other bettors and/or the sports book managers are not informed, or biased due to reliance on previous season's results, then informed bettors might exploit this situation to earn excess returns on bets. Likewise, late in the season, some bettors may better be able to identify which teams will truly play hard when they are out of playoff contention.

These early- and late-season betting strategies have been studied before by Baryla, et al (2007). This study described the potential for early season biases in the NBA, but found that early-season wagers were efficient in terms of sides betting (wagering on a team compared to the point spread) markets but they were biased, with profitable betting strategies possible in totals (over/under) markets. In a similar vein, Borghesi (2007) investigated late-season wagering bias in the NFL, discovering that the often found home underdog bias is mainly due to the performance of home underdogs late in the season. In a slightly different setting, Osborne (2001) found that profitable strategies based on early-season score differentials could be used in the weeks immediately following early-season performances, but dissipates as the season progresses.

We use these ideas as the basis for an empirical test for evidence consistent with the presence of informed traders in the NBA using a sample of NBA games from 2005-06 to 2010-11. In these seasons, data are available on game outcomes, point spreads, and the percentage of bets on the favorite and underdog in each game. We investigate the determinants of variation in the fraction of bets placed on the favored team and how bets on underdogs (home, away, and combined) performed during early-season play and late-season play. Through a combination of regression analysis and betting simulations, we investigate the possibility of informed bettors existing in the NBA wagering market. We find evidence consistent with the presence of informed bettors in this setting: more bets are placed on underdog teams in the first and last two weeks of the NBA season than at other times, and bets on the underdog produce excess returns during the first two weeks of the season. Since the general betting public clearly prefers to bet on favorites, this preference for bets on underdogs and excess returns is consistent with the betting strategies that would be followed by a group of informed bettors. Outcomes in the NBA betting market are consistent with the presence of participant heterogeneity, in the form of informed and uninformed bettors.

The study proceeds as follows. First we present summary statistics and simple regression results relating the percentage bet on the favorite to market factors and the timeframe of the season. After identifying the time frames where informed bettors may be participating, results of simple betting strategies of following these assumed informed bettors are shown for both early and late season sample. The final section discusses the findings and concludes the paper.

Regression Analysis, Results, and Betting Simulations

We analyze data on betting on NBA games. The sample consists of 7,153 usable NBA games (games containing full information about the point spread, game outcome, and betting percentages) from the 2005-06 to 2010-11 seasons. The data were taken from Sports Insights (www.sportsinsights.com) who collect and publish data from a combination of seven on-line, off-shore sports books to estimate how many bettors wager on sides and totals. The seven sportsbooks used to calculate the betting percentages are Sportsbook, BetUS, Carib Sports, Wagerweb, 5Dimes, Sports Interaction, and CRIS. Table I contains sample statistics, the mean, median, and standard deviation for the point spread, absolute value of the point spread, percentage bet on the favorite, and percentage bet on the under.

Table 1: Summary Statistics – NBA Betting Market 2005-06 to 2010-11

Obs = 7153	***Point spread***	***Absolute Value of Point spread***	***Percentage Bet on Favorite***	***Percentage Bet on Underdog***
Mean	-3.34	6.07	59.25	40.75
Median	-4.50	5.50	60.00	40.00
Stand. Dev.	6.15	3.48	12.97	12.97

We use regression analysis to identify the presence of informed and uninformed bettors in the NBA betting market. The idea is to analyze variation in the fraction of bets placed on the favored team in each game. The regression model uses the percentage bet on favored team *i* against underdog *j* in game *t* (Y_{ijt}) as the dependent variable.

$$Y_{ijt} = \beta_0+\beta_1 X_{ijt}+\beta_2 Z_t + e_{ijt} \tag{1}$$

The independent variables are an intercept, a dummy variable for road favorites and the absolute value of the point spread on the game, which are included in the vector X_{ijt}, and week-of-season dummy variables for early- and late-season games, which are included in the vector Z_t. e_{ijt} is an unobservable, mean zero, constant variance random variable capturing other unobservable factors that affect the percentage of bets placed on the favored team in each game. The β_i's are unknown parameters to be estimated. The road favorite dummy is included to account for significant bettor interest in betting on good teams on the road and the absolute value of the point spread is included to account for bettors preferring good teams to poor teams in general. Both of these variables were shown to have a positive and significant effect on bet volume in previous studies of the NBA betting market (Paul and Weinbach, 2008; Paul, Humphreys, Weinbach, 2011). The dummy variables for the weeks of the season are included to identify potential wagering by informed bettors during the early-season or late-season in the NBA. Separate dummies are included for the first two weeks of the season, the second two weeks of the season, the next-to-last two weeks of the season, and the last two weeks of the season.

If informed bettors realize that bets on underdogs represent an attractive betting advantage during the early- and/or late-season, and wager accordingly, these dummy variables should have a negative coefficient and be statistically significant. Again, the general betting public prefers to bet on favored teams, and we assume that the majority of these bettors are uninformed bettors. However, informed bettors will wager on favored teams if they expect these bets will have a positive return. If there is potential for winning strategies, informed bettors should wager until the prices in the market reflect full information. Therefore, whether the effect on the betting percentages lasts for two weeks or a month depends upon the adjustment of market prices (point spreads) and informed bettor actions. The regression results are shown in table II. Coefficients are presented for each variable and t-stats are shown below in parentheses. Stars identify statistical significance with *** representing significance at the 1% level, ** at the 5% level, and * at the 10% level.

Table 2: Regression Results, Dependent Variable is Percentage Bet on Favorite

Variable	Coefficient
Intercept	51.70***
	(138)
Road Favorite Dummy	14.70***
	(44.76)
Absolute Value Point spread	0.595***
	(13.2)
First Two Weeks of Season	-2.589***
	(-4.32)
Second Two Weeks of Season	-1.05
	(-1.60)
Next to Last Two Weeks of Season	-0.884
	(-1.57)
Last Two Weeks of Season	-2.64***
	(-4.16)

The intercept estimate suggests that the favored team receives over 51% of the bets on NBA games. Bettors clearly prefer road favorites in the NBA, as they have been shown to do in other settings, with the road favorite receiving an additional 14.7% of the bets. In addition, bettors prefer favorites in general, implying a preference for the best teams in the league, as each point of the point spread on the favorite increases the percentage bet on the favorite by 0.60%. For example, a 5-point favorite will attract an additional 3% of wagers on the game, while a 10-point favorite will garner 6% more bets. If these hypothetical favorites were road favorites, the 5-point favorite translates into nearly 18% more bets and the 10-point favorite attracts nearly 21% more wagers.

In relation to the early- and late-season dummy variables, in both games in the first two weeks of the season and games in the last two weeks of the season have statistically significant lower bet volume on favored teams. In each case, nearly 2.5% fewer bets were placed on the favorite during these segments of the NBA season. It appears that some portion of participants in the betting NBA market wager on the underdog more often during these periods, despite the general overwhelming support that favorites receive throughout the season.

Given that favorites (and road favorites in particular) are particularly popular with bettors and the balanced book hypothesis does not appear to be valid, as there are clear and easily predictable biases in bet volume, it is worthwhile to formally test if these behavioral biases lead to biased closing prices (point spreads) in the betting market. To test this, simple betting simulations using the full sample of NBA game results (over 7000 observations) are shown in Table III below. The number of winning bets on favorites and underdogs are shown, in addition to the underdog win percentage and the log likelihood ratio tests of a fair bet (win percentage equals 50%) and no profitability (win percentage equals 52.4%) are noted. The null hypothesis of no profitability comes from the nature of the betting market where bettors must wager $11 for each $10 they hope to win. Results are shown for all games, games with road favorites, and games with home favorites. In each case, larger favorites (4+ points) are also shown as a subset of the results.

Table 3: All Games: Full NBA Seasons – ATS Results

Games Analyzed	Favorite Wins	Underdog Wins	Underdog Win Percentage	Log Likelihood Fair Bet	Log Likelihood No Profits
All Games – Home Favorites					
4+ Favorites	1919	1966	50.6	0.569	N/A
All Favorites	2518	2573	50.5	0.594	N/A
All Games- Road Favorites					
4+ Favorites	586	538	47.9	2.050	N/A
All Favorites	1055	1007	48.8	1.118	N/A
All Games – All Favorites					
4+ Favorites	2505	2504	49.9	0.0002	N/A
All Favorites	3573	3580	50.0	0.0069	N/A

As can be seen from the table, the win percentages in each situation are around 50% in all cases. Bets on underdogs win slightly more when they are on the road (home favorite category), while bets on favorites win slightly more when they are on the road (road favorites category). The null hypothesis of a fair bet cannot be rejected in any of the cases.

Given the results on Table III, the null hypothesis of market efficiency cannot be rejected. Simple betting strategies of wagering on favorites or underdogs in different situations do not yield statistically significant excess returns. Despite the clear biases shown in Table I, where bettors clearly prefer to wager on favorites and road favorites, the closing point spread still represents an optimal and unbiased forecast of game outcomes. Therefore, sportsbooks do not appear to operate under the balanced book hypothesis, but appear to set point spreads as optimal forecasts of game outcomes, allowing significant betting imbalances to persist between favorites and underdogs.

To test the notion that bettors, particularly informed bettors, may behave differently early in the season and late in the season, similar betting simulations as those reported on Table III above were constructed for games during the first two weeks of the NBA season and for games during the last two weeks of the NBA season. From Table I, during both of these time periods bettors wagered a higher percentage of bets on the underdog than they did throughout the rest of the season. To test if this represents the actions of informed bettors, we reproduce Table III and show the underdog win percentage and log likelihood ratio tests for a fair bet and for no profitability in each of the these scenarios. Since bettors wager more on underdogs during this timeframe, they could be classified as informed bettors if their strategy yields statistically significant profits. Results for the first two weeks of the season are shown in Table IV and results for the last two weeks of the season are shown in Table V.

Table 4: First 2 Weeks of NBA Seasons – ATS Results

Setting/Situation	Favorite Wins	Underdog Wins	Underdog Win Percentage	Log Likelihood Fair Bet	Log Likelihood No Profits
First 2 Weeks – Home Favorites					
4+ Favorites	128	165	56.3	4.68**	1.82
All Favorites	179	215	54.6	3.29*	0.76
First 2 Weeks- Road Favorites					
4+ Favorites	39	51	56.7	1.60	0.66
All Favorites	75	87	53.7	0.900	0.11
First 2 Weeks – All Favorites					
4+ Favorites	167	216	56.4	6.29**	2.49
All Favorites	254	302	54.3	4.12**	0.84

The results on Table IV suggest some evidence supporting the presence of informed bettors during the early NBA season. During the first 2 weeks of NBA games, bets on underdogs win over 54% of the time. In games involving favorites of 4-or-more points, bets on underdogs win over 56% of the time. Both of these results reject the null hypothesis of a fair bet at the 5% significance level. Statistical significance is also found for the subset of games involving home favorites, where the null of a fair bet can be rejected at the 10% level for all games that meet this criterion, while the null of a fair bet can be rejected at the 5% level where there are home favorites of 4-or-more. Across both subsets of games, home favorites and road favorites, underdogs win around 54% of the time overall, and 56% of the time when they are underdogs by 4-or-more points. Despite win percentages exceeding 54% and 56% in these samples, the null of no profitability cannot be rejected due to the corresponding sample size.

Table 5: Last 2 Weeks of NBA Seasons – ATS Results

Setting/Situation	Favorite Wins	Underdog Wins	Underdog Win Percentage	Log Likelihood Fair Bet	Log Likelihood No Profits
Last 2 Weeks – Home Favorites					
4+ Favorites	165	177	51.7	0.42	N/A
All Favorites	203	221	52.1	0.76	N/A
Last 2 Weeks-Road Favorites					
4+ Favorites	63	47	42.7	2.34	1.06
All Favorites	101	85	45.7	1.38	0.28
Last 2 Weeks – All Favorites					
4+ Favorites	224	228	49.6	0.04	N/A
All Favorites	304	306	50.2	0.01	N/A

Unlike during the first two weeks of the NBA season, bets on the underdog during the last two weeks of the NBA season did not win more than 52.4% of the time. Win percentages are generally around 50% overall for underdogs, with home underdogs performing quite poorly during the closing weeks of the season (winning only 45% of the time), but the results were not found to be statistically significant. In all studied scenarios during the last 2 weeks of the season, the null of a fair bet (and consequently the null of no profits) could not be rejected.

Discussion and Conclusions

Given access to detailed data on gambling markets available from sources like www.sportsinsights.com, analysis of new types of economic behavior of both bettors and bookmakers is now possible. This study develops evidence of the presence of informed bettors in the NBA wagering market. Although economists assume that both informed and uninformed bettors exist in wagering markets, informed bettors are actually quite difficult to identify and little previous research has focused on developing evidence that informed and uninformed bettors exist in betting markets. This does not mean that informed bettors do not exist, however, but given their importance in the study of sports betting markets and market efficiency, it is important to confirm that they exist and to what scale and scope they participate in betting markets.

This study uses detailed data on betting percentages on the favorite and underdog in individual NBA games to determine if betting patterns are consistent with the presence of informed bettors during the early and/or late parts of the NBA season. Sports gambling markets are often thought to contain exploitable inefficiencies early in the season, before full information about team quality is known, and possibly at the end of the season where match-ups often take place between teams fighting for playoff spots or positioning and those out of the playoffs (Humphreys and Soebbing, 2013).

Using betting percentages, we first show that bettors prefer betting on favorites to betting on underdogs, with road favorites being particularly enticing options. The percentage of bets which accumulate on the favorite also increases with the point spread. Therefore, the percentage of bets on the favorite clearly exceeds the percentage of bets on the underdog in most NBA games. This refutes the common idea that book makers operate a "balanced book." Beyond these results, dummy variables for contests occurring in the first two weeks of the NBA season and the last two weeks of the NBA season were shown to have a negative and statistically significant effect on the percentage bet on the favorite. In other words, the percentage bet on the underdog was higher during the first and last 2 weeks of the NBA season, by about 2.5%. If underdogs were more successful during these time periods, this could indicate the presence of informed bettors in the NBA wagering market.

To explore this concept further, the overall NBA betting market was tested for the presence of excess returns to bets on favorites and underdogs. These returns have a nearly 50-50 split in terms of bet outcomes, even with the clear bettor bias toward favorites which was seen in the regression results. These results again clearly reject the balanced book hypothesis, but also show the market to be remarkably efficient. This supports the concept that sports books may choose to ignore many of the clear and predictable public biases and price more as an optimum forecast of game outcomes (Paul and Weinbach, 2005).

Dividing the sample into different segments of the season revealed some interesting results. During the first two weeks of the season, bets on underdogs won over 54% of the time overall, and 56% of the time when the game involved a favorite of 4-or-more points. These results reject the null hypothesis of a fair bet. Given that relatively more wagers were placed on underdogs during this early-season action (even though the vast majority of the bets were still placed on the favorite), this lends some credence to the notion that informed bettors do exist in the NBA betting market. These bettors may have more knowledge of current team strengths and weaknesses at the start of the season, possibly due to lingering biases from the previous season or just from being able to more quickly identify the best and worst teams in the league.

The results for the late-season betting do not reveal the same patterns. Although around 2.5% more bets accumulate on the underdog during the last two weeks of the season, the win percentage of this strategy was not found to be different from 50% as the null of a fair bet could not be rejected. In the subset of home underdogs, the bettors wagering on these teams fared quite poorly, winning only 45-46% of their bets, although these results were not found to be statistically significant due to the general infrequency of home underdogs during NBA campaigns.

Overall, this study develops evidence supporting the existence of informed bettors in the NBA wagering market. Some early season bettors in the NBA appear to recognize the possibility for inflated point spread on successful teams from the previous season who might not be as talented in the current campaign and/or recognize teams that have substantially improved. Although the percentage of these early-season underdog bettors is rather small, it is statistically significant and the win percentage on betting underdogs in the first two weeks is high enough to reject the null of a fair bet and post win percentages exceeding 54% for all games and 56% for games with underdogs of 4-or-more points. Although the percentage bet on the underdogs also increases late in the season (the final two weeks), these underdog bets do not fare as well as the null hypothesis of a fair bet cannot be rejected.

References

Baryla, E., Borghesi, R., Dare, W., and Dennis, R. 2007. "Learning, Price Formation, and the Early Season Bias in the NBA." *Finance Research Letters* 4(3): 155-164.

Borghesi, R. 2007. "The Late-Season Bias: Explaining the NFL Home-Underdog Effect." *Applied Economics* 39(15): 1889-1903.

Brown, W. O. and Sauer, R. D. 1993. "Does the Basketball Market Believe in the Hot Hand? Comment" *American Economic Review*, 83(5), 1377-1386.

Camerer, C. 1989. "Does the Basketball Market Believe in the Hot Hand?" *American Economic Review*, 79(5), 1257-1261.

Humphreys. B.R. and Soebbing, B.S. 2013. "Do Gamblers Think that Teams Tank? Evidence from the NBA". *Contemporary Economic Policy* (31)2: 301-313.

Osborne, E. 2001. "Efficient Markets? Don't Bet on It." *Journal of Sports Economics* 2: 50-61.

Paul, R., Humphreys, B., and Weinbach, A. 2011. "Revisiting the 'Hot Hand' Hypothesis in the NBA Betting Market Using Actual Betting Percentages." *Journal of Gambling Business and Economics*, 5(2): 42-56.

Paul, R. and Weinbach, A. 2008. "Price Setting in the NBA Gambling Market: Tests of the Levitt Model of Sportsbook Behavior." *International Journal of Sport Finance*, 3(3): 2-18.

Paul, R. and Weinbach, A. 2005. "Bettor Misperceptions in the NBA: The Over Betting of Large Favorites and the "Hot Hand". *Journal of Sports Economics* 6(4): 390-400.

Shin, H. S. 1991. "Optimal Betting Odds Against Insider Traders". *The Economic Journal* 101(408): 1179-1185.

Shin, H. S. 1992. "Prices of State Contingent Claims with Insider Traders, and the Favourite-Longshot Bias". *The Economic Journal* 102(411): 426-435.

Comparing Value and Growth Mutual Fund Performance: Bias from the Fama-French HML Factor

Glenn Pettengill and George Chang, Grand Valley State University
C. James Hueng, Western Michigan University

Abstract

We argue that loadings on the *HML* factor cause the Fama-French three-factor model to provide a downward bias of the performance of value portfolios. It is shown that value dominates growth in comparisons of realized risk and returns. When using only the market and the size factors to adjust returns, growth always has significantly negative risk-adjusted returns and value always has positive risk-adjusted returns. Although realized risk is always significantly lower for value portfolios, including *HML* in measuring risk-adjusted returns results in an increase in risk-adjusted returns for growth portfolios and a decrease in risk-adjusted returns for value portfolios.

Introduction

A basic insight of modern portfolio theory is that only systematic risk should be considered in measuring risk of securities considered for inclusion into a well-diversified portfolio. Total risk is an inappropriate measure because a portion of total risk will be eliminated in a well-diversified portfolio. Asset-pricing models measure systematic risk for individual securities by factor loadings on systematic risk factors. Over the last several decades researchers have measured systematic risk with the Fama-French three-factor model, which argues that systematic risk results from exposure to three risk factors. As shown in equation (1) the exposure of a security *i* to systematic risk is measured by the security's loading on: market excess return $(R_m - R_f)$; a portfolio long in value securities and short in growth securities (*HML*); and a portfolio long in small-firm securities and short in large-firm securities (*SMB*):

$$(R_i - R_f)_t = \alpha_i + \beta_{mi}(R_m - R_f)_t + \beta_{Hi}(HML)_t + \beta_{Si}(SMB)_t + \varepsilon_t \qquad (1).$$

In this paper we argue that measuring performance with the Fama-French three-factor model results in a bias against value managers because of the influence of the factor loading on the *HML* factor, β_{Hi}, an estimate of the loading for Security *i* on the *HML* factor. If value securities tend to move together, β_{Hi} will tend to be positive for value securities and portfolios composed of value securities. Likewise, if growth securities tend to move together, β_{Hi} will be negative for growth securities and portfolios composed of growth securities. If the factor loading on the *HML* portfolio is an appropriate measure of systematic risk associated with the *HML* factor, then the value securities have higher systematic risk with regard to this factor than do growth securities. But the impact of the responsiveness of value securities and growth securities on *realized* risk is not linear with the measure of responsiveness to the *HML* factor. And this is the crux of the bias against value securities when measuring portfolio performance.

