The

GETTYSBURG ECONOMIC REVIEW

Volume 7

Spring 2013

Department of Economics
Gettysburg College
Gettysburg, Pennsylvania 17325

The

GETTYSBURG ECONOMIC REVIEW

Volume 7

Spring 2013

Department of Economics
Gettysburg College
Gettysburg, Pennsylvania 17325

The Economics Department and Omicron Delta Epsilon congratulate **Nathan Blyler** and **Kevin Lugo**, winners of the *2013 Dwight D. Eisenhower Society / R.M. Hoffman Family Memorial Prize in Economics*. The Eisenhower/Hoffman Prize is awarded to the economics student writing the best quantitative paper or project with public policy implications.

The Economics Department and Omicron Delta Epsilon congratulate **Nathan Blyler** and **Kevin Lugo**, winners of the *2013 Outstanding Thesis Award*.

The Economics Department and Omicron Delta Epsilon congratulates **Kevin Lugo**, winner of the *2012 Glatfelter Prize*, awarded to one student with junior standing possessing excellent scholarship in the social sciences.

The Economics Department and Omicron Delta Epsilon congratulate **Taylor Smart** for receipt of a *2013 Mellon Grant.*

The Economics Department and Omicron Delta Epsilon congratulates **Marc Franco**, winner of the 2013 *Dr. and Mrs. William F. Railing Fellowship for Faculty-Student Research in Economics*.

The Economics Department and Omicron Delta Epsilon congratulate Nathan Blyler, Nicholas Jesteadt, Kevin Lugo, Ross Nichols, Bradley Wiseman for their induction into Phi Beta Kappa. Phi Beta Kappa celebrates and advocates excellence in the liberal arts and sciences. Its campus chapters invite for induction the most outstanding arts and sciences students at America's leading colleges and universities.

The Economics Department and Omicron Delta Epsilon congratulate the following students for their achievements in the 2012-13 academic year:

Economics Graduation Banner Carrier:	**Nicholas Jesteadt**
2013 Economics Honors Graduates:	**Nathan Blyler**
	Anna Cammisa
	Nicholas Jesteadt
	Kevin Lugo
	Ross Nichols
	Jia Qi Zhou

Omicron Delta Epsilon would also like to thank our outgoing officers, **Ross Nichols** and **Kevin Poplaski** .

CONTENTS

Price Barriers in the Stock Market .....pg 5
by Nathan Blyler

Welfare Incentives and Interstate Migration:
An Analysis of the Migration Decisions of
Poor, Single Mothers .....pg 39
by John Weis

On the Orthodox Nature of Heterodox Income
Distribution Theory.....pg 61
By Ross Nichols

China's Local Government Debt and Economic Growth.....pg 102
By Tom Zhou

Experimental Analysis of Resource Management Policies
Under Assumptions of Resource Migration .....pg 139
by Kevin Lugo

Price Barriers in the Stock Market and Their Effect on the Black-Scholes Option Pricing Model

By Nathan Blyler

Abstract:
The predicted price of an American option by the Black-Scholes (B-S) Option Pricing Model is known to differ from the market price of that option systematically with respect to time to expiration, distance in- or out-of-the-money, and liquidity of the option. We examine the possibility of price barriers in the stock market causing further systemic pricing differences between the market price and B-S predicted price. These differences occur when an option's strike price is near a price barrier and differ in effect and significance depending on the position of the barrier relative to the underlying stocks' price. We find round number price barriers in the stock market are beginning to be internalized into the option market. Additionally, Bollinger bands and Gann levels appear to receive special attention from investors, but do not act as price barriers.

Keywords: Black-Scholes, Price Barriers, Systematic Errors, Option Market

I. Introduction

A. Financial Background

The option market is closely related to the stock market in many ways, but differs in trading practices and uses. Similar to the stock market, the options market relies on speculation of stock price movement over time, but unlike a stock the derivative is only valuable over a set period of time. In this paper we will focus on call options. A call is simply the right to buy a security at a specified price called a strike price. The writer of a call is agreeing to sell the security at the strike price when the buyer exercises the call before a set date. The buyer can choose to let the call expire without ever exercising it, which would normally be done if the price of the security falls below the strike price. Thus, the buyer could simply buy the stock at market price for less than the strike price. In the

case where the strike price is above market price, the call is said to be out-of-the-money, and when the strike price is below the market price, the call is in-the-money. Since an option can move into the money at any time, the price of the option must not only be the difference between the strike price and the current security price to account for immediate exercising (if the call is in-the-money and zero otherwise) but also include a time premium for the chance a call moves to being in-the-money.

Options are appealing to investors for numerous reasons. They are substantially cheaper than buying a security, and are normally settled in cash rather than trading the underlying security at the specified price. This allows investors to speculate on a security's price movement with less initial capital and realize larger percentage returns on their investments. Furthermore, the options market allows for speculation on the volatility of a security's price rather than the direction of its movement. The final use of the options market is to hedge risky positions taken in the stock market. Consider someone who has sold a stock short without owning the stock and has unlimited loss potential. However, if she buys a call option giving her the right to buy the stock at a given price, she can only lose the premium she paid and the difference between the strike price and the price at which she short sold the stock. The use of call options as in this manner led to the most well known option pricing method called the Black-Scholes Model.

In their ground-breaking paper on corporate liabilities, Black and Scholes created a model for pricing call options based on hedging to form a riskless portfolio of stocks and options (Black and Scholes 1973). One of their assumptions was no arbitrage in the market, so any risk-free portfolio should provide a return at the risk free rate, allowing them to find a price for call options given the following inputs: security price, strike price, risk free rate, time to expiration, and volatility of the underlying security's returns. Their predicted

price of a call option is found by the following system of equations, which can be thought of as the risk adjusted probability an option finishes in the money:

$$C_0 = S_0\Phi(d_1) - Xe^{-rT}\Phi(d_2) \quad (1)$$

$$d_1 = \frac{\ln(S_0/X)+(r+\sigma^2/2)T}{\sigma\sqrt{T}} \quad (2)$$

$$d_2 = d_1 - \sigma\sqrt{T} \quad (3)$$

where
C_0 = Current call option value
S_0 = Current stock price
$\Phi(d)$ = The probability a random draw from a standard normal distribution will be less than d.
X = Exercise price
r = Annualized risk-free interest rate
T = Time to expiration in years
 = Volatility of the underlying stock

Part of the reason for the widespread use of this model is the ease with which most of these inputs can be found, as all are easily observable except for volatility. Therefore, pricing an option becomes a question of how much the underlying stock's price will change during the life of the option. The higher the volatility, or movement in price, the higher the probability the stock's price will end above the strike price.

The model is built on other assumptions besides no arbitrage, such as stock prices follow a Brownian motion, disallowing jumps and predictability. Furthermore, the underlying stock is assumed to pay no dividends and the option is assumed to not be exercised until expiration. Additionally, the model assumes there are no transaction costs in the option market. The most important assumption is the variance of the rate of return of the security is constant. While most of these assumptions obviously do not hold in any market, it is important to note that these assumptions only need to hold over the life of an option, not forever, to make the model prediction valid.

One can easily see, based both on the use of options and the derivation of call pricing, there is a close tie between stock markets and options market. Because of this close relationship, we would expect anomalies in security prices to result in option pricing anomalies. A recent finding in behavioral finance involves the existence of price barriers, such as whole number values, in the stock market (Sonnemans 2003, Dorfleitner et. al. 2009, etc.). These price barriers, or levels of support and resistance, are not unbreakable, but have an effect on the movement of stock prices. As a rising stock price approaches a barrier from below, making that barrier a level of resistance, the volatility of the price usually lessens. If a barrier is approached from above, making it a level of support, the stock price's rate of change is again likely to slow (Donaldson and Kim 1993).

B. Paper Outline

The stock market is one of the last strong holds for purely rational actors, with many believers in the efficient market hypothesis even while game theory and psychological microeconomics continue to find flaws in the neoclassical assumption of rationality. The efficient market hypothesis does not allow for price barriers to exist in the stock market; therefore, rational investors should not treat options with strikes near the barriers any different than other options. An anomaly around the barriers would indicate that investors who stand to lose substantial amounts of money do not completely believe the efficient market hypothesis.

Comparing option prices to each other does not allow for analysis as they vary in important ways such as distance from the money, time to expiration, and underlying stock. In order to hold these important factors constant and make meaningful comparisons, we use the 'correct' B-S predicted price of the option, found by regression techniques. Given the existence of price barriers, the B-S model should misprice options that have a barrier between the stock and strike

price because it does not account for the presence of these barriers. In this paper we will examine errors in the B-S model to detect systematic errors around price barriers.

Our findings show round number price barriers found by other studies have begun to be internalized into the options market. This is strong evidence against the efficient market hypothesis from the actors that are said to be completely rational. Our other proposed price barriers, Bollinger bands and Gann levels, do not act as we hypothesize, but also show investors pay particular attention to specific price levels. Thus, the assumptions made by the B-S model result in other systematic errors not yet discussed in the literature. By combining the behavioral finance literature on price barriers and the literature on errors in the Black-Scholes option pricing model, this paper is the first to look at price barriers' effect on the options market. This paper strengthens the argument against efficient stock markets by using an efficient option market. Moreover, it adds to the sizable literature on the accuracy of the B-S and how the market price differs from the predicted price in systematic ways.

In the next section, a detailed background of other literature on errors in the B-S option pricing model caused by the model assumptions is given. This is followed by a brief introduction to the discovery of price barriers in stock markets. The following section describes our methodology of finding the correct B-S price and controlling for known systematic errors. Section four justifies the proposed price barriers and the expected findings. Section five gives a summary of the data used, while section six shows the findings. Finally, section seven makes conclusions given the findings and offers alternative hypotheses to explain the unexpected results.

II. Literature Review

A. Black-Scholes Systematic Errors

As with all mathematical models, the assumptions of the B-S model have faced scrutiny from academics since its introduction in 1973. Problems with the assumptions have emerged in varying degrees from unimportant to creating systematic errors in pricing. To avoid misspecification, known systematic errors must be controlled for when examining the difference between the market and B-S price. Theoretically, an American[1] option is always worth more in the market than exercised, causing many studies to ignore the possibility of early exercising as this violation does not appear to cause systematic errors (Merton 1973, Macbeth and Merville 1979, ect.). Additionally, investors commonly witness jumps in the price of a security from the revealing of new information, violating the assumption of Brownian motion and causing the B-S model to predict prices under the market price (Merton 1976).

The assumption of constant variance in a security's return throughout the life of an option causes significant disparities between observed prices and predicted prices. A common adaptation has been to model variance as an unpredictable stochastic process (Chesney and Scott 1989, Hull and White 1987, Scott 1987, ect.). The results of modeling stock return variance in this way are inconclusive and make predicting the price of an option considerably more difficult. Hull and White include a volatility of volatility term in their model and find the B-S model to over price at- and in-the-money options; however, they note the over pricing is caused by a positive correlation between the price of the underlying stock and volatility. This is a problem, because the correlation between the two is not constant and has been positive some years and negative others (Rubinstein 1978, Schmalensee and Trippe 1978). Others have tried to

1 An American option can be exercised any time between its sale and date of expiration.

model volatility as a function of asset price and time, but failed to achieve better results than attempts to smooth out implied volatility over strike price and time (Dumas et. al. 2002). These failures to create a superior model have allowed the flawed B-S model to remain the standard in option pricing, and as such be using in our analysis of the option market.

Ignoring evidence that stock returns do not have a constant variance over time, simply estimating the expected variance still causes problems as it is the only input not observable in the B-S model. Weighted historical volatility, the simplest of estimators, with recent volatility given the most weight, does not include important factors such as: market volatility, price and volatility correlation, mergers or other large events, and the volatility implied by the options market (Black 1975). In an efficient market, the relevant information would be entirely included in the current market price. Along these lines, multiple authors have found the options market is better at estimating the future volatility of securities returns than historical averages (Black and Scholes 1972, Ncube 1996, Blyler 2012). These studies support our approach to estimating the correct volatility of a stock through the options market.

The largest differences in market and predicted prices occur when an option is far in- or out-of-the-money; however, the direction of these differences is debated. Black finds that far in-the-money options have a market price below the B-S predicted price and far out-of-the-money options have a market price above it (Black 1975). Alternatively, Macbeth and Merville find, on average, in-the-money options are overpriced by the market and out-of-the-money options are underpriced by the market relative to the B-S model (Macbeth and Merville 1979). In partial agreement with the aforementioned authors, Merton finds the market price to exceed the B-S predicted price when the option is both far in- or out-of-the-money and when the option is close to expiration (Merton

1976). Additionally an option's time to expiration has been found to generate discrepancies between the B-S and market prices. Options with a short time to expiration, three months or less, tend to have a market price greater than predicted by the B-S model (Black 1975). Nevertheless, the extent for which in-the-money and out-of-the-money options are mispriced decreases as time to expiration decreases (Macbeth and Merville 1979). Possible explanations for the systematic error accompany each finding; however, no paper we are aware of mentions the possibility of price barriers in the stock market as a reason for systematic error in the B-S option pricing model.

B. Price Barriers

The study of humans' psychological ties to specific numbers and the subsequent relation to the stock market is a somewhat recent development. Academic papers focus mostly on the human affinity for round numbers, with many finding significant price barriers at round numbers (Donaldson and Harold 1993, Sonnemans 2003, Koedijk and Stork 1994, ect.). Price barriers have been found by observing the frequency of specific stock prices and the rate of change in a stock's price around those levels (Donaldson and Harold 1993, Sonnemans 2003). Humans seeing a change in the largest nonzero place holder (e.g. 19.9 changing to 20.0) as a larger jump than an equivalent monetary jump that leaves that number unchanged could be a psychological cause for price barriers in the stock market (Sonnemans 2003). The resistance of breaking a round number is found when a stock closes ending in a 9, as those stocks experience significantly higher levels of selling off than other stocks (Bagnoli et. al. 2006). There is also evidence of price barriers becoming levels of support as trading volume increases after a significant price level is broken (Donaldson and Harold 1993, Huddart 2005). The importance of price barriers in the stock market has been known to

traders for much longer than its discovery in academia. From the early works of W.D. Gann, and possibly before, some traders have attributed their profits to the knowledge of the proportionality of the stock market (Gann 1935).

Not all literature is as supportive of the importance of price barriers in the stock market, although most have found evidence of it to some extent. Emerging markets, possibly because of more rapid growth, do not exhibit strong support for the hypothesis of price barriers (Bahng 2003). Recently, even in more developed markets, such as some in Europe, price barriers were not found to be constant over time. Once the anomalies were recognized, they tended to disappear in accordance with the efficient market hypothesis (Dorfleitner and Klein 2009). Some find that while price barriers exist they are of no use to investors because knowledge of the barriers does not allow investors to predict a stock's return (Koedijk and Stork 1994). Furthermore, automated investing could result in the formation of price barriers because limit orders are usually placed at round numbers, which would account for the clustering of prices and increase in trading volume when a stock price reaches a round number (Chiao and Wang 2009).

This paper examines the effect of price barriers in the stock market on the options market and attempts to identify price barriers in stock prices by anomalies in option pricing. If option prices take all market information into account, including price barriers, then the difference between the market price and the B-S predicted price would vary more near a barrier price. Examining price barriers' effect on the B-S model extends beyond current literature on price barriers, which focuses mainly on locating barriers within stock markets. The paper combines two strands of literature by using similar methodology seen in previous studies on the systematic errors in the B-S model and looking for errors predicted by behavioral finance theories.

III. Methodology

This paper focuses on testing the validity of the Black-Scholes model when accounting for the presence of price barriers through the exploration of systematic errors. Previous literature has found price barriers at round numbers for multiple securities by looking at frequency of a security's price (Donaldson and Harold 1993, Sonnemans 2003, Koedijk and Stork 1994, ect.). As the price of a stock approaches a price barrier the movement of the price slows, allowing for frequency analysis to locate the barriers. This results in the Brownian movement assumption of the B-S model being violated. The violation of this assumption at select price levels should create an error at those levels that is not seen otherwise. Evaluating the difference between the B-S predicted price and the market price is contingent on accurately evaluating the volatility of a security's return to find the correct B-S price.

A. Finding the Correct B-S Price

To accomplish this initial task, we use the methodology of MacBeth and Merville in "An Empirical Examination of the Black-Scholes Call Option Pricing Model." Their analysis is reliant on the assumption that the B-S model accurately predicts an at-the-money call option price. Black notes errors in his model on options far in- or out-of-the-money, and in options with less than 90 days to expiration (Black 1975). To account for this, the model used to estimate implied volatility of a security's return controls for distance from the money using only options with greater than 90 days to expiration.

Taking the B-S implied volatility of an at-the-money option as the true volatility for the underlying security's returns, we run a regression to estimate this volatility. The regression is run on all options traded on that day for a particular security. In total, 252 trading days per security, less the days the particular security

had fewer than 5 different options traded were used. The estimated regression, taken from MacBeth and Merville, is the observed implied volatility of an option's market price regressed on its distance from the money. The model is given by

$$\sigma_{ijt} = \theta_{i0t} + \theta_{i1t} m_{ijt} + \varepsilon_{ijt}, \quad (4)$$

where *i* ranges from 1 to *I* representative of the *I* companies, *t* ranges from January 3, 2011 to December 30, 2011 for each trading day, and *j* ranges from 1 to J, with J≥5, for all different options of company *i* on day *t*. Here, the only control variable *m* is the distance the option is from the money as a percentage of the security's price. More formally,

$$m_{ijt} = \frac{S_{it} - X_{ij} e^{-rT}}{S_{it}} \quad (5)$$

where S is the stock price of company *i* on day *t*, X is the strike price of option *j* of company *i* discounted by the risk free rate back to its present value. This measure is a slight variation on MacBeth and Merville's work, where the difference is taken as a percentage of strike price. The use of call options to hedge positions caused their measure of distance from the money to be severely skewed, but this small variation decreases the skew substantially without drastically changing the results (Blyler 2012).

In the above regression, the intercept is the estimated implied volatility of an at-the-money option. Our assumption states that this estimate is the correct volatility of the underlying security's return and should be used to find the B-S prediction of that security's options at any strike price on the given day. This allows us to find the difference between the market price () and the B-S predicted price given by

$$y_{ijt} = C_{ijt} - C_{BS}(\hat{\theta}_{i0t}). \quad (6)$$

In the above equation, is the market price of option *j* on day *t* of the underlying security and is the B-S predicted price of that call option using the estimated true volatility.

B. Systematic Difference between Market Price and B-S Predicted Price

MacBeth and Merville find the difference between the market price of a call option and the B-S predicted price to be a function of the distance of the option's strike price from the money and the option's time to expiration. MacBeth and Merville's model estimating the difference between the market price and B-S predicted price of a call option is given by

$$y_{ijt} = \alpha_0 + \alpha_1 m_{ijt} + \alpha_2 T_{ijt} + \varepsilon_{ijt} \quad (7).$$

The regression is run separately, not only for each underlying security, but also for different properties of options. Based on previous literature, finding differences in pricing errors between options with short or long times to expiration and options in- or out-of-the-money, each underlying security has four separate regressions to allow different estimates for all possible combinations of options near or far from expiration and in- or out-of-the-money.

MacBeth and Merville's model uses linear variables, but through empirical work a model including a squared term for distance from the money was found to be more appropriate (Blyler 2012). In their 1979 paper, the data is not treated as panel data although options are followed over time, resulting in heteroskedasticity and autocorrelation. More recent advances in panel data analysis have led to an increase in financial data being analyzed using panel techniques to correct for these problems inherent in the data (Petersen 2005, Gow et al. 2010). Previous authors using panel data and similar regression techniques to estimate volatility have suggested the square of distance from the money and the liquidity of both the underlying asset and option contribute to pricing

difference from the B-S option pricing model (Ncube 1996, Feng 2011). To allow analysis over all stocks, we introduce a new dependent variable,

$$y\%_{ijt} = \frac{y_{ijt}}{C_{ijt}} \quad (8).$$

Therefore, MacBeth and Merville's original model is revised so that

$$_{jt} = \alpha_0 + \alpha_1 m_{ijt} + \alpha_2 m_{ijt}^2 + \alpha_3 \ln(vol)_{ijt} + \alpha_3 vol_{ijt} * m_{ijt}^2 + \alpha_4 T_{ijt} + \alpha_5 PBA_{ijt} + \alpha_6 PBB_{ijt} + \varepsilon_{ijt} \quad (9).$$

The variables *y, m*, and *T* are taken from equation 7, and the addition of m^2 comes from our empirical work and other authors (Feng 2011, Blyler 2012). Furthermore, Feng's work on the liquidity effect in the option market makes controlling for the volume of trading of each option on each day appropriate. In his study, volume of trading also affected the curvature of the error caused by distance from the money. In order to account for Feng's finding, is the number of trades an option had on a given day. Its effect on the curvature is controlled for by multiplying , while its overall effect on the error is given in log scale because of decreasing returns. The analysis of this paper focuses on the dummy variables and , representing price barriers above and below the underlying security's current price respectively. Price barriers above the current price, denoted , are expected to have a negative sign because they represent levels of resistance. Alternatively, price barriers below the current price, denoted , are expected to have a positive sign because they represent levels of support.

Previously mentioned price barriers have been found in the stock market by numerous authors; therefore, if the options market is efficient, then these price barriers should impact the market price of options because all available information in reflected in the price of the option. The B-S model assumption of Brownian motion in a security's price does not allow for price barriers to exist. Hence, they cannot be priced into the B-S predicted price. If

price barriers in the stock market are priced into call options, then those options with strikes near the barriers should have differences between the market and B-S predicted price not explained by other independent variables. How price barriers affect a stock, and therefore the corresponding options price, differ depending on the location of the barrier relative to the security's current price.

A price barrier above a security's current price acts as a level of resistance to a price increase. As the price of a security approaches a level of resistance, its rate of change of price is expected to decrease. Furthermore, the probability of the price rising above the price barrier is less than the probability of it rising above an arbitrary level that offers no resistance. A smaller chance of breaking the level of resistance and an expected decrease in the rate of change should lower the price of options with strike prices near the barrier. Out-of-the-money options have a lower probability of finishing in-the-money if the strike price is at or slightly above the barrier. Additionally, finishing far enough in-the-money for the buyer to recoup the B-S predicted option premium is less likely if the strike price is slightly below the barrier.

Alternatively, a price barrier below a security's current price can be seen as a level of support to a price drop. As the price of a security decreases towards a level of support, the absolute rate of change is expected to slow. The properties of a level of support act similarly to a level of resistance, but should result in a higher option price if the strike is near a level of support. Such an option is less likely to fall out-of-the-money before expiration and should command a higher premium from the buyer.

IV. Hypothesized Price Barriers

A. Bollinger Bands

In this section, possible locations for price barriers are presented along with hypothesized reasons and potential implications. The first location for possible price barriers comes from a commonly used financial technical indicator, Bollinger bands. Bollinger bands were made famous by analyst John Bollinger who believed "asking the market what is happening is always a better approach than telling it what to do" (Bollinger 1992). His standard for measuring volatility of a stock started with a simple *n*-period moving average of a stock's price, which evolved into a weighted moving average in some cases, with *k* standard deviations of the stock's last *n* prices added and subtracted from the moving average.

For our analysis the period length is Bollinger's suggested 20 days and the distance from the simple moving average is two standard deviations (Bollinger 1992). Bollinger bands being a measure of historic volatility, an investor could view the bands as price barriers the security's price is unlikely to break if historic estimates of volatility hold. She would then calculate the Bollinger bands at the end of a trading day and use those levels to adjust her valuation of options the upcoming day. Using this strategy, investors would buy options near the bottom Bollinger band and write options near the top Bollinger band, driving prices up and down respectively. Of course, an option's strike price will rarely be exactly at a Bollinger band, meaning some distance must be deemed 'close enough' to the Bollinger band for investors to believe the effects of price barriers would play a role in evaluating the option.

Options have strike prices at $10 intervals if the underlying security's price is greater than $200, $5 intervals if the security's price is between $25 and $200, and $2.5 or $1 intervals if the price less than $25. Notice the strike price interval is never less than 5% of the underlying security's price besides

the extreme cases of very expensive stocks. Working off of this, we deemed an option's strike price 'close enough' to a Bollinger band if

$$|Strike - Bollinger\ band| < .025 * Underlying\ Security's\ Price.$$

This method effective creates a 5% interval around Bollinger band, but does not allow two different strike prices to be 'close enough' to the price barrier. Because of the uncertainty of the differences between options with strike prices directly above and strike prices directly below the price barrier, two separate dummy variables are utilized; however, the expected sign on both is the same depending on the strike being near the upper or lower Bollinger band.

Bollinger bands suffer from a few draw backs that make them less likely price barriers than the other proposed barriers. Investors can use 10 period Bollinger bands as well as vary the number of standard deviations added and subtracted, which would result in differing opinions about the exact location of price barriers. Additionally, Bollinger bands change daily using this formula, meaning an investor using the above strategy would only believe the price barrier existed at that level for a day. Options affected by a barrier for a day would not vary greatly in worth as it is likely they do not expire for many days to come and the barrier would likely shift by then. A more realistic price barrier would stay constant throughout time, or at least for a meaningful length of time.

B. Round Numbers

Most current literature on price barriers focuses on round numbers, meaning integers when a security's price is low, multiples of ten when the price is slightly higher, or multiples of 100 when the price is higher still. Many authors find the existence of price barriers at these numbers in a variety of securities and markets. Barriers at round numbers are the easiest to defend as given because the psychological reasoning behind their existence is a staple of behavioral finance.

Investors seem to weight round numbers more than their mathematical worth if markets are perfectly efficient.

Here the issue that arises is options are often sold at round numbers as explained above. To avoid this problem, barriers, or at least stronger barriers, are hypothesized to exist at round numbers seemingly more important to humans. For most securities, options are sold with strikes at multiples of ten or five because their underlying price is greater than $25. As such, for all securities with a price of $20 or more, the important round number strike prices are chosen to be at multiples of $50. For securities with a price less than $20, important round number strike prices were chosen at multiples of 5. While these numbers are reasonable selections, there is an inherent weakness is simply choosing these barriers rather than investigating all round numbers; however, our methodology would not hold if all round numbers were taken into account.

There are multiple price barriers at any given time and the arguments for round numbers result in barriers switching between support and resistance depending on their position relative to the security's current price. This means a broken level of resistance becomes a level of support and should drive the price of the security higher as investors grow more confident in its performance. Similarly, a broken level of support becomes resistance and should drive the price down even further. This could counteract the differing value given to options with strikes near the barriers. A less risk adverse investor may be willing to pay more for an option with a strike near a level of resistance, knowing if the option moves into the money, it is likely to move further into the money. Additionally, an option with a strike near a level of support may not appear less risky knowing if the option moves out-of-the-money one day it is less likely to move back into the money than other options. The final hypothesized price barrier locations avoid both problems of changing over time and switching between support and resistance.

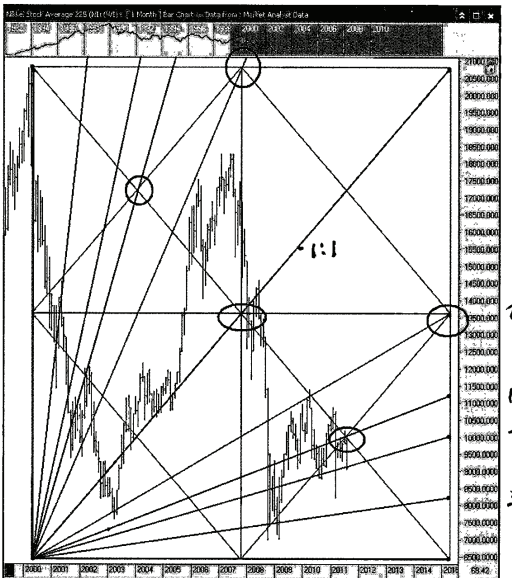
C. Gann Levels

Infamous investor W.D. Gann is known for his unorthodox beliefs about the stock market and the success he had with trading strategies based on seemingly unrelated occurrences. From the effects of planetary retrograde motion to the mathematical properties of geometric shapes, Gann's trading methods are unconventional; however, numerous books have been written by Gann and others about these special proportions in markets (Brown 1999). While skepticism should accompany outlandish claims, the success and popularity of both the books and methods demonstrate many traders know about, if not use, Gann's approach. Knowing investors in the stock market may act differently at special price levels, their expected actions should be priced into the options market. This makes no statement about the logic behind Gann's method; rather, if enough investors believe his approach has merit it becomes a self-fulfilling prophecy.