Posit two portfolios: A value portfolio (Portfolio V) and a growth portfolio (Portfolio G). Assume that Portfolio V has a positive loading on the *HML* factor and that Portfolio G has a negative loading on the *HML* factor. Further, assume that these loadings have the same absolute value. If the estimated factor loadings are equally accurate for Portfolio V and Portfolio G, then the impact on total *realized* variation in response to the *HML* factor from the two securities will be exactly the same.[1] To illustrate, consider equation (2) which shows the variance of a random variable *x* multiplied by a constant *b*:

$$\sigma^2_{b\cdot x} = b^2 \cdot \sigma^2_x \qquad (2).$$

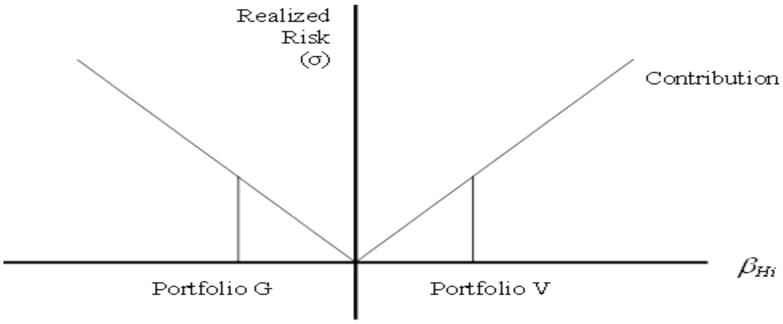
Equation (3) shows the standard deviation of a random variable *x* multiplied by a constant *b*:

$$\sigma_{b\cdot x} = | b | \cdot \sigma_x \qquad (3).$$

Treating the factor loading as a constant and the *HML* factor as a random variable, we have shown that the contribution to realized risk (as measured by the standard deviation in returns) is a linear function of the absolute value of β_{Hi}. Thus, the hypothesized growth portfolio, Portfolio G, and the hypothesized value portfolio, Portfolio V, receive the same contribution to overall risk based on their factor loadings on the *HML* factor. We show this relationship graphically in Figure 1a.

On a realized basis what concerns investors? Is it the amount of risk in a portfolio as measured by a factor loading? Or, do investors care about the amount or risk actually realized? Investors care about realized risk. In this case Portfolio V and Portfolio G will experience the same realized risk. Although the contribution to *realized* risk is the same, risk-adjustment procedures using the three-factor model are biased against the value security.[2]

Figure 1a: Contribution to Realized Risk as a Function of Factor Loading



Equations (4) and (5) illustrate the bias against value securities in determining risk adjusted returns. Equation (4) shows standard calculation for expected return at time *t*, given realized values for the factors of the Fama-French three-factor model.

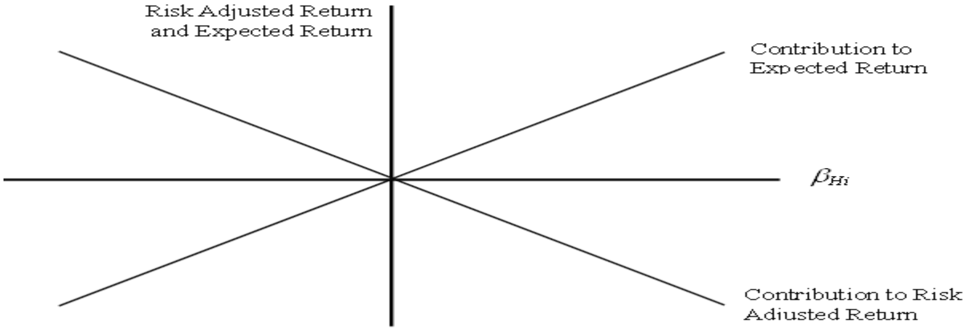
$$E(R_p)_t = RF_t + \beta_{mp}(R_m - R_f)_t + \beta_{Hp}(HML)_t + \beta_{Sp}(SMB)_t \quad (4).$$

Given that the *HML* factor is on average positive, a positive (negative) loading for a value security will, on average, increase (decrease) expected return for the value (growth) security.[3] Holding all other factors constant, for Portfolio V and Portfolio G, as described above, the expected return is greater for the value security than for the growth security despite an equal contribution to *realized* risk. This upward bias in the expected return for the value portfolio will results in a downward bias for risk-adjusted return. Using standard procedures, risk-adjusted return is measured in equation (5), where RAR_{pt} represents the risk-adjusted return for Portfolio *p* at time *t* and RR_{pt} represents the realized return for Portfolio *p* at time *t*.

$$RAR_{pt} = RR_{pt} - E(R_p)_t \quad (5).$$

Assume that in a particular time period the average value for *HML* is zero.[4] Given this assumption and assumption made above that the estimated loadings on Portfolio V and Portfolio G have the same absolute value but different signs, and that the estimated loadings hold, then realize return will be the same for the two portfolios. Further, as argued above, the contribution to realized risk will be the same for the two portfolios. The two portfolios have performed equally well. Because, however, the expected return measured from the Fama-French three-factor model is higher for Portfolio V relative to Portfolio G, the *RAR* for Portfolio V will be lower than the *RAR* for Portfolio G. Thus, measuring the performance of a value (growth) portfolio manager will have a downward (upward) bias. This bias against value and in favor of growth in terms risk-adjusted returns, resulting from the estimate of expected return, is illustrated in Figure 1b.

Figure 1b: Contribution to Expected Return and Risk-Adjusted Return as a Function of Factor Loading



To illustrate our argument, we will compare the performance of portfolios formed from a sample of value and growth mutual funds. In addition, we compare the performance of the Russell 2000 value and growth, representing potential index funds. If the value premium holds in this context the realized return will be higher for the value funds. If the realized risk for the value portfolio is no different or lower than the realized risk of the growth portfolio, then the performance of the value mutual funds ought to be judged to be superior. For our sample we calculate average return, standard deviation of returns, Sharpe Ratios and geometric means. These calculations support superior performance for the value portfolios. Then we calculate the risk-adjusted return for the value and growth portfolios using the Fama-French three-factor model to see if the risk-adjusted return measure corresponds with the judgment provided by realized risk and return measures. To isolate the influence of the loading of the portfolios on the *HML* factor, we first estimate risk-adjusted return using market excess return and the *SMB* factor. This allows us to test the bias from the *HML* factor. Comparing value and growth indexes and value and growth mutual funds, we find using measures of realized returns and realized risk that value outperforms growth. When judgment is made on performance by determining risk-adjusted return with the inclusion of the *HML* factor in the Fama-French model, growth outperforms value. The contribution of this paper is to caution that loadings on the *HML* factor bias judgment against the performance of value fund managers. We do not argue against the use of the Fama-French three-factor model in explaining portfolio or security returns.

The rest of the paper is organized as followed. In the next section we review previous studies with two purposes: First we validate assumptions about the value premium and the loading of value and growth funds on the HML factor. Second we examine previous studies that have compared the performance of value and growth mutual funds. The third section describes our data. The fourth section presents our empirical results. Our conclusions are in the final section.

Previous Literature Dealing with Performance of Value and Growth Mutual Funds

The goal of this paper is to examine the bias in measuring the performance of value portfolios relative to growth portfolios resulting from the loading on the *HML* factor of the Fama-French model. We illustrate this bias by a comparative study of the performance of the value and growth mutual funds. This study assumes the existence of a value premium, implying that the *HML* factor in the Fama-French three-factor model is, on average, positive. The existence of a bias against value securities also depends on the value (growth) mutual funds having a positive (negative) loading on the *HML* factor. Citing previous literature, we confirm both of these requirements. Further confirmation is provided within our on sample data. Also this section reviews previous literature on the comparative performance of value and growth mutual funds.

The Value Premium and Factor Loadings

Our study assumes the existence of a value premium and the tendency for value (growth) funds to load positively (negatively) on the Fama-French *HML* factor. In this short section we seek to confirm both assumptions based on previous research. The value premium argues that value securities, securities with high book-to-market (BtM) ratios or low price-to-earnings (P/E) ratios outperform other securities when raw returns or returns adjusted only for market risk are considered. There is a large literature to support this finding, beginning with Basu (1983) studying the relationship between P/E ratio. These studies have been confirmed in numerous other studies including the discipline altering study of Fama and French (1992). Not only has the value premium been persistently reported, it is also large. From July 1926 through February 2012 this premium has averaged an annualized value of 4.0% according to the Kenneth French website. Thus, one would expect mutual funds holding value stocks to outperform mutual funds that hold growth stocks.

Based on the strong evidence for the value premium, one might logically assume that value mutual funds should outperform growth mutual funds. Clash (1998) argues that investors interested in small-company stocks should invest in small-company value mutual funds on the basis of the superior performance of the Russell 2000 value index versus the Russell 2000 growth index. Superior performance by an index of value stocks relative to an index of growth stocks does not necessarily indicate superior performance by managed value funds relative to managed growth funds. Caution against presuming superior relative performance for value mutual funds from superior relative performance of the value index is advised from the voluminous literature that finds managed mutual funds generally underperform the market. Underscoring this caution, Chan, Chen and Lakonishok (2002, p. 1409) find the evidence of managed mutual fund underperformance “striking” in view of the considerable evidence on market anomalies such as the value premium. We examine evidence from the literature on this question more fully below.

Our argument that a bias exists against value portfolios when measuring performance with a risk-adjusted return calculated by the Fama-French three-factor model assumes that value (growth) portfolios load positively (negatively) on the *HML* factor. As we argued above, a presumption for this relationship exists in the construction of the *HML* factor. Davis (2000) funds with positive (negative) loadings on the *HML* factor as value (growth) funds. In a similar study, Chan, Chen and

Lakonishok (2002) argue that funds with a positive loading on the *HML* factor have a "value tilt." This leaves the question as to whether funds which are identified as value or growth funds load in accordance with this presumption. Our empirical analysis below shows that this is indeed the case.

Comparison of Performance of Value and Growth Mutual Funds

Of the large number of studies evaluating mutual fund performance a few have specifically compared value and growth mutual funds' performance. Shi and Seiler (2002) compare the performance of value and growth mutual funds without the use of an asset pricing model. They gather mutual fund return data return from the Morningstar data base. They identify value and growth funds based on the Morningstar classification and further divide these funds by size categories of large, medium and small. They randomly sample thirty securities from each of the six categories from the bivariate sort, comparing risk and returns over a ten-year period from 08/31/1989 through 08/31/1999. They measure return as the average monthly return as provided by the Morningstar data base over the 10 year period. To standardize returns they measure risk over the same ten-year period. Following Morningstar, they use downside risk, measured as semi-variance, instead of a total risk measure. Shi and Seiler find that the average return is higher for the growth mutual fund in each of the three categories and is significantly higher for the large-firm classification. Consistent with our argument that value funds should not be assumed to have higher risks because of higher loadings on the *HML* factor, they find that in all three size categories the risk is significantly higher for the growth mutual fund sample relative to the value mutual fund sample. They calculate the Sharpe ratio (using semi-variance) for each of the size categories. They argue that, unfortunately, the Sharpe ratio allows no statistical comparison. They find the reward to risk ratio higher for growth firms for large funds but higher for value funds for mid and small cap firms. They conclude that their evidence allows no definitive recommend for investing in growth versus value mutual funds and indicate that results may vary over other time periods (note that their sample covers a significant portion of the dot.com bubble, where one might expect growth mutual funds to do particularly well).

In contrast to the studies cited above, studies comparing the performance of value and growth mutual funds using the Fama-French three-factor model universally conclude that growth outperforms value. Davis (2001) uses data from the CRSP file for the period 1962 through 1998 to compare the performance of value and growth mutual funds. As discussed briefly above, he identifies funds as value (small) or growth (large) funds by factor loading on the *HML* (*SMB*) factor. A positive loading on the *HML* (*SMB*) factor identifies the mutual fund as a value (small-firm) fund. A negative loading on the *HML* (*SMB*) factor identifies the mutual fund as a growth (large-firm) fund. These regressions are conducted over a three-year period. In the following one-year period, performance is measured by regressing monthly returns for each mutual fund against the three factors and determining alpha for each fund. The alphas are averaged for the funds according to a univariate sort on the *HML* or *SMB* factor or a bivariate sort.

Ten portfolios are created from a univariate sort on the loadings on the *HML* factor. Performance is measured for each portfolio by finding the equally-weighted average of the alphas of each fund within the portfolio across the sample period. Portfolios formed from funds with low loadings on the *HML* factor are considered growth portfolios and portfolios form with funds with a high loading on the *HML* factor are considered value portfolios. In general, alphas are positive for the portfolios of growth mutual funds and negative for the portfolios containing value mutual funds. The portfolio containing the mutual funds with the highest loading mutual funds has a significantly negative alpha.

In the bivariate sort, Davis creates nine portfolios based on the value and growth classification and based on the size classification of large, medium and small firms. Each of the classifications is based on the factor loadings on the *HML* and *SMB* factors. In a similar manner, alphas are determined for each of the nine portfolios. For all size classification the value portfolio has a negative alpha and the growth portfolio has a positive alpha.

Davis (2001, p. 25) concludes that "Perhaps the biggest disappointment in the past three decades is the inability (or unwillingness) of funds to capture the value premium that has been observed in common stock returns during the period." We suggest that the premium may exist but that it is obscured by the bias against value inherent in the Fama-French three-factor model. Chan, Chen and Lakonishok (2002) calculate average returns for a using Morningstar data from January 1979 (the inception of the Russell indexes) through December 1997. Although their paper explores a wide variety of issues, they do compare the performance of value and growth mutual funds. Consistent with Davis they identify value and growth funds on the basis of the funds loading on the *HML* factor. Also, they compare value and growth funds for three classifications of size: large cap, mid cap and small cap. They calculate alpha by regressing mutual fund returns against the three-factor model and find average alpha across fund classifications. Unlike Davis, most alphas are positive, but insignificantly different from zero. Consistent with Davis for each of the size categories the alpha for the growth funds is greater than the alpha for the value funds. For the small cap growth fund the alpha is significantly larger than zero. Thus, they find, in contrast to the value premium, growth beats value. We argue that this finding results from the bias in the three-factor model.

Data

To test for bias in measuring the performance of value portfolio managers versus growth portfolio managers, we gather return data on value and growth portfolios. Previous empirical findings as cited above allow the presumption that value securities earn higher raw returns than growth securities. It is not our goal to test whether value beats growth but rather to test for bias in measuring performance of value versus growth in cases where *a priori* findings lead to the expectation that value beats growth. Based on Clash's argument that small-firm investors should favor value funds, we compare value and growth portfolios for small-firm portfolios where the value premium appears strongest. We gather two sets of small-firm value and growth portfolios.

First, using the Bloomberg terminals we gather monthly return data for the Russell 2000 Value Index and the Russell 2000 Growth Index over the period January 1979 (the inception of the Russell 2000 Index) through February 2012. On the basis that index funds are able to reasonably match the performance of large indexes, we treat the returns to the Russell 2000 indexes as returns to realizable portfolios.[5]

To compare the performance of managed growth and value funds, we acquire monthly returns for value and growth mutual funds from the Morningstar database. We obtain data from the inception of the database through February 2012. Morningstar classifies the type of mutual funds as value (growth) if the fund has low (high) valuations [low (high) price ratios and high (low) dividend yields] and slow (fast) growth [low (high) growth rates for earnings, sales, book value, and cash flow]. There are a total of 550,456 observations of monthly returns in the growth fund database and 351,068 observations of monthly returns in the value fund database. Morningstar provides two type classifications for each fund: the first classification types the fund at the start of the period for which Morningstar has data; the second classification types the fund at the end of the sample period.

We cull our Morningstar data on three criteria: 1) We examine only small-firm value and growth funds; 2) We require that the fund category be the same for the first and last observation; 3) To ensure statistically accuracy we require that for any given month to be included in our sample there must be at least 30 individual mutual funds available to investors. There are generally more small-firm growth funds than small-firm value funds. The minimum requirement for small-firm value funds is not met until 1993. Our sample period for the Morningstar mutual fund begins January 1993 and continues through February 2012. During that period there are 111,898 monthly observations for small-firm growth funds and 44,673 monthly observations for small-firm value funds. We also investigate a ten-year subperiod in which the dominance of the value mutual funds is strongest. For each month in our Morningstar sample we create a value superfund containing all the value mutual funds available that month. Likewise, we create a growth superfund containing all the growth mutual funds available that month. In both cases we determine the return to the superfund as the equally-weighted return of all funds in the sample.

We find data for the Fama-French factors from Kenneth French's website, which we gratefully acknowledge.

Empirical Results

Overview

We report empirical results on three different samples: 1) Russell 2000 Growth and the Russell 2000 Value Index over the entire sample period, January 1979 through February 2012, treating these indexes as representing returns from an index mutual fund; 2) The Morningstar value and growth superfunds sample over the entire period for which we have an adequate sample size, January 1993 through February 2012; 3) The Morningstar value and growth superfunds sample over the ten-year period where the performance of value funds relative to growth funds was most dominant, March 2003 through February 2010. For all three of these samples we follow the same procedure. First, we calculate the arithmetic mean and variance, testing for significant difference and calculating a risk-reward ratio. Second, we investigate end wealth of equal endowment placed in both funds. Third, we determine alpha when regressing portfolio returns against the Fama-French factors. To isolate the impact of the *HML* factor we do so in two stages. In the first stage we calculate alpha by regressing returns against market excess returns and the returns of the *SMB* portfolio. In the second stage we add the *HML* factor and observe the change in alpha. In all cases the risk return tradeoff favors the value portfolio. In all cases the addition of the *HML* factor reduces the relative performance of the value portfolio, confirming the bias against the performance of the value portfolio manager.

Comparing the Russell 2000 Value Index and Russell 2000 Growth Index

We use the actual returns on the Russell 2000 Growth Index and the Russell 2000 Value Index to proxy for the returns of hypothetical index mutual funds over the period January 1979 through February 2012. We calculate arithmetic mean returns and variance of monthly returns and test for significance of observed differences. Results are reported in Table 1.

Table 1: Monthly Percentage Returns - Russell 2000 Value vs. Growth Indexes (Jan. 1979 – Feb. 2012)

	Mean	**Variance**	**Sharpe Ratio**
Value	1.016	27.546	19.36%
Growth	0.720	47.649	10.43%
Test-statistic	1.770[a]	1.730[b]	3.271[c]
(*p*-value)	(0.039)	(0.000)	(0.000)

* A *p*-value of 0.000 indicates that the *p*-value is nonzero, but smaller than 0.0005.

a. H_0: $\mu_{value} = \mu_{growth}$; H_A: $\mu_{value} > \mu_{growth}$.

b. H_0: $\sigma^2_{value} = \sigma^2_{growth}$; H_A: $\sigma^2_{value} < \sigma^2_{growth}$.

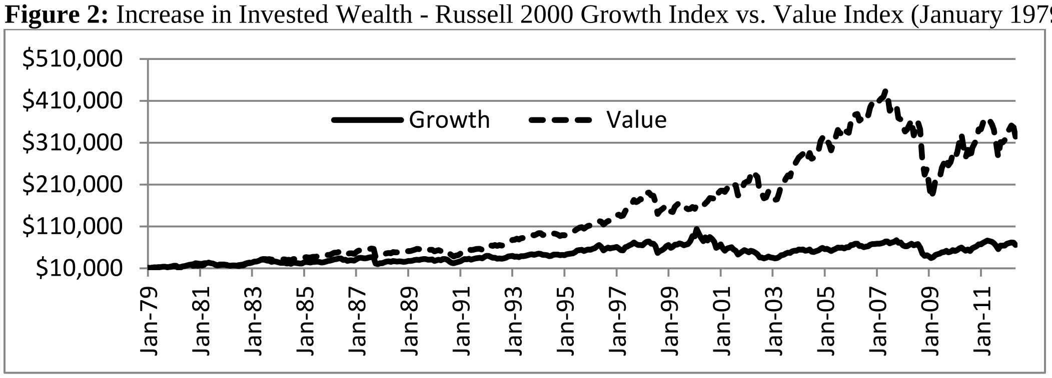
c. H_0: $SR_{value} = SR_{growth}$; H_A: $SR_{value} \neq SR_{growth}$.