Although Gann makes numerous claims, the analysis here will focus on levels created by a method referred to as Gann's wheel.[2] Gann's wheel involves levels of support coming from a pivot high and levels of resistance coming from a pivot low where a pivot is an important price level. Following the methodology described in Brown 1999, we examined the charts of each stock in 2010 to find initial pivot highs and lows of our stocks. These chosen pivots remained until the stock's price fell below the pivot low or rose above the pivot high, at which point the broken pivot switched to the new year to date high or low. From here, Gann describes the angles 45, 90, 120, 180, 240, 270, 315, and 360 degrees of having particular importance. In a Gann wheel, these angles are drawn from the pivot low and pivot high on a plane of price and time until they intersect what Gann calls the square of nine. The time of the intersection is Gann's prediction for when

2 For a more detailed explanation of both how to pick pivot highs/lows and the methodology behind Gann's wheel see Brown, Constance M. *Technical analysis for the trading professional: Strategies and techniques for today's turbulent global financial markets*. McGraw-Hill, 1999.

the stock's price will reach that barrier; however for this analysis we will only focus on the price level at which the intersection occurs. An example of these intersections can be seen in the figure below.



(Brown pg 234, 1999)

These price barriers will remain constant until either the pivot high or pivot low changes, and the important intersections can be calculated by the following equations for a given degree.

$$\text{Resistance level for degree } d = (\sqrt{Pivot\ Low} + \frac{d}{180})^2 \quad (10)$$

$$\text{Support level for degree } d = (\sqrt{Pivot\ High} - \frac{d}{180})^2 \quad (11)$$

Similar to the Bollinger Band price barriers, options are unlikely to be sold with strike prices exactly at Gann's levels. Again, the strike price is deemed 'close enough' to the price barrier if

$$|Strike - Gann\ Level| < .025 * Underlying\ Security's\ Price,$$

creating a buffer area for the strike price to fall.

The drawbacks of Gann's price barriers come from their complexity and often misunderstood fundamentals. Different trading programs calculate Gann levels with built in functions; however, these functions differ in their implementations of Gann's wheel (Brown 1999). If investors are not in agreement on the location of Gann's price barriers, the effect on option prices would not be significant. This could allow traders who correctly estimate Gann's levels to make larger gains or it could push investors away from Gann's method towards a more straightforward method.

V. Data

The data for this paper is called National Best Bid Offer (NBBO) data collected by Options Pricing and Reporting Authority (OPRA). It contains information on the traded call options of 33 companies, most of which currently make up the DOW 30, as well as several important technology firms. The firms and their ticker symbols are shown in the following table.

Company Name	**Ticker Symbol**	**Company Name**	**Ticker Symbol**
Apple	AAPL	International Business Co.	IBM
American Express Co.	AXP	Johnson & Johnson	JNJ
Bank of America Corp.	BAC	JP Morgan Chase and Co.	JPM
Boeing Co.	BAC	The Coca-Cola Co.	KO
Citigroup Inc.	C	3M Company	MMM
Caterpillar Inc.	CAT	McDonald's Corp.	MCD
Cummings Inc.	CMI	Microsoft Corp.	MSFT
Chevron Corp.	CVX	Pfizer Inc.	PFE
E.I. du Pont de Nemours and Co.	DD	Qualcomm Inc.	QCOM
Walt Disney Co.	DIS	Transocean Ltd.	RIG
General Electivc Co.	GE	AT&T Inc.	T
Google Inc.	GOOG	United Technologies Corp.	UTX
The Goldman Sachs Group, Inc.	GS	Verizon Communications Inc.	VZ
Halliburton Co.	HAL	Wal-Mart Stores Inc.	WMT
Hewlett-Packard Co.	HPQ	Exxon Mobil Corp.	XOM
Home Depot Inc.	HD	Standard & Poors 500	SPY
Intel Corp.	INTC		

The data is end of day data from all 252 trading days in 2011 with strike prices ranging from far in-the-money to far out-of-the-money. The times to expirations at the initial sale of the options include one month, three months, six months, and a few times of a year or more. All options are followed until expiration and every day includes the ending spreads, greeks, and trading volumes for the options.

Since the analysis relies on market prices containing all information, only options traded (trading volume greater than zero) on a specific day are included in the analysis. Because the companies involved in the analysis are commonly traded companies, this does not eliminate many options besides those with strike prices too far away from the underlying security's price to be useful to traders, or those with extremely long times to expiration. In order to eliminate data we viewed as likely mistakes, options that were sold for less than the intrinsic value of the option were not included. Additionally, options with less than two days to expiration are not included in the analysis because end of the day data is not reflective of the changes occurring with only a couple days left in the life of the option.

Using regressions to estimate the true volatility of a stock's return creates a weighted average of implied volatility from all of the traded options on that day. In order to assure enough inputs into the weighted average, the true implied volatility is only calculated if options with 5 different strikes or expirations were traded that day. This estimate is then used to calculate the inputs required in the Black-Scholes option pricing model. In order to estimate the annualized risk free rate, the annualized return rate on a three month U.S. Treasury Bill was obtained from the St. Louis Federal Reserve. The time to expiration is obviously not always three months; however, this should not be a problem because annualized Treasury Bill returns only vary slightly with changes in time to maturity from a day to a

year, and the B-S predicted call price lacks sensitivity to the riskless return rate. In fact, MacBeth and Merville note, "…[their] results would be virtually identical had [they] used a single riskless return for a Treasury Bill… ."

The data used differs from MacBeth's and Merville's data in two distinct ways. First, the data includes information on more than five times the number of stocks. Second, many days have over fifty observations used in the regression to calculate the estimated at-the-money volatility. These advantages allow for stronger conclusions because it reduces the likelihood that small sample sizes will cause the results.

VI. Results

Looking at the difference between the market price and B-S predicted price as a percent of the market price allows for comparison of options regardless of the underlying stock. However, our data set does not allow for comparison regardless of distance from the money. Using our formula the percent error cannot go above 100%, but it has no lower bound. When options are in-the-money, y% has a smaller range and a reasonable standard deviation, with summary statistics

Variable	Observations	Mean	Std. Dev.	Min	Max
y%	296554	0.534	12.091	-2622	98

On the other hand, with options out of the money, y% has a much larger range, as seen below:

Variable	Observations	Mean	Std. Dev.	Min	Max
y%	398889	-61.05	154	-16620	100

The extreme difference leads to drastically different coefficient estimates in our regression, which are divided up between in- or out-of-the-money and less or greater than 90 days to expiration. This division is based on the differences in pricing observed by Black, MacBeth and Merville, and others. As such, the

coefficients of the regressions are less important than the significance and sign of the coefficients. In all regressions, observations are demeaned by the underlying stock and the standard errors are clustered by option, allowing for correlation between the errors of the same option sold on different days. Clustering by option, prevents the autocorrelation that plagued previous papers (see MacBeth and Merville 1979, Blyler 2012) and makes the significance tests useable.

Running the regression on options in the money with more than 90 days to expiration yields the following results where positive coefficients mean higher relative market value.

	y%	y%	y%
m'	28.032**	27.899**	27.556**
	(1.123)	(1.138)	(1.148)
m'^2	-34.192**	-34.103**	-33.014**
	(2.293)	(2.304)	(2.286)
month	-0.152**	-0.152**	-0.131**
	(0.011)	(0.011)	(0.011)
volume	-0.316**	-0.319**	-0.334**
	(0.025)	(0.025)	(0.026)
volume * m'^2	0.068**	0.068**	0.080*
	(0.025)	(0.025)	(0.032)
Lower BB Above	-0.202		
	(0.124)		
Lower BB Under	0.373**		
	(0.104)		
Gann Support Above			0.051
			(0.085)
Gann Support Below			0.124
			(0.082)
Round Number Under		0.114	
		(0.180)	
Constant	-3.060**	-3.053**	-3.468**
	(0.202)	(0.200)	(0.214)
Observations	148696	148696	163745
R-squared	0.19	0.19	0.19

Robust standard errors in parentheses
+ significant at 10%; * significant at 5%; ** significant at 1%

The first proposed price barriers, Bollinger bands, have contradicting signs for their coefficients; however, only the coefficient on the variable for options with strikes slightly below the lower Bollinger band is statistically significant. This is promising as it has the sign predicted by our hypothesis that

investors likely view the lower Bollinger band as a level of support and therefore options with strikes slightly below the band are more likely to finish in the money and should have a higher market value than other equivalent options. The results for options with strikes at our selected round numbers also suggest investors may view round numbers as a level of support, but the coefficient is not statistically significant. An option with a strike near a Gann level of support, while yielding coefficients with the sign predicted by our hypothesis, also does not result in a statistically significant increase in the option's price. While the results of the first regression are promising, similar results are needed in order to provide strong evidence for our hypothesis.

The same regressions run on options with less than 90 days to expiration yields certain contradictory results, preventing stronger inferences from being made about levels of support.

	y%	y%	y%
m′	32.181**	32.228**	30.334**
	(1.633)	(1.621)	(2.248)
m'^2	-54.963**	-55.247**	-52.348**
	(3.288)	(3.297)	(3.315)
month	-0.693**	-0.698**	-0.823**
	(0.059)	(0.059)	(0.082)
volume	-0.551**	-0.560**	-0.742**
	(0.035)	(0.035)	(0.053)
volume * m'^2	0.077**	0.080**	0.177**
	(0.022)	(0.023)	(0.050)
Lower BB Above	-0.498**		
	(0.170)		
Lower BB Below	0.496**		
	(0.137)		
Gann Support Above			-0.455**
			(0.163)
Gann Support Below			-0.044
			(0.153)
Round Number Under		0.279	
		(0.245)	
Constant	-2.689**	-2.749**	-3.312**
	(0.246)	(0.237)	(0.407)
Observations	111662	111662	132823
R-squared	0.15	0.15	0.04

Switching from options that are far from expiration to options that are near to expiration not only affected the signs of some of our variables of interest but also their significance levels. Options with strikes just below the lower Bollinger Band once again have an increased value in the market's view and this increase is statistically significant. Once again, options with strikes slightly above the lower Bollinger Band are less valuable relative to the B-S predicted price, but this time the difference is statistically significant. Our hypothesis has no explanation for this phenomenon, but the positive and statistically significant coefficient when the strike is slightly below the lower Bollinger band provides evidence that investors do use the previous day's Bollinger band to predict the stock's future movement and investors believe a stock's price is less likely to fall below the lower Bollinger band than an arbitrarily selected price. A possible reason for this phenomenon is the liquidity of the options market allows investors to trade frequently; therefore, they only try to predict movements for a short time before selling the option.

Price barriers at round numbers appear to have the expected affect in the options market, although the coefficients were not statistically significant in either regression. Price barriers being a relatively new discovery in the stock market, investors may be slow to change their evaluation methods. A smaller number of investors using round numbers as levels of support may be pushing the price of these options higher, but not a significant amount because other investors continue to pull the price downwards.

The result from Gann's levels of support in this second regression is most surprising. The statistically significant negative sign suggests investors pay attention to Gann's levels; however, investors value options with strikes near Gann's levels of support less than other options. The inconsistency between options near to expiration and options far from expiration may be from the time aspect of Gann's wheel that is not incorporated into this analysis. If Gann's wheel

predicts a level of support being reached long after an option expires, investors are unlikely to put weight into Gann's analysis. Our hypothesis does not explain the statistically significant negative coefficient for options with strikes near Gann's level of support with less than 90 days to expiration. It does not appear that Gann's levels of support, at least using the levels our methodology located, are of importance to investors.

Next, the regressions are run on options out-of-the-money; thus we move from level of support to resistance. These coefficients are not comparable to the previous two tables; however, the signs and significance test are still viable. We again begin with options with more than 90 days to expiration and find the following with regards to levels of resistance.

	y%	y%	y%
m′	374.262**	379.530**	358.130**
	(14.297)	(13.867)	(12.802)
m'^2	90.057**	92.373**	85.137**
	(11.765)	(11.922)	(11.329)
month	3.035**	3.062**	3.091**
	(0.191)	(0.195)	(0.176)
volume	1.110**	1.247**	1.090**
	(0.326)	(0.341)	(0.314)
volume * m'^2	0.313	0.314	0.343
	(0.300)	(0.301)	(0.304)
Upper BB Above	7.888**		
	(1.319)		
Upper BB Below	6.609**		
	(1.550)		
Gann Resistance Above			11.594**
			(1.582)
Gann Resistance Below			16.573**
			(1.722)
Round Number Above		-4.877	
		(3.395)	
Constant	-14.314**	-11.259**	-23.244**
	(2.321)	(2.144)	(2.060)
Observations	220929	220929	237111
R-squared	0.27	0.27	0.28

Analyzing out-of-the-money options with strikes near Bollinger Bands, the results are opposite of what our hypothesis suggests. Options with strikes around the upper Bollinger band appear to have a higher relative value to investors. A possible explanation is investors believe the upper Bollinger Band acts as a spring board and once broken investors will buy up the stock, increasing the price significantly past the strike price. As seen in options out-of-the-money, investors seem to impute a higher probability to a stock's price staying within the Bollinger bands, so the increase in price expected if the upper Bollinger band is broken must be large enough to offset this decreased probability of its occurrence. Alternatively, investors may simply put less weight in the upper Bollinger band, believing the market will be bullish and the upper Bollinger band will not be as important.

Options with strikes at round numbers continue demonstrate the expected effects of a price barrier on the option market; however, the coefficient is still not statistically significant. Gann's levels of resistance on the other hand, display similar traits to the upper Bollinger band. Options with strikes near to a Gann level of resistance, whether the strike price is slightly above or below the barrier, have a higher relative value to investors. Since Gann levels do not switch between support and resistance relative to the underlying stock's price, our hypothesis does not explain this occurrence.

Similarly, the regressions on options close to expiration find the following results, supporting the results for levels of resistance from options far from expiration.

	y%	y%	y%
m′	34.360*	48.320**	50.946**
	(17.394)	(17.995)	(15.917)
m'^2	72.511**	80.594**	69.491**
	(17.056)	(18.185)	(13.968)
month	-18.036**	-18.136**	-24.714**
	(1.234)	(1.238)	(1.166)
volume	5.258**	5.945**	0.464
	(0.494)	(0.514)	(0.544)
volume * m'^2	-4.464	-4.846+	-1.421
	(2.755)	(2.878)	(2.145)
Upper BB Above	8.851**		
	(2.269)		
Upper BB Below	27.004**		
	(2.168)		
Gann Resistance Above			-0.052
			(2.525)
Gann Resistance Below			4.728+
			(2.565)
Round Number Above		-7.558	
		(4.723)	
Constant	-24.072**	-15.152**	-26.201**
	(2.647)	(2.502)	(3.086)
Observations	139226	139226	161778
R-squared	0.15	0.14	0.08

In options near to expiration, the only proposed price barrier with a significant effect is the Bollinger band. Again, options with strikes near the upper Bollinger band are of relative higher value to investors, contradicting our hypothesis. The Gann barriers appear to be insignificant when the option is close to expiration for both levels of support and resistance. This is most likely explained by our exclusion of the time component of Gann's wheel.

Options with strikes near our selected important round numbers again have the expected devaluation associated with a level of resistance. Throughout our analysis, round numbers have not had a statistically significant coefficient, but have continually had the predicted change relative to the B-S predicted price.

Overall, the results for round numbers and the lower Bollinger band are promising, but the significant coefficients of the wrong sign do not provide strong support for our hypothesis. Because Bollinger bands and Gann levels have

not been explored as price barriers in the literature, it is not surprising options with strikes near those levels do not act as expected; although, the statistically significant coefficients suggests more research is needed at these levels. The obvious skew caused by taking the difference as a percentage of market price hinders comparison and most literature focuses on absolute difference between the market price and the B-S price. Further exploration into the effect of examining the error as a percentage of market prices is needed before absolute conclusions can be drawn.

VII. Conclusions

A. Price Barrier Hypothesis

We propose three price barriers, Bollinger bands, round numbers, and Gann levels, in the stock market and attempt to find evidence of their internalization in the options market. Options with strikes near levels of support are expected to have higher market prices, whereas the opposite should be true if the strikes are near a level of resistance because the B-S price does not take the lowered probability of breaking a price barrier into account. To find relatively lower and higher prices, the market price is compared to the Black-Scholes Option Pricing Model predicted price, which does not allow for price barriers to exist in the stock market.

We find evidence of systematic deviation from the B-S price at Bollinger bands in options both in- and out-of-the-money. Gann levels of resistance had the opposite effect of what was expected, and the levels of support did not appear to have an effect on the price of options. This suggests investors are more comfortable with conventional measures when attempting to estimate volatility; however, if Gann levels act as price barriers in the stock market, investors could make substantial profits by buying options at levels of support and selling those at

levels of resistance. Testing this proposed trading method is left for future work.

The results for options with strikes near round numbers were inconclusive although there was evidence of them beginning to be treated as levels of support and resistance. Although no coefficient was statistically significant, all had the sign predicted by our hypothesis. Not only does this support the behavioral finance findings of price barriers in the stock market, but also shows efficiency in the options market as option traders recognize patterns in the stock market not allowed by the efficient market hypothesis.

Closer to hypothesized results were found when analyzing options in-the-money, possibly because of the use of options as a hedge. Increased attention to in-the-money options could push the price closer to the fair present value, while out-of-the-money options have a less efficient price. This paper does not attempt to prove the existence of price barriers in the stock market; rather it examines how the options market acts around special price values. Significant results provide evidence that investors do care about certain price levels more than others, although further exploration is needed to completely understand how the option market internalizes investor preferences of these numbers. However, when looking for systematic errors in the B-S model, special price levels should be included with the commonly accepted distance from the money, time to expiration, and liquidity control variables. The discovery of price barriers affecting the options market allows future research on price barriers to occur regardless of a stock's current price.

B. Updated Expected Volatility Hypothesis

Differences between the market price and B-S predicted price of options with strikes near Bollinger bands and Gann levels had differences not explained by previously found control variables, but these differences were not the differences predicted by our hypothesis. The coefficients on the dummy variables being

statistically significant indicate investors do pay attention to these levels, but not in our hypothesized manner. When the Bollinger bands were below the current price, hypothesized to act as levels of support, options just above (inside) the band were valued less, while options just below (outside) the band were valued more. For both Bollinger bands and Gann levels, when the strike was near those proposed barriers and above the strike price the results were opposite the hypothesized result. If those barriers acted as levels of resistance, the coefficient would be negative, but in both cases it was positive and statistically significant.

To explain these divergences from our original hypothesis, we propose an alternative explanation. Bollinger bands and Gann levels may not have any effect on a stock's price, but investors may pay attention to them for other reasons. The B-S predicted price, which is commonly used by investors as a baseline, depends heavily on expected volatility of a stock. When evaluating a stock, investors often look at not only the price, but also technicals such as Bollinger bands. Therefore, an investor may use Bollinger bands and Gann levels as indicators their original estimation of volatility needs to be updated.[3] Updating expected volatility around these levels would result in higher values for the option relative to the B-S price that holds a constant expected volatility estimate. This hypothesis explains the positive coefficient around barriers proposed to be levels of resistance and why levels of support followed our initial hypothesis more closely than resistance. Further research is needed to test the validity of this hypothesis, but it appears as if academic researchers need to obtain more information from real investors before price barriers can be fully explored.

3 Imagine seeing a stock price with Bollinger bands and using its historical volatility to estimate volatility. Since Bollinger bands are 2 SD away from the moving average, if the price breaks out of the bands, then the historical estimate is likely incorrect. At that time it may be best to update expected volatility to a larger value. Therefore, if Bollinger bands or Gann levels are reached, expected volatility expands and options are worth more.

VIII. References

Bagnoli, Mark, Jinyoung Park, and Susan Watts. "Nines in the endings of stock prices." *Available at SSRN 728544* (2006).

Bahng, Seungwook. "Do Psychological Barriers Exist in the Stock Price Indices? Evidence from Asia's Emerging Markets." *International Area Studies Review* 6.1 (2003): 35-52.

Bates, David S. "Empirical option pricing: A retrospection." *Journal of Econometrics* 116.1 (2003): 387-404.

Bates, David S. *Testing option pricing models*. No. w5129. National Bureau of Economic Research, 1995.

Bakshi, Gurdip et al. "Empirical Performance of Alternative Option Pricing Models" Journal of Finance, Dec. 1997

Black, Fischer. "Fact and Fantasy in the Use of Options" Financial Analysts Journal, Jul.-Aug. 1975.

Black, Fischer and Myron Scholes. "The Pricing of Options and Corporate Liabilities" The Journal of Political Economy, June 1973.

Black, Fischer, and Myron Scholes. "Taxes and the pricing of options." *The Journal of Finance* 31.2 (1976): 319-332.

Black, Fischer and Myron Scholes. "The valuation of option contracts and a test of market efficiency." *The Journal of Finance* 27.2 (1972): 399-417.

Blyer, Nathan "Revisiting an Empirical Evaluation of Errors in Black-Scholes Option Pricing." *Working Paper* (2012).

Bollinger, John. "Using Bollinger Bands." *Technical Analysis of Stocks & Commodities* 10 (1992).

Brown, Constance M. *Technical analysis for the trading professional: Strategies and techniques for today's turbulent global financial markets*. McGraw-Hill, 1999.

Chesney, Marc, and Louis Scott. "Pricing European currency options: A comparison of the modified Black-Scholes model and a random variance model." *Journal of Financial and Quantitative Analysis* 24.03 (1989): 267-284.

Chiao, Chaoshin, and Zi-May Wang. "Price Clustering: Evidence Using Comprehensive Limit-Order Data." *Financial Review* 44.1 (2009): 1-29.

Chiras, Donald and Steven Manaster. "The Information Content of Option Prices and a Test of Market Efficiency" Journal of Financial Economics, Spring 1978

Donaldson, R. Glen, and Harold Y. Kim. "Price barriers in the Dow Jones industrial average." *Journal of Financial and Quantitative Analysis* 28.3 (1993).

Dorfleitner, Gregor, and Christian Klein. "Psychological barriers in European stock markets: Where are they?." *Global Finance Journal* 19.3 (2009): 268-285.

Dumas, Bernard, Jeff Fleming, and Robert E. Whaley. “Implied volatility functions: Empirical tests.” *The Journal of Finance* 53.6 (2002): 2059-2106.

Eraker, Bjorn. “Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices” Journal of Finance, June 2004.

Feng, Shih-Ping. “The Liquidity Effect in Option Pricing: An Empirical Analysis.” *International Journal of Business and Finance Research* 5.2 (2011): 35-43.

Galai, Dan and Ronald Masulis. “The Option Pricing Model and the Risk Factor of Stock” Journal of Financial Economics, March 1976.

Gann, William Delbert. “The Basis of My Forecasting Method” 1935.

Garcia, René, Ghysels, Eric and Renault, Eric, The Econometrics of Option Pricing (August 1, 2003). University of Montreal working paper.

Gow, Ian D., Gaizka Ormazabal, and Daniel J. Taylor. “Correcting for cross-sectional and time-series dependence in accounting research.” *The Accounting Review* 85.2 (2010): 483-512.

Huddart, Steven, Mark Lang, and Michelle Yetman. “Psychological factors, stock price paths, and trading volume.” *AFA 2006 Boston Meetings Paper*. 2005.

Hull, John and Alan White. “The Pricing of Options on Assets with Stochastic Volatilities” Journal of Finance, June 1987.

Kim, I. N. “The analytic valuation of American options.” *Review of financial studies* 3.4 (1990): 547-572.

Koedijk, Kees G., and Philip A. Stork. “Should we care? Psychological barriers in stock markets.” *Economics Letters* 44.4 (1994): 427-432.

Latane, Henry A., and Richard J. Rendleman. “Standard deviations of stock price ratios implied in option prices.” *The Journal of Finance* 31.2 (2012): 369-381.

Lintner, John. “The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets” Review of Economics and Statistics, February 1965.

Macbeth, James D. and Larry J. Merville. “An Empirical Examination of the Black-Scholes Call Option Pricing Model” Journal of Finance, Dec. 1979.

Malkiel, Burton G. “Reflections on the Efficient Market Hypothesis: 30 Years Later” The Financial Revew, 2005.

Merton, Robert C. “Option Pricing When Underlying Stock Returns are Discontinuous” Journal of Financial Economics, Jan.-March 1976.

Merton, Robert C. “Theory of rational option pricing.” *The Bell Journal of Economics and Management Science* (1973): 141-183.

Merton, Robert C., and Paul A. Samuelson. “Fallacy of the log-normal approximation to optimal portfolio decision-making over many periods.” *Journal of Financial Economics* 1.1 (1974): 67-94.

Mossin, Jan. “Equilibrium in a Capital Asset Market” Econometrica. October 1966.

Ncube, Mthuli. “Modelling implied volatility with OLS and panel data models.” *Journal of Banking & Finance* 20.1 (1996): 71-84.

Petersen, Mitchell A. “Estimating standard errors in finance panel data sets: Comparing approaches.” *Review of financial studies* 22.1 (2009): 435-480.

Roll, Richard. “An Analytic Valuation Formula for Unprotected American Call Options on Stocks with Known Dividends” Journal of Financial Economics, November 1977.

Rubinstein, Mark. “Nonparametric tests of alternative option pricing models using all reported trades and quotes on the 30 most active CBOE option classes from August 23, 1976 through August 31, 1978.” *The Journal of Finance* 40.2 (2012): 455-480.

Samuelson, Paul A. “Proof that properly anticipated prices fluctuate randomly.” *Industrial management review* 6.2 (1965).

Schmalensee, Richard and Robert Trippi. “Common Stock Volatility Expectations Implied by Option Premia” Journal of Finance, March 1978.

Scott, Louis O. “Option pricing when the variance changes randomly: theory, estimation, and an application.” *Journal of Financial and Quantitative analysis* 22.04 (1987): 419-438.

Sonnemans, Joep. “Price clustering and natural resistance points in the Dutch stock market: a natural experiment.” (2003): 1.

Stein, Jeremy. “Overreactions in the options market.” *The Journal of Finance* 44.4 (1989): 1011-1023.

Thorp, Edward O. “Extensions of the Black-Scholes option model.” *Proceedings of the 39th Session of the International Statistical Institute, Vienna, Austria* (1973): 522-29.

Tobin, James. “Liquidity Preference as Behavior Towards Risk” The Review of Economic Studies, Vol. 25, No. 2 (Feb.,1958), pp. 65-86.

Trippi, Robert. “A test of option market efficiency using a random-walk valuation model.” *Journal of Economics and Business* 29.2 (1977): 93-98.

Wiggens, James. “Option Values Under Stochastic Volatility: Theory and Empirical Estimates” Journal of Financial Economics, December 1987.

Welfare Incentives and Interstate Migration: An Analysis of the Migration Decisions of Poor, Single Mothers

John Weis

Abstract
The purpose of this paper is to examine the role of welfare incentives in the decision to move for poor, single mothers. Using micro-level data from the Panel Study of Income Dynamics (PSID) and other sources, I develop an econometric model that estimates the influence of state welfare benefits on the interstate migration decisions of poor, single mothers, whether that be moving from states with low benefits or to states with high benefits. This study builds upon previous literature concerning interstate migration by considering new methodological approaches and theoretical models. Ultimately, the evidence suggests that while the welfare benefits offered at the current state of residence and those at potential states of residence influence the migration decisions of poor, single mothers, the effects are modest.

I. Introduction

Under The Personal Responsibility and Work Opportunity Reconciliation Act of 1996 (PRWORA), Temporary Assistance for Needy Families (TANF) replaced Aid for Families with Dependent Children (AFDC) as a primary federal cash assistance program. The provisions of TANF obligated states to individually develop their own welfare systems. As a result, there was a rapid and substantial increase in the diversification of welfare policies between states relative to the past (Blank, 2002). These diversified welfare rules and benefit levels potentially provide greater incentives for poor families to move to those states offering greater benefits levels in order to receive more financial aid (De Jong, Graefe, and St. Pierre, 2005). Of those affected by welfare incentives, poor, single mothers comprise the subgroup of the population most likely to be influenced by greater welfare benefits given their comparatively difficult situation (Levine and Zimmerman, 1999). Thus, if a welfare-induced migration phenomenon does exist, it is most likely present in the migration decisions of poor, single mothers.

As such, the purpose of this paper is to investigate the migration decisions of poor, single mothers by estimating to what extent they move from low welfare benefit states to high welfare benefit states.

The economic significance of welfare-induced migration has been well-documented. Assuming the poor can afford the cost of moving, states with lax welfare requirements or large benefit payments could experience considerable influxes of populations largely comprised of poor families (Gelbach, 2004). Having to bear the burden of the additional poor population has concerned states for decades (Meyer, 2000). In response, setting lower benefit levels or stricter eligibility requirements for a state can disincentivize poor families from moving there (Cushing, 2002). However, multiple states attempting to accomplish the above can lead to a 'race to the bottom' wherein states compete to export their poverty burden by offering progressively lower welfare payment benefits (Brueckner, 2000). Evidence of the efforts by states to outcompete other states to avoid poverty populations has been well-documented (see Figlio et al. (1998) and Saavedra (1998)). The efforts by state policymakers have even led to several prominent Supreme Court cases (*Shapiro v. Thompson* [1969] and *Saenz v. Roe* [1999]) in which residency requirements for welfare payment were deemed unconstitutional (Gelbach, 2004). Overall, the results of the economic policy and changes to welfare systems enacted by states have tangible impacts on the well-being of those individuals and families living in poverty. Since the lives of individuals are dependent upon state-specific objectives influenced by welfare-induced migration concerns, the economic significance of welfare-induced migration is clear.

My interest in the research subject developed from two primary sources of personal significance. Firstly, as will be developed later, the literature on the topic is largely inconclusive. This represents an opportunity for someone to make

a significant contribution and have a lasting impact on the discourse surrounding the issue. Additionally, because the research is uncertain, any significant research could have substantial policy implications for state policymakers. Secondly, I have always had a general interest in welfare policy and the effects of said policy (in this case potentially adverse effects). While the purpose of the paper is largely exploratory in nature, exposing potential shortcomings of welfare policy, it also serves as an attempt to promote better practices for policymakers by encouraging a deeper consideration of the consequences of state policies. The hope is that this research paper may, in some way, generate progress in the form of more appropriate policy.

As stated earlier, this paper will investigate the degree to which poor, single mothers move from low welfare benefit states to high welfare benefit states. In the next section, a brief survey of relevant welfare migration literature will be provided. Afterwards, using the theoretical frameworks provided by recent welfare-induced migration research, an econometric model will be developed to analyze the research question from an alternative perspective. In the section that follows, the description of data and data collection will be presented. In the penultimate section, the empirical results of the econometric model will be discussed. Lastly, I will conclude with a brief summary of the study's findings and implications for future research.

II. Literature Review

In developing the Tiebout-Tullock hypothesis of local public expenditures, Tiebout (1956) became one of the first individuals to discuss, albeit indirectly, the potential for welfare-induced migration. While the focus of the original paper was on the role that the mobility of workers played in promoting an efficient allocation of public goods, underlying ideas and concepts are directly

applicable to research on welfare-induced migration. Specifically, the concept that individuals moved to states or regions which offer a bundle of goods that maximize their utility has served as a primary motivating factor in the decision to move.

According to Moffitt (1992), early researchers in the 1970s only had access to aggregated data, making it impractical to identify poverty populations. They were forced to rely on race as a proxy variable for poverty. Furthermore, they failed to account for important state-level differences, such as unemployment rates, in their analyses of the migration decision-making process for families. Therefore, research in the 1970s is generally discounted (Cushing, 2002).

Meanwhile, the research of the 1980s remedied shortcomings of previous work with the aid of microdata which gave them the ability to directly observe poverty populations. In this period, Gramlich and Laren (1984), Friedli (1986), Blank (1988), Cebula and Koch (1989), Peterson and Rom (1989), and Dye (1990) all conducted studies that found strong evidence for the existence of a welfare effect on the decision to move for poor families. Most of the work at this time made use of the newly available Public Use Microdata Series (PUMS) to examine welfare-induced migration. Compelling evidence that suggest opposing conclusions during the same time period is relatively scarce, if not entirely nonexistent.

Of course, if there was such an overwhelming consensus today, I would be doing research on a different topic. In fact, the research conducted since the early 1990s has witnessed a significant split in general opinion. Hanson and Hartman (1994), Borjas (1998), Gelbach (2004), and Snarr (2011) all reached conclusions supportive of a substantial influence of welfare benefits on the decision to move. Conversely, Walker (1994, 1995), Allard and Danziger (2000), Kennan and Walker (2010) failed to find any evidence of a welfare effect on the

decision to move. These two groups represent different extremes with most of the research reaching conclusions that lie somewhere in between these two sides. Specifically, Frey et al (1996), Enchautegui (1997), Schram, Nitz, and Krueger (1998), Levine and Zimmerman (1999), Meyer (2000), Kaestner, Kaushal, and Van Ryzin (2003), Gelbach (2004), and McKinnish (2005, 2007) have all found results suggesting there is welfare-induced migration but that this effect is fairly modest. Unfortunately, direct comparisons are difficult to establish given the drastically different methods employed in literature. As suggested by Snarr (2011), the reason for such large discrepancies in the literature is that there is no strong theoretical justification for any particular model or research design, the literature is largely empirical. Therefore, all of the research is justifiable in its own right, insinuating that more effort has to be taken in establishing a stronger theoretical foundation or basis for the field so that a consensus may be reached.

The principal difference between the empirical studies lies in their research methods, particularly how they decide to measure the decision to migrate between states, and how they decide to model such a decision. For instance, Walker (1994) and Levine and Zimmerman (1999), like early empirical studies, examine population flows into states (inmigration) and out of states (outmigration) for different groups. Meyer (2000) decides to compare differences in welfare participation rates between those who migrate to another state and those who do not. McKinnish (2005, 2007) studies the differences in welfare expenditures at the county-level between border counties of adjacent states. Blank (1988) adopts a less conventional and more time-consuming choice model of location which develops a more theoretical foundation than most research in the literature. Unlike prior work, Frey et. al (1996) treats the decision to move as a sequential decision-making process where individuals first decide if they are going to migrate and then where to migrate using a nested logit framework. While some of these works

reach similar conclusions, the only other thing they have in common as that they all adopt drastically different fundamental approaches to the research question. This notion is consistent for the entire literature of welfare-induced migration.

The second major difference between the studies in the literature that contributes to their differing conclusions is the choice of data employed by the researchers. As with the research methods, there is no strong theoretical foundation supporting any particular data set, especially considering that the choice of data is dependent on the methodology employed. For example, Levine and Zimmerman (1999) and Kennan and Walker (2010) make use of data from the National Longitudinal Study of Youth (NLSY) for examining young, female subpopulations. Breuckner (2000) and Kaestner, Kaushal, and Van Ryzin (2003) use TANF-era data while Schram, Nitz, and Krueger (1998) and Gelbach (2004) utilize the Public Use Microdata Series derived from the U.S. Census. Allard and Danziger (2000) use the Panel Study of Income Dynamics (PSID). Despite all of the different data sets, the U.S. Census is perhaps the most commonly used data source. Enchautegui (1997) uses data from the 1980 U.S. Census, Meyer (2000) uses the 1980 and 1990 U.S. Census data, and Frey et. al (1996) uses the 1990 U.S. Census data. Again, the differences in Censuses used are largely dependent on the individual study and how they decided to examine welfare-induced migration. Ultimately, the differences in methodology and data combine to form two of the prominent factors leading to the general inconclusive nature of the literature.

III. Research Methods and Modeling

Given the preponderance of logit and probit models in the literature[4], the econometric model developed in this study follows the probit framework. The primary advantage of this model over other probabilistic models, specifically the

4 Blank (1988); Frey et. al (1996); Enchautegui (1997); Levine and Zimmerman (1999); Davies, Greenwood, and Li (2001); Cushing (2003); Bailey (2005); Gelbach (2004).

Linear Probability Model (LPM), is that the effect of explanatory variables (e.g. welfare benefit levels) can estimated to dynamically influence the likelihood of an event. Additional benefits include imposing limits on the likelihood of an event (i.e. the probability of an event cannot be greater than 1 or less than 0).

The dependent variable measures whether or not the poor, single mother family moved from one state to another. The explanatory variables in the model are the level of welfare benefits for a family of four offered in the current state of residence as well as the level offered in the previous state of residence. These variables encapsulate the potential welfare effect. The control variables are established by previous literature[5]. I estimate the following probit model:

$$PR(MIGRATED) = \Phi (B_0 + B_1HDAGE + B_2HDEDUC + B_3HDEXPER + B_4NOKIDS + B_5AGEYNG + B_6BLACK + B_7RACEOTHER + B_8WELFLAST + B_9MAXBEN + B_{10}MAXBENLAST + B_{11}UNEMP + B_{12}UNEMPLAST + B_{13}DISTANCE)$$

where

AGEHD = Age of head of household

HDEDUC = Education of head of household

HDEXPER = Experience of head of household

NOKIDS = Number of kids

AGEYNG = Age of youngest child

BLACK = 1 if head of household is black, 0 otherwise

RACEOTHER = 1 if head of household is non-white and non-black, 0 otherwise

WELFLAST = 1 if family was on welfare last year, 0 otherwise

MAXBEN = Maximum welfare benefits for a single-parent family of four in current state

MAXBENLAST = Maximum welfare benefits for a single-parent family of four in previous state

UNEMP = Unemployment rate in current state

UNEMPLAST=Unemployment rate in previous state

DISTANCE = Log of population-weighted greater area circle distance between previous state and contiguous state with highest welfare benefit offering

5 Rogers (1968); Todaro (1969); DaVanzo (1978); Tienda and Wilson(1992); Enchautegui (1997)

Using the previous literature, I generate expectations for the signs of dependent variables in ***Table 1*** below.

Table 1

Variable	Sign
HDAGE	Negative
HDEDUC	Negative
HDEXPER	Negative
NOKIDS	Negative
AGEYNG	Negative
BLACK	Negative
RACEOTHER	Negative
WELFLAST	Negative
MAXBEN	Positive
MAXBENLAST	Negative
UNEMP	Negative
UNEMPLAST	Positive
DISTANCE	Negative

Briefly discussing the expected signs of the coefficient estimates, one would expect that being older, having more kids, and having older kids would all decrease the likelihood of migrating from a convenience standpoint. Being black or another non-white and non-black race would also potentially decrease the likelihood of moving for these families as they generally have less income and, therefore, an inability to afford moving. We would also expect greater education and experience for the head of household to decrease the likelihood of moving because they are aware of the negative impacts moving can have on a family, especially the children. Being the focus of the research question, we would expect that greater welfare benefits of the previous state would have a negative influence on the likelihood of moving (i.e. they wouldn't need to move) and that greater welfare benefits in the current state would have positive influence on the likelihood of having moved (i.e. they moved to get the benefits). The unemployment signs can be interpreted similarly.

IV. Data

1. Panel Study of Income Dynamics (PSID) 1968 -2007

The primary data set used in this econometric analysis will be the Panel Study of Income Dynamics (PSID) from 1968 to 2007. Developed by the Institute for Social Research (ISR) at the University of Michigan in 1968, the PSID is a longitudinal data set comprised of both individual and family level data

collected for a variety of variables every year. While the previous discussion on the uncertainty of data selection in the literature suggested that any sufficient data set can be justified, the U.S. Census seemed to be the most universally accepted. The justification for the decision to use the PSID instead of the U.S. Census is twofold. Firstly, PSID data was readily available and easy to manipulate to generate the necessary sample data. Secondly, migration in the U.S. Census data is indicated by a person living in a different state than five years prior. In effect, migration in each U.S. Census would only account for moves for every five years. In contrast, because the data comprising the PSID is collected every year, one is able to find yearly migration patterns. This allows for a more dynamic analysis than the U.S. Census could provide.

An additional fact to note, the PSID sample data used in the analysis was not the raw data but rather a subset of the data. More precisely, the sample PSID data is comprised of only poor, single females who have been heads of household at some point in their life. This was the best representative sample for examining the decision of poor, single mothers to move. Assuming the decision to move is largely determined by the head of household, being a head of household would be a prerequisite for a mother to be able to make the decision to move. While there are potential issues of specification bias involved with using a specific subset, for the sake of answering the research question, it is necessary. Overall, the PSID data set accounts for all variables in the model except for unemployment rates, welfare benefits, and the distances between states.

Most of the variables from this data set included in the econometric model were used in their raw form. However, the dummy variables for *black* and *raceother* were derived using the variable for race, with white serving as the omitted condition.

2. Bob Moffitt Welfare Guarantee Variables

The first supplemental data set was independently developed by established scholar Bob Moffitt and aggregates a variety of data from a wide array of sources. With respect to this analysis, the data set contains information on unemployment rates of states and the maximum payment benefit for a single-parent, family of four from 1968 to 1996. As a note, the identification of welfare benefits in the model was at the state level instead of the county level, even though welfare benefits vary across counties for some states. While county level information was available for individuals in the PSID, there was no clear procedure to assign the welfare benefits to the counties across time using the Moffitt dataset and the HHS dataset, described below. As such, the welfare benefits levels used for states with varying welfare benefit levels across counties was that of the county with the greatest proportion of the total state population and, therefore, the most likely destination for a mover.

Available at www.econ2.jhu.edu/people/moffitt/ben_doc.pdf

3. U.S. Department of Health and Human Services

The second supplemental data set was developed by the Urban Institute for the Welfare Rules Database. The data set contains maximum payment benefits for a single-parent family of four data from 1996 to 2007.

Available at http://anfdata.urban.org/WRD

4. U.S. Bureau of Labor Statistics

The third supplemental data set used consisted of the historical tables of average state unemployment rate for each state from 1996 to 2007.

5. Greater Circle State Distances

The final supplement data provides data on the log of the population-weighted greater circle distance between any two states. The population weights are used to determine population centers of states while the greater circle

measures incorporate the curvature of the earth in the calculation of distances between these population centers. In the context of this analysis, the *distance* variable measures the distance between the state of residence in the previous year and a potential state, determined by finding the contiguous state with the greatest welfare benefits offerings. This variable will serve as a proxy variable for the moving costs associated with interstate migration.

Descriptive statistics of the dependent variables, derived and described previously, were also generated as shown in ***Table 2*** below.

Table 2

Variable	Mean	Std. Dev.	Min	Max
HDAGE	31.20968	11.60559	0	99
HDEDUC	6.905973	9.568182	0	99
HDEXPER	17.61352	33.955	0	99
NOKIDS	1.596788	1.523294	0	11
AGEYNG	3.516814	4.255007	0	17
BLACK	0.752964	0.4313111	0	1
RACEOTHER	0.0472085	0.2120959	0	1
WELFLAST	0.3798314	0.4853713	0	1
MAXBEN	385.2062	179.7597	60	1064
MAXBENLAST	380.7177	177.7997	55.02	1064
UNEMP	6.542549	2.066404	2.1	15.5
UNEMPLAST	6.541871	2.105017	2.1	15.5
DISTANCE	4.583395	3.234206	0.384506	11.2152

Given prior knowledge, none of the variables seem particularly strange, although 99 years for the age of head, education and experience is relatively concerning. Further inspection of the data reveals that there are only a few observations with this problem, mostly from the earlier years of the dataset when there was likely more measurement error.

V. Empirical Results

Estimating the econometric model, I obtained the regression output shown in ***Table 3*** below.

Table 3

Variable	Coefficient	t-Score	*P*-value	Marginal Effects
HDAGE	-0.038255	-5.47	0.000*	-0.0016894
HDEDUC	-0.0013423	-0.31	0.759	-0.0000593
HDEXPER	0.0047558	1.81	0.070**	0.00021
NOKIDS	-0.0741016	-1.69	0.091**	-0.0032724
AGEYNG	-0.0229523	-1.52	0.129	-0.0010136
BLACK	-0.131427	01.06	0.291	-0.006228
RACEOTHER	-0.0319729	-0.15	0.882	-0.0013707
WELFLAST	-0.5434885	-4.35	0.000*	-0.024173
MAXBEN	0.0012908	2.36	0.018*	0.000057
MAXBENLAST	-0.0013529	-2.42	0.015*	-0.0000597
UNEMP	-0.2134598	-3.26	0.001*	-0.0094266
UNEMPLAST	0.1289763	2.05	0.040*	0.0056957
DISTANCE	0.019423	1.25	0.212	0.008577
N = 3001	**Prob > chi2(13) = 0.0000**	**Psuedo R2 = 0.1567**	**Log likelihood = -375.42839**	

Before interpreting the regression output for the probit regression, it was important to run an alternative regression model (LPM) to allow for the testing of common regression issues. Issues of heteroskedasticity were present in the model, namely in the *hdage*, *hdeduc*, and *hdexper* variables, as evidenced by the Breusch-Pagan Test and White's Test for heteroskedasticity and accompanying graphical analyses. In response, one could employ robust standard errors. Additionally, problems of multicollinearity where found, using variable inflation factors, particularly between *unemp* and *unemplast*, and *maxbenlast* and *maxben.* This is unsurprising since these variables are strongly collinear (only approximately 5

percent of the total observations actually migrate to another state). Since dropping or transforming these key variables would negate the analysis hitherto, nothing will be done, although one should remain mindful of the negative consequences of multicollinearity. Lastly, we find evidence of model misspecification using the Ramsey RESET test. The most obvious functional form issue lies in the omission of a squared term for *distance* whose addition may improve the model as the effect of *distance* on the likelihood of moving may increase at a decreasing rate. For further information on these tests, refer to the ***Appendix***.

Enacting the remedies for these regression issues, most notably the addition of a distance-squared term, I generated the regression output depicted in ***Table 4*** below.

Table 4

Variable	Coefficient	t-Score	*P*-value	Marginal Effects
HDAGE	-0.0391325	-5.52	0.000*	-0.0016446
HDEDUC	-0.0019966	-0.45	0.656	-0.0000839
HDEXPER	0.0051451	1.95	0.051**	0.0002162
NOKIDS	-0.0814076	-1.82	0.068**	-0.0034214
AGEYNG	-0.0225078	-1.46	0.143	-0.000946
BLACK	-0.0696176	-0.55	0.582	-0.0030258
RACEOTHER	0.0122366	0.06	0.955	-0.0005202
WELFLAST	-0.5492922	-4.36	0.000*	-0.0232821
MAXBEN	0.0012441	2.28	0.023*	0.0000523
MAXBENLAST	-0.0010184	-1.79	0.074**	-0.0000428
UNEMP	-0.220425	-3.36	0.001*	-0.009264
UNEMPLAST	0.1242484	1.97	0.049*	0.0052219
DISTANCE	0.2361678	3.07	0.002*	0.0099256
DISTANCESQ	-0.017081	-2.89	0.004*	-0.0007179
N = 3001	**Prob > chi2(13) = 0.0000**	**Psuedo R2 = 0.1666**	**Log likelihood = -371.02267**	

All of the coefficient estimates with a single asterisk were found to be statistically significant at the 0.05 significance level while those with a double asterisk were found to be statistically significant at the 0.10 significance level.

All of the signs of the coefficient estimates on the variables were found to be in accord with prior expectations except for the sign of the coefficient estimate on *hdexper*, which was presumed to have the same sign as *hdeduc* (i.e. positive). The exact cause of this discrepancy is difficult to discern, but *hdeduc* and *hdexper* moving in opposite directions does not seem entirely intuitive. Overall, all of the coefficient estimates were found to be statistically significant at the 0.10 significance level except for those on the education of the head of household, the race dummies, and the age of the youngest child. However, these coefficient estimates being statistically insignificant does not necessarily impact the research question in any substantial manner.

Interestingly, we find that coefficient estimate on *maxben* is significant at the 0.05 significance level while the coefficient estimate on *maxbenlast* is significant at the 0.10 significance level. This suggests that, given the sign of these coefficients, as the benefit level of the previous state decreases, the probability of moving increases and that as the benefit level of the current state increases, the probability of having moved increases. Taken in conjunction, these results suggest that the population of poor, single mothers move from states with low welfare benefits to states that offer high welfare benefits. In effect, this is evidence that the phenomenon of welfare incentives may exist for poor, single mothers, incentivizing them to move to other states in order to receive higher benefits.

Further analysis of the magnitude of the marginal effects of the coefficient estimates is necessary to determine whether the welfare effect is economically significant. In order to accomplish this analysis, I use the marginal effects generated by the regression analysis to estimate the effect of certain benefit levels on the likelihood of moving to another state for poor, single mothers. More precisely, I develop a general range of effects for both benefit variables, ranging from two standard deviations below the respective mean benefit level to

two standard deviations above said mean. With respect to the *Maxben* variable, I calculate a predicted increase in the likelihood of moving for poor, single mothers ranging from 0.134 percent to 3.895 percent at benefit levels of 25.69 dollars and 744.73 dollars, respectively. Similarly, with respect to the *Maxbenlast* variable, I calculate a predicted decrease in the likelihood of moving for poor, single mothers ranging from 0.108 percent to 3.151 percent at benefit levels of 25.12 dollars and 736.32 dollars, respectively. Thus, the role of welfare benefits in the migration decisions of poor, single mothers ranges from fairly inconsequential to somewhat substantial. Still, even at their most substantial, the influence on the decision to move is relatively minimal. Overall, these results largely agree with the majority of the prior literature on welfare incentives which found a statistical significant but economically modest role for welfare incentives.

Additionally worth discussing, adding the *distancesq* term to the regression analysis made both the coefficient estimate on it and *distance* statistically significant at even the 0.01 significance level. Clearly, being as the purpose of these variables was to control for the cost of moving, one would imagine that they would play a significant role, even if they were not the best proxy for moving costs. Taken in conjunction, these variables suggest that as the distance between the current state and a potential state with greater benefits increases, the probability of moving increases but at a decreasing rate. This does not make theoretical sense as, presumably, greater distances would decrease the likelihood of moving because of greater moving costs. A better proxy for moving costs, given the likely measurement error in this variable, may alleviate this inconsistency.

Furthermore, we find that the coefficient estimates on the unemployment variables were statistically significant at the 0.05 significance level. This evidence suggests that employment opportunities play an important role in the decision to

role for poor, single mothers. Given their difficult situation, greater employment opportunities, as measured by the unemployment rate of states, would reasonably provide incentives for moving. Even so, the marginal effects of these coefficient estimates are relatively small when considering that a one percentage increase in the unemployment rate of the previous state would increase the likelihood of moving by less than 1 percent, as predicted by the model.

Given the diagnostic tests run, there is some concern that these results may not be valid but, beyond basic amendments, any significant changes to the model may itself invalidate the ability of the model to address the research question. While issues of heteroskedasticity and multicollinearity are concerning, however, they would not necessary nullify any of the conclusions of the model, but they do suggest that the probit framework may not be the ideal model for answering the research question, at least in the incarnation used in this analysis. Recently, the literature on welfare incentives in migration decisions has begun to adapt optimal choice models wherein an individual must consider multiple locations across the country before deciding where to move. Comparatively, the model used in this analysis was relatively simplistic, which may not be ideal given that the decision to move can be very complex. Furthermore, these issues may arise as a result of poor identification of explanatory variables, although the problem does not seem as severe as it was prior to several amendments.

VI. Conclusions

In order to investigate how the migration decisions of poor, single mothers migrate were influenced by the welfare benefits offered by states, I chose to estimate a basic econometric probit model. Using prior research to determine relevant control variables and data from a variety of sources, I tested my model and found evidence for a modest influence of welfare incentives on the decision

to move for poor, single mothers. Most of variables included in the model were found to be statistically significant with appropriate signs at the designated significance level, suggesting that poor, single mothers do move from states with low welfare benefits to those with high welfare benefits.

These results are generally in accord with the majority of the literature on welfare incentives and interstate migration, which finds modest roles for welfare incentives in the decision to move. The policy implications of this study, and those that found similar conclusions, are vast. Specifically, the foundation of a 'race to the bottom' between states is largely unfounded if welfare incentives do substantially influence decisions to move. As such, the only effect of competing to offer lower benefits may be to simply worsen the standard of living for poor populations rather than export those populations. Ultimately, while state legislators should consider welfare incentives when changing welfare policy, it should not be a primary consideration. Further research into the 'race to the bottom' phenomenon, how much state legislators consider welfare incentives when crafting policy, and the potential impacts of the latter on the poverty population are all advents for further research. Within the scope of the literature, further consideration should be given towards developing a more theoretical foundation so that a greater consensus in the literature can be obtained, at the very least establishing theoretical preference to certain datasets or methodologies. Otherwise, it will remain impractical to compare the conclusions of these widely differing analyses. In addition to further consideration to these topics, the econometric model may be improved by utilizing an optimal choice framework, better indentifying explanatory variables, and obtaining better measurements for certain variables such as moving costs. If done, one may be able to obtain even better evidence for the moderate role of welfare incentives in the decision to move, especially as relates to poor, single mothers.

References

Allard, Scott and Sheldon Danziger. 2000. "Welfare Magnets: Myth or Reality?" *Journal of Politics*, 62: 350-368.

Bailey, M.A. 2005. "Welfare and the Multifaceted Decision to Move." *American Political Science Review*, 99(1): 125-135.

Blank, Rebecca. 1988. "The Effect of Welfare and Wage Levels on the Location Decisions of Female-Headed Households." *Journal of Urban Economics*, 24: 186-211.

Blank, Rebecca. 2002. "Evaluating Welfare Reform in the United States." *Journal of Economic Literature*, 40: 1105-1166.

Borjas, George 1998: "Immigration and Welfare Magnets." NBER Working Paper No. W6813,

Brueckner, Jan. 2000. "Welfare Reform and the Race to the Bottom: Theory and Evidence." *Southern Economic Journal*, 66: 505-525.

Cebula, Richard and James Koch. 1989. "Welfare Policies and Migration of the Poor in the United States: An Empirical Note." *Public Choice*, 61: 171-176.

Cushing, Brian. 2002. "The Role of Welfare and Space in the Migration of the Poor." *Regional Research Paper 2002-16*, Regional Research Institute, West Virginia University.

DaVanzo, Julie. 1978. "Does Unemployment Affect Migration? Evidence from Micro Data." *Review of Economics and Statistics*, 60: 504-514.

Davies, Paul, Michael Greenwood, and Haizheng Li. 2001. "A Conditional Logit Approach to U.S. State-To-State Migration." *Journal of Regional Science*, 41: 337-360

De Jong, G., D. Graefe, and T. St. Pierrre 2005. "Welfare Reform and Interstate Migration of Poor Families." *Demography*. 42: 469-97.

Dye, Richard E, and Therese J. McGuire. 1998. "Block grants and the sensitivity of state revenues to recession." *Proceedings of the National Tar Association,* 15-23.

Enchautegui, Maria E. 1997. "Welfare Payments and Other Economic Determinants of Female Migration." *Journal of Labor Economics* 15(3): 529-554.

Figlio, David N., Van W. Kolpin, and William Reid. 1997. "Do states play welfare games?" *Journal of Urban Economics.*

Frey, William, Kao-Lee Liaw, Yu Xie, and Marcia Carlson. 1996. "Interstate Migration of the U.S. Poverty Population: Immigration 'Pushes' and Welfare Magnet 'Pulls'." *Population and Environment*, 17: 491-536.

Snarr, Hal W., Dan Friesner and Mark L. Burkey. 2011. "Unintended migration consequences of US welfare reform." *Economic Analysis and Policy* 41(3): 233-251

Gelbach, Jonah. 2004. "Migration, the Lifecycle, and State Benefits: How Low is the Bottom?" *Journal of Political Economy*. 112(5), 1091-1130.

Gramlich, Edward M., and Deborah S. Laren. 1984. "Migration and Income Redistribution Responsibilities." *The Journal of Human Resources*, 19: 489-511.

Kaestner, R., N. Kaushal, and G. Van Ryzin. 2001. "Migration Consequences of Welfare Reform." Working Paper 8560. National Bureau of Economic Research, Cambridge, MA

Kennan, J., and J. R. Walker. 2010. "Wages, Welfare Benefits and Migration." *Journal Of Econometrics* 156(1): 229-238.

Levine, P. B. and D. J. Zimmerman. 1999. "An Empirical Analysis of the Welfare Magnet Debate Using the NLSY." *Journal Of Population Economics* 12(3): 391-409.

McKinnish, Terra. 2005. "Importing the Poor: Welfare Magnetism and Cross-Border Welfare Migration." *The Journal of Human Resources*, 40: 57-76.

McKinnish, Terra. 2007. "Welfare-Induced Migration at State Borders: New Evidence From Micro-data." *Journal of Public Economics*, 91: 437-50.

Meyer, B. 2000. "Do the poor move to receive higher welfare benefits?" Unpublished manuscript.

Moffitt, Robert (1992): "Incentive Effects of the U.S. Welfare System: A Review." *Journal of Economic Literature*, 30: 1-61.

Peterson, Paul E., and Rom, Mark C. 1990. "Welfare Magnets: A New Case for a National Standard." Washington, DC: Brookings Institution.

Saavedra, Luz Amparo. 1998. "A Model of Welfare Competition with Evidence from AFDC." Working Paper, University of Illinois.

Schram, S., L. Nitz, and G. Krueger. 1998. "Without Cause or Effect: Reconsidering Welfare Migration as a Policy Problem." *American Journal of Political Science* 42: 210–30.

Tiebout, C., 1956. "A pure theory of local expenditures." *Journal of Political Economy*, 64.

Tienda, Marta, and Wilson, F. D. 1992. ''Migration and the Earnings of Hispanic Men.'' *American Sociological Review*, 57: 661–78.

Todaro, Michael. 1969. ''A Model of Labor Migration and Urban Development in Less Developed Countries.'' *American Economic Review*, 59: 138–48.

Walker, James. 1994. "Migration Among Low-Income Households: Helping the Witch Doctors Reach Consensus," Working paper.

Walker, James. 1995 "Low Income Migration During a Decade of Impoverishment." Working paper.