The average monthly return to the value index of 1.02% is 0.30% higher than the average return of 0.72% for the growth index. Using a standard paired *t*-test we find this difference to be significant at the 5% level (for a one-tail test). The annualized difference is 3.60%. The contrast in the variance between the two return series is stronger. The monthly return variance for the portfolio of growth mutual funds is more than 70% higher than the monthly return variance for the portfolio of value mutual funds. The difference in the level of return variation between the value and growth portfolios is highly significant according to a standard F-test. Thus, the value funds achieve higher returns with lower realized risk. The Sharpe Ratio is 19.36% for the value index fund and 10.43% for the growth index fund. We use the methodology suggested by Opdyke (2007) to compare the Sharpe Ratios. The difference in the Sharpe Ratios is highly significant. Clearly the value index fund outperforms the growth index fund.

This difference in the variation in the two return series suggests that the comparisons of the arithmetic means understate the real difference in wealth creation between the two return series. This understatement results as the greater variation in the growth index returns causes greater fluctuations in the base used to calculate returns for the growth index, inflating the returns for the growth portfolio. This bias suggests the need to examine the difference in the geometric mean between the two series.[6] It is, after all, the geometric mean that determines the difference in the end value of an investment. Certainly this difference is the focus of an investor's concern. Therefore, we postulate the investment of $10,000 in each of the two index funds and determine the end wealth of that investment held through the end of our sample period, February 2012.

As shown in Figure 2, the difference in mean returns clearly understates the real difference in returns for an investment in the value index fund and the growth index fund. An investment of $10,000 in the growth index fund at the start of the sample period, January 1979, results in an end wealth of $65,382 at the conclusion of our sample period, February 2012. An investment of $10,000 in the value index fund at the start of the sample period, January 1979, results in an end wealth of $324,205 at the conclusion of our sample period, February 2012. The value investor has a startling five times the accumulated wealth at the end of the period! The difference in the geometric mean between the two samples is much greater than the difference in the arithmetic mean. The geometric mean of monthly returns for the value index is 0.87%. In contrast the geometric mean of the monthly returns for the growth index is 0.47%.

Figure 2: Increase in Invested Wealth - Russell 2000 Growth Index vs. Value Index (January 1979 - February 2012)



Clearly the value index fund outperforms the growth index fund. The value fund has a significantly higher arithmetic mean return while experiencing lower realized risk.[7] The risk to reward ratio (arithmetic mean return / standard deviation of returns) for the value index fund is almost twice as high as the reward to risk ratio for the growth index fund. An investor in the value fund would amass five times the wealth of the growth investor. Because the value index fund achieves a higher return with lower realized risk, an appropriate measure of risk-adjusted return ought to favor the value index. The result we find from applying the three-factor model contradicts this expectation.

Table 2 presents the results of regressing the value and growth index returns against the Fama-French factors. Panel A shows the results of regressing returns against market excess return and the *SMB* factor only. The estimated risk-adjusted return is positive for the value index, but insignificantly different from zero. The risk-adjusted return for the growth index is significantly negative. Thus, based upon the results from the two-factors, the comparative measures are consistent with the very real advantage of the value index.

Table 2: Factor Loadings and Risk-Adjusted Returns - Russell 2000 Value and Growth Index (Jan. 1979 – Feb. 2012)

	Risk-Adjusted Return (alphas) (*p*-value)	**Market Beta (*p*-value)**	**Size Beta (*p*-value)**	**Value Beta (*p*-value)**
(A) Excess Market and *SMB* Loadings				
Growth	-0.575%	1.213	1.092	---
	(0.000)	(0.000)	(0.000)	
Value	0.016%	0.909	0.761	---
	(0.851)	(0.000)	(0.000)	
(B) Excess Market, *SMB* and *HML* Loadings				
Growth	-0.470%	1.147	0.878	-0.193
	(0.000)	(0.000)	(0.000)	(0.000)
Value	-0.286%	0.994	0.680	0.596
	(0.000)	(0.000)	(0.000)	(0.000)

* A *p*-value of 0.000 indicates that the *p*-value is nonzero, but smaller than 0.0005.

In Panel B we introduce the influence of the *HML* factor. The growth index loads negatively on the *HML* factor reducing expected return and increasing risk-adjusted return. The value index loads positively on the *HML* factor increasing expected return and decreasing risk-adjusted return. As a result of the inclusion of the *HML* factor the value index now has a significantly negative risk-adjusted return. If the asset pricing model seeks to relate expected return to risk, the growth index should not have a lower expected return because of a non-zero exposure to the *HML* factor. Because, as shown above, a growth portfolio with a negative loading on the *HML* factor and a value portfolio with a positive loading on the *HML* factor with the same absolute value will provide the same contribution to total risk. An increase (a decrease) in the risk-adjusted return for the growth (value) portfolio when the *HML* factor is added results from model bias rather than the higher (lower) systematic risk of the value (growth) portfolio.

Comparing Performance of Morningstar Small-Firm Value and Growth Mutual Funds: January 1993 and February 2012

Returns to the Russell 2000 Value Index dominate returns to the Russell 2000 Growth Index. The value index has a significantly higher realized return and a significantly lower realized risk. If index funds had been available over the entire period, as they are now, an investor holding the value index fund would have held a clearly superior investment. We now examine returns from managed growth and value funds with data collected from the Morningstar data base. We begin our sample in January 1993, because the number of managed value funds before that date are insufficient to provide statistical accuracy. Results reported in Table 3 below show that managers of the small-firm value mutual funds fail to earn the average return reported in the Russell 2000 Value Index. The managers of the small-firm growth mutual funds, on the other hand, earn a return slightly higher than the average for Russell 2000 Growth Index. As a result, although the average return for the small-firm value mutual fund portfolio is larger than the average return for the small-firm growth mutual fund portfolio, the difference in the mutual fund portfolios is statistically insignificant. The annualized difference in monthly returns is 0.67%.

The portfolio of small-value mutual funds does maintain a reward to risk dominance over the portfolio of small-firm growth mutual funds, as the value portfolio has both a higher mean and a smaller variance. The difference in variance in monthly returns is very similar to the difference reported in the Russell 2000 Index and is statistically different. Value investing in the small-firm mutual funds provides a significantly lower level of risk than experienced by investing in small-firm growth mutual funds. The Sharpe Ratio for the superfund of value portfolios is 18.08% compared to the Sharpe Ratio of 12.75% for the superfund of growth mutual funds. These differences, however, are not statistically significant.

Table 3: Monthly Percentage Returns: Equally-Weighted Portfolio of Small-Firm Growth vs. Small-Firm Value Mutual Funds (January 1993 - February 2012)

	Mean	**Variance**	**Sharpe Ratio**
Value	0.893	24.397	18.08%
Growth	0.837	43.067	12.75%
Test-statistic	0.228[a]	1.765[b]	1.346[c]
(*p*-value)	(0.410)	(0.000)	(0.180)

* A *p*-value of 0.000 indicates that the *p*-value is nonzero, but smaller than 0.0005.

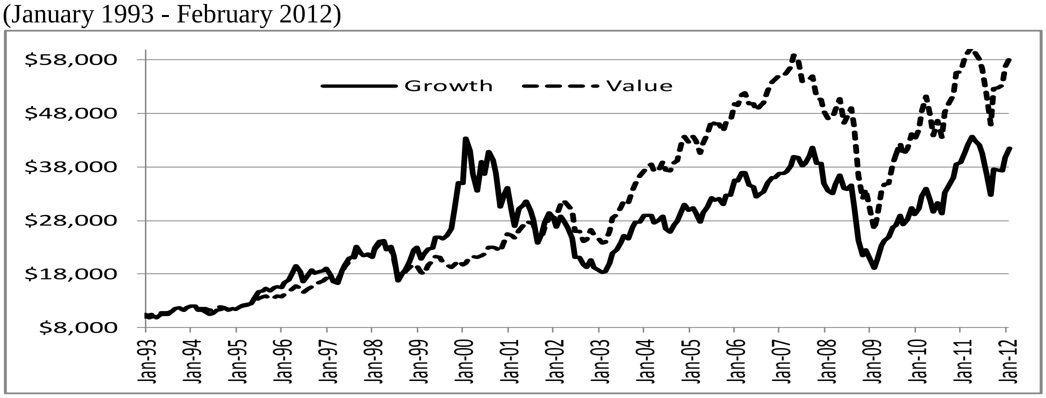
a. H_0: $\mu_{value} = \mu_{growth}$; H_A: $\mu_{value} > \mu_{growth}$.

b. H_0: $\sigma^2_{value} = \sigma^2_{growth}$; H_A: $\sigma^2_{value} < \sigma^2_{growth}$.

c. H_0: $SR_{value} = SR_{growth}$; H_A: $SR_{value} \neq SR_{growth}$.

As with the comparison of the Russell 2000 Indexes the risk to reward ratios of the superfunds suggest that the difference in the geometric mean should be greater than the difference in the arithmetic mean, resulting in a larger difference in end wealth than suggested by the comparison of arithmetic mean returns. Figure 3 shows the increase in value of a $10,000 investment at the start of January 1993 through the end of February 2012. An investor in the superfund portfolio of small-firm value mutual funds would enjoy an end wealth more than 40% higher than would an investor in the superfund portfolio of small-firm growth mutual funds. The end value of $58,277 for the small-firm value mutual fund portfolio associates with a geometric mean of 0.77% monthly or a 9.64% annualized return. The end value of $41,461 for the small-firm growth mutual fund portfolio associates with a geometric mean of 0.66% monthly or an 8.21% annualized return. The pattern in the value fund portfolio mimics closely the pattern of the Russell 2000 Value Index. The portfolio of growth mutual funds shows a very high rate of growth during the later period of the dot.com bubble. Strangely, the Russell Growth Index did not follow a similar pattern, emphasizing the divergence in performance of the growth index and the growth mutual fund portfolio.[8]

Figure 3: Increase in Invested Wealth- Portfolio of Small-Firm Growth Funds vs. Portfolio of Small-Firm Value Funds (January 1993 - February 2012)



Because of the significantly lower risk and the higher geometric mean return for the superfund of managed value mutual funds, one would expect calculation of risk-adjusted returns to show a superior performance by value mutual funds. We apply the same test procedures to the managed superfunds as we did to the Russell 2000 indexes. The results are shown in Table 4.

Table 4: Factor Loadings and Risk-Adjusted Returns - Small-Firm Value and Growth Mutual Funds (January 1993 - February 2012)

	Risk-Adjusted Return (alphas) (*p*-value)	Market Beta (*p*-value)	Size Beta (*p*-value)	Value Beta (*p*-value)
(A) Excess Market and *SMB* Loadings				
Growth	-0.182%	1.077	0.830	---
	(0.097)	(0.000)	(0.000)	
Value	0.101%	0.860	0.339	---
	(0.495)	(0.000)	(0.000)	
(B) Excess Market, *SMB* and *HML* Loadings				
Growth	-0.109%	1.057	0.774	-0.180
	(0.294)	(0.000)	(0.000)	(0.000)
Value	-0.137%	0.925	0.522	0.588
	(0.110)	(0.000)	(0.000)	(0.000)

* A *p*-value of 0.000 indicates that the *p*-value is nonzero, but smaller than 0.0005.

Regression results are similar to those for the Russell indexes. When using only the excess market return and SMB factors, the risk-adjusted returns for the growth superfund is significantly negative at the 0.10 level. The risk-adjusted return for the value superfund portfolio is positive but insignificantly different from zero. Still, the risk-adjusted return for value is superior to the risk-adjusted return for growth. The comparisons are different when the *HML* factor is added. The value superfund loads positively on the *HML* factor. This high loading causes the expected return for the value portfolio to increase and the risk-adjusted return for the value portfolio to decrease. Indeed the risk-adjusted return for the value superfund is now found to be negative, just outside the margin for significance. The loading on the growth superfund is negative suggesting that sensitivity to the *HML* factor reduces the risk of the superfund portfolio. But, as shown by realized return, the growth superfund has greater risk than value. An investor choosing between value and growth should not select growth because the negative loading on *HML* reduces risk.

Comparing Performance of Morningstar Small-Firm Value and Growth Mutual Funds: March 2000 and February 2010

We have established the existence of a value premium in the index data and in the mutual fund data. The value premium is most evident in the terminal wealth and geometric mean calculations. Especially for the mutual funds comparisons the lower variance in monthly returns for the value portfolio, the lower risk, creates this dominance. Our goal is to test for the effectiveness of the asset pricing models in adjusting raw returns to measure the performance of mutual fund managers. Results from applying asset pricing models to control for risk ought to be consistent with the return and risk differentials that we report above. In the total sample for the Russell index fund and the Morningstar mutual fund positive risk-adjusted returns for the value portfolios became negative risk-adjusted returns with the introduction of the *HML* factor. We seek to determine the strength of this influence by conducting the test in a period where the dominance of the value funds is particularly strong.

On an *ex catherda* basis we decide to select a ten-year period where the value dominance is particularly strong. We select the Sharpe Ratio as the variable with which to identify dominance. Beginning with December 2002 (the first month for which we have ten years of data) we calculate the arithmetic mean and standard deviation of monthly returns for a ten-year period separately for the portfolio of value and the portfolio of growth mutual funds. We then use these values to determine the risk to reward ratio. We repeat the process monthly using ten-year rolling samples. After calculating the risk to reward ratio we find the difference in the ratios between the two portfolios. We select the ten-year period with the highest difference in the Sharpe Ratio as the period where value is the most dominance. The ten-year period meeting this requirement is March 2003 through February 2010.

Table 5 reports the arithmetic mean return and variance of monthly returns for the value and growth superfunds. The mean monthly return for the value portfolio of 0.82% is significantly higher than the mean monthly return for the growth return of -0.07% at the 1% significance level. As in the other two samples, the realized risk is much lower for the value portfolio. The difference is highly significant. The Sharpe Ratio for the value portfolio is 15.34% and this ratio is negative for the growth portfolio. The difference in the Sharpe Ratio is highly significant. This comparison predicts a much higher end wealth for funds invested into the portfolio of value mutual funds

Table 5: Comparison of Monthly Percentage Returns - Equally-Weighted Portfolio of Small-Firm Growth vs. Small-Firm Value Mutual Funds (March 2000 - February 2010)

	Mean	**Variance**	**Sharpe Ratio**
Value	0.822	28.690	15.36%
Growth	-0.074	44.167	-1.12%
Test-statistic	2.751[a]	1.539[b]	2.969[c]
(*p*-value)	(0.003)	(0.000)	(0.003)

* A *p*-value of 0.000 indicates that the *p*-value is nonzero, but smaller than 0.0005.

a. H_0: $\mu_{value} = \mu_{growth}$; H_A: $\mu_{value} > \mu_{growth}$.

b. H_0: $\sigma^2_{value} = \sigma^2_{growth}$; H_A: $\sigma^2_{value} < \sigma^2_{growth}$.

c. H_0: $SR_{value} = SR_{growth}$; H_A: $SR_{value} \neq SR_{growth}$.

The end wealth from investing \$10,000 in the Value superfund over the period March 2000 through February 2010 is \$22,465. In contrast, the end wealth from investing \$10,000 in the Growth superfund over the same is \$6,983. Holding the value fund produces more than three times the wealth of holding the growth fund. The geometric mean of the monthly returns for the value index is 0.68% as compared to -0.30% for the growth index. The monthly difference corresponds to an annual difference of 12.42%. (Figure omitted for space consideration.)

This final examination of the *HML* factor occurs in a period where the portfolio of value funds experiences and clearly superior performance. Table 6 reports the risk-adjusted returns and the factor loadings for these regressions. The results differ only by degree from previous findings. As shown in Panel A, during this period, regressing returns of the mutual fund portfolios on the market excess return and the *SMB* factors results in positive excess returns for the value portfolio and negative excess returns for the growth portfolio. In both cases the risk-adjusted return is significantly different from zero at the 1% significance level. Introduction of the *HML* factor increases (decreases) expected return and decreases (increases) risk-adjusted return for the value (growth) portfolio of mutual funds. This result occurs because the factor loading on the *HML* factor is positive for the value portfolio and negative for the growth portfolio. The risk-adjusted return for the growth portfolio remains negative and the risk-adjusted return for the value portfolio remains positive. The risk-adjusted return for the value portfolio, however, is no longer significant.

Table 6: Factor Loadings and Risk-Adjusted Returns - Small-Firm Value and Growth Mutual Funds (March 2000 - February 2010)

	Risk-Adjusted Return (alphas) (*p*-value)	**Market Beta (*p*-value)**	**Size Beta (*p*-value)**	**Value Beta (*p*-value)**
(A) Excess Market and *SMB* Loadings				
Growth	-0.414%	1.061	0.734	---
	(0.007)	(0.000)	(0.000)	
Value	0.585%	0.883	0.307	---
	(0.010)	(0.000)	(0.000)	
(B) Excess Market, *SMB* and *HML* Loadings				
Growth	-0.341%	1.058	0.711	-0.078
	(0.030)	(0.000)	(0.000)	(0.000)
Value	0.030%	0.910	0.483	0.601
	(0.111)	(0.000)	(0.000)	(0.000)

* A *p*-value of 0.000 indicates that the *p*-value is nonzero, but smaller than 0.0005.

Summary of Empirical Results

We compare the performance of value and growth portfolios across three samples. In all three cases the arithmetic mean is higher for the value portfolio than the growth portfolio (in two of the cases the difference is statistically significant). In all three cases realized risk as measured by return variance is significantly lower for the value portfolio. Because of the impact of the significantly lower variance and the higher arithmetic mean, the value portfolios have higher Sharpe Ratios. The value portfolios also have much higher geometric means consistent with a much higher end wealth for given a level of investment as compared to a growth portfolio.

We submit that efficient measurement of risk-adjusted return also ought to favor the value portfolio. Based on the arguments developed in the introduction, however, we expect a bias in measuring the risk-adjusted returns against the value portfolio and in favor of the growth portfolio. We illustrate this bias by calculating risk-adjusted returns using market excess returns and the *SMB* factor and then recalculating the risk-adjusted returns including the *HML* factor. In all cases the addition of the *HML* factor increases the risk-adjusted return for the growth portfolio and decreases the risk-adjusted return for the value factor. The impact tends to be stronger for the value portfolio than the growth portfolio. Table 7 summarizes the results from adding the *HML* factor. The bias against the value portfolios is clearly evident.

Table 7: Impact of Adding the *HML* Factor in Measuring Risk-Adjusted Returns

Sample	Portfolio	Two Factor Measurement of Risk-Adjusted Returns	Three Factor Measurement of Risk-Adjusted Returns
Russell 2000 Indexes 01/1979 – 05/2012	Growth	Significantly Negative	Increased but Still Significantly Negative
	Value	Insignificantly Positive	Significantly Negative
Small-Firm Morningstar Funds 01/1993 – 02/2012	Growth	Significantly Negative	Insignificantly Negative
	Value	Insignificantly Positive	Insignificantly Negative
Small-Firm Morningstar Funds 03/2000 – 02/2010	Growth	Significantly Negative	Increased but Still Significantly Negative
	Value	Significantly Positive	Insignificantly Positive

Conclusion

This paper relies on the insight that systematic risk is the appropriate measurement of risk for selecting securities for portfolios but that realized risk is the appropriate measure of risk for portfolios on an *ex post* basis. Thus, appropriate measures of systematic risk ought to effectively predict realized risk. We argue that measuring systematic risk by the factor loading on a portfolio that is long in value securities and short in growth securities (the well-known *HML* factor of the Fama-French three-factor model) does not meet this criterion. Value securities tend to load positively on the *HML* factor and growth securities tend to load negatively on the *HML* factor. The higher the value of the loading the more "systematic" risk is assigned to the security by the Fama-French three-factor model. But, it is the absolute value of this factor loading that impacts realized risk. Thus, when risk-adjusted returns are determined on the basis of the factor loadings rather than the absolute values of the factor loadings a bias occurs against managers of value portfolios in judging performance.