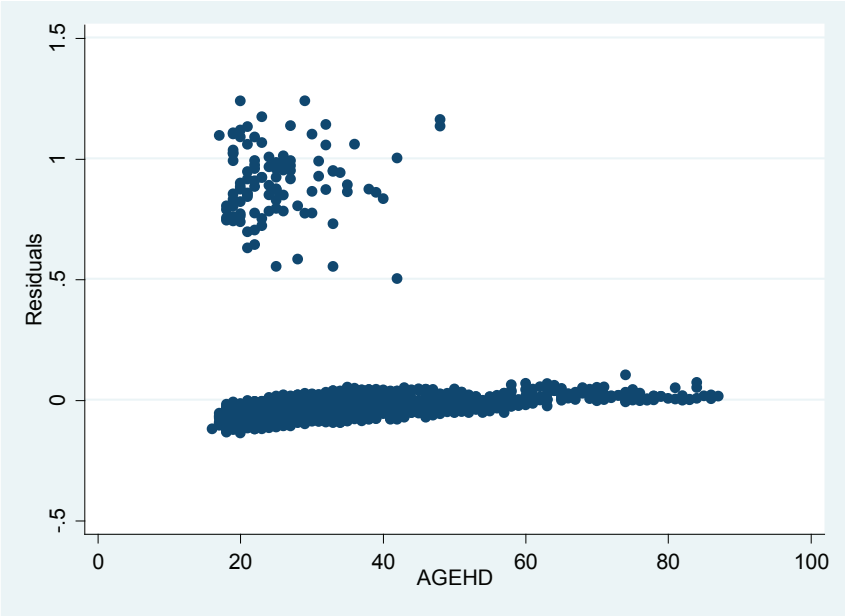
Appendix

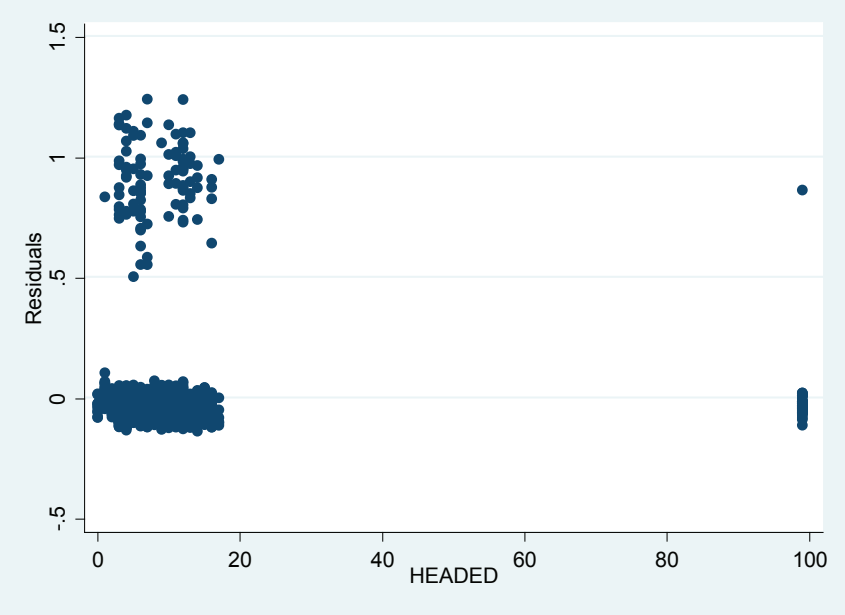
Tests for Heteroskedascity

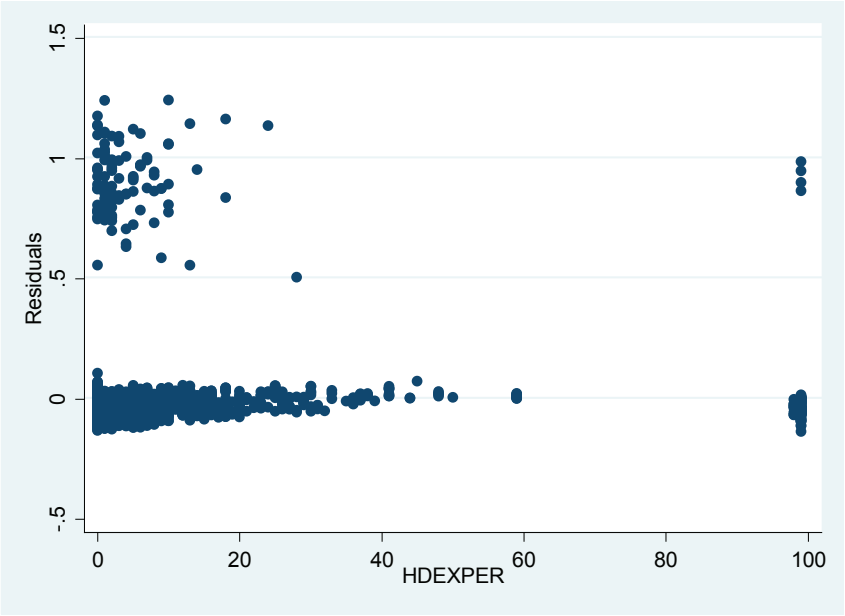
```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of chgstate
         chi2(1)      =   200.41
         Prob > chi2  =   0.0000


White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity
         chi2(100)    =   1814.64
         Prob > chi2  =    0.0000
Cameron & Trivedi's decomposition of IM-test
---------------------------------------------------
              Source |       chi2     df      p
---------------------+-----------------------------
  Heteroskedasticity |    1814.64    100    0.0000
            Skewness |     452.91     13    0.0000
            Kurtosis |      84.66      1    0.0000
---------------------+-----------------------------
               Total |    2352.22    114    0.0000
---------------------------------------------------
```

Graphical Analyses for Heteroskedasticity of Relevant Variables







Test for Multicollinearity

Variable	VIF	1/VIF
UNEMPLAST	27.66	0.036159
UNEMP	27.62	0.036208
MAXBENLAST	23.18	0.043146
MAXBEN	23.15	0.043195
BLACK	1.47	0.682390
WELFLAST	1.28	0.778561
RACEOTHER	1.28	0.782762
NOKIDS	1.27	0.785496
AGEHD	1.17	0.853040
DISTANCE	1.15	0.865918
HDEXPER	1.14	0.876884
AGEYNG	1.09	0.915698
HDEDUC	1.04	0.963181
Mean VIF	8.65	

On the Orthodox Nature of Heterodox Income Distribution Theory

By Ross Nichols

Abstract
The goal of this paper is to show that orthodox and heterodox theories of personal income distribution developed in the mid-twentieth century are effectively identical, despite their claims to the contrary. While segmented labor market theory contends that neoclassical theories of personal income distribution, such as human capital theory, ignore the impact of social institutions on the labor market, human capital theory actually implicitly incorporates them. Social institutions are, therefore, just as important in the orthodox approach to personal income distribution. Yet, while this is the case, the heterodox perspective is valuable because of the stress it places on social institutions, the importance of which is not always explicitly recognized in human capital theory.

Introduction:

A.B. Atkinson titled his 1996 Presidential Address to the Royal Economic Society, "Bringing Income Distribution in from the Cold." His rationale for doing so was that income distribution in the twentieth century was studied mainly through the lens of development economics, and that it therefore neglected the theory behind income distribution itself (Atkinson, 1997, 299). He believed that economists placed too much emphasis on studying the effects of income distribution at the cost of failing to attempt to understand the causes of income distribution. The study of income distribution had diverged from explaining "how the economy works" (ibid., 299). Atkinson urged the need for economics to collaborate with other social sciences to incorporate the importance of social norms into income distribution analysis because "a subject so central to social science as income distribution is one that we [economists] cannot solve on our own, and…a receptiveness to outside ideas [is] a sign of a discipline in

good health" (ibid., 318). Building an adequate income distribution theory thus required an interdisciplinary approach. But what Atkinson, as a representative of contemporary mainstream economics, failed to recognize was that great strides had been made in income distribution theory during the twentieth century. A renaissance had occurred several decades earlier that was of paramount importance to the theory of income distribution. Theories of personal income distribution emerged during the mid-twentieth century.

A study of the rebirth of income distribution theory naturally raises questions concerning the timing of this renaissance. One of the main goals of classical political economy was to explain how income was distributed between classes. During that time, a tension developed between the classical theory of income distribution and its Marxian critique that stemmed from the implications Ricardo and Marx drew from their respective class theories of income distribution. Renewed vigor of theoretical work on income distribution occurred in the mid-twentieth century and led to a seemingly much more decisive split between neoclassical economists and their critics. Whereas the theoretical foundations of Ricardian and Marxian theories of income distribution were similar, heterodox economists argued that the logic underpinning neoclassical theories of income distribution was flawed.

Sahota (1977) found human capital theory to be one of the most complete theories of personal income distribution developed by the neoclassicals. Human capital theory drew closely upon the theory of marginal productivity to explain how income was distributed between individuals rather than social classes. Instead of analyzing income distribution within the context of capitalists and laborers, the focus of income distribution theory shifted to an analysis of how income was distributed between labor and capital as factors of production. Segmented labor market theory was introduced as a neo-Marxian critique of

human capital theory.[6] It was based on the argument that sociopolitical forces sorted workers into distinct and rigid labor markets. Individual incomes were thus largely the product of class relations in the workplace. Segmented labor market theory argued that "the law of one price will not prevail in labor markets, even in the long run" (Rebitzer, 1993, 1411). The rebirth of interest in income distribution thus appeared to generate a wide split between competing theories of personal income distribution. Segmented labor market theory criticized human capital theory for neglecting the role of social institutions in the distribution of personal income.

Closer analysis of human capital theory and segmented labor market theory, however, reveals that these two theories are essentially equivalent. Rebitzer (1993) cites several instances of neoclassical economists incorporating the notion of segmented labor markets into their work, but the relationship between human capital theory and segmented labor market theory is much closer than previous literature implies. Therefore not only is it important to study the rebirth of income distribution in the mid-twentieth as a reminder to contemporary economic discourse of the theoretical work done on income distribution during this time, the mid-twentieth century also produced a heated debate between neoclassicals and neo-Marxists concerning the explanation of the distribution of personal income. And while Rebitzer acknowledges that these opposing views were reconciled to an extent, this paper will show that human capital theory and segmented labor market theory are even more fundamentally similar than Rebitzer suggests. In short, a better understanding of how income distribution theory was "brought in from the cold," and the implications associated with this revival, is needed.

6 Both neo-Marxists and neo-institutionalists advocate forms of segmented labor market theory. While Osterman, Rebitzer and Piore are referenced in the text, they are considered neo-institutionalists. Since the purpose of this paper is to compare segmented labor market theory with human capital theory rather than examine the various forms of segmented labor market theory, the two groups are not differentiated.

The following analysis is divided into six sections. The first section examines the origins of income distribution theory in classical and Marxian economics, and show that classical political economy never attempted to develop a theory of personal income distribution. In the second section, I discuss how Clark's theory of marginal productivity served as a bridge to the development of theories of personal income distribution in the mid-twentieth century. In the third section, I analyze the circumstances that contributed to the rise of theories of personal income distribution and present the theoretical foundations of human capital theory. I also examine the criticisms of human capital theory. The fourth section focuses on the critique of human capital theory developed by radical political economy in an effort to show how segmented labor market theory emerged as an important heterodox explanation of personal income distribution. The fifth section shows that orthodox and heterodox explanations of personal income distribution share close theoretical foundations and are thus effectively identical explanations for the distribution of personal income. The only discernible difference between the two is the emphasis placed on social institutions. In the final section, I briefly summarize my findings.

I. Classical and Marxian Theories of Income Distribution

Writing at the height of classical political economy, Ricardo believed that income distribution was so important to political economy that economics should be defined as "an enquiry into the laws which determine the division of the produce of industry amongst the classes who concur in its formation" (Ricardo, 1951, 278). In short, Ricardo contended that the distribution of income was the machine that drove the economy. Explaining the rate of profit was one of the main goals of Ricardo's *Principles*.

Ricardo argued that "the produce of the earth – all that is derived

from its surface by the united application of labour, machinery, and capital, is divided among three classes of the community" (Ricardo, 1821, v). The profits that accrued to capitalists were the product of capitalist relationships with both landlords and laborers. Ricardo accepted the Malthusian theory of population and its attendant assumption that rapid population growth would drive wages down to the subsistence level. Profits were thus governed by a socially and morally[7] determined level of subsistence. Yet while profits were determined by the subsistence wage level, the subsistence wage and technology level was in turn determined by "the quantity of labour requisite to provide necessaries for the labourers, on that land or with that capital that yields no rent" (ibid., 128). In effect, the profits of capitalists were determined by the extent to which infertile land was being used to grow the crops on which workers subsisted.

There was thus a tendency for the rate of profit "to fall; for, in the progress of society and wealth, the additional quantity of food required is obtained by the sacrifice of more and more labour" (ibid, 120). As society progressed, increasingly less fertile land came under cultivation which subsequently increased the amount of rent collected by landlords. This made food more expensive, forcing capitalists to pay higher wages and keep a smaller portion of their revenue. Furthermore, a high rate of profit attracted outside capital so competition depressed profit rates, but "a fall in the general rate of profits is by no means incompatible with a partial rise of profits" (ibid., 119). Capitalists could therefore experience positive profits despite declining rates of profit due to counteracting influences such as increased demand.

Karl Marx agreed with the general framework of class income distribution theory laid out by Ricardo. He too believed that there was a tendency for the rate of profits to fall and wages to fluctuate around the subsistence level. Non-wage

7 Ricardo argues that the natural wage in a country "essentially depends on the habits and customs of the people" in a given country (Ricardo, 91). That is, the level of sustenance that is acceptable in one country is not necessarily the same in all countries. If a worker lives in a society with high wants, their wage will be higher than if they lived in a society of simpler means.

income was accepted as a residual measurement. Marx, however, thought two important aspects of class income distribution needed to be addressed. First, he differentiated between labor and labor power. Labor power "is to be understood [as] the aggregate of those mental and physical capabilities existing in a human being, which he exercises whenever he produces a use-value of any description" (Marx, 1867, 167). Marx believed that labor power was a commodity sold by laborers. Whereas labor power was the potential to do work, "labour-power in use is labour itself" (ibid., 177). Labor was thus the act of realizing the ability to work. This distinction was important because it is the foundation for his concept of surplus value. Humans had a capacity to work much greater than the amount of labor we need to expend to replenish this potential. Marx referred to the portion of the working day needed to earn enough for subsistence "'necessary' labor time" (ibid., 217). Any labor expended beyond this point generated surplus value for capitalists.

Capitalists therefore paid laborers just enough to ensure their subsistence but extracted an excessively large amount of labor compared to the amount of labor necessary for subsistence. Marx also argued that wages tended towards subsistence, but his rationale for why this occurred was different from Ricardo's. Marx rejected the Malthusian theory of population as an explanation for the tendency of wages to fluctuate around the subsistence level. He instead believed that an "industrial reserve army" of unemployed workers was maintained by capitalists in order to foster competition between laborers and prevent wage increases (ibid., 632). Furthermore, the presence of an industrial reserve army resulted from the historical evolution of capitalism. As capitalism progressed, capitalists accumulated increasingly more capital. The majority of this accumulation was ever more productive physical capital. Demand for variable capital (i.e. labor) therefore grew at "constantly diminishing rate" because it

became a smaller and smaller portion of total capital (ibid., 629). According to Marx, downward pressure on wages was thus not a natural occurrence but rather a phenomenon specific to the historical conditions of the capitalist mode of production.

Classical and Marxian income distribution theories shared similar foundations. While Marx sought to improve upon Ricardian theory, the biggest difference between the two was the nature of capitalism. Ricardo argued that since capitalists would only produce commodities for which there was sufficient demand, capitalists would naturally seek to accumulate capital up to the point where profits equaled zero (Ricardo, 340). A decreasing rate of profit was attributable to higher wages made necessary by the diminishing marginal productivity of agriculture. Accumulation was thus portrayed as a self-regulating phenomenon that ensured stability in capitalism. Ricardo acknowledged the existence of intense competition between capitalists to accumulate capital, but did not believe that capital necessarily became concentrated within an ever smaller group of capitalists.

Marx rejected the notion that accumulation was governed by the rate of profit. He contended that accumulation was necessary for capitalist survival in the capitalist mode of production:

> "the development of capitalist production makes it constantly necessary to keep increasing the amount of capital…and competition makes the immanent laws of capitalist production to be felt by each individual capitalist, as external coercive laws. It compels him to keep constantly extending his capital, in order to preserve it" (ibid., 592).

Accumulation fostered vicious competition between capitalists, the losers in which did not survive. Marx believed that as capital became concentrated in a shrinking group of capitalists, capitalists would exert a greater degree of exploitation of the

working class. Ultimately, this desire for accumulation would contribute to the downfall of capitalism after a critical mass of degraded proletariats was reached, sparking a popular revolt against the capitalists (ibid., 763). While Ricardo asserted that the desire for accumulation could wane peacefully, Marx depicted a much more urgent picture of capital accumulation.

Obstacles to an Analysis of Personal Income Distribution

There are two important reasons why Ricardo and Marx never attempted to develop a theory of personal income distribution. Class conflict caused a rise in class consciousness that hindered the ability of Ricardo and Marx to analyze income on a personal level. It simply did not make sense for Ricardo and Marx to study how personal income distribution when people increasingly identified themselves as members of a class rather than as individuals. This is exemplified in the controversy surrounding the Corn Laws, where capitalists and landlords competed to attain supremacy in the public arena. Whichever class prevailed was effectively able to dictate how income was distributed between classes. Capitalists and landlords viewed themselves as part of a greater class movement rather than individual members of society. In addition, the labor force of the time was relatively more homogenous than it is today. This made it difficult to differentiate workers or treat them as individual economic agents. Ricardo and Marx, at least from an economic perspective, would have had difficulty distinguishing between individual members of a given class.

Ricardo and Marx lived in an era when class conflict was at the center of policymaking in Europe. Class distributions of income reflected economists' perspectives on class conflict. The controversy over the British Corn Laws demonstrated the tension between landlords and capitalists that persisted during the rise of classical political economy. Landlords favored implementing the Corn Laws because protection from foreign competition allowed landlords to charge

higher rents. Capitalists wanted to keep food prices as low as possible because the subsistence level depended upon the price at which workers could feed themselves in order to restore their labor power. Cheaper food meant higher profits for capitalists. While enacting the Corn Laws seemed to favor the landlords and their repeal appeared to cater to the interests of the urban capitalists, class conflict was much more complex than a dispute between capitalists and landlords. Conflict also arose between members of the same social class as class distinctions blurred.

One instance of intra-class tension arose between established landlords and newly landed capitalists. Many established landlords favored legislation that

> "served to reinforce the status of the existing elites of both town and countryside by re-emphasizing the notion that the prosperity of the various classes which composed the same interest group was primarily affected not by one another but by a rival interest group [the urban capitalists]," (Moore, 1965, 544).

The Corn Laws thus strove to maintain the status quo of class relations in British society. That hierarchy survived mainly on the mutual interest among well-established landed and capitalist elites to preserve traditional British social structure. Yet as industry expanded in Great Britain, the formerly mutual interests between rural and urban elites diverged. Capitalists sought cheap food to keep production costs down but established, or hereditary, landlords wanted prices to remain high. Hereditary landlords sought to maintain prosperity through high agricultural prices, but an "arriviste" class of newly landed elites who had made a fortune in the cities strove to reap profits by implementing innovative farming techniques that increased crop production (ibid., 551). Social standing of older landed elites was further undermined by capitalists who purchased land in order to accumulate more capital. As the pace of capital accumulation accelerated and landed capitalists subsequently gained influence, older landlords felt their authority in the public arena begin to wane.

The eventual repeal of the Corn Laws ultimately benefited the capitalist class because foreign competition lowered food prices, but Sir Robert Peel's official justification for presenting the necessary legislation was "to extricate the kingdom from the social dilemmas" that arose from the incessant class disputes in European society that dominated the era (ibid., 560). Thus while the repeal of the Corn Laws had economic implications, it was more focused on achieving social harmony. Peel wanted to encourage the hereditary landed elite to shift their focus from prices to output, and thus develop an entrepreneurial outlook similar to capitalists. Unfortunately, this legislation was misguided in that it "failed to recognize the impossibility of commercializing the status of the landlord without also commercializing the status of the tenant" (ibid., 559). Urban capitalists emerged as the victors from repeal of the Corn Laws because legislators failed to comprehend the nature of the rural hierarchy properly. The controversial nature of the Corn Laws and their repeal was representative of the divide within classes that existed during the rise of classical political economy and the critiques that quickly followed. Since class conflict dominated European society, income distribution was viewed as a class-based issue rather than one pertaining to individuals.

The relative homogeneity of the labor force also inhibited the ability to differentiate between individuals. Reich, Gordon, and Edwards argued that before the era of monopoly capitalism began in 1890, production was governed by the rules of competitive capitalism (Reich, Gordon, and Edwards, 1973, 360). Production was heavily standardized, primarily took place in factories, and involved many simple tasks. Capitalists favored this type of production strategy because strong competition disincentivized the extra expenses associated with the types of specialized training that accompanied monopoly capitalism. Monopoly capitalism was characterized by the production of differentiated good by specialized labor. Furthermore, the dramatic rise in population and movement of unskilled workers

into concentrated urban areas ensured that a large supply of easily substitutable labor was readily available for capitalists to draw from (Brown, 1977, 72).

Increasing class consciousness and a relatively homogenous labor force prevented Ricardo and Marx from examining the distribution of income on a personal level. It did not make sense for them to study personal income distribution when public debates such as those surrounding the Corn Laws were class-based. Individuals were viewed more as members of a social class rather than unique economic agents. The structure of industry during this era made it even more difficult to study society on an individual level. Production often centered on simple, repetitive tasks so labor was easily interchangeable.

II. Marginal Productivity Theory

Ricardo first introduced the marginal principle to economic theory. He argued that as less fertile land came under cultivation, the rent on more fertile land increased (Ricardo, 1821, 60). John Bates Clark endeavored to expand Ricardo's marginal principle in two important ways. In his *Distribution of Wealth*, Clark generalized the principle of substitution to include all factors of production, and proposed that "*the pay of labor in each industry tends to conform to the marginal product of social labor employed in connection with a fixed amount of social capital, as such*" (Clark, 1899, 116; emphasis his). Although Clark articulated an explanation of how the natural rates of profits and wages were determined endogenously, he maintained the classical assumption that the endowments of capital and labor were naturally determined. Consequently, while the theory of marginal productivity inspired later theories of personal income distribution, it is itself more a theory of factor demands than one of income distribution.

Clark developed his theory of marginal productivity as an analogue to Ricardo's explanation of rents. He imagined a "universal field for employment"

that included all workers (ibid, 110). Those who had access to the central field of fertile land or a sufficiently stocked store naturally had access to higher levels of productivity. Workers located farther from this central field had to choose from lower quality employment. In the zone of indifference, employers earned zero profit and thus stopped hiring. Competition ensured that all employers hired employees up to this point. The analogy of a universal field for employment is important because of an important aspect Clark omitted from his discussion: he did not explain how workers were placed throughout the field, or if it was possible for them to move from their initial position. Their position was determined naturally. During the rebirth of income distribution theory in the coming decades, this assumption motivated human capital theory and segmented labor market theory to explain how workers moved within the "universal field of employment."

The theory of marginal productivity also maintained the classical assumption of homogenous labor. Clark acknowledged that skilled workers were more productive than unskilled workers, but argued that all labor could be measured in the same units of labor (ibid., 63). He therefore assumed that all labor could be reduced to a common denominator, which minimized the importance of skill differential. A skilled worker could be replaced by two or more less-skilled workers. Although Clark maintained these important classical assumptions, he departed from the classical theory of income distribution in several important ways.

Foremost among Clark's critiques of classical political economy was the argument that Ricardian economics was an endeavor that "was really studying a static…world with no complete idea of its nature," which he addressed by relaxing the assumption of a static economy (ibid., 69). Clark believed that economic theory needed to reflect the dynamic nature of the world. *The Distribution of Wealth* can thus be viewed as his attempt to complete Ricardo's work. While natural law remained a governing principal in that the productivity of workers was

naturally determined, "social economic dynamics" such as increasing population and wealth, technological innovation, introduction of new products, and labor flow between groups were introduced (ibid., 73). Clark realized that the state of the world in the present was not the same as it was in the past, and that it would also be different in the future. Even though wages naturally tended to their natural level, the dynamic nature of society had to be taken into account. The biggest contribution Clark made to income distribution theory was that the determination of wages adhered to natural law because they were equal to an exogenously determined marginal productivity, while granting that these exogenous forces changed over time. To put it concisely, "what we have to see is how static laws operate in a dynamic state" (ibid., 403).

It is clear that Clark drew upon Ricardo when formulating his theory of marginal productivity, yet there is also evidence that he did more than simply apply the marginal principle to include all factors of production. Ricardo held the proportion of capital to labor constant; his theory of rents, therefore, assumed varying fertility of the soil. Clark varied the factors of production separately (ibid., 163). While Ricardo assumed technology to be unchanging, the interchangeability between labor and capital in *The Distribution of Wealth* implied that technological change was an important aspect of production. This is an important implication because it allowed human capital theory to maintain that labor was both dynamic and highly substitutable, a necessary condition for the assumption of perfect competition in the labor market underpinning human capital theory.

The Distribution of Wealth inspired future work on personal income distribution theory through the introduction of the idea that workers were paid according to their marginal productivity. Clark touted his work as "an inspiring vista for future advances" in economic theory (ibid., 75). What he did not foresee, however, were future attempts to treat social economic dynamics as endogenous

to the explanation of personal income distribution. Relaxing the assumption of natural skill determination served as the impetus for numerous theories of personal income distribution. Whereas Clark took the distribution of factor endowments as given, future work on income distribution sought to explain how marginal productivity was determined. Theories of personal income distribution can be seen as an extension of marginal productivity theory. While marginal productivity closely identified with classical political economy, Clark's work effectively broke the hegemony of class income distribution in economic analysis.

III. Theories of Personal Income Distribution

Circumstances Contributing to the Rise of Theories of Personal Income Distribution

One of the reasons economic thought focused upon theories of personal distribution during the renaissance of income distribution in the mid-twentieth century was the widespread study of the causes of discrimination. Becker's *The Economics of Discrimination* (1957) was one of the pioneering works on the subject. Becker argued that it was possible to conduct an economic analysis of the effects of discrimination because "if an individual has a 'taste for discrimination,' he must act *as if* he were willing to pay something either directly or in form of reduced income, to be associated with some persons instead of others" (Becker, 1971 (1957), 14). Employers associated non-pecuniary costs of production with minority employees. This "taste" resulted from prejudice and ignorance, and varied both temporally and spatially.

Although Becker believed that discrimination in the labor force existed, he also believed the forces creating discrimination were dynamic. The level of discrimination present in the work place could thus change over time. To test this, he measured the change in average occupational position for both whites

and minorities from 1910-1950. Relative occupational position was measured by comparing the income for skilled, semiskilled, and unskilled minorities with the average incomes for their respective white counterparts. Becker found that the relative occupational position of minority workers had remained stable over time (ibid., 140). Discrimination had therefore not decreased. An absolute increase in income for minorities did not necessarily imply an increase in their position relative to whites because the incomes of white employees increased as well.

The Economics of Discrimination served as an impetus for numerous studies examining the impact of discrimination. Rayack (1961) and Gilman (1965) criticized Becker's arguments about the persistence and magnitude of discrimination in the workplace, respectively. Rayack argued that Becker's conclusion of an unchanged level of discrimination towards blacks, as measured by their income relative to whites in the first half of the twentieth century resulted from the erroneous construction of his occupational index (Rayack, 1961, 210). Rayack believed Becker did not account for the fact that blacks were heavily concentrated in semi-skilled and unskilled professions. After generating an occupational index that factored in this characteristic, Rayack showed that income for blacks had, in general, increased more than it had for whites so that by 1957 the occupational position of blacks had increased by 34 percent relative to 1900 (ibid., 211). He also contended, however, that this increase did not reflect a decrease in discrimination. The increase was instead due to increased demand for labor, and any sustained increase in occupational position was "substantially, a function of the tightness of the labor market" (ibid., 214). Becker could thus be correct in his assessment that the level of discrimination in the labor force had remained unchanged throughout the twentieth century and simultaneously incorrect in his belief that the position of blacks in the workplace had improved neither absolutely nor relatively.

Gilman (1965) provided a further critique of *The Economics of Discrimination*. He believed that the impact of discrimination on minority unemployment was significantly smaller than initially thought (Gilman, 1965, 1080). This evidence challenged the notion that minorities were targeted in the hiring-firing process. Gilman drew a conclusion similar to Rayack (1961) by suggesting that discrimination was most evident in wage rigidity. Minorities experienced greater wage rigidity and thus higher unemployment rates because "the greater the pressure in an occupation or region for nonwhite-white wage equality, the greater will be the gap between equilibrium and actual wages, and the greater will be the reduction in employment opportunities for nonwhite relative to white workers" (ibid., 1091). Minimum wage laws and unions keep the actual wage above the equilibrium wage. The greater this disparity, the fewer employment opportunities for minority workers there would be.

Regardless of the extent to which discrimination existed, one reason it endured in the labor market was imperfect information. Gathering information on potential employees was costly, which made it difficult for minorities and females to show that they were equally as skilled as their white male counterparts. Arrow (1971) claimed that minorities and women were paid less than equally skilled white male employees because "skin color and sex are cheap sources of information" (Arrow, 1971, 25). Employers had preconceived notions of the productivity of women and minorities, and imperfect information in the labor market allowed these prejudices to persist. There was less incentive for female and minority workers to make the investments necessary to increase their productivity because no amount of investment could outweigh the cheap information provided by their skin color or gender (ibid., 29). Therefore while a minority or female worker and a white worker could begin with the same productive potential, the latter would be more likely to realize this potential and thus enjoy a better occupational position.