We illustrate this bias with three samples of portfolios containing small-firm value and growth securities. In all three samples, the value portfolio has a higher arithmetic mean and a decidedly higher geometric mean. The difference in the geometric and arithmetic mean of the value and growth portfolios results from differences in realized risk of the value and growth portfolios. In all three samples the value portfolios have significantly lower realized risk. But, risk-adjustment based on the factor loading on the *HML* factor within the Fama-French three-factor model assigns greater risk to the value portfolio. In all three cases the effect of the *HML* factor is to increase the risk-adjusted return for the growth portfolio and decrease the risk-adjusted return for the value portfolio. The impact of the *HML* factor causes the Fama-French three-factor model to be biased in measuring the performance of value portfolios.

Acknowledgment

The authors would like to thank Jocelyn Evans and other participants at the Academy of Economics and Finance 2013 for their helpful comments and discussion. Professor Pettengill gratefully acknowledges support from the E. Seidman chair.

Notes

[1] We recognize that each variable may experience different impacts on variation from the other factors in the Fama-French three-factor model, but any such variation is not endogenous to the model.
[2] Notice that a similar bias would exist for small firm securities relative to the *SMB* factor, but that the bias will be largely missing from the excess market return factor as the vast majority of securities will load positively on market excess return. Thus, this particular bias is generally missing from the Capital Asset Pricing Model (CAPM). Textbook often use gold stocks as an example of a negative beta security. Causal observations of the expectations of gold investors suggest that these investors are expecting returns commensurate with risk higher than that of the risk-free asset.
[3] In determining expected return the influence of the factor loading will depend on the sign of the factor in a particular period. In application of the CAPM the expected return of high beta securities will be lower than the expected return for low beta securities in down markets (see Pettengill, Sundaram and Mathur (1995)), and, according to the Fama-French three-factor model, the expected return for securities with high loadings on the *SMB* factor will be lower than the expected returns for securities with low loading on the *SMB* factor when the *SMB* factor is negative, everything else constant.
[4] Of course the *HML* factor is on average positive and under our other assumptions the value portfolio would outperform the growth portfolio. The same bias, however, remains.
[5] Among the funds providing matching returns to the Russell indexes are the iShares Russell 2000 Value ETF and the iShares Russell 2000 Growth ETF.
[6] For the sake of clarity we offer an example. Assume two assets with a given return series: Asset A has a return series that consists of returns of -10%, 10%, and 3%; Asset B has a return series that consists of returns of -50%, 50% and 21%. The return series for A has a arithmetic mean return of 1% and a standard deviation of 10.15%. The return series for B has a arithmetic mean return of 7% and a standard deviation of 51.45%. The higher arithmetic mean return for series B inaccurately suggests a greater return for investing in B at the cost of a greater risk. In fact, if $100 was invested in Asset A the end wealth would be $101.97. A similar investment in Asset B would result in an end wealth of $90.75. The comparative geometric means of 0.65% for Asset A and -3.18% for Asset B more correctly reflect the return to an investor.
[7] The difference in scale in Figure 2 obscures the greater variation around the mean return for the growth index. When the log of wealth is plotted the greater variation for the growth index is clearly evident. Corresponding log figures for all three graphs of invested wealth are available from the authors.
[8] We hypothesized that herding on the part of managed growth funds may be responsible for this divergence.

References

Basu, S. "The Relationship Between Earnings Yield, Market Value and Return for NYSE Common Stocks: Further Evidence." *Journal of Financial Economics*, 1983, pp. 129-156.

Chan, K., Chen, H., and Lakonishok, J. "On Mutual Fund Investment Styles." *Review of Financial Studies*, Winter 2002, pp. 1407-1437.

Clash, J. "The Wallflower Strategy." *Fortune*, February 2, 1998, pp.113-117.

Davis, J., "Mutual Fund Performance and Manager Style." *Financial Analysts Journal*, Jan-Feb. 2001, pp. 19-27.

Fama, E. and French, K. "The Cross Section of Expected Stock Returns." *The Journal of Finance*, 1992, pp. 427-465.

Fama, E. and French, K. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics*, 1993, pp. 3-56.

Fama, E. and French, K. "Multifactor Explanations for Asset Pricing Anomalies." *The Journal of Finance*, 1996, pp. 55-84.

Opdyke, J. "Comparing Sharpe ratios: So where are the p-values?" *Journal of Asset Management*, 2007, 8-5, pp. 308-336.

Pettengill, G., Sundaram, S. and Mathur, I. "The Conditional Correlation Between Beta and Returns." *Journal of Financial and Quantitative Analysis*, March 1995, pp. 101-116.

Shi, W. and Seiler, M. "Growth and Value Style Comparison of U.S. Stock Mutual Funds." *American Business Review*, January 2002, pp. 25-32.

The Relationships between Gasoline, Agricultural Feedstocks, and Exchange Rates: A Cointegration Analysis

Abbes Tangaoui and Michael Farmer, Texas Tech University

Abstract

Over the last few years, there has been a growing interest on biofuels as clean energy alternatives to fossil fuels to reduce the US reliance on foreign oil, limit the greenhouse gas emissions and promote rural development. In this paper, we use daily settlement futures prices from 2002 to 2012 to investigate price relationships between energy market, feedstock markets, and the US foreign exchange rate using Johansen cointegration methodology. Our results suggest that gasoline, agricultural feedstocks and exchange rate are only weakly cointegrated during the year 2010.

Introduction

Over the few past decades, the landscape of the energy market has witnessed major and profound mutations. The oil shock of the 1970s generated serious concerns about energy dependence and this shifted interest in biofuels as energy alternatives to the major imported transportation fuels. During the 1980s and 1990s, global warming gave new impetus to use biofuels as alternatives to heavy GHG fossil fuels; so the emphasis on using agricultural output directly as a transportation fuel accelerated. Consequently, feedstock prices, such as corn, experienced a sharp increase seemingly in parallel to energy prices. These changes have raised expectations about the linkages between agricultural commodity prices and transportation fuel prices. Understanding the interplay between agricultural and the energy prices, both in the long and short run, constitutes a valuable tool for policy making.

Our study aims to shed light on the relationship between agricultural commodity and energy markets. We investigate the long and short-run dynamics in prices of selected feedstocks, gasoline and the US foreign exchange rate. The rest of the paper is organized as follows. In section II we give a brief overview of the literature. Section III describes the data and the methodology used in this study. Section IV is dedicated to results and discussion. Conclusions are in section V.

Literature Review

The growing interaction between food and energy needs has been the subject of intensive academic and public debate during the last decade. Agriculture is increasingly conceived of as an important sector to provide clean energy in addition to its traditional role of providing food. The neat separation between energy and agriculture markets that prevailed for many decades is being questioned. A simplistic representation of the traditional linkages between agricultural and energy prices is that higher energy prices increased the costs of agricultural production, thereby lowering farm profits and raising agricultural commodity prices the next season. With great asset fixity in agriculture, the subsequent price rise was likely modest as nearly the same amount was still produced in the short run. A long run secular change to higher energy prices however would induce some structural adjustment in agriculture.

Now as the linkages between agriculture and energy markets have been tightening. The advent of directed crops, which reorient some portion of high energy commodity crops such as corn and sugar cane to produce transportation fuels, may pair energy and agricultural commodity prices more closely. For those directed crops, prices may move together quite powerfully; or the shift in farmland from some crops to directed crops able to take advantage of this tight correlation may be strong enough that production increases overall in those crops in the long run may moderate positive price correlation. As there are many scenarios possible depending on the relative price elasticities of agricultural inputs and what is now several new output markets, the search for evidence in the price trends themselves has been on-going to unpack the short term and long term comovements in prices of different agricultural commodities and transportation fuels.

Most academic research has focused on both short and the long run linkages terms by applying different tools of statistical and quasi-deterministic analyses. The models generally fall into structural models that rely on general or partial equilibrium and econometric models

Campiche et al. (2007) use vector error correction model to study the covariability between crude oil prices and corn, sorghum, sugar, soybeans, soybean oil, and palm oil prices from 2003 to 2007, they find corn and soybean prices were only cointegrated during the period 2006-2007, but were not cointegrated in the period 2003-2005. They also find that crude oil prices do not adjust to changes in the corn and soybean markets. Furthermore, the authors find that soybean and palm oil prices were, through the biodiesel market, more correlated to the crude oil prices than corn in the 2007 time period. Yu et al. (2006) analyze the cointegration and causality between crude oil prices and vegetable oils used in biodiesel production, including soybean, sunflower, rapeseed, and palm oil, using weekly data extending from the first week of January of 1999 to the fourth week of march 2006, they find that shocks in petroleum prices have an insignificant influence on the variations in edible oil prices.

Yu et al. (2006) analyze the cointegration and causality between crude oil prices and vegetable oils used in biodiesel production, including soybean, sunflower, rapeseed, and palm oil, using weekly data extending from the first week of January of 1999 to the fourth week of March 2006. They find only one cointegration relationship between crude oil prices and the vegetable oil prices. With respect to causality, they find that edible oil markets have strong linkages in contemporaneous time. Furthermore, they show, using variance-covariance decomposition and the Impulse Response Function, that edible oil were only affected by their own shocks in the short run. However, they find that shocks in the soybean oil prices to be a significant factor influencing the uncertainty of edible oil prices in the long-run, while shocks in petroleum prices do not have any significant influence on the variations in edible oil prices. The authors argue that the influence of crude oil price on edible oils will grow if high oil prices continue and edible oils become an increasing source of biodiesel.

Harri, Nally and Hudson (2009) conduct a cointegration analysis to investigate the relationship among crude oil, exchange rate and five others agricultural commodities, including corn, soybean, soybean oil, cotton and wheat. Using monthly data from January 2000 to September 2008, they find that crude oil and corn cointegrate starting mid-2006; the authors explain the crude oil-corn cointegration relationship by the increasing use of corn for ethanol production. They also find that crude oil, exchange rate and corn markets were interrelated.

More recently, Serra, Zilberman, Gil, and Goodwin (2010), investigate the relationship between ethanol, corn and crude oil markets. They use an Exponential Smooth Transition Vector Error Correction Model (ESTVECM). They find one cointegration relationship in which ethanol price was the only endogenous variable in the system, and positively related to the exogenous prices of corn and crude oil. The authors explain this finding by the fact that ethanol is crude oil substitute in the fuels markets, and therefore, an increase in the crude oil prices is not expected to compromise the competitiveness of the ethanol industry within US. However, they note that an important increase in the Brazilian ethanol prices in the order of magnitude of 60% of the growth rate of crude oil prices would cause the U.S. ethanol to be less competitive in international markets. With respect to corn, the authors explain the positive relationship of ethanol to corn in the error correction model by the fact that corn is direct input in the production of ethanol, and consequently ethanol-based demand drives corn prices up, which, in turn raises the ethanol prices. However, they note that, given the limited area dedicated to corn production at least in the short run, prices of corn are expected to rise with the expansion of the ethanol industry, which constitutes a limitation to the U.S. ethanol industry. They, further, use a Generalized Impulse Response Function (GIRF). They find that a shock to both crude oil and corn prices when the system is far from the equilibrium causes changes in ethanol prices in the same direction; with ethanol response to crude oil price shocks slower and smaller than ethanol response to corn.

Serra, Zilberman, and Gil (2011) conduct a similar study for crude oil, ethanol and sugar cane in Brazil using weekly data from July 2000 to February 2008. They use a maximum likelihood estimator that jointly estimates an error correction term (ECT) and volatility spillover. They find one cointegration relationship between ethanol, crude oil and sugarcane, in which ethanol price was the only endogenous variable and positively related to the weakly exogenous oil and sugar prices. They explain this cointegration relationship as a long-run parity of ethanol industry to be in equilibrium with respect to crude oil and sugarcane markets (i.e. to obtain normal profits). Similarly to the U.S. case, they explain the positive relationship of corn to oil and sugar by the fact that sugarcane is a feedstock input to the Brazilian ethanol industry, and that ethanol and oil in Brazil operate as substitutes. They also argue that in the short-run, ethanol prices were driven by their own dynamics as well as by oil and sugar lagged prices changes. Their volatility spillover analysis suggests that all markets considered in their study were affected by their own volatilities, and directly and indirectly affected by past turbulences in the other markets. They further use an Impulse Response Function to study shocks across the three markets, both for the mean and volatility prices. Their results suggest that an increase in crude oil prices leads the system towards a new equilibrium characterized by higher ethanol prices.

Zhang et al. (2009) employ cointegration analysis, Vector Error Correction and multivariate BEKK-GARCH models to investigate long-run relationships and volatility linkages of the US markets of ethanol, oil, gasoline, corn and soybean. They use weekly wholesale prices, for the selected commodities, from the last week of March 1989 through the first week of December 2007. They do not find any long-run equilibrium among the fuel prices and the agricultural commodities prices. Furthermore, they find that agricultural commodities prices volatility impact the fuel prices volatility.

Du, Yu, and Hayes (2011) study the volatility spillover of crude oil price to corn and wheat prices. Using Bayesian Markov Chain Monte Carlo methods on weekly data, from November 16, 1998 to January 26, 2009, they find evidence of volatility spillover among crude oil, corn and wheat after the fall 2006. The authors conclude that these prices volatility transmission was probably due to the expansion of the ethanol market.

Data and Methodology

We use daily settlement futures prices from the New York Mercantile Exchange (NYMEX) for gasoline, corn, soybean, soybean oil, sugar and foreign exchange rate(the US Dollar index), which is a weighted average of the value of the U.S. Dollar against the currencies of a group of major US trading partners. The time period spans from August 2002 to august 2012.

As a first step, we test for cointegration between gasoline and the selected feedstocks prices to get insights about any long-run relationships between these markets. The concept of cointegration suggests that two non-stationary processes at the level, but stationary at their first difference, might have a stationary linear combination, when such a relationship exists, the two series are said to be cointegrated (Engel and Granger, 1987). The Johansen multivariate cointegration procedure (Johansen, 1988) and (Johansen and Juselius, 1990) is widely used to test for long-run and short term dynamics among time series data. The Johansen cointegration procedure can be conducted based upon the following error-correction model (VECM) formulation:

$$\Delta Y_t = \Pi Y_{t-1} + \Sigma\, \Gamma_i \Delta Y_{t-1} + \varphi D_t + \varepsilon_t \quad (1)$$

Where, Yt is a (nx1) vector of non-stationary variables to be examined for cointegration, Π, Γ_i and φ_i are matrices of parameters to be estimated, D_t is a vector of deterministic regressors (constant, trend, dummies), and ε_t are random errors. The lag length k selection is based on minimum value of information criteria. The cointegration relationships are examined by evaluating the rank r of the matrix Π. If the rank r is zero, then all variables are non-stationary and not cointegrated. If $0<r<n$, then there exist r cointegrating vectors, and if the matrix Π is of full rank, then all variables in the vector Y_t are stationary.

The Johansen procedure involves the estimation of the matrix Π, through an unrestricted vector autoregressive (VAR) model, and tests restrictions implied by the reduced rank of the matrix Π. Two likelihood ratio tests are used to test for long-run relationships (Johansen and Juselius, 1990), the trace test and the maximum eigenvalue test:

$$\lambda_{trace} = -T\, \Sigma \ln(1 - \lambda^2) \quad (2)$$

$$\lambda_{max}(r,r+1) = -T\, \Sigma \ln(1 - \lambda_{r+1}) \quad (3)$$

Where, T is the number of the observations and λ_i are the estimates of the ordered eigenvalues obtained from the estimation of the matrix Π. The trace statistics tests the null hypothesis of at most r cointegrating vectors against a general alternative hypothesis of more than r cointegrating vectors. The maximal eigenvalue tests the null hypothesis of r cointegrating vectors against the alternative of r+1 cointegrating vectors.

Results

Unit Root Tests

The first step in our data exploration was a series of unit root tests. The Augmented Dickey-Fuller (ADF) and the Phillips-Perron (PP) test are used to check for the existence of unit roots in gasoline[1], corn, soybean, soybean oil prices and the exchange rate series. The (PP) test has the advantage over the (ADF) test, in that the (PP) test is robust to general forms of heteroskedasticity in the error term. Another advantage of the (PP) test is that the user does not have to specify a lag length.

These tests were first carried out in the natural logarithms of the price series. The (ADF) and the (PP) tests reveal that all series are non-stationary. These tests were again carried out in first differences and all of the series have been identified to be first-difference stationary. The lag length selection was based on the Akaike Information Criteria (AIC).

Cointegration tests

We use Johansen cointegration test to check whether the variables in the present study cointegrate. The AIC was used to set the lag length for the cointegration relations, a lag length of two for all prices was appropriate. We start by applying the test recursively from the year 2002 and move forward by adding a year at a time. We found a consistent cointegration relationship between all variables during the 2010 time frame.

Table 1 displays the cointegration tests results for the trace test and the maximum eigenvalue test respectively. Both tests reveal that the variables have one cointegration relationships. However, the maximum eigenvalue test is significant at the 5% level, while the trace test reveals only one weak cointegration relationship at the 10%. The presence of a cointegration relationship among gasoline, feedstock markets and the foreign exchange rate, during the year 2010 time frame, suggests most notably that the market of energy and feedstock market were not cointegrated prior 2010. However, the ethanol boom in 2006 and the increasing use of corn for large scale ethanol production, as well as the important use of soybean, soybean oil and sugar for the production for biodiesel, have had an important impact on the traditional gasoline market, and consequently, a few year later after the ethanol boom, the feedstocks used in the production of biofuels (ethanol and biodiesels) and gasoline enter a long- term period of mutual interactions starting in 2010.

Table 1: Cointegration Tests Results

		Trace test				***Maximum test***	
Statistics	***10%***	***5%***	***1%***	***Statistics***	***10%***	***5%***	***1%***
98.77	97.18	102.14	111.01	42.17	37.45	40.30	46.82
56.60	71.86	76.07	84.45	23.96	31.66	34.40	39.79
32.64	49.65	53.12	60.16	13.68	25.56	28.14	33.24
18.96	32.00	34.91	41.07	9.82	19.77	22.00	26.81
9.14	17.85	19.96	24.60	6.93	13.75	15.67	20.20
2.2	7.52	9.24	12.97	2.20	7.52	9.24	12.97

Weak Exogeneity

In a cointegrated system, a variable is weakly exogenous if no useful information is lost when this variable is conditioned without specifying its generating process. In other words, if a variable does not respond or responds with a very low speed to the discrepancy from the long-run equilibrium relationship, it is weakly exogenous. The weak exogeneity test follows a χ^2 distribution and is used to identify the weak exogeneity effect of each variable to the others. Table 2 shows that none of the prices considered in this study is weakly exogenous, which means that after a short term deviation from the long run equilibrium, prices tend to adjust relatively quickly to the equilibrium and, therefore, there is a long-run Granger causality among these prices.

Table 2: Weak Exogeneity Tests

Variable	χ^2	***p-value***
Gas	29.87	0.00
Corn	27.34	0.00
Soy	30.32	0.00
Soybean Oil	28.29	0.00
Sugar	25.83	0.00
Exchange Rate	29.91	0.00

The Vector Error Correction model

The presence of a cointegration relationship among the price variables, as revealed by the Johansen tests, sets the stage for the estimation of a vector error correction model (VECM). For a set of non-stationary variables, the VECM has the advantage of estimating short-term dynamics, while taking into account the long-run equilibrium between these non-stationary variables, which, de facto, eliminates the problem of spurious relationships.

In order to estimate a VECM for our price variables, we have to determine first the appropriate number of lags in each equation of differenced log prices. Different lag lengths were tested using the likelihood ratio (LR) test supplemented by the AIC information criterion, a lag length of two was appropriate for all price series. Table 3 shows results from the estimation of the Error Correction model, the long-run equilibrium is represented by the error correction term (ECT), and short term dynamics are the first differenced series.

An examination of table 3 reveals that all the coefficients of error correction term are significant, except those for sugar and the foreign exchange rate. The fact that the coefficients of the ECT for sugar and the foreign exchange rate are not significant, confirms the weak cointegration relationship we found earlier.