Marxists viewed discrimination differently. While neoclassical economists treated the origins of discrimination as exogenous to the capitalist system of production, Marxists believed that discrimination was perpetuated endogenously. Capitalists implemented various forms of discrimination as a means to prevent camaraderie among workers (Reich, Gordon, and Edwards, 361). Employers exploited ethnicity, race, and sex to ensure competing factions of workers who would not compromise capitalist hegemony. They hired groups of rival nationalities to antagonize each other. Jobs were "race-typed" and women were paid less than men as a means of forcing these workers to accept submissive roles in society (ibid., 362). As capitalism evolved beyond a relatively homogenous labor force, capitalists stoked race conflicts and other forms of social unrest to ensure their continued perch atop the social hierarchy.

Such an emphasis in academia on discrimination encouraged development of theories of personal income distribution because discrimination was fundamentally based on the notion that not all workers were the same. Employers assumed that white male workers were superior to other workers even if the "intrinsic identities" between workers were equal (Arrow, 1971, 28). The presence and influence of discrimination required that laborers no longer be viewed as homogenous members of a social class but instead be considered individual economic agents. When studying income distribution, economists acknowledged that individuals faced different environments and constraints that influenced their position in the labor force. There was thus a need to conduct economic analyses on a personal level. Becker foreshadowed the rise of human capital theory in *The Economics of Discrimination* by mentioning that a relationship existed between economic capacity and "the capital invested in [people] through education" (Becker, 1971 (1957), 112).

The notion of heterogeneity in the labor force was an important

implication of the discussion on discrimination. Incorporating a heterogeneous labor-force into income distribution analysis became necessary because people could no longer be sorted into broad categories such as laborers and capitalists. Reich, Gordon, and Edwards (1973) studied heterogeneity within the labor market from a historical perspective. They asserted that the captains of industry sought to capture control over product and factor markets because they had been relieved of the competitive pressures inherent in the previous stage of capitalism (Reich, Gordon, and Edwards, 1973, 361). In order to establish themselves in product markets, capitalists of the new age of capitalism had to differentiate themselves from their competitors in order to survive. Yet while product differentiation conferred the indirect benefits of monopoly capitalism, Reich, Gordon, and Edwards pointed out another, more sinister and explicit motive for promoting heterogeneity in the labor force that accompanied the rise of monopoly capitalism.

Reich, Edwards and Gordon argued that capitalists encouraged a shift away from homogenization in the labor force "to break down the increasingly unified worker interests that grew out of the proletarianization of work" (Reich, Edwards and Gordon, 1973, 361). They believed that a homogenous labor force fostered a sense of unity among the workers that threatened the consolidation of power in the capitalist class. A strategy of "divide and conquer" was therefore needed to quash any semblance of solidarity in the labor force (ibid., 361). Thus while heterogeneity of the labor-force may have arisen with the evolution of capitalism, it was perpetuated by the capitalists as a preventative measure against class cohesion amongst laborers.

There were also studies conducted within a neoclassical framework that showed evidence of heterogeneity of the contemporary labor force. Gallaway (1967) found that although workers responded positively to earnings, distance acted as a deterrent to job mobility. Workers thus did not always move to regions

paying higher wages (Gallaway, 1967, 465). The decision to forego a higher income was rational because the uncertainty due to imperfect information created trade-off costs between distance and earnings. Workers faced a higher degree of uncertainty with longer distances because distance acted as an "information filter which inhibits the flow of labor market knowledge between areas" (ibid, 472). Trade-off costs became increasingly large with distance so workers were less likely to move in order to gain a marginal increase in income. Wages therefore did not necessarily equalize across regions; differences in incomes for identical jobs could persist.

Gallaway (1967) also found evidence that labor was not easily substitutable. Trade-off costs were not limited only to distance, skill also acted as a barrier to entry for employment in an industry (ibid., 471). Workers from some industries faced more restrictive barriers to entry than workers from other industries because labor was specialized. A wide range of trade-off costs existed across industries: workers in professions with higher trade-off costs embodied less transferable skills as a result of extended parochial training (ibid., 472). Workers who thus received highly specialized training were not able to find alternate employment outside of their chosen industry and earn an income comparable to the one they received in their former industry. Specialization of skills in the labor force therefore greatly affected the ability of workers to switch professions. Both interregional and inter-industry heterogeneity therefore challenged the classical assumption of homogeneity within the labor force.

Analyses of discrimination and labor force heterogeneity facilitated a renaissance of interest in income distribution theory. They signaled that labor could no longer be viewed through the perspective of classical political economy. Workers could no longer be viewed as easily interchangeable. But while discrimination and heterogeneity of the labor force identified the need for

a reevaluation of income distribution, these studies did not develop theories of personal income distribution on their own. Human capital theory and segmented labor market theory offered competing explanations of how discrimination and heterogeneity within the labor force influenced personal income distribution. For instance, discrimination encouraged segmented labor markets, but "discrimination itself does not create the segmentation" (Harrison and Sum, 1979, 698). Human capital theory sought to "single out individual investment behavior as a basic factor in the heterogeneity of labor incomes" (Mincer, 1970, 6). Theories of personal income distribution that emerged during the rebirth of thought on income distribution were thus influenced by the prominent issues of the mid-twentieth century.

Human Capital Theory

Human capital theory operated under the assumption that individuals decided to invest in training or education that allowed them to obtain the skills that made them more productive and consequently determined their income (Becker, 1962, 9). The distribution of personal income could thus be explained through an analysis of the distribution of human capital among participants in the labor force. Human capital theorists argued that demand for this training and education was determined by the marginal rate of return on investment, and that its supply was determined by the volume of funds available for an individual to acquire training or investment (Mincer, 1970, 18). Wage rates were market prices that reflected the relative scarcity or surplus of different types of labor. This explanation of the personal distribution of income was thoroughly grounded in neoclassical economics.

Jacob Mincer, Theodore Schultz, and Gary Becker were the pioneers of human capital theory. All three acknowledged that human capital was a broad term that included components such as physical health and psychological

well-being, but they agreed that training was the most important type of human capital formation. Mincer (1958) constructed a model examining the effect of investment on human capital under the assumption of rational choice. He argued that individuals chose the amount of training they wanted based on their perceived learning capacity (Mincer, 1958, 286). People with greater learning capacities chose to acquire more training and enter professions requiring more training. Furthermore, earnings within a profession fell along a "life-path" where older workers earned more than younger workers (ibid., 288). Mincer suggested that workers gained experience the longer they worked in a profession, which increased their productivity and income. Professions that required more training also paid higher salaries because they required longer postponement of earnings (Mincer, 1970, 7). Higher incomes were thus partly compensation for the shortened period during which those who received the training enjoyed returns on their investment. Personal income distribution was determined by the initial decision of how much training to acquire and how much on-the-job training a worker obtained in their chosen profession.

Schultz (1961) built upon Mincer's work by contending that the decision to invest in human capital was influenced by the expected return on investment. He asserted that while "any capability produced by human investment becomes part of the human agent and hence cannot be sold; it is nevertheless 'in touch with the marketplace' by affecting the wages and salaries the human agent can earn" (Schultz, 1961, 8). Similar to Mincer, Schultz concluded that people who benefited the most from investment in human capital were the most likely to invest the greatest amount in training. Schultz's analysis diverged from Mincer's, however, in the type of training studied. Schultz focused on measuring the returns to formal education because he believed that the exact role of on-the-job training in modern industry was not adequately understood (ibid., 10). He

argued that formal education had taken over a significant portion of the training and preparation traditionally acquired through on-the-job training arrangements such as apprenticeships. Schultz chose to study how the stock of education in the labor force affected economic growth. He asserted that a more educated labor force was a more productive labor force and found that between 1900 and 1956, the stock of education in the labor force grew twice as fast as the stock of reproducible capital (ibid., 11). It was therefore greater educational attainment, and subsequently higher levels of human capital, that drove American economic growth in the first half of the twentieth century.

Perhaps the most comprehensive promotion of human capital theory advocated during the rebirth of income distribution theory was *Human Capital and the Personal Distribution of Income* by Becker (1967). He constructed a model of income distribution similar to Mincer's by incorporating the assumptions of rational choice and variable life-paths of earnings. Furthermore human capital was discussed mostly in the context of educational attainment. But rather than simply reviewing earlier work, Becker also wanted to expand "our rudimentary knowledge of the forces generating income distributions" (Becker, 1967, 12). Therefore while Mincer and Schultz identified the causes of the skewness of income distribution, Becker undertook to explain them better.

Becker first summarized two special cases of the distribution of human capital. Under the "egalitarian" approach, he assumed that all people faced the same demand conditions for human capital and that income was determined by the supply of opportunities to invest in human capital faced by individuals. In short, the egalitarian approach proposed equal capacity to benefit from investment in human capital, and that differences in environment determined the distribution of human capital (ibid., 13). Income variances could be explained by family wealth, subsidies and factors such as luck that shifted the supply curve for human capital

outward. The "elite" approach was essentially the opposite. It presumed supply conditions to be identical and that demand for human capital was determined by the amount of investment in training and education (ibid., 16). More able workers, for instance, were more likely to invest in human capital and thus have a higher demand for it.

Becker believed that in reality, social institutions influenced both the supply and demand for human capital (ibid., 24). Students with greater natural ability not only had greater demand for human capital, their exceptional capacities also made them likelier to attend better schools and make them more attractive scholarship applicants. Legislation aimed at eradicating poverty shifted the supply curve of human capital for less wealthy people outward, thereby reducing the cost of investment. Through his analysis Becker provided a comprehensive explanation for how investment in human capital determined the personal distribution of income.

In the twentieth century, economists began to shift their focus from away from explaining the class distribution of income and towards analyses of the distribution of income among individuals. This change was motivated by studies on economics of discrimination and increasing heterogeneity of the labor force. Clark's theory of marginal productivity first broached the notion of disaggregating classes, and served as the foundation for human capital theory, the most influential neoclassical theory of personal income distribution that emerged during the renaissance of income distribution theory in the mid-twentieth century. Mincer, Shultz, and Becker developed human capital theory to explain how the marginal productivity of workers was determined, and thus develop a more comprehensive understanding of the relationship between wages and marginal productivity.

Criticisms of Human Capital Theory

Three basic criticisms arose in response to human capital theory. One group of economists cited various econometric issues with studies that measured the impact of investment in human capital on output. Griliches (1977) asserted that an "ability" problem and the possible influence of optimizing behavior on schooling decisions by individuals had not been addressed in models constructed in the framework of human capital theory. Ability accounted for the possibility that a given level of investment in human capital yielded varying returns depending on the person. Including an ability variable in empirical analyses, however, proved troublesome because it was difficult to measure (Griliches, 1977, 6). Optimization of schooling decisions was troublesome for human capital theory because such behavior was based on anticipated future earnings. Calculating the optimal level of schooling or on-the-job training implicitly required strong assumptions about individual behavior, which Griliches argued models of human capital theory failed to recognize. For instance, while there was initially a positive relationship between age and experience, older workers also reached a point where they became less productive than younger workers (ibid., 14). Human capital theory therefore implicitly assumed infinite life, even though it argued that jobs requiring more training required higher compensation due to a shorter working life. Furthermore, he also stated that since optimal schooling decisions were based on anticipated earnings, any difference between ex-post and ex-ante incomes increased the correlation between the schooling and residual terms in a model measuring income (ibid., 13).

Blaug (1976) was also critical of the econometric viability of human capital theory. Measuring the effect of on-the-job training on income was especially problematic because the various aspects of on-the-job training were not adequately defined. He contended that human capital theory did not differentiate

between increased productivity from costless learning-by-doing and costly self-investment (Blaug, 1976, 839). Human capital theory therefore did not provide sufficient measures of on-the-job training by limiting itself to general and specific training. Blaug faulted human capital theory models for producing significant, unexplained differences in returns to investment in different types of human capital as well. These discrepancies were due to the neglect of variations in educational quality and the existence of an "overtaking point" (ibid., 838). The benefits of schooling grew over time rather than being fully realized immediately after schooling was completed. Blaug attributed these measurement errors to the overly ambitious nature of human capital theory. He believed the focus of human capital theory was too broad, making it difficult to determine "what hypothesis is being tested" (ibid., 832).

Advocates of the screening hypothesis constituted another group critical of human capital theory. They asserted that the assumption of perfect information in labor markets was unrealistic. Economic theory needed to reflect the high degree of imperfect information employers faced when reviewing job applicants. Supporters of the screening hypothesis agreed that human capital theory was correct in that individual incomes were determined by the level of investment in human capital, but they developed a different explanation for how human capital affected income. The link between human capital theory and the principle of marginal productivity did not explain personal income distribution because a "diploma serves primarily as an imperfect measure of performance ability rather than as evidence of acquired skills" (Arrow, 1973, 193). The screening hypothesis essentially argued that education separated more able workers from less able workers. Income was determined by ability, not productivity. Stiglitz (1975) contended that educational screening occurred naturally in society because it was the "byproduct" of providing knowledge and career direction to students in

schools (Stiglitz, 1975, 294). Bright students were identified by teachers who then passed this information along the educational chain. School was used primarily to sort students into levels of ability, not to develop skills to make more productive workers, as human capital theory suggested.

Several important assumptions served as the foundation for the screening hypothesis. Most important among them was the existence of inherent market failure within the labor markets stemming from the lack of knowledge and cost of obtaining information about potential employees (Taubman and Wales, 1975, 112). College diplomas were used by employers as a proxy for ability because they offered quick insight into the skills and capabilities embodied in applicants. As a result, the supply of labor for high-paying occupations is restricted to the well-educated (ibid., 118). Stiglitz argued that access to information also affected the decision making of job applicants. Individuals decided how much education to invest in based on their perception of their ability. Risk-averse people therefore chose to forego the chance of being screened as a below-average worker even if they were highly capable (Stiglitz, 287). Yet while educational screening was imperfect, Stiglitz cautioned against forbidding employers to practice it. Screening would still occur, it would merely change forms. Forcing employers to rely solely on on-the-job screening would make screening more expensive and reduce output (ibid., 291). Everyone would be left worse off.

The screening hypothesis made a compelling case against the limitations of human capital theory, but the former faced scrutiny on theoretical grounds as well. In the development of his model supporting the screening hypothesis, Arrow (1973) acknowledged that "employers cannot measure ability directly, and there is no reason to suppose that the economist is going to do any better" (Arrow, 216). Screening was based on the assumption that people have differing levels of ability and that the more able use educational attainment to signal this.

Yet if ability in general cannot be measured, it is nearly impossible to determine its distribution. Layard and Psacharopoulos (1974) also conducted an empirical study based on Arrow's work and obtained results challenging the predictions of the screening hypothesis. For instance, they found that dropouts and students who completed their coursework earned similar rates of return (Layard and Psacharopoulos, 1974, 991). That challenged the notion put forth by the screening hypothesis that a bachelor's degree signaled to employers a more capable worker than an applicant who completed only some coursework. Although the screening hypothesis provided a neoclassical alternative to human capital theory, it failed to unseat human capital theory as the primary explanation of personal income distribution in orthodox economics.

IV. Radical Political Economy

Radical political economy criticized human capital theory from a heterodox perspective. Its supporters were less concerned with the empirical issues of human capital theory, instead choosing to focus on the fundamental perception of production in neoclassical economics. While neoclassical economics shifted focus to the functional aspect of production with the introduction of the theory of marginal productivity, radical political economy asserted that the social aspect of production was the primary determinant of personal income. Marxists criticized human capital theory for artificially resolving the inherent class conflict associated with capitalist systems of production by considering every worker a capitalist (Bowles and Gintis, 1975, 74). In fact, radical political economy questioned human capital theory's definition of capital. Learning could only be a form of capital if it allowed workers to go into production on their own (ibid., 79).

Radical political economists challenged the notion that workers were paid according to their marginal productivity because they believed that the

structure of capitalist firms encouraged the de-skilling of their workers (Rebitzer, 1993, 1401). Essentially this meant that, rather than encouraging workers to acquire more training in order to become more productive, they actually preferred workers to embody only a minimum level of human capital. As a result they became discouraged and the mundane nature of their work prevented them from increasing, or even maintaining, their productivity. This is a contemporary re-statement of the inevitable alienation of workers by capitalists espoused by Marx. Human capital theory argued that higher productivity levels caused higher wages. Radical political economy effectively argued the opposite: lower wages caused productivity to fall. Workers embodied a natural endowment of "human capital" that was augmented by schooling and training, but the economic return to these investments was governed by the extent to which these same workers legitimated the authority of firms over their employees (Bowles and Gintis, 1974, 80). Schools were important, but not in the way human capital theory proposed. The main goal of the education system was "to prepare students by developing attitudes appropriate to the political position they can be expected to occupy within firms" (Rebitzer, 1993, 1403). Marginal productivity was therefore unimportant to an analysis of personal income distribution.

Since radical political economy criticized the theoretical foundations of human capital theory, it can be viewed as a critique of the orthodox theory of personal income distribution. Radical political economy was "not ready to reduce the school system's economically relevant activities to screening and labeling" (ibid., 75). Imperfect information did exist in the labor market, but education was not used as a signaling device to help more capable workers of signaling their level of ability. The screening hypothesis was thus not a sufficient explanation for the distribution of personal income distribution either. While radical political economy acknowledged neoclassical economics' contribution to the theory of

personal income distribution, its explanations of personal income distribution were incorrect. Marxist dissatisfaction with neoclassical theories of personal income distribution led to the formation of their own theory of personal income distribution: segmented labor market theory.

Segmented Labor Market Theory

While human capital theory explained differences in personal income levels through a neoclassical perspective, segmented labor market theory sought to explain personal income distribution in a Marxian framework. In this framework, "political and economic forces within American capitalism have given rise to and perpetuated segmented labor markets, and … it is incorrect to view the sources of segmented labor markets as exogenous to the economic system" (Reich, Gordon, and Edwards, 359). Instead of contending that the personal distribution of income was determined by the functional aspect of production, proponents of segmented labor market theory argued that it was mainly the result of the social aspect of production. Wages and productivities applied to the jobs themselves, rather than the individual workers occupying those positions (Harrison and Sum, 1979, 694). Similar to the divide in classical political economy between Ricardo and Marx, a split between neoclassicals and a group influenced by Marxism developed during the renaissance of income distribution theory in the mid-twentieth century.

Segmented labor market theorists differentiated between primary and secondary labor markets. The primary labor market comprised firms with market power, sustainable sources of income, and the ability to pay above-subsistence wages (Harrison and Sum, 1979, 689). Firms in this sector could afford to pay for training for their employees because their market power allowed them to pass some of the cost to consumers. Firms also invested in human capital to increase the productivity of their workers. Consistent with Marxism, members of a relatively

small group whose size was maintained by rigid entry requirements earned high incomes while most workers toiled away in unattractive and unfulfilling jobs. And since firms in the primary sector invested significant resources in training and physical capital, jobs in this sector required stable working habits. The primary labor market thus provided relatively stable employment. Due to higher wages and greater stability, jobs in this sector were highly valued. Poor people were not excluded from the primary segment merely for being lazy or lacking the capacity for human capital necessary for entry into the primary segment. The institutional framework of the capitalist system artificially restricted entry into the primary labor market (ibid., 694). Poor people were poor mostly due to employment prospects restricted to the secondary labor market.

Jobs in the secondary sector were much less secure. Unstable product demand prevented firms from ensuring long-term employment. Furthermore, production processes in this segment were labor-intensive and involved simple or repetitive tasks so workers were interchangeable. Stable working habits were discouraged as a result and there was little opportunity for career advancement. The secondary labor market was connected to the primary labor market through such means as subcontracting but "many adults are unable to escape from it and spend much or all of their lives there" (ibid., 690). Secondary workers thus played an important role in the economy because the primary sector was dependent upon their employment for such services as subcontracting, but they were also expendable. In effect, the secondary labor market was the modern Marxists' equivalent to the reserve army of the unemployed. Competition kept wages low among a group of people without which the economy could not function.

Labor market segmentation went beyond this general distinction between primary and secondary sectors. Piore (1972) further segmented the primary labor sector into upper and lower tiers. Workers in the upper tier held management jobs,

which conferred higher status and pay, and greater economic security. There were relatively high turnover rates but this was attributed to career advancement rather than termination. Members of the lower tier of the primary sector were regularly employed blue collar workers. Workers adhered to a strict set of work rules that were predicated upon the hegemonic relationship between worker and supervisor. Each segment had a different "mobility chain" that signified the opportunity for career advancement (Piore, 1972, 6). Workers in the upper tier had the most opportunity for promotion, the lower tier was more rigid, whereas there was little chance for advancement in the secondary labor market.

Race was cited as one of the primary causes of labor market segmentation. Segmentation by race arose from "certain jobs that are 'race-typed,' segregated by prejudice and labor market institutions," (Reich, Gordon, and Edwards, 360). This was compounded by geographic separation of employment opportunities, which hindered the flow of labor. Harrison (1972) found that underemployment and poverty persisted in urban ghettoes, due to a lack of economic opportunity for minorities. Minorities were considerably limited to a selection of "typically urban" jobs that prevented workers from fully recognizing their productive potentials (Harrison, 1972, 811). There were thus jobs that were often associated with specific races located in specific geographic areas. According to segmented labor markets, social institutions were the most important factor in determining the distribution of personal income. Workers with higher levels of productivity did earn higher incomes, but the access to investments that increased productivity was determined by social institutions prevailing at the time.

Segmented labor market theory presented a heterodox alternative in the proliferation of theories of personal income distribution in the mid-twentieth century. While human capital theory offered an explanation for personal income distribution within a marginal productivity framework, segmented labor market

theory stressed the importance of social institutions to the determination of income distribution among individuals. It therefore appeared that two polemic approaches to the analysis of personal income distribution developed during the renaissance of income distribution analysis.

V. Reconciling Human Capital Theory and Segmented Labor Market Theory

Previous work has been done comparing the relationship between mainstream orthodox economics and the heterodox alternative. Rebitzer (1993) argued that "radical and mainstream neoclassical labor economics have exerted an important influence on the other" (Rebitzer, 1396). A relationship could therefore be drawn between the two, despite their presumably opposed theoretical foundations. But despite making this connection, Rebitzer treats neoclassical and radical approaches to labor market segmentation separately (ibid., 1412). Thus while there was significant overlap between the competing theories of personal income distribution, they were still distinct. A closer examination of the human capital theory and segmented labor market theory literature, however, reveals an even greater degree of overlap than suggested by Rebitzer, so much so that they effectively become identical.

The treatment of class conflict by human capital theory was one of the major points of contention segmented labor market theorists had with the orthodox approach to personal income distribution. Proponents of segmented labor markets argued that human capital theory "formally excludes the relevance of class and class conflict to the explication of labor market phenomena" (Bowles and Gintis, 75). One of the fundamental underpinnings that segmented labor market theorists built upon when responding to human capital theory was the argument that human capital theory artificially resolved class conflict by making everyone a capitalist. Segmented labor market theorists believed that the persistent

significance of sorting workers into groups "is neither explained nor predicted by orthodox theory" (Reich, Gordon, and Edwards, 359). Although segmented labor market theory championed itself as a theory of personal income distribution that had corrected the theoretical flaws in human capital theory, it was not the dramatic departure from orthodox theory promised. Reconciliation between these competing theories is therefore not only possible, but relatively straightforward.

It is not entirely correct for critics of human capital theory to argue that it "predicts that labor market differences among groups will decline over time" (ibid., 359). Human capital theorists implicitly acknowledged the importance of social institutions when explaining how individuals decided on the amount of human capital to invest in. Accumulation of human capital was restricted by "legal and other obstacles to financing investment in human capital" (Becker and Chiswick, 1966, 359). Important social institutions such as inheritance of property income and availability of scholarships and loans were crucial factors that workers had to consider when making their investment decisions regarding human capital. Since the type of employment available to applicants was determined by the amount of training and education they embodied, social institutions had direct influence on the personal distribution of income in human capital theory. The influence of social institutions was thus an integral part of human capital theory and served to perpetuate differences between groups of workers. To argue that differences between groups would decline over time necessarily implies a convergence of social institutions that allowed everyone equal opportunity to invest in human capital, and that everyone benefited equally from this investment.

More importantly, segmented labor markets themselves are implicitly developed in human capital theory. The type of profession available to workers, according to human capital theory, depended on their investment in human capital. Once people decide how much to invest, they are effectively sorted into labor

market segments according to the level of training they have acquired. Workers then remain in their assigned segment. Human capital theorists develop their concept of segmented labor markets by referring to the stream of income as the life-path of earnings determined by investment in human capital. The notion of mobility chains held by segmented labor market theorists articulated a similar progression. Furthermore, both theories speak of stations as points along the life-path of earnings or mobility chain.

Mincer (1958) effectively argued that workers were more unlikely to switch professions in other industries and job categories once they decided how much to invest in human capital because life-paths of earnings existed for each occupation. Jobs could be sorted by "occupational rank" (Mincer, 1958, 288). Higher occupational rank led to higher income and greater social standing, which resulted in steeper life-paths of earnings. Workers gained experience and became more productive the longer they remained in a profession and moved up their respective life-paths of earnings (ibid., 287). There was therefore little incentive for workers to switch professions. They would not be able to obtain a job of higher occupational rank because they invested too little in human capital and were thus unqualified for such positions. Switching to a profession of lower occupational rank was irrational because doing so required sacrificing a steeper life-path of earning for a flatter life-path of earning, as well as movement down the new life-path of earning to reflect the lost experience from switching professions. Segmented labor market theory therefore essentially reiterated the notion of segmented labor markets developed by human capital theory.

Entry into stations occupying the primary labor market of segmented labor market theory was similarly regulated in human capital theory. Segmented labor market theory argued that "the educational system does much more than produce human capital," (Bowles and Gintis, 1975, 78). Segmented labor market

theorists believed that the existence of meritocratic society that gave members of one group power over another group. While there was small window for young workers to escape the secondary labor market, most did not. Only those who obtained "bridge" jobs in such fields as metal-working could facilitate inter-segment mobility (Osterman, 1975, 514). Workers who were unable to enter the primary labor market early were trapped in the secondary labor market. Human capital theory developed rigid barriers to entry to labor markets prior to the development of segmented labor market theory. The belief that entry into professions was regulated by investment in human capital implied the existence of a meritocratic society from a human capital theory perspective. Furthermore, windows of opportunity were also an important aspect of human capital theory. Once individuals decided how much human capital to invest in, their window of opportunity effectively closed because they had sorted themselves into their respective professions.

There is thus a common perception of segmented labor markets and a labor market hierarchy. Both human capital and segmented labor market theories of personal income distribution promoted the notion of a rigid stratification of the labor market in which "each occupation is seen as a set in a stratum of society defined by income and way of life" (Brown, 1977, 118). Workers will be sorted into their stations at a fairly early age. Segmented labor market theory argues that this sorting is governed by social institutions. Human capital theory argues that it is determined by the level of investment made in human capital, which itself is influenced by social institutions. But regardless of how this sorting is actually conducted, both human capital theory and segmented labor theory argue that its effects are lasting. Once the sorting process is completed, it is uncommon for workers to move between segments. Thus contrary to a gradual decline in differences between groups of workers, human capital theory implies that there

is a tendency for them to endure, and even strengthen, over time. A worker sorted into a lower segment of the labor market was not likely to move into the upper segment and subsequently propagate the benefits of greater opportunity for investment in human capital upon future generations.

It is therefore difficult for supporters of segmented labor market theory to criticize human capital theory when the implications of each theory of personal income distribution are so similar. Human capital theory argues that segmented labor markets are the result of individual choice. People face constraints on the amount of human capital they can invest in, but within these constraints people are free to choose any level of investment they want. While segmented labor market theory also shows how labor market segmentation persists, it treats the origins of labor market segmentation as exogenous: the prevailing social institutions determined where in the labor market hierarchy an individual ultimately ended up. The largest theoretical difference between the two is thus a matter of endogeneity. This is ironic because Bowles and Gintis cite "the assumption of exogenously determined individual preferences" as a shortcoming of human capital theory while failing to recognize that segmented labor market theory is endogenous only to the extent that it explains how segmented labor markets reproduce themselves from an exogenously determined origin (Bowles and Gintis, 81).

The above criticisms are based on the argument that human capital theory does not account for the importance of social institutions in the determination of an individual's income. Yet as the discussion above demonstrates, this argument can be clearly and concisely refuted. Human capital theory implicitly incorporates social institutions into its theoretical framework, and does develop a notion of segmented labor markets, but it is not manifested in the same way as segmented labor market theory. The influence of social institutions is much more implicit in human capital theory because human capital theory implicitly

factors them into individual decisions to invest in human capital. Therefore a more accurate assessment of human capital theory is that human capital theory assumes that individual incomes are determined by the level of investment made by individuals, given their unique constraints which are the product of social institutions. Segmented labor market theory makes these constraints much more rigid, diminishing the impact of individual choice. This, however, does not preclude the undeniable similarities underpinning the theoretical frameworks of orthodox and heterodox theories of personal income distribution, respectively.

VI. Conclusion

To argue that the late-twentieth century ushered income distribution in from the cold implies that its importance to economic theory had faded and needed to be revived. It is important to recognize the renaissance of income distribution theory in the mid-twentieth century. Atkinson failed to acknowledge the important contributions of human capital and segmented labor market theories of personal income distribution, instead choosing to cite development economics as the primary focus of income distribution in the twentieth century. In reality the twentieth century saw a great debate emerge over competing theories of personal income distribution, a debate that appeared to represent a wider gap than the one between classical and Marxian class theories of income distribution.

Ricardo and Marx developed class theories of income distribution to explain the class conflict that dominated European society during the rise of classical political economy. Each acknowledged the importance of the rate of profit to class distributions of income, but from this common point of emphasis their analyses diverged considerably. The divergence in class theories of income distribution was due primarily to competing beliefs about the sustainability of the capitalist system of production. Ricardo believed a falling rate of profit regulated

capitalist accumulation: capitalists were competitive but could coexist. Marx, on the other hand, deemed capitalism unsustainable. The capitalist thirst for accumulation was driven by an eat-or-be-eaten urgency. Accumulation meant survival and as capital became more concentrated, capitalist oppression grew to such an extent that it would help spark a proletariat uprising.

Neither Ricardo nor Marx attempted to develop a theory of personal income distribution because the labor force at the time was relatively homogenous, and they were living during an era of intense class competition, as evidenced by the Corn Laws. The primary concerns facing economists during the era of classical political economy were thus much more related to classes than individuals. The dominance of class theories of income distribution persisted until Clark introduced the notion that factors were paid according to their marginal productivity. He did not explain how marginal productivity was determined, however, choosing instead to merely state that there was a relationship between the two. Neoclassical economists developed theories of personal income distribution that sufficiently explained how marginal productivity influenced income. Human capital theory emerged as an important theory of personal income distribution but faced many critics. One group of critics, heavily influenced by Marxism, argued that income distribution on the individual level was determined by the social institutions of capitalist society. They developed segmented labor market theory in response to human capital theory and it appeared that an even greater divide over income distribution theory surfaced than the one between Ricardo and Marx.

Growing interest in discrimination and increasing heterogeneity of the labor force also encouraged the development of theories of personal income distribution. In short, the consensus that all labor was easily substitutable, which had persisted since classical political economy, began to break down. Economists began operating under the assumption that laborers were not easily

interchangeable. Drawing upon the limitations of Clark's *The Distribution of Wealth*, numerous theories of personal income distribution were developed. Among them, human capital theory and segmented labor market theory emerged as the most influential orthodox and heterodox theories, respectively.

Yet upon closer examination, the presumed differences between human capital theory and segmented labor market theory can be reconciled to a large extent. Segmented labor market theory faults human capital theory for ignoring the role of social institutions and a segmented labor-force in the distribution of personal income. Human capital theory, however, implies that labor markets are segmented, and that barriers to entry in the labor market are largely due to social institutions. Workers were sorted according to the amount of human capital they embodied. Investment in human capital, which was constrained by the environment produced by social institutions, thus acted as a barrier to entry. The most significant difference between the two is the amount of freedom in workers' initial employment decisions. Human capital theory argues that segmented labor markets are essentially a product of individual choice while segmented labor market theory asserts that segmented labor markets are pre-assigned.

Not only were there numerous important contributions to income distribution theory during the period it was supposedly out in the cold, a close connection between human capital and segmented labor market theories of personal income distribution developed. Segmented labor market theory was essentially a modification of human capital theory, placing more explicit emphasis on the role of social institutions in the explanation of individual income distribution. Although this distinction may seem nuanced, it is important to recognize that segmented labor market theory is not entirely redundant. Segmented labor market theory stresses the importance of social institutions while human capital theory acknowledges their influence much more implicitly. Professor Atkinson may have

had a valid argument that income distribution theory had become too focused on development economics, but the twentieth century was the most important era of income distribution analysis since classical political economy.

Bibliography

Arrow, Kenneth. (1971). "The Theory of Discrimination." *Princeton University Working Paper 30A*, 1-31.

Arrow, Kenneth. (1973). "Higher Education as a Filter." *Journal of Public Economics*, vol. 2 no.3 (July): 193-216.

Atkinson, A.B. (1997). "Bringing Income Distribution In From The Cold." *The Economic Journal*, vol. 107 (March): 297-321.

Becker, Gary. 1971, (1957). *The Economics of Discrimination*. Chicago: The University of Chicago Press, 2nd edition.

Becker, Gary. (1967). *Human Capital and the Personal Distribution of Income: Am Analytical Approach.* Woytinski Lecture No. 1. Ann Arbor: University of Michigan, Institute of Public Administration.

Becker, Gary . (1962). "Investment in Human Capital: A Theoretical Analysis." *Journal of Political Economy*, vol. 70 no.5 (October): 9-49.

Becker, Gary and Barry Chiswick. (1966). "Education and the Distribution of Earnings." *The American Economic Review*, vol. 56 no. 1/2 (March): 358-369.

Blaug, Mark. (1976). "The Empirical Status of Human Capital Theory: A Slightly Jaundiced Survey." *Journal of Economic Literature*, vol. 14 no. 3 (September):827-855.

Bowles, Samuel and Herbert Gintis. (1975). "The Problem with Human Capital Theory – A Marxian Critique." *The American Economic Review*, vol. 65 no. 2 (May): 74-82.

Brown, Henry Phelps. (1977). *The Inequality of Pay*. Berkeley: University of California Press.

Clark, John Bates. (1899). *The Distribution of Wealth: A Theory of Wages, Interest and Profit*. New York: Macmillan.

Gallaway, Lowell. (1967). "Industry Variations in Geographic Labor Mobility Patterns." *The Journal of Human Resources*, vol. 2 no.4 (Autumn): 461-474.

Gilman, Harry. (1965). "Economic Discrimination and Unemployment." *The American Economic Review*, vol. 55 no. 5 (December): 1077-1096.

Griliches, Zvi. (1977). "Estimating the Returns to Schooling: Some Econometric Problems." *Econometrica*, vol. 45 no. 1 (January): 1-22.

Harrison, Bennett. (1972). “Education and Underemployment in the Urban Ghetto.” *The American Economic Review*, vol. 62 no. 5 (December): 796-812.

Harrison, Bennett and Andrew Sum. (1979). “The Theory of ‘Dual’ or Segmented Labor Markets.” *Journal of Economic Issues*, vol. 13 no. 3 (September): 687-706.

Layard, Richard and George Psacharopoulos. (1974). “The Screening Hypothesis and the Returns to Education.” *Journal of Political Economy*, vol. 82 no.5 (September/ October): 985-998.

Marx, Karl. (1867). *Capital*, vol. 1. Edited by Friedrich Engels. Translated by Samuel Moore and Edward Aveling. Moscow: Progress Publishers, 1965.

Mincer, Jacob. (1970). “The Distribution of Labor Incomes: A Survey With Special Reference to the Human Capital Approach.” *Journal of Economic Literature*, vol. 8 no. 1 (March): 1-26.

Mincer, Jacob . (1958). “Investment in Human Capital and Personal Income Distribution.” *Journal of Political Economy*, vol. 66 no. 4 (March): 281-302.

Moore, D.C. (1965). “The Corn Laws and High Farming.” *The Economic History Review*, vol. 18 no. 3: 544-561.

Osterman, Paul. (1975). “An Empirical Study of Labor Market Segmentation.” *Industrial and Labor Relations Review*, vol. 28 no. 4 (July): 508-523.

Piore, Michael. (1972). “Notes for a Theory of Labor Market Stratification.” *MIT Department of Economics Working Paper*, no. 95 (October): 1-40.

Rayack, Elton. (1961). “Discrimination and the Occupational Progress of Negroes.” *The Review of Economics and Statistics*, vol. 43 no. 2 (May): 209-214.

Rebitzer, James. (1993). “Radical Political Economy and the Economies of Labor Markets.” *Journal of Economic Literature*, vol. 31 no. 3 (September): 1394-1434.

Reich, Michael, David Gordon and Richard Edwards. (1973). “A Theory of Labor Market Segmentation.” *The American Economic Association*, vol. 63 no. 2 (May): 359-365.

Ricardo, David. 1821, (1817). *On The Principles of Political Economy and Taxation*. London: John Murray Publishers, 3rd ed.

Ricardo, David. (1951). *The Works and Correspondence of David Ricardo*, vol. VIII, (ed. Sraffa). Cambridge: Cambridge University Press.

Sahota, Gian. (1978). “Theories of Personal Income Distribution: A Survey.” *Journal of Economic Literature*, vol. 16 no. 1 (March): 1-55.

Schultz, Theodore. (1961). “Investment in Human Capital.” *The American Economic Review*, vol. 51 no. 1(March): 1-17.

Stiglitz, Joseph. (1975). “The Theory of “Screening,” Education, and the Distribution of Income.” *The American Economic Review*, vol. 65 no. 3 (June): 283-300.

Taubman, Paul and Terence Wales. (1975). “ Education as a Screening Device.” *Education, Income and Human Behavior*. Edited by F. Thomas Juster. NBER: 95-122.

China's Local Government Debt and Economic Growth

By Tom Zhou

Abstract
This paper explores the impact of China's local government debt on economic growth. This analysis, based on a panel of 31 provinces over 14 years, takes into account a broad range of economic growth determinants as well as various estimation issues including heteroskedascity and omitted variable. The empirical results suggest an inverse relationship between China's local government debt and economic growth, controlling for other determinants of growth: on average, a 10 percentage point increase in the debt-to-GDP ratio is associated with a slowdown in annual real per capita GDP growth of around 0.27 percentage points per year

I. Introduction:

The financial crisis of 2007-2008, which originated in the United States, was considered by many economists to be the worst financial crisis since the Great Depression of the 1930s. It not only dragged down many financial institutions in the U.S., but it also triggered the most severe economic contraction in many foreign countries. Even though China, an export-led growth country, has enjoyed double digit growth for the past ten years, China's economy was also threatened by economic contraction in the West. Fortunately, China's economic growth soon reclaimed lost ground thanks to a 4 trillion Yuan (USD 586 billion) fiscal stimulus package aimed to upgrade infrastructure and transportation and modernize different industries (Yerxa, 2011). Although the stimulus package received lot applause for boosting growth and reviving the economic contraction, it raised the leverage ratio for China's government. Most importantly, the credit loosening due to the stimulus plan has increased debt burdens for local Chinese governments. As a matter of fact, the collapse of many private lending facilities in Wen Zhou, an eastern city in Zhejiang Province, which was well known for its exporting business, and other major cities in other provinces drew a lot of public

attention to the health of China's public debt especially at provincial level (Caixin Magazine online, December, 2011).

Even though the worries that some local governments were not able to pay back their debts after series of default events in many local cities in 2010 were intensified, it is still not clear whether China's local government debts has a significant impact on China's economic growth. Most of current studies on China's local government debts rely on news analysis and observation. For instance, the debt problem in Wen Zhou in 2010 drew a lot of attention from domestic and foreign news media. Bloomberg, one of the most influencing financial media, released a news report saying that, "China's first audit of local government debt found liabilities of 10.7 trillion yuan ($1.7 trillion) at the end of last year and warned of repayment risks, including a reliance on land sales" (Bloomberg news, 2011). Professional economists and fiscal pundits such as Nouriel Roubini, an economics professor at New York University's School of Business, start publishing paper to question the health of local government's fiscal policy. Some of studies may have a convincing inside story to conclude that the current borrowing level carried by local government has detrimental effect on China's fiscal sustainability. However, whether or not the growing amount of debts held by local governments has negative effect on China's economic growth is still unclear at this stage.

Recently, there have been a lot researches focusing on the relationship between government debt and economic growth. Also, many literatures identify the determinants of China's economic growth. Among all the literatures focusing on the impact of government debt on economic growth, Kumar and Woo (2010) did a distinguished job by providing an empirical analysis on the impact of high public debt on long-run economic growth based on a panel of advanced and emerging economies over four decades. Besides focusing on government debts, they also include a broad range of determinants of growth. Their empirical results suggest

an inverse relationship between government debt level and economic growth, controlling for other determinants of growth. Besides the international evidence of impact of government debts on economic growth, Lin (2003) identifies the components of China's government debt, which comprises domestic and foreign debt. He concludes that risk of foreign debt is very low for China since China's foreign exchange reserves are much higher than total foreign debt outstanding. For China's domestic government debt, it is more complicated because domestic debt consists of explicit fiscal deficits, local government debt, state banks' non-performing loans, and fiscal subsidies made to social security funds (Lin, 2003). Lin (2003) attributes the local government debt to the tax system adopted in 1994. Since local government cannot get enough revenue from tax due to the shared tax collecting with central government, they have been running deficits since 1994. Lin (2003) concludes that if current tax system remains unchanged, local government debts accumulated from fiscal deficits will cause a serious trouble.

However, what the two literatures above and other literatures focusing on government debt and economic growth in China have not identified is the impact of China's local government debt. Most importantly, what Lin (2003) did not emphasize is that the true cause of local government debt after the 1994 tax reform comes from local financing platform, which this paper will delve into later. As a result, this paper intends to fill up the research gap by answering the main question of the paper: does local government debt in China have significant impact on China's economic growth? In order to answer the question, this paper will contribute to the empirical literature on the relationship between china's local government debt and economic growth by first providing a comprehensive historical analysis on local fiscal policy and implementation plan in China and then constructing a panel regression model by using cross provincial dataset. This regression model will include provincial GDP per capita growth from 1994 to

2011 as dependent variable and local debts and other determinants of economic growth from 1994 to 2011 as independent variables. And finally, this paper will run regressions to test whether local government debt has significant impact on China's economic growth.

The results, based on a range of econometric techniques such as ordinary least squares (OLS), fixed effect (FE) and random effect (RE), suggest an inverse relationship between local government debt and economic growth, controlling for other determinants of growth: on average, a 10 percentage point increase in the debt-to-GDP ratio is associated with a slowdown in annual real per capita GDP growth of around 0.27 percentage points per year.

The rest of this paper is organized as follows: Section 2 has four main parts. Part one provides an economic theory of government debt. Part two reviews the empirical evidence of government debt. Part three reviews China's fiscal history and fiscal structure at local government. Part four highlights the debates from economists and reiterates this paper's key contributions. Section 3 introduces the modeling and data. Section 4 discusses and interprets the regression results. Section 5 concludes with further research suggestions.

II. Literature Review

A. Economic theory of government debt

Before going too deep into China's local government debt problems, two critical questions are raised: why do high debt levels affect future growth, and how does high debt level affect future growth? From a conventional view, the government's debt policy affects the economy both in the short run and in the long run. Government debt and deficit is interrelated because a country with a large debt have difficulty to finance annual deficits through more borrowing and, consequently, will be more likely to raise tax to collect more revenue. Therefore,

we begin analyzing the effect of debt on economic growth by discussing the short-run effects of budget deficits.

In the short run, if the government creates a budget deficit by holding spending constant and reducing tax revenue, this policy will raise households' current disposable income. Based on the Keynesian view, the increases in income and wealth boost household spending on consumption goods and thereby, raise the aggregate demand for goods and services. In the long run, suppose that the government holds spending constant and reduces tax revenue, this policy creates a budget deficit and decreases public saving assumed the economy follows this identity:

$$□ + (□ - □) = □ + □□□$$

Where S is private saving, T is taxes, G is government purchases of goods and services, I is domestic investment, and NFI is net foreign investment. The left side of this equation is national saving, which is the sum of private and public saving while the right side indicates save funds for investment at home and abroad. This identity is also a description of the market for loanable funds (Elmendorf and Mankiw, 1998).

If we adopt the conventional view that private saving rises by less than public saving falls, then national saving declines. After national saving declines, total domestic investment may decline as well because the sum of private and public saving equals to the sum of investment and net export. Reduced domestic investment will cause domestic capital stock to shrink, which implies lower output and income. The marginal product of capital will be higher after total capital becomes less and less. The rising marginal product of capital causes the interest rate to rise. Meanwhile, labor productivity would decrease, thereby reducing the average the average real wage and total labor income. (Elmendorf and Mankiw, 1998) As a result, the lack of domestic investment and the reduction

in productivity will drag down the total output in the economy, and thereby slowdown the economic growth.

Despite the short term and long term effect of government's debt policy, debt financing is the center of financial system nowadays. Countries cannot flourish without borrowing. When individuals have the ability to borrow to expand their consumption level, the overall demand in the economy becomes stronger. When business owners have the ability to borrow to expand their firms, they are expanding the capacity of the economy to generate more jobs and demands. However, past experiences teach us that we cannot borrow forever because high leverage ratios create instability. As debt level increases, borrowers' ability to repay becomes more sensitive to drop in asset prices and increase in interest rate because once the prices of their collaterals drop, they are no longer considered as creditworthy. After a series of default happens, lenders start to lose faith in the market and to scale back their lending. Once the entire credit line breaks, consumption and investment fall. Afterward, high unemployment rate and insufficient aggregate demand will drag the real economy down. The real economy will experience an even worse economic downturn if the level of debt before crisis is enormous (Cecchetti et al, 2009).

B. Empirical Evidence of Government Debt

In order to see whether government debt problem poses a significant effect on China's economic growth, we have to find an accurate model that can capture the effect of debt on economic growth. Rogoff and Reinhart (2010) develop a model using data from 105 countries consisting of both advanced and emerging economies from the past twenty years to analyze the relationship between the public leverage ratio (debt-to-GDP ratio) and average GDP growth. Based on their calculation, they come up with a threshold for debt-to-GDP ratio. When the debt-to-GDP ratio rises above 90 percent, median growth rates fall by

one percent, and average growth falls considerably more. Also, they find that the thresholds for both advanced and emerging economies are similar even though emerging economies will more likely face a worse economic downturn when debt-to-GDP ratio exceeds 90 percent. Moreover, emerging markets face a lower threshold for external debt, which is usually denominated in a foreign currency. When external debt in emerging economy exceeds 60 percent of GDP, annual growth declines by about two percent (Reinhart and Rogoff, 2010). In addition, high debt level not only limits country's long run economic growth, but it also limits the effectiveness of fiscal policy.

However, the model developed by Rogoff and Reinhart does not take reverse causality into account. High debt can potentially lower the future economic growth, but low economic growth also can cause debt to grow in the long run. As a result, analyzing government debt requires a more comprehensive model because economic growth is determined by many factors ranging from political structures to cultural aspect. If a model cannot fully cover all the determinants of economic growth, the impact of debt on economic growth is biased. In a paper written by Robert Barro, Determinants of Economic Growth in a Panel of Countries (1997), he provided a framework to analyze the factors that determine the economic growth rate in different countries. He derives his model from the hypothesis from the neoclassical growth model that poorer countries typically grow faster per capita at initial level and thereby catch up with the richer countries. The convergence hypothesis of neoclassical growth model implies that the growth rate of real per capita GDP during a time period would tend to be inversely related to the level of real per capita GDP in the initial year. In this model,

$$\square = \square(\square, \square^*)$$

Where g is the growth rate of per capita output, y is the current level of per capita output, and is the steady-state level of per capita output. If the current level of

per capita output, y, is high, the growth rate of per capita output, g, will diminish for a given steady-state level of per capita output,. If holding the current level of per capita output constant, the growth rate of per capita output will increase with a rise in the steady-state level of per capita output, because the improvements in external conditions such as government regulations, law, and market openness, are beneficial to the long-run growth of the economy. The steady-state level of per capita output is determined by a number of factors comprising social, economic, cultural and demographic. In addition, the steady-state level of per capita output is also determined by political factors such as the degree of political freedom, the extent of market efficiency, and size of government expenditure (Barro, 1996).

Barro's model provides a frame work to analyze the determinants of economic growth even though he does not take government debt into account. To test whether government debt has significant effect on economic growth empirically, debt has to be added as a new variable into the model. Recently, the empirical literature on debt and economic growth has grown. But many of those growth regressions used in the literature have many shortcomings. Instead of focusing on a few socioeconomic variables that are statistically significant, they tend to include as many variables as possible to mitigate omitted variable bias. Even though they may correct omitted variable bias, they commit multicollinearity errors where two or more predictor variables in a multiple regression model are highly correlated. As a result, it is better to focus on a core set of explanatory variables that have shown a strong and consistent relationship with economic growth. If there are some extra variables that are important to explain economic growth, we can include them in the model while keeping the same core variables.

Specifically, the findings of Kumar and Woo (2010) select the core sets of growth determinants, which are consistent with Barro's model. Besides selecting the core sets of growth determinants, Kumar and Woo utilizes a variety

of estimation methodologies, such as pooled OLS, robust regression, between estimator, fixed effects panel regression, and system GMM (SGMM) dynamic regression to capture the effect of government debt on economic growth. Since there are many sources of bias that can result in inconsistent estimates of the coefficients in panel regressions ranging from omitted-variables bias to endogeneity, using different estimation methodologies will mitigate them. The econometric results from Kumar and Woo (2010) suggest an inverse relationship between initial debt and subsequent growth after controlling for other determinants of growth: a 10 percentage point increase initial debt-to-GDP ratio is associated with a slowdown in annual real per capita GDP growth of around 0.2 percentage points per year. On average, a 10 percentage point increase in initial debt is associated with a decline of investment by about 0.4 percentage points of GDP. Cecchetti et al (2009) use similar methodology analyzing data from OECD to conclude that high debt is bad for growth. When public debt is above 85 percent of GDP, further increases in debt may begin to have a negative impact on growth. More specifically, a 1 percentage point increase in corporate debt is associated with an approximately 2 basis point reduction in per capita GDP growth. A 1 percentage point rise in household debt-to-GDP is associated with a 2.5 basis point reduction in growth.

While the cross-country empirical results from Kumar and Woo (2010) and Cecchetti et al (2008) conclude the negative effect of government on economic growth despite of the size of economy, they have not emphasized the importance of local government debts especially in China. Even though a lot of previous literature helps explain and identify the determinants of China's economic growth, only a few of them start to emphasize the importance of China local government debt on China economic growth in recent years. However, finding the variables that best explain China's economic growth is still very challenging due to the uniqueness of China's economic structure. Most importantly, just like what

Lin(2003) mentions in his paper, "No one knows exactly how large is the local government debt. It is believed that township government debt is widespread and severe." Because of this, finding the correct variable to represent China's local government debt is a very challenging task for this paper. Therefore, the first step before constructing a model that explains the effect of China local government debt on China's economic growth is to review the history of China's fiscal policy at provincial level.

C. China's fiscal history and local government's fiscal structure

China's debt problems are much more complicated than developed countries because China's unique political system yields a different fiscal policy implementation plan. Before going to empirical analysis of China's local government debt, we have to review the history of fiscal policy and the fiscal structure of China's local government.

After Deng Xiaoping introduced the market economy to China in 1980, China's economy started to take off. Between 1980 and 1999, the average growth rate of China's GDP was around 9.5 percent. However, the robust growth of China's economy did not increase the central government's revenue because the tax system implemented in 1980 was not efficient enough to generate revenue for central government. The budgetary revenue-to-GDP ratio decreased from 28.4% in 1979 to 12.6% 1993. Most importantly, the central government revenue-to-total revenue ratio decreased from 46.8% in 1979 to 31.6% in 1993 (Zhu, 2007). The insufficient revenue in central government diminished its authority in the implementation of fiscal policy. The central government was sometimes forced to borrow money from other local governments between 1980 and 1990 (Zhu, 2007).

One of the main reason that central government could not collect enough revenue to support its expenditure was China's centralized fiscal system, which relied on local government agencies to collect revenues for transfer to

the national treasury. Each local government had an assigned duty to collect a certain amount of tax by the end of each year. In return, the central government assigned re-transferring revenues to local governments' budgets. Because such a policy was basically a part of socialist planned economy, where everyone eats from the same pot, local governments had no incentives to promote the local economy. In addition, in order to address the insufficient revenue, China's central bank, People's Bank of China, had to print more money to ensure that there was enough money circulating around the economy. Such an aggressive money printing policy eventually resulted in the risk of high inflation. Many economists and policymakers in China during that time started to worry about the central government's fiscal authority, and called reform in the tax system. However, because of different political interests, the tax reform did not take place until 1994 after a series of fiscal crises in central government (Zhu, 2007).

To address problem of inefficient fiscal policy, China's Prime Minister, Zhu Rongji, who was an economist himself and a devout follower of Milton Friedman, decentralized China's tax system in 1994 by introducing a new tax system called the Tax Sharing System although he was under a lot political pressure from different interest groups. Before 1994 reform, there were three categories of taxes levied in China: the industrial and commercial tax, tariff, and the agriculture tax (Zhu, 2007). The three main taxes were controlled by central government. The new tax sharing system introduced in 1994 basically separated the tax revenue collected by central and local governments. The tax revenue (not tax legislation), according to the new sharing system, is divided in a way that some taxes are exclusively assigned to the central level, some are assigned exclusively to the local level, and some taxes are shared between both levels based on a fixed ratio. For central government, the tax revenue comes from taxes such as consumption tax, tariffs, vehicle acquisition tax, and business tax. For local government, the

tax revenue comes from taxes such as urban maintenance and construction tax, vehicle and vessel usage license tax, and orientation adjustment tax on investment in fixed asset. Since VAT (Value-added Tax) and income tax are the two largest tax revenues, they are divided between the central and local level. For VAT, the central level takes 75% while the local level takes 25%. For both individual and Enterprise Income Tax, the central level takes 60% and local level takes 40%. The new tax sharing system basically left local government no choice but to generate as much tax revenue as it could to support its annual expenditure level. As a result, the new tax sharing system provides incentives for local governments to generate more tax revenues by promoting their local economies (Zhu, 2007).

However, although the new tax system was designed to generate more revenue for the central government, it also limits the range of tax revenues collected by local government. From 1960 to 1985, local governments had budget surpluses. From 1986 to 1993, even though local government started experiencing deficits in some years due to the fixed asset investments, they soon recovered the "lost ground" in following years. But the tax reform in 1994 changed everything. Local government budgets jumped from a surplus of 6.1 billion yuan in average in 1993 to a deficit of 172.7 billion yuan in average in 1994 (Lin, 2003). Basically, starting in 1994, almost every local government from 31 provinces had a budget deficit each year. In addition, since China is currently undergoing massive infrastructure spending in different provinces after 1994, the local governments have to find another way to come up with more funds besides tax revenue to fund their building projects.

However, no matter how fast a local economy can develop, its economic capacity cannot expand without enough budget revenue to support. Since even the richest provincial government such as Guangdong province, one of the first provinces to experience market reform in 1980 had a hard time to collect enough

revenue to finance their annual budget deficit. Moreover, because the tax and budget reforms enacted in 1994 prohibit local governments from issuing bonds, the local governments could not simply issue debt to finance their projects. After the 1997 Asian Financial Crisis, the local governments came up with a new way to generate more funds. The law forbids the local governments to issue debts directly, but it doesn't prohibit them to generate funds indirectly by setting up an entity called the Local Financing Platform (LFP) (Walter and Howie, 2011).

These LFPs provide non-tax revenue to local governments to fund their operating budgets. Their sources of fund come from bank loans. But the question is how those LFPs can secure bank loans that local governments cannot obtain directly. In order to circumvent restrictions on local government borrowing from bank, local governments create corporations that act as the borrowers to obtain loans from banks. These corporations consist of State-Owned Enterprises (SOE) and local incorporated investment companies. Sometimes they combine together to form a financial entity such as a private equity or a fund management company. The mechanism of the local government funding process through Local Financing Platform is not hard to comprehend. Basically, local governments contribute lands to local financing platforms and provide tax subsidies for them. The LFPs have lands as collaterals to borrow money from banks. Meanwhile, the LFPs use the lands they obtained from local governments to develop infrastructure projects such as high ways, railways, and airports. One the one hand, these infrastructure projects is beneficial to those LFPs because they solely control the profits from these projects. For example, China Railway Company was one of the biggest LFPs in northern China before the head of this company was arrested in 2010 because of corruption. They controlled 99% of railways in northern China. More than half of the profits made from ticket selling and cargo shipping goes to this company. They not only used the profits to develop more railways, but they also started

developing real estate projects in Beijing, Tianjing, and other major metropolitans in China to expand their financial influence. On the other hand, these projects developed by LFPs are beneficial to local economy. When a new railway or real estate project starts, they need more workers to work and more raw materials such as cement and steel to build bridges and airports, and thereby increase total industrial outputs at local level. That is why some provinces such as Guangdong, the biggest provinces in southern China, and Zhejiang, the biggest provinces in eastern China, experienced accelerating economic growth in recent years.