From a short run dynamics perspective, with the exception of soybean, feedstocks used in the biofuels production do not seem to have any short term impact on the gasoline market. On the contrary, and unlike most of the feedstocks, foreign exchange rate seems to have a short term influence on the gasoline market in the US, this is because the US gasoline market reacts actively to the short-term-volatility that characterize the international market of energy.

Table 3: Vector Error Correction Model Estimates

Variable/Equation	ΔGas_t	$\Delta Corn_t$	ΔSoy_t	ΔSoybean Oil_t	$\Delta Sugar_t$	ΔExchange $Rate_t$
ΔECT_{t-1}	-0.005***	-0.005**	-0.005*	-0.001	0.008*	-0.001
	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)
ΔGas_{t-1}	0.012	-0.011	-0.005	0.017	0.005	-0.006
	(0.022)	(0.015)	(0.013)	(0.013)	(0.016)	(0.004)
$\Delta Corn_{t-1}$	-0.052	0.045	-0.044***	-0.014	0.030	-0.012***
	(0.037)	(0.026)	(0.023)	(0.023)	(0.023)	(0.007)
ΔSoy_{t-1}	0.098***	0.025	0.044	0.024	0.063	0.020**
	(0.054)	(0.039)	(0.033)	(0.033)	(0.039)	(0.010)
ΔSoybean Oil_{t-1}	0.012	-0.082**	-0.057***	0.013	-0.002	-0.003**
	(0.051)	(0.036)	(0.031)	(0.037)	(0.010)	(0.010)
$\Delta Sugar_{t-1}$	0.043	0.018	0.014	0.019	-0.101*	0.003
	(0.029)	(0.021)	(0.018)	(0.018)	(0.021)	(0.006)
ΔExchange $Rate_{t-1}$	0.216***	-0.028	-0.024	-0.047	-0.004	-0.020
	(0.115)	(0.082)	(0.070)	(0.070)	(0.083)	(0.022)
Normality test	:33365.36					
Skeweness	: 169.62					
Serial Correlation Test	:33169.54					
ARCH test	: 4984.64					

*significant at 1% level, ** significant at 5% level, *** significant at 10% level

Conclusion

As the biofuels industry continues to expand at a high pace, the relationship between energy and agricultural feedstock markets tightens. Although our study reveals that these markets are only weakly integrated over our sample period, and this is because the biofuel industry is still in its infant stage, expectations remain important concerning strong markets integration between energy and agricultural markets in the future.

Notes

1. We use Gasoline prices instead of Crude Oil prices because, although the Augmented Dicky Fuller, and Phillips and Perron tests reveal that Oil prices over our sample period were non-stationary, the Zivot and Andrews test, on the contrary, revealed that oil prices were trend-stationary with a broken trend. Therefore, we dropped Oil prices from our analysis. Results for the Zivot and Andrews test are available from the author upon request

References

Campiche, Jody L., Henry L. Bryant, James W. Richardson, and Joe L. Outlaw. 2007. " Examining the Evolving Correspondence between Petroleum Prices and Agricultural Commodity Prices." *Paper presented at the American Agricultural Economics Association Annual Meeting*, Long Beach, Ca, July 29-August 1, 2007.

Dickey, David A., and Wayne A. Fuller. 1979. "Distribution of the Estimators for Autoregressive Time Series with a Unit Root." *Journal of the American Statistical Association* 74: 427-431.

Du, Xiaodong., Cindy L. Yu , Dermot J. Hayes. 2011."Speculation and volatility spillover in the crude oil and agricultural commodity markets: A Bayesian analysis" *Energy Economics* 33(2011) 497-503

Engel, Robert F., and Clive WJ. Granger. 1987. "Cointegration and Error Correction Representation, Estimation, and Testing." *Econometrica* 55: 251-276.

Harri, Ardian., Lanier. Nally, and Darren. Hudson. 2009. "The relationship between Oil, Exchange Rates, and Commodity Prices." *Journal of Agricultural and Applied Economics* 41(2009):501-510.

Johansen, Soren. 1988. "Statistical Analysis of Cointegrating Vectors." *Journal of Economics Dynamics and Control* 12: 231-254.

Johansen, Soren., and Katarina.1990. Juselius. "Maximum Likelihood Estimation and Inference on Cointegration with Applications for the Demand for Money." *Oxford Bulletin of Economics and Statistics* 52: 169-210.

Phillips, Peter CB., and Pierre. Perron. 1988. "Testing for a Unit Root in Time Series Regression." *Biometrika* 75(2): 335-346.

Serra, Teresa., David. Zilberman,Jose M. Gil, and Barry K. Goodwin. 2010. "Price Transmission in the US Ethanol Market." In D. Zilberman, M. Khanna, and J. Scheffran, eds. *Handbook of Bioenergy Economics And Policy* 4New York, NY: Springer New Yoek, vol.33 of *Natural Resource Management and Policy*,chap.5, pp.55-72.

Serra, Teresa., David. Zilberman, Jose M. Gil. 2010. "Price Volatility in Ethanol Markets." *The European Review of Agricultural Economics* 38(2): 259-280.

Yu, Tun-Hsiang., David A. Bessler, and Sephen W. Fuller. 2006. " Cointegration and Causality Analysis of World Vegetable Oil and Crude Oil Prices." *Paper presented at the American Agricultural Economics Association Meeting*, Portland, OR, July 23-26, 2006.

Zhang, Zibin., Luanne. Lhor, Cesar. Escalante, and Michael. Wetzstein. 2009. "Ethanol, corn, and soybean Price Relations in a Volatile Vehicle-Fuels Market." *The European Review of Agricultural Economics Energies* 2: 320-339.

Zivot, Eric., and Donald WK. Andrews. 1992. "Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis" Journal of Business & Economic Statistics Vol. 10, No. 3.

South Africa's Trade Functions with India and Brazil: The Relevance of the India-Brazil-South Africa Dialogue (IBSA)

Ranjini Thaver, Stetson University

Abstract

This research study investigates South Africa's trade with Brazil and India with reference to the India-Brazil-South Africa Dialogue (IBSA) of 2003. The study employs the bounds testing approach to cointegration. Results indicate that a long run cointegration relationship exists between exports and imports and their determinants in both South Africa-India and South Africa-Brazil trade. Estimates of the long-run and short-run elasticities of imports and exports meet theoretical expectations, and are significant in South Africa-Brazil trade, but not in South Africa-India trade. Surprisingly IBSA has not affected trade between South Africa, Brazil, and India significantly in the short and long-run.

Introduction

While South Africa is deemed one of the fastest growing emerging market economies in the world, it manifests serious trade deficiencies in the global economy: it is import dependent on capital and intermediate goods from the west; its share of exports in international markets is declining, as is its foreign reserves; and it suffers from a negative balance of payments. Unlike most other developing countries, these problems are primarily a result of the rigidities imposed by the apartheid state through the distorted allocation of resources and the racialization of the economic structures of accumulation at the expense of efficiency considerations.

With the end of apartheid, the newly established government liberalized the economy, executing a series of strategic trade policies, among them, promoting privatization, loosening exchange controls, and reducing tariffs and export subsidies. At the same time, it increased efforts to reduce its trade dependency on the US and Europe by forging intra-African trade. In 1996, it implemented the Growth, Employment, and Redistribution policy (GEAR) intended to increase its trade positioning in the global economy. GEAR was reinforced in 2001 by the adoption of the New Partnership for Africa's Development, NEPAD, which embodied an alliance of key African leaders in a quest to integrate African markets. In 2003, South Africa, along with middle-power emerging economies, Brazil and India, agreed to expand economic growth, employment, social development and increase standards of living, by creating a trilateral alliance among them. Economically, this would encourage trade among them with the idea of reducing trade with the traditional superpowers. Hence, they came together to form the India-Brazil-South Africa Dialogue Forum, or IBSA. The driving force behind IBSA was that together they would serve as the southern hemisphere's trade and economic powerhouse. Despite the good intentions of IBSA, there exists no empirical estimation of SA's import and export demand functions with India and Brazil, let alone studies investigating the success of IBSA in enhancing trade among these countries. Undertaking these tasks is the objective of this study.

We proceed in the next section with a brief review of the literature on import and export demand functions. We then specify our theoretical model and explain the data used for estimation. Thereafter we explain and discuss the empirical results, and the final section concludes the paper.

Literature Review

Given the importance of understanding how import and export demand changes with economic and political conditions, numerous studies have estimated these functions of countries on all continents. Earlier studies, focusing on developed countries, used ordinary least squares (OLS) methods to estimate these functions (McKenzie and Brooks, 1997; Gafar, 1995; Giovanneti, 1989; Thursby, 1988; Akhter and Hilton,1984). However, researchers have questioned the use of OLS methods in analyzing import and export demand functions precisely because this method assumes that the time series data used in such estimates are stationary (Thursby, 1989). However, macroeconomic time series are typically non-stationary, as observed by the high serial correlation between successive observations, and therefore results from OLS studies may cause serious spurious regression problems (Modeste, 2011) leading to unreliable results.

To overcome these problems, the last decade has witnessed a surge of studies employing refined econometrics methods to estimate the import demand and export demand functions for different countries. Hibbert, Thaver, and Hutchinson (2012), for example, investigate Jamaica's aggregate import demand function with the United States and the United Kingdom from 1996 to 2010. Their findings indicate that in trade with the US, income has a lower negative elasticity in the short run than the long run, whereas relative prices is more elastic in the short run. In Jamaica's trade with the UK, GDP is less elastic in the short run than in the long run, while relative prices adjusts much faster in the short run. Grullon (2012) employs the bounds testing approach to cointegration to examine the Dominican Republic's export demand function between 1960-1984 and 1985-2005. Two different export demand functions areestimated under different economic regimes, and both reveal a cointegrating relationship between the dependent variable and its regressors, foreign income, and relative prices. They also reveal that during the autarkic period, exports are relatively price elastic to changes in relative prices, whereas during economic liberalization, exports are price inelastic, and devaluation does not influence economic growth significantly. Ozturk and Kalyoncu (2009) contribute to the empirical discussion on the impact of exchange rate volatility on the trade of six countries including South Korea, Pakistan, Poland, South Africa, Turkey, and Hungary during1980-2005. They draw on the Engle-Granger (1987) residual-based cointegration technique, and their results show that there is a long-run equilibrium relationship among real exports, relative prices, real foreign demand, real exchange rates, and exchange rate volatility. The signs of the coefficients meet theoretical expectations for Pakistan and South Korea, but not for any other country in their study. They find further that real foreign income has the greatest impact on exports. Razafimahefa and Hamori (2005), in comparing the aggregate import demand function of Madagascar with Mauritius for the period 1960-2000, show that Madagascar's long-run income elasticity is higher than that of Mauritius, but their long-run relative price elasticities are almost equal and highly elastic. Tsionas and Christopoulos (2004) in examining the import demand function of five industrial countries including France, Italy, the Netherlands, the UK, and the US, reveal significant effects from relative prices and incomes, as well as significant short-run effects from temporary shocks.

Very few efforts have estimated the import and export demand functions of Sub-Saharan African countries. Among them, Ekanayake and Thaver (2011) utilize Pesaran's bounds testing approach to study the impact of exchange rate volatility of the top ten export industries from the US to South Africa between 1990 and 2010. Their results confirm a cointegrated relationship between exports of seven of the ten industries (machinery, passenger vehicles, aircraft and spacecraft, optical and medical instruments, organic chemicals, cereals, and plastic) and their associated regressors, while electrical machinery, mineral fuel and oil, and miscellaneous chemical products are not cointegrated. Gumede (2000) estimates the import demand function for South Africa from 1972-1997. His results reveal a long-run significant income elasticity of import demand, but short-run elasticities are less significant. Thaver and Ekanayake (2010), upon estimating South Africa's aggregate import demand function from 1950 to 2008, explicitly investigate the impact of apartheid, in particular international sanctions against the apartheid government, on South Africa's aggregate imports. Their results reveal that apartheid exerted a negative influence on its import demand function in the short run only, while international sanctions affected imports positively in the short-run, but negatively in the long-run.

To date, we are aware of no study that has been undertaken to analyze the effects of the IBSA on trade between South Africa-India (SA-India) and South Africa-Brazil (SA-Brazil), which is the objective of this paper.

Theoretical Model and Data Sources

Earlier time-series studies on import and export demand functions assumed that economic data are stationary, displaying constant means and/or variances over time. Hence, the classical linear regression model was the most commonly utilized method of empirically testing import and export demand functions. This simplistic assumption led to misspecification of models and spurious regressions, in which variables, though unrelated, sometimes appeared to be statistically significant, leading to forecasting errors (Hendry & Juselius, 2000). In reality economic data are influenced by such factors as multilateral economic arrangements, technological change, political chaos, legislation, nominal and real economic growth, and other structural changes in the economy, leading to non-stationary processes, or stochastic trends.

More recently scholars have developed sophisticated time series models that allow for the elimination of non-stationarity by applying transformations in a process called cointegration. The rationale behind the theory of cointegration is that while the dependent variable and its regressors may be individually non-stationary, they will nonetheless tend to move together over time, rendering a linear combination of them stationary, with the dependent and independent variables cointegrated (Engle and Granger, 1987).

One method of cointegration, the bounds testing approach developed by Pesaran, Shin, and Smith (2001), has been widely used by scholars, among them, Grullon (2012), Thaver, Ekanayake, and Plante (2012), Modeste (2011), Hye and Mashkoor (2010), Thaver and Ekanayake (2010), Ghorbani and Motalleb (2009), Jeon (2009), Acharya and Patterson (2005), Ang, (2007), Arize (2005), Kasman and Kasman (2005), Chang, Ho and Huang (2005), and Das (2003). The bounds test method is considered more appropriate for economic models because it is suitable whether the independent variables are purely stationary (i.e. I(0)), purely integrated of order 1 (unit-root) (i.e. I(1)), or mutually cointegrated. Furthermore, unlike other models such as by Engle-Granger (1987) that require all variables to be integrated of the same order, this model bypasses the pre-steps of determining the order of integration of the underlying regressors prior to testing for cointegration (Pesaran, *et al.*, 2001; Hendry and Juselius, 2000). As Ssekuma (2011) suggests, making an error during a pre-step can lead to unreliable and invalid results. In addition, unlike the Engle-Granger (1987) and Johansen (1991) models, the bounds testing method to cointegration is robust for small and finite samples (Kurihara, 2003), which is the case in this study with only 70 observations.

In order to estimate South Africa's export and import demand functions with Brazil and India separately, we employ Pesaran *et al*. (2001) bounds testing approach to cointegration within the autoregressive distributed lag (ARDL) framework. To do so, we first specify Equation (1) and Equation (2) respectively.

$$\ln REXP_t = \varphi_0 + \varphi_1 \ln RGDP_t + \varphi_2 \ln RER_t + \varphi_3 \ln VOL_t + \varphi_4 D_{1t} + \varepsilon_t \quad (1)$$

$$\ln M_t = \beta_0 + \beta_1 \ln Y_t + \beta_2 \ln RP_t + \beta_3 \ln FR_t + \beta_4 VOL_{1t} + \beta_5 D_{1t} + \varepsilon_t \quad (2)$$

Where, in period t,

Ln is the natural log;
$REXP_t$ is the real export volume;
$RGDP_t$ is the real GDP of SA's partner country;
RER_t *is* the real exchange rate, measured as the ratio of the domestic price to the foreign price, multiplied by the nominal exchange rate between the foreign and the domestic currencies;
VOL_t is real exchange rate volatility, measured as the moving sample standard deviation of growth rate of RER_t, as shown by Equation (3) with $m = 4$ is defined as the moving average of order 4;

$$VOL_t = \left[\frac{1}{m} \sum_{i=1}^{m} (\ln RER_{t+i-1} - \ln RER_{t+i-2})^2 \right]^{1/2} \quad (3)$$

D_{1t} is a dummy variable representing IBSA in 2003;
M_t is by real import volume, and is measured as nominal imports divided by SA's import price index;
RP_t is by relative price of imports, and is calculated as the ratio of import price to domestic price;
FR_t is by real foreign exchange reserves, which divides the nominal FR by the GDP deflator;
Y_t is South Africa's real GDP, which divides the nominal GDP by the GDP deflator; and
ε_t is the white-noise disturbance term.

According to economic theory, the expected coefficient signs in Equation (1) and Equation (2) are:

$\varphi_1 > 0$: An increase in SA's trading partner's GDP will have a positive impact on exports to that country;
$\varphi_2 > 0$: Appreciation of RER leads to depreciation of SA's currency, which will in turn increase exports;
$\varphi_3;\beta_4$?: *VOL* is the source of exchange-rate risk and has certain implications on the volume of international trade. As such, we do not have prior expectation of the signs of φ_3, and β_4;
$\varphi_4,\beta_5 > 0$: The effect of IBSA on SA's import and export demand functions with its trading countries is the objective of this study. We hypothesize that φ_4,and β_5 are positive;
$\beta_1 > 0$: An increase in SA's GDP will have a positive impact on its imports;
$\beta_2 < 0$: An increase in the relative price of imports discourages imports; and
$\beta_3 > 0$: higher real foreign reserves tend to encourage imports.

The expected signs of φ_1, φ_2, φ_3, β_1, β_2, and β_3 are borne out in empirical results by numerous scholars. Among them are, Grullon (2012), Hibbert et al.(2012), Thaver et al. (2012), Modeste (2011), Hye and Mashkoor (2010), Thaver and Ekanayake (2010), Ghorbani and Motalleb (2009), Jeon, (2009), Rehman (2007), Narayan and Narayan (2005), Chang, Ho and Huang (2005), Dash (2005), Razafimahefa and Hamori (2005), Bredin, et al. (2003), Tang (2002, 2003), and Dutta and Ahmed (1999).

Equations (1) and (2) may in turn be expressed in the form of an Error Correction Model (ECM) following Pesaran, *et al.* (2001) as in Equations (4) and (5). The symbols ln, M_t, $REXP_t$, *t*, $RGDP_t$, Y_t, RER_t, RP_t, RFR_t, VOL_t, and ω_t are explained in (1) and (2) above, Δ is the first difference operator.

$$\begin{aligned}\Delta \ln M_t = \alpha_0 &+ \sum_{i=1}^{n} \beta_i \Delta \ln M_{t-i} + \sum_{i=0}^{n} \gamma_i \Delta \ln Y_{t-i} + \sum_{i=0}^{n} \delta_i \Delta \ln RP_{t-i} + \sum_{i=0}^{n} \eta_i \Delta \ln FR_{t-i} + \sum_{i=0}^{n} \nu_i \Delta \ln VOL_{t-i} + \\ &+ \alpha_1 D_{1t} + \lambda_1 \ln M_{t-1} + \lambda_2 \ln Y_{t-1} + \lambda_3 \ln RP_{t-1} + \lambda_4 \ln FR_{t-1} + \lambda_5 \ln VOL_{t-1} + \omega_t\end{aligned} \quad (4)$$

$$\Delta \ln REXP_t = \alpha_0 + \sum_{i=1}^{n} \omega_i \Delta \ln REXP_{t-i} + \sum_{i=0}^{n} \psi_i \Delta \ln RGDP_{t-i} + \sum_{i=0}^{n} \kappa_i \Delta \ln RER_{t-i} + \sum_{i=0}^{n} \pi_i \Delta \ln VOL_{t-i} + \\ + \nu_1 D_{1t} + \theta_1 \ln REXP_{t-1} + \theta_2 \ln RGDP_{t-1} + \theta_3 \ln RER_{t-1} + \theta_4 \ln VOL_{t-1} + \omega_t \tag{5}$$

Pesaran et al's (2001) bounds testing approach is based on two procedural steps, namely establishing cointegration, and thereafter, estimating the short-run and long-run coefficients of the cointegrated model.

The first step to determine cointegration involves restricting the lagged variables in Equations (4) and (5) to equal zero. We then use the *F*-statistic or Wald test to test the null hypothesis of no cointegration, as shown below for the import demand function (Equation (4)):

$$H_0 : \lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = 0 \tag{6}$$

against the alternative hypothesis of cointegration:

$$H_1 : \lambda_0 \neq \lambda_1 \neq \lambda_2 \neq \lambda_3 \neq \lambda_5 \neq 0 \tag{7}$$

This test is performed for SA-Brazil and SA-India imports, as represented in Equation (4). The equivalent test of the null and alternative hypotheses of Equations (6) and (7), this time with $\theta's$, are performed on the SA-Brazil and SA-India exports model using Equation (5).

The asymptotic *F*-statistic has a non-standard distribution and depends on whether the variables included in the ARDL model are integrated, on the number of determinants in the model, and on the sample size. The computed F-value is compared with Pesaran et al.'s (2001) two critical bounds, I(0) at the lower end, indicating pure integration, and I(1) at the upper end, indicating unit root. If the *F*-value exceeds the upper critical bounds value, H_0 is rejected signaling cointegration among the independent variables. If it falls below the critical bounds values, we fail to reject H_0, and if it falls within the critical values, the result is inconclusive.

Once cointegration has been established, the second step involves estimating the short-run and long-run coefficients of the cointegrated model.

To estimate our model, we use quarterly data from January 1995to June 2012. The data series are taken from the International Monetary Fund's *International Financial Statistics Yearbook (2011)*. The data series include SA's nominal imports and exports, its import and export price indexes, nominal exchange rates, foreign exchange reserves, and SA's, Brazil's and India's real GDP, GDP deflator, and consumer price index, CPI (2005=100). Nominal imports (exports) in *Rands* are deflated by South Africa's import price index (export price index) (2005 = 100) to obtain the real import (export) variable. To convert Y_t into real terms, we divide it by South Africa's GDP deflator (2005 = 100). The relative price of imports series is constructed as the ratio of South Africa's trading partner to South Africa's CPI. To obtain the real foreign reserves series, we deflate the nominal foreign exchange reserves series by the GDP deflator.

Empirical Results

While the bounds testing model does not require unit root tests, it is based on the assumption that the variables are either I(0) or I(1). To determine robust results however, we have implemented the unit root tests of Kwiatkowski *et al*. (1992), the KPSS. Based on those results, we conclude that all variables are stationary in first differences, confirming that all variables do indeed comply with the underlying assumptions of the bounds testing model.

Cointegration among Variables

As Table 1 reveals, the calculated *F*-statistics for SA's exports and imports with India and Brazil are above their critical upper bound values at varying levels of significance. For example, in the case of SA's import demand function with Brazil, the computed *F*-value of 11.044 exceeds the upper critical bounds value of 4.68 at the 1% level. Thus, H_0 is rejected signaling cointegration among the independent variables. Similarly, in the case of SA's export demand function with India, the computed F-value of 3.895 exceeds the upper critical bounds value of 3.52 at the 10% level. Following this reasoning, we can conclude that in each of the four cases, cointegration exists between the dependent variable and its determinants, leading us to reject the null hypothesis. Our next step then, is to estimate the long-run and short-run partial elasticities, which we carry out in the next section.

Table1. Cointegration Results - South Africa's Export and Import Demand Functions with Brazil and India

Critical value bounds of the F-statistic: intercept and no trend:

	10 percent level		5 percent level		1 percent level	
k	I(0)	I(1)	I(0)	I(1)	I(0)	I(1)
4	2.45	3.52	2.86	4.01	3.74	5.06
5	2.26	3.35	2.62	3.79	3.41	4.68

EXPORTS Calculated F-Statistic: K=4

Brazil:	F_X(X| GDP, RER, VOL, D_1)	**4.430****
India:	F_X(X| GDP, RER, VOL, D_1)	**3.895***

IMPORTS Calculated F-Statistic : K=5

Brazil:	F_M(M| Y, FR, RP, VOL, D_1)	**11.044*****
India:	F_M(M| Y, FR, RP, VOL, D_1)	**11.2179*****

*Note: This table shows the results of the ARDL bounds test for cointegration. Critical values are taken from Pesaran, Shin, & Smith (2001), Table CI(iii) Case III, p. 300). k is the number of regressors. ***, **, * indicates significance at the 1 %, 5 %, and 10 % levels, respectively.*

We test the accuracy of our model specifications using several diagnostic tests. The Durbin-Watson (D-W) and Breusch-Godfrey LM tests check for autocorrelation, or in other words, for similarity between observations in the residuals. The augmented Dickey-Fuller test is a unit root test that is essential for cointegration analysis. The Ramsey RESET tests for correct functional form, serial correlation, if they are normally distributed, and homoskedastic. The Jarque Bera test examines if the errors are normally distributed. Finally, the adjusted R^2 determines if the dependent variable is explained by the independent variables within the model. Results of these tests (see Appendix A) indicate that our model for SA-Brazil is robust, but the SA-India model is less so. Since the RESET test indicated incorrect model specification, and unexplected coefficient signs indicated omitted variable bias, we tested several other model specifications. But our models are based on economic theory, and follows other empirical studies, so we decided to leave our original model specification as it was. In any case, our model specifications in the case of SA-Brazil indicated correct functional forms and no omitted variable bias, reinforcing our decision to keep the original model specification.

Long-Run Elasticities of Export and Import Demand

The second step after establishing cointegration, according to Pesaran et al., is to estimate the long-run partial elasticities of the dependent variable with respect to each regressor in both the exports and imports models for SA-India and SA-Brazil. Results are reported in Table 2, Panel A and Panel B. Most coefficients are statistically significant and conform to theoretical expectations in SA-Brazil exports and imports, but not for SA-India trade, in particular, exports. Moreover, in SA-India exports and imports, the coefficient signs for RER, RGDP, and RP are opposite to theoretical expectations, reinforcing the potential for omitted variable bias as revealed by the RESET test. Interestingly, exports to Brazil are overwhelmingly positively determined by Brazil's real GDP, where a 1% increase in RGDP yields a 2.95% change in real exports from SA to Brazil, holding all other variables constant. However, imports from Brazil decreases by 1.12% when SA's real GDP increases by 1%. The latter may be due to SA's very strong import substitution policies. Furthermore, in SA-Brazil and SA-India trade, the estimates for D_1 indicate that IBSA has not had a significant long-run impact on SA's export demand and important demand functions with its trading partners.

TABLE 2. Long-run Elasticities – South African Exports to, and Imports from, Brazil and India

Panel A: Dependent variable: ln REXP

	Brazil		India	
Explanatory Variables	**Coefficient**	***t*-statistic**	**Coefficient**	***t*-statistic**
Constant	-48.33**	-2.31	60.42	0.29
$\ln RER_t$	0.65*	1.81	-0.89	- 0.24
$\ln GDP_t$	2.95***	2.78	-1.83	- 0.21
$\ln VOL_t$	0.21	2.52	-0.86	- 0.86
D_{1t}	-0.06	-0.20	2.05	0.88
$\bar{R}^2$	0.42		0.36	

Panel B: Dependent variable: lnM

	Brazil		India	
Explanatory Variables	**Coefficient**	***t*-statistic**	**Coefficient**	***t*-statistic**
Constant	36.50***	4.31	-37.56	-1.61
$\ln FR_t$	0.53***	3.61	1.73***	3.90
$\ln RP_t$	-0.23	-0.54	8.62*	1.95
$\ln VOL_t$	0.15**	2.38	-1.43***	-4.48
$\ln Y_t$	-1.12***	-3.60	-1.18	-1.16
D_{1t}	0.17	1.00	-0.99	-1.44
$\bar{R}^2$	0.61		0.63	

*Note: This table shows SA's long-run elasticities of the estimated export and import demand functions with the Brazil and India. *** and ** indicate statistical significance at the 1% and 5% level, respectively.*

Short-Run Elasticities of Export and Import Demand

We next examine the dynamic short-run causality among the relevant variables by estimating the error correction model (ECM) specified in Equations (4) and (5). In the ECM, the movement of any of the regressors in time *t*, is related to the gap of that same regressor in time *t-1* from its long-run equilibrium. This step integrates the reality that economic and political forces are in constant flux so that trade functions are seldom in equilibrium. The lagged error correction term (ECM_{t-1}) is important for the cointegrated system as it allows for adjustment back to long-run equilibrium after a shock in the system. The ECM_{t-1} results are shown in Table 3. Using Hendry's general-to-specific method, four lags of each regressor are at first included, and then in subsequent step-by-step iterations, insignificant variables are eliminated.

As seen in Table 3, the error correction term, ECM_{t-1}, which gauges the rate at which import demand adapts to changes in the regressors in the short run before returning to its long run equilibrium level, conforms to theoretical expectations; it is statistically significant at the 1% level with the expected negative sign. The coefficient for ECM_{t-1} for SA-Brazil exports shown in Table 3, Panel A is -0.44, indicating that once the model in Equation (5) is shocked by changes in one of the export demand determinants, convergence to equilibrium adjusts at a speed of 44% in the first year. Import changes adjust at a 54% rate, and adjustments in both models are higher than in SA-India exports (14%) and imports (38%).

Table 3, Panel A reveals that in the short run, in both SA-Brazil and SA-India, the partial elasticities of most regressors with respect to export demand are statistically significant, at either the 1% level (-0.45), 5% level (0.10), or the 10% level (23.42). Interestingly, RGDP is highly elastic in both SA-Brazil (a 1% change in RGDP leads to a 6.67% change in exports) and SA-India (23.42), but the elasticity GDP coefficient in SA-India is 3.5 times more than in SA-Brazil exports. The same cannot be said of the short-run elasticities of import demand due to changes in GDP, with values hovering around unit elasticity (Table 3, Panel B). In addition, as in the case of the long run, the estimates for D_1 indicate that IBSA has not had a significant short-run impact on SA's export demand and important demand functions with its trading partners.

TABLE 3: Error-Correction Model and Short-run Elasticities –SA's Exports and Imports with Brazil and India

Panel A: Dependent variable: $\Delta lnREXP_t$

	Brazil		India	
Explanatory Variables	**Coefficient**	**t-statistic**	**Coefficient**	**t-statistic**
Intercept	0.00	0.00	0.00	0.00
$\Delta lnREXP_{t-1}$	-0.25*	-2.32	-0.29**	-3.12
$\Delta lnRER_{t-1}$	0.22	1.03	-2.39**	-2.82
$\Delta lnRGDP_{t-1}$	6.67*	2.18	23.42*	2.52
$\Delta lnVOL$	0.10**	3.25	-0.45***	3.81
D_t	-0.03	-0.68	0.28	1.81
ECM_{t-1}	-0.44***	-4.34	-0.14***	-3.90
$\bar{R}^2$	0.4563		0.3914	

Panel B: Dependent variable: ΔlnM

	Brazil		India	
Explanatory Variables	**Coefficient**	**t-statistic**	**Coefficient**	**t-statistic**
Intercept	0.00	0.00	0.00	0.00
$\Delta \ln M_{t-i}$	-0.20*	-2.32	-0.26***	-3.60
$\Delta \ln FR_{t-i}$	0.64***	4.96	-2.43***	-5.54
$\Delta \ln VOL_{t-i}$	-0.07*	-2.51	-0.33***	-3.55
$\Delta \ln RP_{t-i}$	0.97***	4.96	10.00***	-3.55
$\Delta \ln Y_{t-i}$	0.91***	4.62	1.17	1.65
D_t	0.10	1.94	-0.378***	-3.50
ECM_{t-1}	-0.54***	-7.76	-0.38***	-7.73
$\bar{R}^2$	0.6436		0.6542	

*Note: This table shows the results of the short-run elasticities of the error-correction model.***, **, and * indicate statistical significance at the 1% 5%, and 10% level, respectively.*

Conclusion

This paper utilized the bounds testing approach to cointegration developed by Pesaran et al. (2001) to empirically estimate the export and import demand functions of South Africa with its trading partners, Brazil and India within the context of the three countries' commitment to south-south trade integration to facilitate the development of the poorer countries in the globalized world. Results indicate that IBSA has not significantly affected trade between SA and its IBA counterparts, Brazil and India. This means that IBSA has had far more rhetorical power than the ability to effect real economic change.

Our results confirm a unique cointegration relationship between exports and its regressors, and between imports and its regressors for both SA-Brazil and SA-India trade relations. This permitted us to estimate the short-run and long-run elasticities of the respective export and import demand functions. We determined that most elasticity coefficients are statistically significant and conform to theoretical expectations in SA-Brazil exports, but not for SA-India exports in the short and long runs. Contrary to expectations, our model revealed that the IBSA trade treaty in 2003 seem to have no significant long-run and short-run impact on SA's export demand and important demand functions with its trading partners. Neither does it affect the short run adjustment process.

Our results further reveal that foreign income is highly elastic in both SA-Brazil and SA-India exports, the elasticity GDP coefficient in SA-India 3.5 times more than SA-Brazil export demand. These results are much higher than other studies reveal, but this may be because other studies have focused more on the developed world while this study focuses on developing countries. Once the system is shocked, convergence back to equilibrium takes place quite slowly in the exports model at -0.44 in the SA-B case, indicating that 44% of the adjustment occurs within the first period. In the SA-I export case, convergence to equilibrium is even slower, with only 14% of the correction occurring within the first period. The convergence in the imports model is significantly better than in the exports model. This means that imports from Brazil and India adjust much faster to shocks in the system than do exports to these countries.

Acknowledgements

The author is grateful for constructive comments and suggestions of the participants at the 2013 annual conference of the Academy of Economics and Finance. These have been incorporated into the paper. This research study was financially supported by XXXX 2011 Office of Academic Affairs Summer Professional Development Grant, for which the author is also grateful.

References

Akhtar, M. and R. Spence Hilton. 1984. "Effects of Exchange Rate Uncertainty on German and U.S. Trade." *Federal Reserve Bank of New York Quarterly Review*, Vol. 9, pp. 7–16.

Akinlo, A. E. 2008. "Cost function analysis of import demand for Nigeria." *Applied Economics*, 40, 2911-2940.

Aliyu, S.U.R. 2010. "Exchange rate volatility and export trade in Nigeria: an empirical investigation." *Applied Financial Economics*, 20(13), 1071-1084.

Arize, C. A., T. Osang, and J. D. Slottje. 2005. "Exchange-rate Volatility in Latin America and its Impact on Foreign Trade." September 2005. Available at: http://faculty.smu.edu/tosang/pdf/latin.pdf

Athukorala, P., and Riedel, J. 1996. "Modeling NIE exports: Aggregation, quantitative restrictions and choice of econometric methodology." *Journal of Development Studies*, 33(1), 81.

Baak, S. 2004. "Exchange Rate Volatility and Trade among the Asia Pacific." *International University of Japan*, 26 April. Available at: http://repec.org/esFEAM04/up.29293.1080736850.pdf

Bardsen, G. 1989. "Estimation of long-run coefficients in error correction models." *Oxford Bulletin of Economics and Statistics*, 51, 345-50.

Bredin, D., Fountas, S., and E. Murphy. 2003. "An empirical analysis of short-run and long-run Irish export functions: Does exchange rate volatility matter?" *International Review of Applied Economics*, 17(2), 193-208.

Boug, P., and Fagereng, A. 2010. "Exchange rate volatility and export performance: a cointegrated VAR approach." *Applied Economics*, 42(7), 851-864.

Department of Trade and Industry, South Africa. 2011. Retrieved July 17, 2012 from http://www.thedti.gov.za.

Das, S. K. 2003. "Effects of Exchange Rates Volatility on International Trade: An Empirical Analysis." Vanderbilt University, March 2003, Manuscript.

Dutta, D., and Ahmed, N. 2004. "An aggregate import demand function for India: A cointegration analysis." *Applied Economics Letters*, 11, 607-613.

Edwards, L. 2001. "Trade and the structure of South African production, 1984-97." *Development Southern Africa*, 18(4), 471-491.

Ekanayake, E.M. and Thaver, Ranjini L. 2011. "The Impact of Dollar-Rand Volatility on U.S. Exports to South Africa." *International Journal of Business and Finance Research*, Vol. 5, No. 3, pp. 73-85, 2011. Available at SSRN: http://ssrn.com/abstract=1879422

Engle, R.F. and C.W.J. Granger. 1987. "Co-integration and error correction representation, estimation and testing." *Econometrica*, 55: 251-276.

Gafar, J. 1995. "Some Estimates of the price and income elasticities of import demand for three Caribbean countries." *Applied Economics , 27* (11), 1045-1048.

Giovannetti G. 1989. "Aggregate imports and expenditure components in Italy: an econometric analysis." *Applied Economics , 21*, 957-971.

Ghorbani, M. and Motalleb, M. 2009. "Application Pesaran and Shin Method For Estimating Irans' Import Demand Function." *Journal of Applied Sciences , 9* (6), 1175-1179.

Grullon, S. 2012. "Determinants of the Dominican Republic's Export Demand Function: Results from the Bounds Test for Co-integration." *The International Journal of Applied Economics and Finance, 6: 109-116*

Gumede, V. 2000. "Import performance and import demand functions for South Africa." *Department of Trade and Industry.* Retrieved March 17, 2011 from Department of Trade and Industry's website, http://www.thedti.gov.za/.

Hibbert, K., Thaver, R., and Hutchinson, M. 2012. "An Empirical Analysis of Jamaica's Import Demand Function with the US and the UK." *International Journal of Business and Finance Research* , 6 (1), 11-22.

Hondroyiannis, G., Swamy, P. A. V. B., Tavlas, G., and Ulan, M. 2008. "Some further evidence on exchange-rate volatility and exports." *Review of World Economics, 144*(1), 151-180.

Hoque, M. M. and Yusop, Z. 2010. "Impacts of trade liberalisation on aggregate import in Bangladesh: An ARDL bounds test approach." *Journal of Asian Economics*, 21, 31-52.

Hye, Q. M. and Mashkoor, M. 2010. "Import Demand Function For Bangladesh: A Rolling Window Analysis." *African Journal Of Business Management* , *4* (10), 2150-2156.
International Monetary Fund. World Economic Outlook. Retrieved June 10, 2012 from http://www.imf.org/external/pubs/ft/weo/2011/01/weodata/index.aspx
Johansen, S. and Juselius, K. 1992. "Testing structural hypotheses in a multivariate cointegration analysis of the PPP and the UIP for UK." *Journal Of Econometrics, 53*(1/2/3), 211-256.
Kabundi, A. 2009. "Synchronization between South Africa and the U.S.: A structural dynamic factor analysis." *South African Journal of Economics*, 77(1), 1-27.
Kasman, A. and A. Kasman. 2005. "Exchange Rate Uncertainty in Turkey and its Impact on Export Volume." *METU Studies in Development*, Vol. 32 (June), pp. 41–58.
Kurihara, Y. 2003. "APEC: International trade and output." *Pacific Economic Review*, *8*(3), 207-217.
Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. and Shin. Y. 1992. "Testing the null hypothesis of stationary against the alternative of a unit root." *Journal of Econometrics*, 54: 159-178.
McKenzie, M. and Brooks, R. 1997. "The impact of exchange rate volatility on German-US trade flows." *Journal of International Financial Markers, Insitutions and Money, 7,* 73-87.
Liu, A. and Saal, D. S. 2001. "Structural change in apartheid-era South Africa: 1975-93." *Economics Systems Research,* 13(3), 235-257.
Matsubayashi, Y. and Hamori, S. 2003. "Some international evidence on the stability of aggregate import demand function." *Applied Economics*, *35*(13), 1497-1504.
McKenzie, M. D. and R. Brooks. 1997. "The Impact of Exchange Rate Volatility on German–US Trade Flows." *Journal of International Financial Markets, Institutions and Money*, Vol. 7, pp. 73–87.
Modeste, N. C. 2011. "An Empirical Analysis of the Demand for Importsin Three CARICOM Member Countries: An Application of the Bounds Test for Cointegration." *Review of Black Political Economy,* 38: 53-62.
Mwega, F. M. 1993. "Import demand elasticities and stability during trade liberalization: A case study of Kenya." Journal of African Economies**,** 2, 381-416
Narayan, S., and Narayan, P. K. 2005. "An empirical analysis of Fiji's import demand function." *Journal of Economic Studies*, 32(2), 158-168.
Ngandu, S. 2008. "Exchange rates and employment." *South African Journal of Economics*, 76(S2), S205-S221.
Ngandu, S. 2009. "The impact of exchange rate movements on employment: The economy-wide effects of a rand appreciation." *Development Southern Africa*, 26(1), 111-129.
Ozturk, I. and Kalyoncu, H. 2009. "Exchange rate volatility and trade: an empirical investigation from cross-country comparison." *African Development Review*, *21*(3), 499-513.
Pesaran, H., Shin, Y., and Smith, R. J. 2001. "Bounds testing approaches to the analysis of level relationships." *Journal of Applied Econometrics*, 16, 289-326.
Razafimahefa, I. F., and Hamori, S. 2005. "Import demand function: some evidence from Madagascar and Mauritius." *Journal of African Economies*, 14(3), 411-434.
Read, D. Y., and Parton, K. A. 2009. "Economic deregulation and trade liberalization in Kenya, Tanzania and Uganda: Growth and Poverty." *Journal of Economic Issues* (M.E. Sharpe Inc.), 43(3), 567.
Saayman, A. 2010. "A panel data approach to the behavioural equilibrium exchange rate of the ZAR." *South African Journal of Economics*, 78(1), 57-77.
Schnabl, G. 2008. "Exchange rate volatility and growth in emerging europe and east asia." *Open Econ Rev, 20,* 565-587.
Senhadji, A. 1998. "Time-series estimation of structural import demand equations: A cross-country analysis." *IMF Staff Papers*, 45, 236-268.
Ssekuma, Rajab. 2011. "A Study Of Cointegration Models With Applications." Master's Thesis, University of South Africa, 2011.
Statistics South Africa. 2011. Retrieved June 11, 2012 from http://www.statssa.gov.za/.
Tang, T. C. 2002. "Determinants of aggregate import demand in Bangladesh." *Journal of Bangladesh Studies*, 4, 37-46.
Tang, T. C. 2003. "An empirical analysis of China's import demand function." *China Economic Review,* 14, 142-163.
Tang, T. C. 2004. "Does financial variables explain the Japanese import demand? A cointegration analysis." *Applied Economics Letters*, 11, 775-780.
Thaver, R. and Ekanayake, E. M. 2010. "The impact of apartheid and international sanctions on South Africa's import demand function: An empirical analysis." *International Journal of Business and Finance Research* , 4 (4), 11-22.
Thaver, R. L. Ekanayake, E. M., and Plante, D. R. 2012. "An Estimation of the Impact Of GEAR and NEPAD on South Africa's Disaggregated Import Demand Function with Nigeria." *International Journal Of Business & Finance Research (IJBFR)*, *6*(2), 69-79.