While the LFPs start developing projects, they secure more loans for local governments to finance their deficits. Since local governments need to pay wages to their employees, to provide funding for public schools, and to pay for healthcare payments to local government staff, the amount of revenue they get annually cannot cover their expenditure level due to the 1994 tax reform. After LFPs get loans from banks that are owned by the state by providing lands as collateral, local governments no longer experienced insufficient funds. Also, local governments can invest in those LFPs, which issue primary shares, to become one of the largest shareholders who can get dividend payment every year. Meanwhile, the LFP can issue municipal bonds backed by local government and sell them to bank and municipal bond investors. Then the cash obtained from bonds sell goes to local government to finance their budget deficits and other expenditures. Thus, the local financing platform serves as an agent to pass the loans from bank to local government (Walter and Howie, 2011). On the surface, this funding process between local governments and LFPs are mutually beneficial. Local governments do not have to beg money from central government in Beijing to finance their budget deficits and infrastructure projects. LFPs can get the lands for their projects at an artificially low price. Also, they can pay fewer taxes than other corporations especially private corporations.

Theoretically, the whole funding process should work perfectly well under a high economic growth because the platform's capacity for borrowing money from banks depends on the value of the collateral, which is mostly land. Because LFPs use the lands to develop real estate projects such as office buildings, resident houses, and shopping malls and other infrastructure projects, the price of the land might triple due to the increase in demands for housing or rental business. The repercussion of the increase in price of this land will spread to lands nearby. If the economic growth does not slow down in the near future, the price of lands will increase for a long period. The more valuable the land, the more money the LFPs can borrow from banks. Most importantly, those loans backed by the lands will not turn sour if the price of land is still very attractive for investment.

However, once the economic growth starts to slow down, it will be unclear whether the loans obtained by LFP to fund local government budget deficits and expenditures become a Chinese style of "subprime mortgage crisis". Recently, because of global economic slowdown, China cannot immune from the global repercussion. China's GDP growth went down from 14.2% in 2007 to 9.6% in 2008 due to the global financial crisis. Even though the 9.6% growth is still faster than a lot of countries, the 4.6% drop of GDP growth still had a lot impact on China's economy especially at housing sector (World Bank Data). Based on the data from Global Property Guide, house prices in China rose rapidly from 2000 to 2008, primarily driven by low interest rate and cheap credit (Global Property Guide). The skyrocketing house prices were partly caused by speculator who wished to make a huge fortune from selling their homes at higher price. Since it was cheap and easy to obtain a home loan, anyone can become a speculator. Most importantly, China was still lack of regulation in homeowner loan market. Many people who did not have a good credit history could easily take out a loan to buy house. As the economy started showing sign of weakness, the speculators pulled

out from the housing market. The lack of demand for houses caused the house prices to drop at the beginning of 2009.

But economic slowdown was not the primary reason causing home price to drop in China. In order to ease public complaint over the skyrocketing house prices in Beijing, Shanghai, and other major metropolitan areas due to the speculative activities before 2009, the government adopted some housing market-cooling measures in April 2010. For instance, the down payment for first-time buyers' mortgages was increased to 30% from 20%, while for second homes down payment rose to 60% from 50%. New property taxes were introduced in Shanghai and Chongqing between 0.4% and 0.6% in Shanghai (Global Property Guide). These government measures caused house prices to drop substantially in the last quarter of 2011. After house prices dropped, the price of land also decreased. Since those loans obtained by LFPs from banks are backed by lands as collaterals, the sudden drop in house prices will cause those loans to sour.

Theoretically, under the contract between borrower and banks, if the price of land drops, the borrower has to use other methods such as liquidation to pay back what it owes to the banks and bond investors. However, that was not the case in China. According to an article from Bloomberg News, written by Henry Sanderson and Michael Forsythe, "China Cities Value Land at Winnetka Prices with Bonds Seen Toxic", many local governments tried to overstate the land value by three to four times the actual value in order to secure more loans from banks. In one case, local officials were not even able to the specify the exact location of the specific property securing loan, reportedly stating, "it's somewhere north of town, I don't exactly known where. It's like the land outside the city, you know, with the big piles salt" (Bloomberg News, July 13, 2011). This investigation done by Bloomberg News clearly indicates that the attitude by local government officials towards repayment of the loan was disturbingly noncommittal and dismissive.

Recent data from China's statistical yearbook indicates that the growth of loans in each province has already reached an unprecedented level. Also, most of loans shown in China's statistical yearbook are one-year loans. Also, since only state-owned banks have authority to issue loans in China, the close relationship between state-owned banks and LFPs, which are mainly the cover identities of state-owned enterprises in each province, gives local governments a back channel to take out more loans no matter how turbulent the overall economy around the world is. Therefore, given by the evidence provided by Bloomberg try to pay back the loans they ask LFPs to borrow in previous years.

As a matter of fact, not many people before 2008 paid too much attention to the transaction between local governments and LFPs and most importantly, the amount of debt existed under local government's balance sheet, because China was still on an accelerating growth before 2008. But after the growth correction in 2009, local government's unique funding process drew both applause and criticism from different economists after China experienced economic slowdown at the beginning of 2010. One the one hand, one group of economists thinks financing through LFP is less transparent and less manageable than direct government borrowing. The opaque business agreement between LFP and local government leads many economists to question the accuracy of the official reported number of total loan that local governments borrow each year. Also, they worry that the enormous amount of loan accumulated each will have a negative impact on China's economy. On the other hand, some economists think that borrowing from LFP to finance infrastructure projects is beneficial to the economy because these projects will ultimately contribute to long-term economic growth. If the local government has ability to generate enough revenue to pay off its loans from previous years, LFP borrowing will not hurt the whole economy.

The pro sides of local government borrowing think because China

is still a developing country, the space for future growth is enormous. A well-known economist, Justin Lin who is a Chinese economist and former Chief Economist and Vice President of the World Bank recently said in his speech, *Beyond Keynesianism and The New "New Normal"*, that China is not going to collapse because of global recession since China's potential economic growth is still strong enough to compensate short run economic shock. Unlike the U.S. and other western advanced economies, China is still undergoing a transition from a poor country to a developed country. The large gap between rural area and urban area provides China a valuable opportunity to continue its infrastructure projects to reach urbanization. According to Lin, China's fiscal expansionary policy during the past ten years has raised government's debt-to-GDP ratio, but if adding central government debt and LFP loans together, the total debt-to-GDP ratio is around 40 percent. Comparing this number to other countries such as Japan and Greece, China's current debt-to-GDP ratio is still relatively healthy. In addition, not only does China have a fine fiscal condition, China's high household saving rate and 3 trillion U.S. treasury holding will provide adequate fund to expand its economy (Lin, 2011).

Also, according to a report written by a group of economists from Credit Suisse, most of loans borrowed by local government financing platform is largely used in infrastructure projects. As shown in the graph, urban infrastructure and transport projects accounted 62% of the loans and land purchases accounted 10.6%. Infrastructure and transport projects used more than 70% of the total loans borrowed by LFPs. Even though it seems that the amount of loans is enormous, local government's assets are increasing from these investments. Unlike the public debt used to pay for social pension fund and other benefits in Southern Europe, local government's debts in China are utilized to accumulate a large amount of fixed assets to increase the total output. Thus, if China's economic growth is

steady over the next ten years, according to Credit Suisse, the debt problems at local level should not be problematic.

However, even though Lin's and Credit Suisse's arguments that China's robust future growth mitigates government's debt burden seems to be reasonable, many independent research departments and economists cast doubts on his argument. In a special report done by Bloomberg, China's banks didn't fully report the total loans they lent to local government financing platform. For instance, Industrial and Commercial Bank of China (ICBC), second largest state-owned banks in China, reported in 2011 that they include the total loans borrowed by 10,000 LFPs across the entire country in bank's public report. However, Bloomberg sent its own investigation team and found that ICBC only includes 113 LFPs, or about 2% of those reported by the banks. The hidden loans, according to Bloomberg, sometimes contained toxic assets, which came from failed real estate and infrastructure projects from 2008 to 2011. Also, with prices dropping dramatically in China's real estate market in 2011, many local governments were struggling to sell lands to raise more cash. Meanwhile, the decrease in land price causes the price of collaterals held by LFPs to drop. As a result, they were not able to obtain enough loans from banks for local governments (Bloomberg news, 2011).

In more scary news reported by Financial Times, it stated that some local governments are having troubles paying their wages in their public servants because of inadequate cash and loans:

> For example, in the Shandong Province capital Jinan, not a single developer bid for nine of the 11 plots offered by the city in early November. The two plots that sold went for bottom-line prices.
>
> A city with a serious land market crash is Guangzhou, where in November some 32-plots failed to sell. In some cases, auctions were suspended by the city government, which blamed poor market conditions.

> These plots were supposed to generate about 18.7 billion yuan for Guangzhou's city government, representing some 29 percent of the planned land sale revenues written into the 2011 fiscal budget. Asking prices averaged 5,584 yuan per square meter of floor space (Financial Times, 2011).

Therefore, the decrease in residential real estate construction diminishes local government's ability to pay back their debts through land sales.

In addition to the scary China's local government debts reported by Financial Times and Bloomberg, Nouriel Roubini, a professor from NYU's Stern School of Business, criticized that the high borrowing by China's local governments eventually would become another "subprime mortgage" (Reuters news, 2011). His reasoning behind the prediction of China's debt problem was based on his field study in 2011. He stated that even though local governments have spent billions of dollars to upgrade city's transportation, housings, and schools, many investments cannot generate enough revenue to pay back those debts local governments borrow. He gave an example of high speed Maglev train between Shanghai and Hangzhou, another major city in eastern China. The governments of Shanghai and Hangzhou spent billions of dollars to develop the high speed train line. The initial investment was estimated around 300 million USD. Although the new high-speed train has already cut the travelling time between the two cities from four hours to less than one hour, the ticket is so expensive that frequent travelers can just take regional jet with much cheaper price. As a result, most of the trains were just half-empty. Roubini stated that the lack of passengers will reduce the profit generated from the high-speed train line and diminish local governments' capacities to pay back their debts (Reuters news, 2011) (Project-Sydicate, 2011).

The high-speed train line between Shanghai and Hangzhou is just a small corner of China's investment story. Many local governments, according

to Roubini, were developing the similar sumptuous projects, which could not generate enough revenue to pay back debts they borrowed. He worried if local governments continue the excessive investment, China's local government debts will become a new debt crisis, which will drag the whole country down.

D. Summary

At this stage, it is still not clear whether China's local government debts pose a significant threat to China's economic growth although the some economists suggests that China is facing a serious debt crisis. On the surface, the debt problem at local level in China seems to be manageable since China is still on the trend of high economic growth regardless of recent financial crisis. However, the data from China's Statistical Yearbooks from 1994 to 2011 and investigations conducted by two authoritative news agents, Bloomberg and Financial Times, indicate that the borrowing levels at local governments are still too high. The question that whether the high level of debts held by local governments through local financing platforms has significant impact on China's economic growth remains unanswered. Thus, the purpose of the rest of this paper is to provide a quantitative analysis to answer the remaining question left from the previous literature.

Before going deep into quantitative analysis, we still need to find the right variable to represent local government. Since it is impossible to find exact numbers representing China's local government debts at this point, we have to find a proxy number to represent China's local government debts. In fact, based on the previous literature, we can identify the correct variable for debt. Since this paper centers around the loans borrowed by local governments through local financing platforms, the amount of loans from all 31 provinces that report on China's statistical yearbook every year can be used as proxy variable to represent local debts. Even though it is possible that many China's local governments are willing

to hide the actual debts they owe to the banks, the amount of annual loans from each province shown in statistical yearbook, conducted by central government's banking regulation agencies and other statistical departments in Beijing, cannot cover the truth because they have to fully analyze the health of state-owned banks, which are public companies that list their shares in both mainland and Hong Kong stock exchanges. Also, given the evidence provided by Credit Suisse's report that more than 70% of the debts borrowed by local financing platforms are short term loans, which are set to mature by end of 2013, we can conclude that most of LFPs are not going to pay back these debts. Moreover, we also have evidence from Bloomberg that most local governments have not even paid attention with the amount of debts they owe for a long time and, most importantly, they do not have plans to pay them back. Thus, we can conclude that the annual short-term loans from each province shown on China's statistical yearbook are the best proxies to represent China's local government debt.

III. Model Specification

Based on the findings in cross-country growth literature and discussion on China's local fiscal history, we adopt a cross-provincial panel regression model to investigate the impact of local government debt on economic growth across thirty-one provinces in China. The panel spans 17 years from 1994 to 2011.

The baseline panel regression specification is as follows:

$$y_{it} = \alpha + \beta_{i,t} x + \gamma z_{i,t} + \varepsilon_{i,t},$$

Where t denotes the end of a period; i denotes province; y is the growth rate of real provincial income per capita; is an unobservable error term; is a vector of economic variables; is local government debt (in percent of GDP).

Since Kumar and Woo (2010) and Barro (1996) uses a similar panel regression model to analyze the impact of government debt on economic growth,

this paper will select a core set of explanatory variables that have been identified in previous literatures. The variables X in the baseline model specification are as follows: human capital, to reflect the notion that provinces with an abundance of knowledge are more likely have a greater chance of surpass other provinces in the domestic competition; social-economic factors such as population growth, inflation rate, provincial government expenditure growth, State-owned enterprise industrial output growth, and personal disposable income growth.

The econometric model introduced above is actually based on the neoclassical model, which identifies the factors of long-run economic growth. In the neoclassical model, labor and capital are the two key components of this model. Later on, technology also plays a key role in determining long-run economic growth. But new technology requires advanced knowledge in order for the technology to become fully applicable. As a matter of fact, knowledge is acquired through a long duration of education. The higher the education, the more efficient a worker will perform in the real world. Recently, more and more empirical and theoretical research illustrate that human capital plays a key role in determining economic growth. Nelson and Phelps (1966) use the experience of United States agriculture to support the evidence that better education leads to higher production level. The better educated farmer is quicker to adopt new technologies through different channels. He understands how the new technologies can help him to increase his output level since his education gives him ability to assess different ideas, and hence less likely to make errors. As a result, he will achieve a higher production level than other farmers.

Mincer (1981) and Barro (1996) identifies that investment in school education is a very essential component of human capital. Employers are willing to pay higher wages to workers who spend a lot of time and money to acquire proper skills and experience that required by the job. In addition, both Mankiw et

al (1992) and Barro (1996) include educational-attainment, which is the amount of students graduating from middle school, in his model to test whether education level has impact on economic growth. The empirical result from Barro (1996) indicates that an increase by 0.3 year of education increase the per-capita GDP growth by 0.2 percentage points per year. Mankiw et al (1992) use the percentage of working-age population that is in secondary school as a proxy to measure human capital. The empirical result from Mankiw et al (1992) also confirms the strong correlation between education attainment and economic growth. Countries with a higher level of education grow faster for a given level of initial per capita GDP. After China reopened universities after ten years of Cultural Revolution, the enrollment in higher institutions started to increase. Wang and Yao (2003) use different level of school enrollment rate as a proxy for human stock. They find that the accumulation of human capita was very quick from 1952 to 1999 and the empirical result confirms that it contributed significantly to grow and welfare. Thus, in this paper, we use the growth of higher education enrollment as an indicator of human capital.

Since this model has to use other economic determinants as control variables, we have to take in account of the other interesting characteristics reflecting on China's economic growth. Even though the theoretical growth studies have no conclusive evidence regarding the relationship between economic growth and trade openness, most empirical studies illustrate that a country with a more open trade policy will likely experience high economic growth. Barro (1991) Yanikkaya (2003) and Rodriguez and Rodrik (2001) use empirical evidence to show that there is a significant correlation between a country's trade growth rate and economic growth. An improvement in trade growth stimulates and expands domestic output. If a country uses its comparative advantage to produce product with lower opportunity cost and trade with another country, it can result in more

efficient use of its scarce resource and reach higher standards of living (Ball and McCulloch, 1996). Most coastal cities in China have achieved high levels of economic growth due to high trade volume since Deng Xiaoping opened China's economy to the world in 1989. Thus, trade growth is definitely an important variable in this model. In this paper, we calculate the growth of trade (sum of import and export) in each province starting in 1994.

Based on economic theory, unexpected inflation has three main costs to the society. First of all, unexpected inflation leads to wealth redistribution. In particular, the wealth of creditors may pass to debtors after unexpected inflation. Second, uncertainty about future price levels is likely to distort consumption, saving, borrowing and investment. In addition, variability of inflation has some significant impact on wages. For instance, variable inflation causes risk-averse workers to increase wage levels incorporated with premium, which serves as insurance to against rising price in order to maintain the current lifestyle. This will push up both nominal and real wages. Increase in nominal wage adds up the production cost of producers, and thereby causes the overall price level in the society to increase because producers are demanding higher price levels to compensate their costs (Briault, 1995) (Sidrauski, 1967). However, unexpected rising price in consumer goods may cause people to scale back their consumption. Once aggregate demand falls, investment will decrease. Thus, theoretically, high inflation has negative effect on economic growth. However, empirically, Barro (1996) only finds some country experience economic downturn after inflation level reaches certain level. According to Barro (1996), for a given rate of inflation, the variability of inflation has no significant relationship with economic growth. In this paper, we are not going to concentrate on this variable since inflation rate in the model above is only a control variable.

High population growth has been found to have a detrimental effect

on growth. If the population is growing, a part of the economy's investment is allocated to provide capital for new workers (Barro, 1996). The empirical evidence, provided by Barro (1996), confirms that the relationship between population growth and economic growth is negative. In this paper, population growth is only used as a control variable in the model. We still expect the coefficient of this variable to be positive.

The last two economic variables, industrial output growth and infrastructure investment ratio, are very important variables to explain China's economic growth. After the market reform in 1980, China's economy started to take off mainly due to urbanization and investments in fixed assets. In this paper, we use these two variables as control variables.

IV. Data and facts

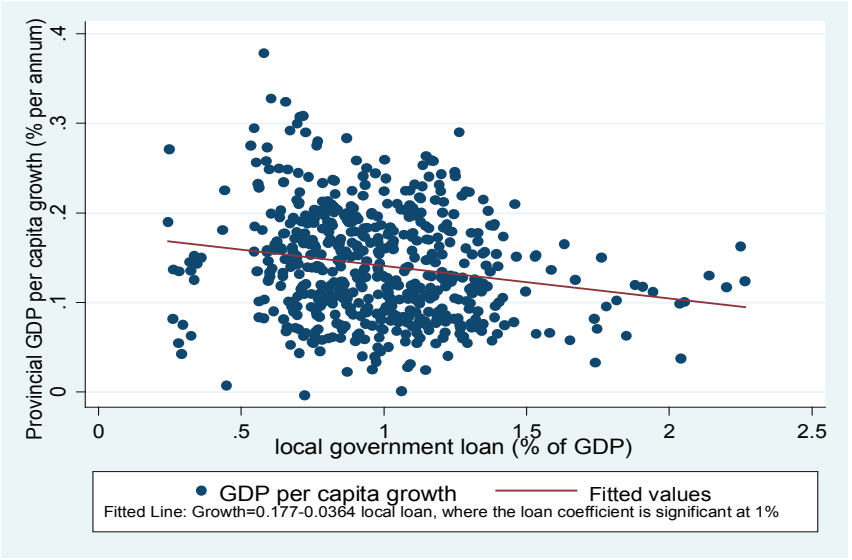
Since it is hard to find accurate data on China's local government debt due to the opaque governance in China, it is necessary to find a proxy number for the amount of debt at local level. Because provincial short-term loans, as noted in previous section, are the best variables to represent local government debts, we will use this as the proxy for debt. Based on the table 2, the growth of provincial loans from state-owned banks especially in Northern provinces has increased dramatically over the past eleven years due to boom in housing market. As we can see from the table, the total amount of short-term loans borrowed by Shanxi, a resource-rich province in Northern China, was only 802 million yuan (133 million US dollar), but in 2010, the amount of short-term loans went up to 9634 million yuan (1605 million US dollar). The 110% increase in short-term loans in Shanxi province over the past thirteen years is just a tip of iceberg among all the local governments in China. Provinces in Southern China such as Hubei, as shown in Table 3, experienced a 200% increase in short-term loans over the past 17 years.

Thus, provincial short-term loan is the best proxies to represent local government debt (China's statistical yearbook).

Another proxy used in this paper is educational achievement, which represents human capital. Since many previous theoretical models of economic growth such as those of Barro (1991), Mincer (1981), and Nelson and Phelps (1966) use educational achievement to serve as a proxy for human capital, this paper will emulate the same proxy used in those studies. Data from China's annual statistical yearbook indicates that student enrollments in higher education have increased dramatically since the end of Cultural Revolution in 1979. For instance, the total of students enrolled into higher institutions in Anhui, one of the poorest provinces in China in 1979, was only 294000. Thirty-two years later, the number of higher institution students reached 990,000 in 2011. Not only has Anhui experienced this tremendous increase in human capital, but other provinces, which had the same economic growth level as Anhui in 1979, also have enlarged its human capital over the past 30 years. The increase in stock of human capital in most provinces in China indicates that the local governments have invested a lot of their funds to improve educational level in China. As a result, this paper will use student enrollments in higher education as a proxy to represent human capital.

Other key variables such as population growth, inflation rate, trade growth, government expenditure, disposable income growth, State-owned enterprise output growth, and infrastructure investment in percent of GDP are obtained from the China's Statistical Yearbook from 1994 to 2011. The main analysis is based on a panel of 31 provinces for the period 1994-2011. The data for infrastructure investment implies that China's economy is strongly dependent on government's infrastructure investments such as bridges and highways. For example, Beijing experienced a 17% annual increase in infrastructure investment from 1997 to 2011. At the same time, Beijing's annual real GDP growth reached

almost 14% before 2007. The fast economic growth in Beijing is also happening in other provinces especially those are heavily dependent on infrastructure investments.



Data on local government debt and growth clearly show that there is a negative correlation between government debt and growth of real per capita GDP. Figure shows a scatter plot of local government debt against subsequent growth of real per capita GDP over seventeen-year periods in the full sample. According to the OLS fitted line, the coefficient of local government debt is -0.036. Taken at face value (i.e., ignoring the potential endogeneity problem, and not controlling for other growth determinants), it suggests that a 10 percentage point increase in local government debt-to-GDP ratio is associated with a slowdown in per capita GDP growth of 0.36 percentage points. This result is quiet consistent with the similar result shown in Kumar and Woo (2010).

V. Estimation Strategy

As noted above, this paper intends to fill up the research gap by analyzing the impact of China's local government debt on economic growth empirically. However, one of the most challenging obstacles facing this paper is to find an array of potential economic determinants, which can be used as control variables to explain economic growth. Even though Barro (1996) identifies an array of potential economic determinants by regressing output growth on an array of potential determinants, many variables he used in the paper cannot address the concern of robustness. By solving this issue, recent studies such as Sala-i-Martin et al (2004) focus on selecting the core set of growth determinants.

Sala-i-Martin et al (2004) uses Bayesian Averaging of Classical Estimates, which constructs estimates by averaging ordinary least squared coefficients across models, to conduct regressors selection, which can best explain economic growth. Of all 67 explanatory variables, they find 18 to significantly and robustly partially correlated with economic growth and another three variables to be marginally related. Among the 18 variables, there are a few economic variables such as share of government consumption in GDP, primary school enrollment, trade openness, and the average price of investment goods. The rest of the variables relates to socio-political factors such as fraction of population Confucian, which used to explain the economic growth in East Asia.

This paper also consider a variety of estimation methodologies, such as pooled OLS, random effect (RE), fixed effects (FE) panel regression, and GLS regression. Each of estimation method has a trade-off. Although one method corrects one particular econometric problem, it sometimes can lead to a different type of bias. For example, the slope coefficients from pooled OLS regression sometimes have the expected signs and the value is reasonably high. It also assumes that the slope coefficients of the X variables are all identical for all the

provinces. Obviously, there are restricted assumptions for OLS regression. Thus, the pooled regression may distort the true picture of the relationship between Y and the X's across the 31 provinces in China. One way to take into account the individual characteristic of each province is to let the intercept vary for each company but still assume that the slope coefficients are constant across provinces (Gujarati, 2003).

In addition, OLS regression sometimes encounters the presence of heteroscedasticity because the usual OLS method does not follow this strategy and thus does not make use of the "information" contained in the unequal variability of the dependent variable Y. In this case, generalized least squares (GLS) takes such information into account explicitly and is therefore capable of producing estimators that are best linear unbiased estimators (BLUE) (Gujarati, 2003).

VI. Empirical results

This section reports econometric results estimating cross-province GDP per capita growth rates in China. Five regressions are run on 557 observations with 17 years of time span. Among the 31 provinces in China, autonomous regions inhabited by different ethical groups such as Yunnan and Guangxi are under direct control of central government. Also, municipalities such as Shanghai and Beijing are under the authority of central government.

The main results for local economies are presented in Table 1. Columns 1-5 show that the coefficients of local government debt are negative and are significant at the 1 percent level besides Column 1, with their values ranging from -0.022 to -0.047 across the various estimation techniques. The ordinary least squares (OLS) estimation in Column 1 suggests a 10 percentage points of GDP increase in local government debt is associated with a slowdown in growth in real GDP per capita of around 0.24 percent per year. The random effect (RE) and

general least square (GLS) in Columns 4 and 5 yield the same result. Even though the FE estimate of debt coefficient in Column 3 is somewhat larger than other estimates, it is still significant at 1 percent level. Also, the goodness of fit is quite promising, with an adjusted ranging from 0.63 for FE estimation to 0.8 for OLS estimation.

The coefficients on other explanatory variable (trade growth, government expenditure growth, disposable income growth, industrial output growth and investment in infrastructure to GDP ratio) are of the expected sign and mostly significant at 1% percent level across different estimation techniques. However, the coefficient for inflation is inconsistent with Barro's theory that inflation has negative effect on economic growth because in China's case, inflation tends to have positive effect on economic growth. Besides OLS estimation, all other estimations for inflation are significant at 1%.

After running test on heteroskedasticity given by Breusch-Pagan test, which tests the null hypothesis that the variance of the residuals is homogenous. Since the p-value is very small in this OLS regression, we would have to reject the hypothesis and accept the alternative hypothesis that the variance is not homogenous. Besides test on heteroskedasticity, we also conduct test on Multicollinearity. We use VIF, variance inflation factor, to test whether the OLS has strong multicollinearity. As a rule of thumb, a variable whose VIF values are greater than 10 may merit further investigation. In this case, the VIF and tolerance (1/VIF) value for yeardum2 is worrisome. However, since yeardum2 is a dummy variable control for year, the high VIF is negligible.

At last, model specification is final test we conduct on this OLS regression. A model specification error can occur when one or more relevant variables in the model. If relevant variables are omitted from the model, the common variance they share with included variables may be wrongly attributed to those variables,

the error term is inflated. We first use a link test command performs a model specification link test for single-equation models. We will be looking at the p-value for **_hatsq,** which stands for the variable of squared prediction**.** If the model is specified correctly, the squred predictions should not have much explanatory power. That is we would not expect the variable of squared prediction to be a significant predictor if the model is specified correctly. In this case, the p-value for _hatsq is 0.256. It indicates that the linktest has failed to reject the assumption that the model is specified correctly. Thus, it seems to us that we do not have a specification error. But now, we will conduct another test before we jump to the conclusion. After running ovtest, the test result indicates there is small probability that the model has omitted variables since p-value, 0.0483, is close to 0.05. The model may have specification bias, but both ovtest and linktest indicate the bias is not significant enough to be considered as problem.

Since the dataset for this regression model is a panel data, we have also run a panel regression with fixed effects. The result is similar to OLS regressions, but more variables become statistically significant. The coefficient of debt is negative and significant at 95 percent of confidence interval. Also, the coefficient of -0.026 for FE is still in a similar range with OLS. Besides running fixed effect regression, we also consider random effect regression. After running a Hausman test to test decide between fixed or random effects. The null hypothesis is that the preferred model is random effects vs. the alternative the fixed effects. It basically tests whether the unique errors are correlated with the regressors, the null hypothesis is they are not. Since the p-value is less than 0.05, we can conclude that we should consider fixed effect regression as the best choices for panel regression.

After running heteroskedasticity test for FE, the test results still indicates that the model has strong heteroskedasticity. Also, the Lagram-Multiplier test for serial correlation also indicates that the model has strong serial correlation. As a

result, in order correct both heteroskedasticity and serial correlation, general-least squares model is considered. The coefficient for debt in GLS regression is still significant with a correct sign.

VII. Conclusion

Given the empirical evidence on the impact of debt on economic growth for a panel of 31 provinces in China over the period of 1994-2011, we can conclude that local government debt in China does have significant impact on China's economic growth. The results, based on a range of econometric techniques, suggest an inverse relationship between debt and economic growth, controlling for other determinants of growth: on average a 10 percentage point increase in the debt-to-GDP ratio is associated with a slowdown in annual per capita GDP growth of around 0.23 percentage points per year.

Although econometric model cannot provide a concrete result that China has already in a fiscal crisis, the negative correlation between debt-to-GDP ratio and economic growth suggests that if local governments in China continue the current spending path without considering any future consequences, local government debt eventually will follow the path of Greece and other countries with high government debts.