Thaver, R. L. and Ekanayake, E. M. 2010. "The Impact of Apartheid and International Sanctions on South Africa's Import Demand Function: An Empirical Analysis." *International Journal Of Business And Finance Research*, *4*(4), 11-22.
Thompson, L. 2000. *A History of South Africa*. New Haven, Connecticut: Yale Univ. Press.
Thursby, J.G. 1988. "Evaluation of Coefficients in Misspecified Regressions with Application to Import Demand." *The Review of Economics and Statistics*, 70(4), 690-695.
Truett, L. J. and Truett, D. B. 2003. "A cost function analysis of import demand and growth in South Africa." *Journal of Development Economics*, 70, 425-442.
Tsionas, E. G. and Christopoulos, D. K. 2004. "International evidence on import demand." *Empirica*, 31, 43-53.

Appendix

Appendix Table A: Results of Diagnostic Tests for the Selected ARDL Model

Panel A: Exports

	Brazil		India	
Explanatory Variables	**Coefficient**	**p-value**	**Coefficient**	**p-value**
Durbin Watson Test	2.02	0.42	1.90	0.22
Breusch-Godfrey Test	1.51	0.22	1.98	0.11
RESET Test	5.23	0.01	9.60	0.00
Augmented Dickey-Fuller Test	-2.85	0.23	-5.12	0.01

Panel B: Imports

	Brazil		India	
Explanatory Variables	**Coefficient**	**p-value**	**Coefficient**	**p-value**
Durbin Watson Test	1.96	0.33	2.23	0.71
Breusch-Godfrey Test	0.37	0.83	2.34	0.07
RESET Test	1.70	0.19	20.33	0.00
Augmented Dickey-Fuller Test	-3.08	0.14	-3.32	0.07

Appendix Table B: Correlation Matrices for SA-Brazil Exports and Imports

Panel A: Export Variables –SA-Brazil

	REXP	RER	VOL	RGDP
REXP	-			
RER	0.488	-		
VOL	0.098	0.419	-	
RGDP	0.820	0.110	-0.164	-

Panel B: Export Variables –SA-India

	REXP	RER	VOL	RGDP
REXP	-			
RER	-0.521	-		
VOL	-0.106	0.008	-	
RGDP	0.897	-0.700	-0.046	-

Panel C: Import Variables – SA-Brazil

	RIMP	RFR	VOL	RP_CPI	RGDP
RIMP	-				
RFR	0.362	-			
VOL	0.296	-0.308	-		
RP_CPI	-0.834	-0.357	-0.046	-	
RGDP	-0.526	0.187	-0.568	0.295	-

Panel D: Import Variables – SA-India

	RIMP	RFR	VOL	RP_CPI	RGDP
RIMP	-				
RFR	0.864	-			
VOL	-0.216	-0.112	-		
RP_CPI	-0.309	-0.055	0.046	-	
RGDP	0.882	0.371	-0.173	-0.314	-

Revisiting the Twin Deficits Hypothesis: Evaluating the Impact of External Financing on US Twin Deficits: 1977-2009

Charles Tibedo, University of Southern Mississippi

Abstract

The twin deficit hypothesis, which contends that the current account decreases in response to a fiscal shock, has resurfaced since the U.S. has simultaneously experienced persistent and growing budget and current account deficits. However, empirical estimates of the direction and magnitude of this relationship have varied considerable. Alternatively, this study tests if the twin deficit effect is due to variation in the composition of deficit financing, rather than the deficit itself. The findings suggest a strong inverse relationship between foreign acquisitions of U.S. securities and the percent change in the current account relative to GDP.

.

Introduction

Proponents of the twin deficit hypothesis claim that fiscal tightening will be necessary to reduce the trade deficit, but the assertion of a direct one-to-one or causal linkage may be premature. This premature linkage stems in part from disagreement on the exact channels through which fiscal shocks affect movements in the current account balance, yet it has also proven difficult to consistently estimate both the magnitude and direction of the correlation between the deficits. It is generally accepted that budget deficits directly impact the current account through changes in the level and composition of consumption and investment. However, this paper suggests that the current account deficit responds contemporaneously to budget deficits. Furthermore, this paper tests the hypothesis that the proportion of the budget deficit financed through external rather than internal liabilities (foreign vs. domestic savings) is the factor which determines the presence and magnitude of correlation between the deficits.

Determining how the current account responds to fiscal shocks depends on whether one ascribes to the conventional hypothesis or the Ricardian equivalence hypothesis. Conventional theory states that in an open economy, while increases in the budget deficit may lead to appreciation of nominal interest rates, the primary channel through which the deficit manifests is through relative price differentials arising between domestic and foreign goods. On the other hand, the Ricardian equivalence hypothesis holds that rising budget deficits do *not* lead to an increase in the real-interest rate (i.e. there are no 'wealth effects' on private consumption as private savings increase in response to a reduction in public savings) and should have no effect on the current account. Further difficulties in estimating the magnitude and direction of the relationship between fiscal shocks and the current account balance arise when the observed dynamics of the response of the latter to the former may be dependent upon whether the short run or long run effects are being estimated.

Although much attention has been paid in the literature to the twin deficits hypothesis, there is little devoted to the effect of an increase in external liability on the magnitude and direction of responses of the current account to variations in the composition of debt financing. Of the studies that have been conducted, most examine the level of total external liability as a portion of the total debt, rather than measure the effects of changes in the percentage of the budget deficit financed through foreign purchases of U.S. securities. Thus, this study aims to test whether the magnitude and direction of the current account response to an increase in the budget deficit depends on the composition of the financing of these deficits, specifically on the approximate percentage change in the level of foreign purchases of U.S. Treasury debt securities.

Literature Review

Of the studies that claim to support the twin deficit hypothesis, the relationship is often only shown to be a weak positive correlation. For example, Baxter (1995) shows that an increase in the budget deficit results in a decrease of the current account balance by .5% GDP, regardless of whether the increase was due to raised taxes or increased government spending. Erceg, Guerrieri, and Gust (2005) also find that an increase in the budget deficit leads to a reduction in the trade balance, although the impact is milder by their estimate (.15% - .12% of GDP). Kim and Roubini (2008) suggest one potential reason for this inconsistency in empirical tests of the TD hypothesis is that endogenous movements between the current account and budget deficit are not properly captured within existing models. Endogeneity can explain why such weak correlations, and at time paradoxical results, are found.

Consumption, Investment and the Persistence of Fiscal Shocks

Ricardian Equivalence theory claims that consumers internalize the government's inter-temporal budget constraint, and that an increase in federal spending (whether through deficit financing or tax increases) will in the long run have the same net effect on the level of private consumption, due to 'complete' or one-period consumption smoothing. Barro (1974), who is seen as the originator of this 'Ricardian' theory, stated that a shift in the ratio of debt/expenditure or tax/expenditure would not significantly impact aggregate demand, real interest rates, or capital formation (investment). This is because he assumed capital markets were perfect and there were no intergenerational effects on net-wealth.

Neoclassical economic theory suggests that consumption smoothing makes consumption respond less than fully to output shocks as higher output raises consumption less than dollar for dollar. The effects of shocks decay over time, and temporary shocks therefore lead to an increase in the current account as inter-temporal consumption smoothing occurs through asset accumulation. On the other hand, more permanent shocks have no effect on the CA because expectations (perfect-foresight) lead to full adjustment or 'smoothing'. An unanticipated shock leaves permanent output to fluctuate *more than* current output resulting in higher levels consumption, implying a current account deficit will result (Obstfeld & Rogoff, 1996).

The basic idea is that the more persistent productivity shocks are, the lower the current account is (more of a reduction). If there is substantial serial correlation in productivity shocks then growth should have (p=1), and there should be increasingly larger effects on consumption as households fully adjust (Deaton, 1992). Even with p close to 1, the slow rate at which output shocks dissipate over time makes it difficult to establish them as stationary or non-stationary. The consumption response is thus quite sensitive to the rate at which output shocks die out (Obstfeld & Rogoff, 1996).

Positive productivity shocks raise the expected path of future output, and also induce investment by raising the expected return to capital. This tends to worsen the current account via a) domestic residents borrow abroad to finance additional investment, and b) productivity increases affect saving. Glick and Rogoff (1995) claim that country-specific productivity shocks have a negative impact on the current account, while global productivity shocks have no significant impact. They test the hypothesis that in a permanent productivity shock investment rises and saving falls, so the current account effect is larger than the investment effect. They reach the conclusion that investment responds more than consumption to a positive productivity shock. Bussiere et al (2005) test whether there is evidence supporting a contemporaneous effect of cyclically adjusted primary budget deficits on the current account, which clearly calls into question earlier findings suggesting that country specific productivity 'shocks' are primarily responsible for observed changes in the current account.

Mankiw (2000) assumes a non-Ricardian view that budget deficits do affect consumption levels, and that not all households respond the same to budget deficits (increased ratio of debt/expenditure); while a fraction spend their disposable income within the period (non-Ricardian) the rest of the population consumes its permanent income (Ricardian), and therefore smooth resources inter-temporally. Assuming different weights for each type of consumer as suggested by Mankiw (ibid), Bussiere et al (2005) test the idea that after accounting for the effect of productivity shocks on the current account, the primary budget deficit will have a significant negative impact on the current account. Regarding the hypothesis that the primary budget deficit has a significant negative effect on the current account, after accounting for the effect of country-specific productivity shocks, they find that the size of the effect depends upon the weight of the spender types within a population.

Therefore, the effect that productivity shocks would have to an open economy's net level of saving depends on the ratio of saver to spender types within the economy. Relaxing the assumption of full Ricardian equivalence is really just an assumption about the static behavior of consumption and investment in response to a positive fiscal shock. However, this does not take into account the degree of openness to trade or the effect that the expected persistence of the fiscal shock has on that ratio, as Corsetti and Muller (2006) suggest are vital to understanding the transmission mechanism of fiscal shocks.

The Openness Factor

Corsetti and Muller (2006) study the effects of fiscal shocks one the trade balance (primary component of the current account) in four advanced economies using a VAR analysis, and conclude that the hypothesized negative effect of fiscal expansions on the current account depends on two factors: countries' openness to trade, and the persistence of the fiscal shock. They contend that both the Mundell-Fleming and inter-temporal models ignore the existence of non-homogenous goods and don't differentiate between government expenditures on domestic versus imported goods. Specifically, they claim that the Mundell-Fleming model is inadequate as it treats the rate of return on investment to (marginal product of) the capital stock as exogenous, which they alternatively claim is endogenously determined. The authors also claim that the inter-temporal approach ignores the impact of relative price differentials on investment decisions through resulting changes the real interest rate and the return to investment.

Their findings suggest that fiscal deficits can lead to improvements in the trade balance, if the shock to government spending is temporary and the economy is 'sufficiently closed'. In a relatively closed economy, fiscal expansions can lead to an increase in the real interest rate, which will cause a reduction in private investment but also induce an increase in private savings, as long as there is partial Ricardian behavior. Furthermore, if the shock is viewed as temporary, households will not adjust their consumption and home bias towards domestic goods will drive an *improvement* in the current account. On the other hand fiscal shocks will likely lead to a worsening of the trade balance if the economy is 'relatively open' and the shock is expected to last. For the U.S., which they denote as 'relatively closed', they estimate that fiscal shocks have limited effects on the current account, while the response of private investment to fiscal shocks according to their findings is greater than conventional theory would suggest.

In sum, Corsetti and Muller (2006) state that the key determinant of the effect of fiscal shocks on the current account is the return to investment (or marginal product of capital) that is endogenously determined by the openness of the country, and on the degree to which the prices of domestic goods rise in response to government spending. The more persistent the shock to an economy is, assuming it is relatively open, the greater the improvement in the current account balance. The more persistent a shock is when the economy is relatively closed, the more the country will face a tradeoff between reductions in capital stock and a worsening of the current account.

Relative Price Differentials and the Real Exchange Rate, or Not?

The neo-classical (Perotti and Blanchard, 2002) approach would state that a negative wealth effect of the expansion should generate a reduction in consumption with either no or a positive effect on real exchange rates. The latter effect depends on whether the economy is relatively open or relatively closed, as the demand for domestically produced goods is increasing in the degree of home bias and thus an appreciation of the real exchange rate will follow positive shocks. These findings are not in line with most observed data on the RER, which shows that real depreciations tend to follow expansionary fiscal shocks.

Alternatively, Mundell-Fleming based models of open economies may predict the observed consumption increase, but fail to predict the observed decrease in the real exchange rate (Ravan et *al*, 2007). This hypothesized appreciation is contingent upon higher relative returns to capital due to assumed increases in interest rates, which in turn attract foreign capital inflows and influence interest rates through a money demand function. Even if staggered price adjustments or other nominal rigidities are added as in the New Keynesian models, government-spending increases induce appreciations of the real exchange rate in the long run.

Aristovnik & Ljubljana (2010) found that real exchange rates should theoretically appreciate in response to a fiscal shock, but this requires the assumption of high price elasticity of inter-temporal substitutability of goods. Alternatively, Kim and Roubini (2008) suggest that in the event there exists a low price elasticity of inter-temporal substitutability of goods, budget deficits could actually initiate a depreciation, and thus actually cause an improvement to the current account. They employ the use of Vector Auto-Regression models to test examine the validity of conventional theories, which predict a positive relationship between budget deficits and current account deficits, or the 'twin deficit hypothesis.' Specifically, they test the hypotheses that after controlling for business cycle effects on output variations, an exogenous shock to the government budget deficit leads to a) a decrease in the current account, and b) an appreciation of the real exchange rate. They assume that an exogenous government budget deficit shock is the portion of such deficits not explained by variations in output caused by the business cycle. Their analysis suggests however that twin divergence, rather than twin deficits, fit the observed data from the U.S. in the flexible-exchange rate period. Specifically, they argue the data supports the notion that in response to a fiscal shock, a decrease in private investment improves the current account, while fluctuations in the nominal exchange rate rather than relative price differentials drive a real depreciation of the exchange rate.

'Crowding in'

Corsetti, Meier, and Muller (2009) suggest that accounting for implied future-spending restraints (debt-stabilizing policy) brings the New Keynesian model's predictions of the responses of consumption and the real exchange rate more in line with observed increases and depreciations, respectively. The mechanism they suggest is the 'deep habit incentive' that exists during expansions for firms to lower markup costs *relative* to foreign firms. The effect is to induce the crowding in of consumption and depreciation of the real exchange rate to follow declines in long-term interest rates, despite any increase in government spending. This occurs due to an increase in domestic real wages and a resulting substitution of consumption for leisure. Hence, an increase in consumption and a real depreciation would be seen in the presence of sticky domestic prices (no relative price changes), which causes interest rates to remain lower than they would under fully flexible prices (full pass-through). Since output then 'adjusts' less to relative price differentials during a fiscal expansion, short-term interest rates will also rise by less than expected under fully flexible prices or pass-through.

The Role of International Capital Flows

Ito and Clower (2012) suggest that current account balances should become increasingly divergent across countries as the world becomes increasingly financially integrated. They argue that the degree of global imbalances will continue toward ever greater extremes, with the result that there will be greater variance in measures of the CA over time. Feldstein and Horioka (1980) argue that easier access to international financial markets can help delink domestic saving and investment (Faruqee and Lee, 2009). Kim and Roubini (2008) use a VAR approach to isolate the individual response functions, or estimate the dynamic relations between 'exogenous' fiscal shocks, the trade balance and real exchange rate. The proportion of change in the budget deficit due to output shocks is estimated to be 50% or higher, as recessions induce increased deficits and booms induce improvements in the deficit. Output shocks in their study are decomposed into business cycles and deficit induced shocks. The deficit-induced shocks produce interesting results: an appreciation of the real exchange rate and an increase in the current account. However, it may be necessary to apply caution in assuming that current account imbalances are the 'natural' result of market forces, and consider the potential that empirical estimates of the twin deficits hypothesis may suffer from some type of misspecification. Indeed, Obstfeld (2012, 39) notes that "contrary to a complete markets or "consenting adults" view of the world, current account imbalances, while very possibly warranted by fundamentals and welcome, can also signal elevated macroeconomic and financial stresses…They need not be the benign result of advances in market efficiency, as is sometimes claimed. Valuation changes in NIIPs, while possibly important in risk allocation, cannot be relied upon systematically to offset the changes in national wealth implied by the current account."

Theoretical Framework

Government spending increases are assumed to be on both traded and non-traded goods. This negative effect of increases in the percentage of government debt purchased by foreigners occurs because of the 'offsetting' effect on domestic consumption (it increases). The financing of debt essentially finances the national savings deficiency produced by a low U.S. savings rate and allows U.S. private consumption to remain higher than it would without the offsetting capital inflow. Instead of facing significantly reduced domestic investment through a crowding-out effect, foreign financing of government spending allows investment levels remain constant. However, future output potential will fall as these foreign owned U.S. productive assets will not generate income for the foreign economy, even though this will also produce employment opportunities that will help to keep domestic consumption patterns high.

This effectively means that higher degrees of foreign financing of the budget deficit today will lead to higher levels of consumption but also lower levels of private domestic savings, and therefore domestic investment, in the future, implying a persistent deficit in the current account even in the face of growing demand for traded goods. By implication, the growth rates of the U.S. have remained high due to higher investment, but that investment has been due to foreign financing which allows further worsening of the current account, implying that the future tax burdens faced by society will be increasingly large as the portion of GDP spent on servicing those debts will rise and the increase in output spurred by the foreign capital inflow returns less and less to future GDP (ability to offset increasing external debt servicing payments).

For simplicity, it is assumed that every dollar of present debt purchased by a foreign entity allows a domestic resident to a) maintain current consumption or increase (non-Ricardian), particularly if the shock is expected to be temporary, b) offset their own investment, and thus choose to either save more or consume 'more' (than they would have had no foreign purchases been made). If no foreign purchases of domestic debt had been made, then the Ricardian effect would be forced and full adjustment would need to happen almost within a single time period, or the government would become insolvent, without resorting to hyperinflationary monetary behavior. Thus, foreign financing of the budget deficit does produce consumption (allows an increase) and investment (allows a 'domestic' decrease but an overall increase) effects that result in an exchange rate appreciation, most likely through higher productivity and domestic demand (particularly with home bias). An increase in the real interest rate would also induce a shift in the costs of financing transport for trade-able goods if sticky prices hold (thus hurting the potential gain through exports, as opposed to imports which would be expected to increase if the foreign* economy (U.S.) price level or nominal interest rate went up, as cheaper goods would be substituted). This process would take a while to build up however, and is not deterministic or mechanistic, which is what allows a 'contemporaneous' linear model to be estimated (assuming contemporaneous erogeneity holds).

If Ricardian equivalence holds in a relatively closed economy, particularly when fiscal shocks are highly persistent (or expected to be), then the predicted effect would be no net effect on capital formation, interest rates, or aggregate demand.

Alternatively, according to the Mundell-Fleming model, real interest rates should respond to the fiscal expansion, causing capital inflow and appreciation. Kim and Roubini (2006) find that the exchange rate appreciates leading to an improvement of the current account in the short-run. However, if the budget deficit is financed through debt, then the proportion of that debt that is financed by foreigners should offset the amount of 'Ricardian' equivalence that takes place (leading to the excess demand predicted by the Mundell-Fleming model). In other words, domestic output will expand due to the fiscal shock, and excess demand as predicted in the Mundell-Fleming models will occur, but private saving will not fall as much as traditionally predicted by these models.

The hypothesized effect of an increase in external debt relative to the level of the budget deficit (foreign holdings of U.S. Treasury Securities) is that the home country's ability to maintain higher than equilibrium levels of consumption and investment, even in the face of a persistent shock, increases with foreign bond holdings of U.S. debt. Corsetti and Muller (2006) conclude that the current account is increasing in the persistence of the shock, as the persistent shock to aggregate demand puts upward pressure on prices and thus raises the return to capital and propensity to invest. For a relatively closed economy with home bias where investment responds strongly to fiscal shocks and private saving increases, it is possible to have a period of sustained deficits (which would conventionally cause deteriorating trade balances) where investment is financed by foreign central banks and private investors while domestic consumption remains stable (as if the shock was only temporary).

Such a scenario could explain the twin deficits observed in the U.S. in the latter part of the 2000's. A strong level of home bias would only accentuate real interest rate differentials, the latter assumed by Corsetti and Muller (2006) to drive the *relative* investment and saving decisions of households in the domestic and foreign economies. To the extent the budget deficit is financed, increases in private investment would drive a current account reduction, even though private investment would initially be discouraged in the home economy *relative* to the foreign economy in the face of rising RIR differentials following a fiscal shock. The hypothesis tested here is that growth in the level of external liabilities due to deficit financing has a negative effect on the current account (relative to GDP).

Model

Initially, lags were included to capture previous changes in the budget deficit and foreign holdings of U.S. debt securities, but this approach was later modified by taking first differences of the independent variables and the dependent variable instead (% change in current account). The serial correlation implied by persistent and 'positive' output shocks on the current account is negated by first differencing each of the variables that an output shock would affect. Initially output shocks were included in the model, as measured by the squared output deviation from trend (and even without the square); the variable was found to be highly insignificant. The problem is most likely in the construction of the variable, as the business cycle 'output effects' on the current account identified by Roubini (2008) are endogenously determined by net changes in export and import shares of GDP, which are in turn responding to changes in the real interest rates and relative price levels. Certainly these factors influence the absolute level of the current account level and thus its size relative to GDP, as well as its direction and rate of change. The final model, presented below, was constructed to capture 'contemporaneous' effects (measured over annual time periods) hypothesized to exist between the current account and budget deficits as well as changes in the composition of financing those deficits. The model includes as independent variables the ratio of exports and imports to GDP and changes in the real exchange rate and real interest rate constant.

The dependent variable was constructed as the change in the current account as a percentage of GDP. This is similar to measuring a change in the rate of growth in the current account. Thus, the coefficients on the independent variables capture the rate of change in the relative size of the current account relative to GDP. Note that there is no presumption in the model that a current account deficit exists. Rather, the variable is constructed so that an *increase* in the size of the current account relative to GDP could equally represent a movement towards an increased deficit or surplus, and vice versa. The independent variables will therefore capture both the direction and magnitude of relationships, presuming that these functions are constant rather than non-linear.

Changes in log-level of foreign purchase of U.S. debt securities are expected to have a negative effect on the current account. Although this is a level variable, the log represents the belief that it operates in a trending manner, which is captured by a constant average growth rate (assuming that the appetite for such debt has not relinquished, relatively speaking, despite fluctuations in world demand levels since the 2008 global recession). By taking the log it also allows the interpretation of the relations between foreign purchases and the growth rate of the current account to be framed in terms of elasticity. Equation 1 expresses the final model.

$$\Delta CA\%GDP_t = ß_0 + ß_1\Delta BD\%GDP_t + ß_2\Delta EX\%GDP_t + ß_3\Delta IM\%GDP_t + ß_4 LnTRS_t + ß_5\Delta RIR_t + ß_6\Delta RER_t \qquad (1)$$

Hypotheses

A rise in the relative size of the budget deficit to GDP should negatively impact the current account balance (alternative), contrary to the Ricardian prediction of no impact due to offsetting private savings increases. The intent of this study is to demonstrate that a negative impact is observed precisely because the non-Ricardian behavior is being financed through foreign acquisition of U.S. Treasury Securities. Similarly, the impact of an increase in the rate of growth of foreign acquisitions of U.S. Treasury Securities is tested under the hypothesis that it will have an inverse relationship with the rate of change in the current account (relative to GDP).

An increase in exports relative to GDP will lead to an expected improvement in the current account balance. Due to the way the dependent variable is measured, this can result in a the current account moving from a negative percentage of GDP to a positive percentage, which would lend support to the idea that export growth can be a significant contributor to achieving external balance. Similarly, for changes in imports relative to GDP, an increase in imports should lead to an expected worsening of the current account balance, which would lend support to various ideas such as relative price differentials and shifts in demand, productivity differentials, or perhaps the Harrod-Balassa-Samuelson effect. Regardless, neither of these variables is of interest and here it is chosen as a control, to account for trends within the current account balance that are due to factors other than the budget deficit that influence the export and import sectors. Nevertheless, the predicted directions are important for the theoretical framework surrounding the effect of an increase in foreign holdings of U.S. Treasury Securities to hold. Specifically, estimating the impact of a change in exports and imports separately assumes that demand for imports is independent of a change in demand or supply of exports, which is consistent with the idea of nominal sticky prices in tradable goods and services data suggesting purchasing power parity doesn't hold.

An increase in the Real Interest Rate is hypothesized to cause a negative shift in the current account balance relative to GDP, either 'shrinking' the surplus, or worsening the deficit. Alternatively, an increase in the Real Exchange Rate is hypothesized to cause a positive shift in the current account balance, as the value of exports (regardless of volume effects) will rise and cause an immediate improvement in the current account holding foreign demand fixed (which in this study, is assumed to be exogenously determined and in the error term). Factors affecting foreign demand that also affect the real interest rate and the real exchange rate are not assumed to zero mean or constant effects on foreign demand, thus the error term is not expected to follow an autoregressive conditional heteroskedastic process.

Data

All data were obtained from the Office of Congressional Budget Research and from the Historical Tables of the Office of Management and Budget of the White House, and the U.S. Bureau of Economic Analysis. The years were from 1977 to 2009 for the United States. Data on the following variables were taken: Current account (as % of GDP), Real Interest Rate, Real Exchange Rate, Exports (as % of GDP), Imports (as % of GDP), Market Value of Foreign Holdings of Treasury Securities (Total Annual Foreign Purchases).

Estimation and Results

The overall stochastic process is assumed to be stationary, after de-trending through first differencing most series and taking the logarithm of another series. The following independent variables in the model are transformed by first differences (Budget Deficit, RIR, RER), as each is suspected of following an upward or downward linear trend. Exports, and Imports, as a percentage of GDP, are not assumed to follow any clear linear or nonlinear trend. The natural log of Foreign Purchases of U.S. Treasury Securities was taken as the process showed evidence of an exponential trend (constant average growth rate). A Ramsay test for functional form misspecification was also run to confirm these results. The F statistic with (18, 7) degrees of freedom was 0.84 with a $p > F = 0.6418$, resulting in a failure to reject the null hypothesis that the model is correctly specified as linear. The OLS regression results are reported in the first row of each variable in Table 1.

Assuming that the explanatory variables do not meet the strict erogeneity assumption, a simple robust test for serial correlation in the error term (regress u_t on x_j, u_{t-1}) revealed that p=.195 (se=.276, t=0.70, p=.488) and therefore a failure to reject the null hypothesis of no serial correlation in the error term as there is little significance on the coefficient for the lagged residual. Additional tests for autoregressive components in the model that would violate the stability condition needed for weak dependence were then tested via the Breusch-Godfrey test and the Durbin Watson alternative test. Both tests were unable to reject the null hypothesis of no serial correlation, indicating that the model was appropriately specified after de-trending, with the assumptions regarding stationary and weak dependence being met.

As there was little evidence of serial correlation in the error terms, the use of Prais-Winsten and Cochrane-Orcutt estimation would produce invalid standard errors for the coefficients, so neither was used. However, as there may be an arbitrary form of serial correlation, or heteroskedasticity, Newey-West (NW) standard errors were computed and are shown in bold in the second row of each variable in Table 1. Compared with t heteroskedastic-robust OLS estimates, the NW OLS estimates have higher standard errors, but these have a negligible impact on the size of their partial effects and level of statistical significance. As the process has been determined to be a trend-stationary, weakly dependent I(0) process and the model does not suffer from serial correlation, the NW estimators are robust to arbitrary forms of heteroskedasticity and can be used to make inferences about the entire population (stability of estimated relationshp over time).

Table 1: Comparison of robust OLS and Newey-West coefficients

Variable	***Coefficient***	***Standard Error***	***T-Statistic***	***P-value***
Constant	4.368*	1.027	4.25	.000
	3.562**	**1.442**	**2.47**	**.021**
ΔBD%GDP	-.104*	.015	-6.57	0.00
	-.095*	**.024**	**-3.95**	**.001**
ΔEXP%GDP	1.319*	.039	33.36	.000
	1.325*	**.070**	**18.78**	**.000**
ΔIMP%GDP	-.870*	.040	-21.69	.000
	-.906	**.051**	**-17.62**	**.000**
LnTRS	-.781*	.103	-7.56	.000
	-.705*	**.133**	**-5.28**	**.000**
ΔRIR	-.104*	.020	-5.10	.000
	-.103*	**.034**	**-2.99**	**.006**
ΔRER	.010*	.003	2.85	.009
	.012**	**.005**	**2.22**	**.036**

*significant at the 1% level, **significant at the 5% level

Analysis

The model a good fit with approximately 99% of the variation in the first difference (rate of change) of the current account as a % of GDP explained by the model. Further, Newey-West estimators which are robust to more arbitrary forms of heteroskedasticity also result in estimated coefficients that retain their expected signs and decrease only slightly in magnitude. Holding other factors constant a 1% increase in the Budget Deficit to GDP ratio causes a .095% decrease in the current account to GDP ratio, which would bring about a *decrease* in a current account 'surplus', or an *increase* in a current account 'deficit'. The estimated effect is also in line with the findings of Erceg, Guerrieri, and Gust (2005) who find a weak positive correlation between the two deficits of between .15% and .12% of GDP). The estimates are considerably lower than those reached by Baxter (1995) who shows that an increase in the budget deficit results in a decrease of the current account balance by .5% GDP.

The primary variable of interest, however, was the log of Foreign Purchases of U.S. Treasury Securities. Holding other factors fixed, a 1% increase in annual foreign purchases of treasury securities leads to a .78% decrease in current account to GDP ratio, which would bring about a *decrease* in a current account 'surplus', or an *increase* in a current account 'deficit'. The data reveal that annual increases in the proportion of budget deficits financed via foreign securities purchases of between 5 and 15% are quite common, which would suggest this factor plays an important role in the persistency of the current account deficit despite the theoretical prediction of the non-existence of a crowding out effect.

These findings support earlier studies that report a modest inverse link from budget deficits to current account deficits, however the coefficient on foreign purchases of U.S. debt securities reveals an almost one to one inverse relationship exists between a change in this factor and a change in current account (as % of GDP). As the measures employed in this study capture percent changes in the relative levels of the BD and log of foreign securities acquisitions, these can be interpreted as measures of elasticity due to the way the dependent variable was measured; as the percent change in the current account relative to GDP.

Results suggest a positive elasticity between growth in exports (as % of GDP) and the size of the current account (as % of GDP), while there is a negative and marginally inelastic relationship between growth of imports (as % of GDP) and the current account (as % of GDP). This indicates that export growth may be a more important 'remedy' or policy factor when considering how to best reduce the massive imbalances in the U.S. budget and trade balance simultaneously. This study assumes however that instead of seeing a 'crowding out' effect due to an increase in domestic interest rates, what the increase in foreign purchases of U.S. securities has artificially propped up levels of domestic consumption and investment by distorting the domestic money market equilibrium interest rate. Even though fiscal spending has increased and domestic consumption and investment should not decrease specifically in response to this spending, this boost to aggregate demand is not to the degree necessary to reduce the growing size of the trade deficit in a permanent manner. What this implies is that strength and persistence of the current account deficit is due to the lack of adjustment to the presence of the foreign savings in the domestic money market (i.e. interest rate adjustment), which in turn weakens the conventional links between the budget deficit and current account via the interest rate. Such an interpretation could be consistent with evidence presented by Corsetti and Muller (2006) who find that the more persistent a fiscal shock to an economy is when the economy is relatively closed (like the U.S.), the more the country will face a tradeoff between reductions in the capital stock and a worsening of the current account.

Conclusion

The twin deficit hypothesis contends that a cyclical worsening of the current account accompanies a fiscal shock. This hypothesis lost credibility by the early 1990's. However, as time is the great equalizer, it is curious indeed that a re-examination of the twin deficit hypothesis has resurfaced in scholarship as the U.S. and other major economies have simultaneously experienced persistent and growing budget deficits (or surpluses) and current account deficits (or surpluses). The co-movement of these two phenomena during the 2000's has created peak interest once again in the possibility of a direct relationship between fiscal shocks and the current account.

There are a wide variety of issues with making this claim, but in general proponents of the twin deficit hypothesis claim that fiscal tightening (via spending cuts or taxes) will be necessary to reduce the trade deficits and global imbalances currently present in the international economy. However, the assertion of a direct one-to-one or causal linkage may be premature. This study suggests that one element missing from these analyses is the treatment of external liabilities as the factor that generates and sustains current account deficits, and acts as the mechanism through which the 'twin-deficits' phenomenon appears. That is, to the degree that the government finances its deficit with foreign savings, savings (investment) and consumption patterns in the United States will not 'adjust', as predicted by proponents of consumption-smoothing (full or inter-temporal) theories. Thus, there is no reason to suspect the prevalence of any interest rate effects within the United States that such theories would presume result from the government 'competing' with private investors for investment funds. This has the effect of 'preventing' the conventional predicted effect of the interest rate rising in response to an increase in the demand for domestic funds. Transferring this into an open economy model, implies that the real exchange rate will not shift, and no relative price effects on foreign vs. domestic goods is only due to the composition of how the government spends these borrowed funds, and not to downward movements in the exchange rate, which one would normally expect to occur when increasing interest rates result in a decline in aggregate demand.

What this means is that as the U.S. economy increases its external liabilities as a proportion of our budget deficits, this at least appears to be permanent fiscal shock, preventing the predicted consumption smoothing effects that lead to so called 'crowding out' of investment. Thus, foreign savings act as a gross substitute for domestic investment, as it is the same as if the public borrowed funds to invest from abroad, yet at domestic interest rates rather than foreign interest rates. The resulting 'distortion' or imbalance in global savings patterns, leads to the observation that in the U.S. when budget deficits increase, investment does not have to decline in such a scenario. The net effect of the influx of foreign savings on aggregate demand then depends upon a number of factors, including the strength of the multiplier effect as well as the composition of where the government chooses to spend or invest those funds. Another implication of these findings is that we should see an aggravation of the 'twin-deficits' effect in the situation where there is an increase in the size and foreign component of budget deficits in the U.S. However, this study does not test for such non-linear possibilities, such as estimating structural breaks in current account responses to increasing injections of foreign savings into the U.S. economy via Treasury Security purchases. Nor does it attempt to dynamically test the relationship proposed in this study. Yet, such dynamics should be the topic of further study on the effect of foreign savings on the current account, specifically testing the dynamics implied by these finding that neither the real exchange rate nor real interest rate are key drivers of the persistency of the trade deficit in the United States since the 1980's.

References

Aristovnik, Aleksander and Sandra Djuric. 2010. "Twin deficits and the Feldstein-Horioka puzzle: a comparison of the EU member states and candidate countries." MRPA Paper 24149, Germany: University Library of Munich

Barro, Robert. 1974. "Are Government Bonds Net Wealth." *Journal of Political Economy* 82(6), 1095-1117

Baxter, Marianne. 1995. "International Trade and Business Cycles." In *Handbook of International Economics Vol.* 3, edited by G. Grossman and K. Rogoff. Amsterdam: Elsevier Publishing.

Bussiere, M., M. Fratzscher and G. Muller. 2005. "Productivity shocks, budget deficits and the current account." European Central Bank Working Paper No. 509.

Corsetti, G., and G. Muller. 2006. "Twin deficits: Squaring theory, evidence and common sense." *Economic Policy* 48: 598–638.

Corsetti, G., A. Meier, and G. Muller. 2009. "Fiscal stimulus with spending reversals." C.E.P.R. Discussion Papers 7032.

Deaton, A. 1997. "Life-Cycle Models of Consumption: Is the Evidence Consistent with the Theory?" In *Advances in Econometrics Vol.* 2, edited by T. F. Bewley. New York: Cambridge University Press.

Dornsbusch, R. 1976. "Expectations and Exchange Rate Dynamics", Journal of Political Economy 84(6): 1161–1176.

Erceg, C., L. Guerrieri and C. Gust. 2005. "SIGMA, a New Open Economy Model for Policy Analysis." Manuscript, Board of Governors of the Federal Reserve System.

Glick, R. and K. Rogoff. 1995. "Global versus country-specific productivity shocks and the current account." *Journal of Monetary Economics* 35: 159-192

Faruqee, H. and J. Lee. 2009. "Global Dispersion of Current Accounts: Is the Universe Expanding? IMF Staff Papers 56(3): 574-595.

Feldstein, M. and C. Horioka. 1980. "Domestic Saving and International Capital Flows." *The Economic Journal* 90: 314-329.

Ito, H. and E. Cowell. 2012. "The Persistence and Determinants of Current Account Balances: The Implications for Global Rebalancing." ADBI Working Paper No. 400.

Kim, S. and N. Roubini. 2006. "Twin Deficit or Twin Divergence? Fiscal Policy, Current Account, and Real Exchange Rate in the U.S." *Journal of International Economics* 74(2): 362-383.

Kim, S. and N. Roubini. 2008. "Twin deficit or Twin Divergence? Fiscal policy, Account, and Real Exchange Rate in the U.S." *Journal of International Economics* 74(2): 362-383.

Mankiw, G. 2000. "The savers-spenders theory of fiscal policy," *American Economic Review* 90(2): 120-125

Obstfeld, M. and K. Rogoff. 1996. *Foundations of International Macroeconomics*. Cambridge, MA: MIT Press

Obstfeld, Maurice. 2002. "Does the Current Account Still Matter?" *NBER Working Paper* 17877.

Perotti, R. and O. Blanchard. "An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output." *The Quarterly Journal of Economics* 117(4): 1329-1368.

Ravan, M., S. Schmitt-Grohé and M. Uribe. 2007. "Explaining the effects of government spending on consumption and the real exchange rate." NBER Working Paper 13328.