Table 1. Panel Regression-Growth and Local Government Debt, 1994-2011

Sample: 31 Provinces in China
Dependent Variable : Provincial Real per Capita GDP Growth

Explanatory Variables	OLS (1)	Panel Regression (2)	FE (3)	RE (4)	GLS (5)
Population growth	-0.024	-0.045	-0.023	-0.045	-0.045
	-0.037	-0.047	-0.048	-0.047	-0.046
Debt to GDP ratio	-0.022*	-0.027**	-0.042**	-0.027**	-0.027**
	-0.009	-0.005	-0.011	-0.005	-0.005
Inflation rate	0.018	0.205**	0.219**	0.205**	0.205**
	-0.084	-0.05	-0.05	-0.05	-0.049
Trade growth	0.013*	0.041**	0.042**	0.041**	0.041**
	-0.006	-0.006	-0.006	-0.006	-0.006
Higher institution student enrollment growth	-0.033	-0.117**	-0.129**	-0.117**	-0.117**
	-0.036	-0.021	-0.021	-0.021	-0.021
Government expenditure growth	0.077**	0.032	0.02	0.032	0.032
	-0.022	-0.024	-0.025	-0.024	-0.024
Disposable income growth	0.161**	0.523**	0.507**	0.523**	0.523**
	-0.041	-0.029	-0.029	-0.029	-0.028
Industrial output growth	0.027**	0.032**	0.027*	0.032**	0.032**
	-0.01	-0.012	-0.012	-0.012	-0.012
Investment in infrastructure to GDP ratio	0.082**	0.050**	0.053**	0.050**	0.050**
	-0.015	-0.012	-0.013	-0.012	-0.012
Constant	-0.012	0.082**	0.103**	0.082**	0.082**
	-0.017	-0.01	-0.015	-0.01	-0.01
Observations	557	557	557	557	557
R-squared	0.8		0.63		
Number of Province Name		31	31	31	31
Standard errors in parentheses					
+ significant at 10%; * significant at 5%; ** significant at 1%					

Explanatory Variables	OLS (1)	Panel Regression (2)	FE (3)	RE (4)	GLS (5)
Population growth	-0.024	-0.045	-0.023	-0.045	-0.045
	-0.037	-0.047	-0.048	-0.047	-0.046
Debt to GDP ratio	-0.022*	-0.027**	-0.042**	-0.027**	-0.027**
	-0.009	-0.005	-0.011	-0.005	-0.005
Inflation rate	0.018	0.205**	0.219**	0.205**	0.205**
	-0.084	-0.05	-0.05	-0.05	-0.049
Trade growth	0.013*	0.041**	0.042**	0.041**	0.041**
	-0.006	-0.006	-0.006	-0.006	-0.006
Higher institution student enrollment growth	-0.033	-0.117**	-0.129**	-0.117**	-0.117**
	-0.036	-0.021	-0.021	-0.021	-0.021
Government expenditure growth	0.077**	0.032	0.02	0.032	0.032
	-0.022	-0.024	-0.025	-0.024	-0.024
Disposable income growth	0.161**	0.523**	0.507**	0.523**	0.523**
	-0.041	-0.029	-0.029	-0.029	-0.028
Industrial output growth	0.027**	0.032**	0.027*	0.032**	0.032**
	-0.01	-0.012	-0.012	-0.012	-0.012
Investment in infrastructure to GDP ratio	0.082**	0.050**	0.053**	0.050**	0.050**
	-0.015	-0.012	-0.013	-0.012	-0.012
Constant	-0.012	0.082**	0.103**	0.082**	0.082**
	-0.017	-0.01	-0.015	-0.01	-0.01
Observations	557	557	557	557	557
R-squared	0.8		0.63		
Number of Province Name		31	31	31	31
Standard errors in parentheses					
+ significant at 10%; * significant at 5%; ** significant at 1%					

	Beijing	Tianjin	Hebei	Shanxi	Inner Mongolia	Liaoning	Jilin	Heilongjiang
1994	1,428.95	927.05	1,310.16	802.06	674.37	2,412.02	1,099.77	1,396.60
1995	1,779.05	1,113.95	1,578.21	1,223.11	819.87	2,881.80	1,302.76	1,636.80
1996	2,082.83	1,357.38	1,894.66	1,420.11	1,002.98	3,556.74	1,608.84	1,909.80
1997	2,720.68	1,502.91	2,372.15	1,524.98	1,172.17	3,980.29	1,913.61	2,381.10
1998	3,326.57	1,629.12	2,795.20	1,741.79	1,318.75	4,439.65	2,118.79	2,702.90
1999	4,007.76	1,825.26	3,038.32	1,909.21	1,364.17	4,833.76	2,580.41	3,103.94
2000	5,944.60	1,863.60	2,933.19	2,453.15	1,340.74	5,195.56	2,651.18	3,145.10
2001	7,205.99	2,159.86	3,098.89	2,408.40	1,470.75	5,597.40	2,828.25	3,358.60
2002	9,230.78	2,519.04	3,488.18	2,903.18	1,649.78	6,247.40	3,057.70	3,624.00
2003	11,343.28	3,426.02	3,854.72	3,552.29	1,924.13	7,222.30	3,288.87	3,981.30
2004	13,577.45	3,821.38	6,152.20	4,016.12	2,239.76	7,753.00	3,435.02	4,038.90
2005	15,335.50	4,722.38	6,415.20	4,328.90	2,588.57	7,958.05	3,401.30	3,658.50
2006	15,486.90	5,106.94	7,411.88	4,788.51	3,205.19	9,117.20	3,870.30	3,971.90
2007	17,360.20	6,131.63	8,397.82	5,394.47	3,767.74	10,403.88	4,306.00	4,256.40
2008	19,431.08	7,277.46	9,453.30	5,960.33	4,527.86	11,794.60	4,835.89	4,532.70
2009	24,805.10	10,513.44	13,123.80	7,814.74	6,292.52	15,549.60	6,234.70	5,988.30
2010	28,748.10	12,864.75	15,755.74	9,634.32	7,919.47	18,689.80	7,205.94	7,230.50
2011	33,367.00	15,242.20	18,144.00	11,169.35	9,727.70	21,621.00	8,122.41	8,548.70

	Hubei	Hunan	Guangdong	Guangxi	Chongqing	Sichuan	Guizhou	Yunnan
1994	1,396.82	1,263.11	4,339.76	835.53	596.96	2,218.14	421.07	684.68
1995	1,750.41	1,494.03	5,495.69	1,055.67	755.39	2,804.16	513.41	924.67
1996	2,273.20	1,880.94	6,319.77	1,203.41	913.93	3,359.54	610.51	1,194.51
1997	2,979.33	2,123.00	8,195.83	1,423.48	1,156.13	2,948.15	761.88	1,496.95
1998	3,500.36	2,274.41	9,523.56	1,516.49	1,358.61	3,151.74	840.63	1,713.97
1999	3,528.13	2,408.36	10,934.76	1,719.19	1,611.68	3,924.16	899.83	1,824.04
2000	3,493.91	2,403.39	11,636.25	1,613.25	1,881.29	4,053.46	1,064.82	1,987.83
2001	3,787.25	2,787.92	13,093.72	1,764.05	1,871.98	4,498.55	1,212.23	2,173.45
2002	4,312.79	3,227.46	15,206.62	1,941.07	2,244.72	5,158.76	1,403.92	2,418.48
2003	5,000.74	3,796.31	18,190.85	2,320.66	2,774.81	5,910.59	1,714.04	2,955.57
2004	5,377.43	4,258.03	19,491.77	2,759.65	3,246.28	6,475.92	2,020.04	3,398.29
2005	5,649.67	4,590.03	23,261.21	3,056.86	3,719.52	6,743.00	2,303.90	3,987.58
2006	6,430.44	5,173.87	23,182.16	3,595.25	4,388.28	7,833.32	2,696.11	4,803.51
2007	7,496.46	6,037.40	26,776.12	4,287.79	5,131.69	9,200.93	3,128.63	5,671.66
2008	8,465.64	6,989.42	30,224.01	5,066.68	6,320.81	11,163.39	3,569.27	6,594.33
2009	11,659.37	9,369.81	38,893.06	7,268.41	8,766.06	15,680.33	4,656.50	8,779.63
2010	14,136.58	11,303.76	47,191.56	8,867.52	10,888.15	19,129.79	5,747.50	10,568.78
2011	16,395.39	13,186.68	53,411.83	10,408.54	13,001.39	22,033.21	6,841.92	12,114.59

Bibliography

Beamish, Freya. “Korea, Hong Kong, and Taiway: and the winner is?” *Lombard Street Research Montly Review* 279 (2012): 24.

Choyleva, Diana. “China’s “shadow” banking shenanigans.” *Lombard Street Research Monthly Review* 272 (2011): 22.

—. “China’s banks- a disaster waiting to happen.” *Lombard Street Research Monthly Review* 276 (2012): 26.

—. “China’s challenge: letting the ‘invisible hand’ boost consumption.” *Lombard Street Research Monthly Review* (2010): 31.

Dalio, Ray. *An In-Depth Look at Deleveraging*. Monthly presentation. London: Bridgewater, 2012.

Eichengreen, Barry, Donghyun Park and Kwanho Shin. “When Fast Growing Economies Slow Down: International Evidence And Implications for China.” *NBER Working Paper Series* (2011): 1-50.

Elmendorf, Douglas W. and N. Gregory Mankiw. “Government Debt.” *Handbook of Macroeconomics* (1998): 1-73.

Fan, Bonai, et al. “Government Waste in China: Situation, Causes, and Countermeasures.” *JSTOR* (2010): 684-708.

Forsythe, Michael. “Bloomberg .” 18 December 2011. *China Debts Dwarf Official Data With Too-Big-to Finish Alarm.* 09 December 2012.

Mackenzie, Kate. “Financil Times.” 21 December 2011. *China’s worried local government.* 09 December 2012.

Reinhart, Carmen M. and S. Kenneth Rogoff. “Growth In A Time of Debt.” *NBER Working Paper Series* (2010): 1-26.

Roubini, Nouriel. “Project Syndicate.” 13 September 2012. *Fiddling at the Fire* . 9 December 2012.

Shirakawa, Masaaki. “Deleveraging and Growth: Is the Devloped World Following Japan’s Long and Winding Road?” *Lecture at the London School of Economics and Political Science*. London: Bank of Japan, 2012. 24.

Standard & Poor’s Ratings Services. “Asia-Pacific Banking Outlook in 2012.” *S&P Monthly Outlook* (2012): 10-68.

Stein, Gabriel. “Beijing’s false analogy with Japan.” *Lombard Street Research Monthly Review* 257 (2010): 13.

Taylor, Michael. “Is a Japanese sovereign deby crisis inevitable?” *Lombard Street Research Montly Review* 280 (2012): 16.

—. “Korea and Japan-reflating and deflating.” *Lombard Street Research Monthly Review* 249 (2009): 15.

Walter, Carl E. and Fraser J. T. Howie. *Red Capitalism: The Fragile Financial Foundation of China’s Extraodinary Rise*. Singapore: John Wiley & Sons, 2011.

Woo, Jaejoon and S. Manmohan Kumar. “Public Debt and Growth.” *IMF Working Paper* (2010): 1-47.

Yerxa, Gavin. “Project Firefly.” 11 December 2011. *Is local government debt a serious threat to the Chinese state.* 9 December 2012 .

Zhu, Ze. *Essays on China’s Tax System*. PhD Thesis. Rotterdam: Erasmus University , 2007.

Renewable Resource Extraction:
Experimental Analysis of Resource Management Policies Under Assumptions of Resource Migration

Kevin Lugo

Abstract
This paper presents research using a spatially explicit and dynamic common pool resource experiment to compare renewable resource extraction behavior between four treatments combining (1) open access and sole ownership institutions with (2) mobility and non-mobility of the renewable resource. The primary purpose of this research is to test the theory that introducing resource mobility into a sole ownership regime will remove the incentive for subjects to maximize the resource, instead causing them to revert to the myopic strategy predicted for the open access regime. I also test the hypothesis that open access firms are indifferent to resource dispersal. The results show that efficiency is unaffected by dispersal but the behavior of sole owners differs between dispersal conditions. Extraction requests increase at a faster rate under dispersal, fewer tokens remain unextracted in the final period, and some subjects show strategic behavior resulting in greater than 100% efficiency. This is a pilot study that presents preliminary evidence of a behavioral change. The results are subject to experimental factors such as subject misperceptions of linearity and statistical significance suffers from a small subject pool.

1. Introduction

With population growth and economic development continually increasing world demand for natural resources, policy makers must have a robust understanding of the physical and economic factors affecting resource extraction. In 1987 the World Commission on Environment and Development, also known as the Brundtland Commission, published *Our Common Future*. This report introduced the concept of sustainable development and elevated natural resource extraction to a prominent place in policy discussions. Among the many natural resources, renewable resources are of particular interest. These resources are extremely important and include resources like fish and timber, cornerstones of major commercial industries, as well as surface water and aquifers, essential inputs for agriculture and human consumption.

Many renewable resources can be described as common-pool resources (CPRs) because (1) extraction produces an externality, reducing the resource stock and increasing the extraction cost of other firms and (2) restricting access, while not impossible, is difficult and costly. Over 40 years ago Garret Hardin (1968) presented a pasture shared by a number of herdsmen as a classical example of a CPR. According to Hardin, the profit-maximizing herdsman decides whether or not to add an additional herd based on a simple cost-benefit analysis. Because the marginal benefit of the additional herd accrues exclusively to the herdsman but the marginal cost is an externality divided between every herdsman, he chooses to add another herd. Unfortunately, when every herdsman pursues "his own best interest in a society that believes in the freedom of the commons," the outcome is "ruin [for] all." This is the tragedy of the commons. More generally, Mancur Olson argued in *The Logic of Collective Action* that rational actors seeking their own self-interest will not act in the common good without a separate incentive (1965). Works such as these have played a pivotal role in developing policy interventions meant to modify economic incentives.

One such intervention is property rights. In the absence of property rights, CPRs operate under an open access regime where there is unrestricted entrance to the market and therefore potential for the tragedy of the commons. In his 1960 article "The Problem of Social Cost," Ronald Coase presented transferable property rights as an efficient solution to externalities under a specific set of conditions. Assigning sole ownership is one extension of this principle and has proven a solution to a number of CPR problems (Hilborn et al, 1995; Johannes, 1978). For natural resources sole ownership is often established by granting a single firm sole extraction rights within a given geographic area. To assume that the benefits of sole ownership applies to this type of allocation regime is to implicitly assume that the sole owner controls every aspect of that resource.

This paper examines the effects of relaxing this assumption by considering the case of a mobile renewable resource.

Imagine an island nation that is physically isolated from other countries by a large expanse of sea and that has access to a fishery resource within its exclusive economic zone (EEZ)[8]. Assume the fish in this zone are incapable of crossing the physical barriers separating the island from other nations. In this situation the country is not only the sole owner of extraction rights, but also the sole owner of the resource. On the other hand, consider countries bordering the Mediterranean Sea. While these countries also have their own EEZs, the geography allows fish to travel across EEZ borders. The countries in this situation have sole ownership of the extraction rights within their respective EEZ. However, the ability of the resource to move between EEZs prevents any one country from having sole ownership of the resource itself. In this example Mediterranean countries continue to face a common-pool resource and the associated tragedy of the commons.

Resource mobility and spatial attributes are the key factors that create this scenario. Economic research on renewable resources has only recently begun to consider spatial dynamics and resource mobility, and there is a significant lack of experimental work on the subject. This paper presents an experimental methodology that examines the effect of spatial dynamics and resource dispersal on the outcomes of open access and sole ownership regimes. This is, to the best of my knowledge, the first study to do so. To examine this effect I recruit undergraduate students to participate in a dynamic CPR experiment modified to include spatial dynamics and resource mobility.

Subjects are placed into groups of three and tasked with making token extraction decisions from either three common zones (in the open access

8 An exclusive economic zone (EEZ) is an area consisting of all waters within 200 nautical miles of a country's coastline. A country has exclusive fishing rights within this EEZ.

treatments) or one private zone (in the sole ownership treatments). At the end of each period a logistic growth function is applied to remaining tokens to simulate natural resource growth, and this growth is added to the zone's previous token stock. I introduce resource mobility in the form of simplified, density-dependent dispersal that equalizes the token endowment across zones prior to the start of each period. I examine the open access regime and the sole ownership regime both with and without resource dispersal, for a total of four treatments. Payoffs are directly proportional to the total number of tokens extracted by each subject.

The primary purpose of this research is to test the theory that introducing resource mobility into a sole ownership regime will remove the incentive for subjects to maximize the resource, instead causing them to revert to the myopic strategy predicted for the open access regime. I also test the theory that open access firms are indifferent to resource dispersal because the myopic strategy exhausts tokens in the first period, making resource dispersal irrelevant. The main findings are that:

1. Average efficiencies are unaffected by resource dispersal.

2. The behavior of sole owners differs when dispersal is introduced.

    2.1. Dispersal causes requests to consistently increase every period, whereas without dispersal requests appear to follow a cyclical pattern.

    2.2. Under dispersal, increasing requests cause fewer tokens to remain in the final period.

    2.3. Dispersal allows many subjects to achieve greater than 100% efficiency, suggesting that some subjects capitalized on opportunities to harvest tokens dispersed from other zones.

3. Open access behavior is not significantly affected by dispersal

The results show that the research design could be improved by respecifying parameters and extending the experiment length in order to capture long-term behavioral adaptation to dispersal. In this sense the results also highlight some shortcomings of the experimental methodology which will be discussed in detail later.[9] However, as a pilot study of mobile renewable resource policies, this research shows that sole owners facing dispersal do not immediately follow the dominant strategy of myopic extraction.

The paper is organized as follows. Section 2 provides a literature review of renewable resource policy, experimental findings from CPR experiments, and justification for the experimental methodology. Section 3 develops the CPR model and explains the experimental design. Section 4 presents the hypotheses. Section 5 presents the results. Section 6 discusses the findings and section 7 briefly concludes and makes recommendations for future research.

2. Literature Review

2.1 Renewable Resource Policies

There is substantial literature on renewable resource extraction in the both economics and the physical sciences. Fisheries have been studied extensively in the last century and have probably generated the most policy-related research. Modern fishery management policies began after WWII with open access regulations that dictated the number and species of fish that could be caught as well as how, when, and where fishing could occur (Wilen, 1999). During this time new research began connecting biological understandings of fishery stock dynamics to economic behavior and suggested that economic incentives could be manipulated to achieve targeted policy outcomes (Beverton and Holt, 1957;

9 This was a pilot study with limited resources, time, and funding. A number of parameters showed differences across treatments but were not significant at the 5% level. See the results and discussion for details.

Gordon, 1954; Schaefer, 1957; Scott, 1955). Now commonplace, the concept of maximum sustainable yield (MSY) was developed as a policy goal to maximize both economic yield and biological growth (Schaefer, 1957).

The exclusive economic zones created in 1982 by the United Nations Convention on the Law of the Sea introduced limited entry techniques to fishery management by excluding foreign fishers from very large domestic fishing markets (Wilen, 1999).[10] In contrast to the open access regulations already in place, limited entry techniques attempt to constrain overexploitation by limiting the number of firms that participate in an industry. At the most extreme, a limited entry technique might establishment a government agency, nonprofit, or private firm as the sole owner of a renewable resource. Economic theory of the fishery suggests that sole ownership will result in a more socially optimal outcome than an open access regime. In open access an individual firm receives the market's average revenue rather than its own individual marginal revenue, thus incentivizing an over-allocation of effort (Gordon, 1954; Schaefer, 1957). In contrast, sole owners maximize profit by producing where marginal cost is equal to marginal revenue. This corresponds to the optimal level of extraction that lowers effort allocation and maximizes net economic yield (Gordon, 1954) (Figure 1).

When a sole owner is the only producer in a market it can increase its profit by charging a monopoly price higher than the competitive equilibrium. However, many renewable resources exist over a wide geographical area, making it possible to establish many sole owners of distinct resource patches. This prevents any one firm from establishing monopoly power and increasing the market price. Under this assumption, a sole owner who maximizes the present value of a fishery would find equilibrium closer to the social optimum than would firms in competitive equilibrium

10 Prior to this convention, countries' exclusive rights consisted of only their territorial waters within 12 nautical miles from their coastline. Through this act exclusive rights were extended to 200 nautical miles

(Scott, 1955). Sole ownership has proven one of the most effective institutions in promoting sustainable extraction of some renewable resources (Hilborn et al, 1995). In Palau, Micronesia, control of fishing rights by chiefs helped maintain fishing stocks and allowed mutually beneficial fishing transactions in the vein of the Coase theorem (Johannes, 1978). Experimental evidence has shown that sole owners are able to achieve high levels of efficiency in resource extraction decisions when given adequate information (Hey et al, 2008).

2.2 Common Pool Resource Experiments

CPR experiments typically find that self-interested behavior prevents collective action from following an optimal path. Walker, Gardner, and Ostrom (1990) run an experiment where subjects repeatedly choose to invest tokens in either a fixed-return investment or a CPR with decreasing marginal returns. Subjects consistently overinvest in the CPR and even create negative returns when given a large enough endowment. A further study with probabilistic destruction finds that while most subjects appear to play "safe" strategies, a few myopic subjects overinvest and create early termination and low efficiency for the group (Ostrom et al, 1992).

Herr, Walker, and Gardner (1997) use the same basic design to compare performance between time independent and time-dependent treatments. They find that time dependency increases myopic behavior and exasperates the tragedy of the commons. High availability and low marginal extraction costs during initial periods cause subjects to over-extract early on, thus reducing future profits. This coincides with the "fall down" seen in emerging forestry and fishery industries where extractors face a dramatic drop in yield once they harvest all the original and unsustainable stock (Hilborn et al, 1995). The consequences of overcapitalization based on initial harvests are discussed elsewhere in the literature (Johannes, 1978; Moxnes 1998, 2000; Rouwette et al, 2004; Walker et al, 1990; Wilen, 1999).

Economics research has largely ignored spatial distributions until the 1960s and 1970s when biologists began examining discrete spatial distributions (Wiens, 1976). Levin (1976) develops a general mathematical model of populations in patchy environments and discusses density-dependent dispersal and the creation of uniformity across patches. Sanchirico and Wilen (1999) develop a differential equations model of a patchy fishery system that includes spatially distributed effort allocation that adjusts according to resource dispersal patterns. Schnier (2009) adapts the CPR design used by Walker, Gardner, and Ostrom (1990) to examine the sink-source spatial dispersal model discussed in Sanchirico & Wilen (1999). He finds that the spatial component decreases average net returns because subjects consistently over-extract in the more plentiful source CPR. On the other hand, a study of groundwater extraction with a spatial component that increases private costs relative to external costs found that subjects were less likely to behave myopically with the spatial component (Suter et al, 2012).

2.3 Why Use an Experimental Methodology

The experimental methodology lends itself well to research on renewable CPRs for several reasons, one being the challenges facing empirical studies. Data on renewable resources are often limited and rarely accurate due to the inherent difficulty of measurement (i.e. fish stocks are impossible to observe directly and many fisheries rely on informal logbooks for economic data). Estimating economic parameters that rely on biological data is therefore extremely difficult to do with confidence. Most Schaefer production functions of fisheries typically overestimate carrying capacities and maximum sustainable yields which can lead to potential fishery collapse (Zhang & Smith, 2011). Research directly analyzing policy must use extra caution because management policies are sensitive to measurement error and increasing levels of measurement error increase sensitivity to stochastic variation (Moxnes, 2003). Poor estimation has overinflated our perception of the

health of resources and has played a major role in the collapse of some potentially sustainable resources (Hilborn et al, 1995).

Experimental methods can solve some of these problems. Researchers can design a system with known biological parameters to mimic any type of resource environment. If desired, information about the resource can be withheld and stock signals to subjects can be intentionally blurred to simulate real-world uncertainty. Renewable resource extraction is extremely complex because it relies on economic and biological variables that are not only endogenously determined, but also impacted by innumerable exogenous effects such as weather and macroeconomic variables. Experiments use random assignment to control for these confounding variables that are difficult to control and, in some cases, impossible to observe (Angrist & Pischke, 2010; Leamer, 2010). By isolating the institutional and environmental changes from these confounding factors, this research attempts to discover the causal effects of dispersal on outcomes and behavior within open access and sole ownership policies.

3. Model & Experimental Design

3.1 The Model

This research uses a standard Gordon-Schaefer model of fishery extraction (Gordon 1954; Schaefer 1957). This model defines stock, extraction cost, and resource harvest as functions of effort. The resource is modeled as having logistic growth where growth is dependent on the resource's intrinsic growth rate and stock density relative to the carrying capacity of the environment (Figure 2).[11] N represents stock, K carrying capacity, and r intrinsic growth rate. Maximum sustainable yield (MSY) occurs where growth and catch are maximized; in a standard logistic growth function this happens at one half the carrying capacity (Schaefer, 1957).

11 Carrying capacity is determined by factors such as food, competition, predation, etc. Intrinsic growth rate represents the per-capita growth rate of a population.

The current research investigates a resource inhabiting three distinct patches referred to as "zones." I use a standard logistic growth function to model resource growth within each zone and I use a simple density-dependent dispersal that equalizes population densities across zones to model resource mobility (Figure 3). Growth & stock are modeled as:

Growth: $$F(N_{it}) = rN_{it}(1 - \left(\frac{N_{it}}{K_i}\right) \quad (1)$$

New Stock(no dispersal): $$N_{i(t+1)} = N_{it} + F(N_{it}) \quad (2)$$

New Stock (dispersal): $$N_{i(t+1)} = \frac{\sum_{i=1}^{I}[N_{it} + F(N_{it})]}{I} \quad (3)$$

In this design the total number of periods (T) is set equal to 5 while the number of zones (I) is set equal to 3. The carrying capacity (K) is 100 tokens and the intrinsic growth rate (r) is set to 1. N_{it} represents stock in zone I at time t, and the experiment begins as a virgin resource meaning N_o=K=100. There is no harvesting cost and the discount rate is assumed to be zero. Token harvests (x) are positively and directly proportional to cash payments by σ. Thus payoffs to subject n can be represented by:

Payout: $$\pi_n = \sigma \sum_{t=1}^{T} (\sum_{i=1}^{I} x_i) \quad (4)$$

I develop two benchmark strategies for evaluating the results of the experiment. Substituting into equation (1) and solving for gives a growth maximizing stock level of 50 tokens. In this environment a rational agent thinking dynamically would maximize profit by reducing the stock in all available zones to the optimal level of 50 tokens in the first period. The agent would then harvest only the new growth in periods 2-4 to return the stock to the growth-maximizing level. In the final period the agent would extract all remaining tokens. Thus the maximization strategy (MAX) is to request (50, 25, 25, 25, 75) tokens which achieves 100%

efficiency (Figure 4). The second benchmark is extremely simple and considers a rational agent who thinks only in terms of present conditions. Such an agent would request all the tokens from each zone in the first period, leaving no tokens for future consumption. The MYOPIC strategy is thus (100, 0, 0, 0, 0) with an efficiency of 50%. The logic behind this benchmark is explained in the next section.

3.2 Experimental Design

A total of 113 subjects were recruited to participate in this experiment.[12] Subjects were undergraduate students of Gettysburg College randomly recruited through email. They were told they would receive $5 for participating in the experiment and would be able to earn an additional cash payment based on their performance in the experiment. Payoffs are designed such that the average subject was projected to receive approximately $10 for 20 to 40 minutes of participation.

The experiment uses the z-Tree experimental software.[13] The basic design follows a standard CPR experiment wherein subjects are placed into groups of three and must make individual token extraction decisions for five periods. After each period any remaining tokens within a zone grow according to the logistic growth function described in equation (1). The tokens available in the next period are calculated using equation (2) if the treatment does not include dispersal or equation (3) if dispersal is present. After each period subjects see a summary of the results from that period as well as the tokens available in the next period. Each stage of the periods is set to a nonbinding timer to encourage timely action.

There is one zone per subject. Extraction requests need not be whole numbers but may never exceed the endowment available in the zone. Tokens extracted by a subject are credited to her account and, at the end of the experiment,

12 See Appendix B for details on recruitment.

13 Fischbacher, Urs. "z-Tree: Zurich Toolbox for Ready-made Economic Experiments." *Experimental Economics* 10, no. 2 (2007): 171–178.

are exchanged for dollars at a known rate of $0.05 per token. There are a total of four treatments: (1) open access with no dispersal (OAND); (2) open access with dispersal (OAD); (3) sole ownership with no dispersal (SOND); and (4) sole ownership with dispersal (SOD).

In the open access treatments subjects are free to request tokens from all of the three common zones. If the group request from a zone is less than the zone's endowment, each subject receives a number of tokens equal to her request. If the group request from a zone is greater than the zone's endowment, each subject receives a number of tokens equal to her proportion of the group request**.** Period summaries display the group withdraw, the number of tokens received by the subject, and the number of tokens available in the next period for each zone. In the sole ownership treatment each subject may only request tokens from her own private zone. Likewise, period summaries only display the number of tokens received and the number of tokens available in the next period for the subject's own zone.

An important component of the design is that subjects are given both the exact growth function as well as a table of growth possibilities.[14] The experimenter reads the description of growth and dispersal aloud and gives subjects time to review the instructions before beginning the experiment. This availability of information is not representative of real-world renewable resources. However, giving subjects this information removes the confounding affect of adaptive management strategies subjects would need to use to determine the optimal stock level and extraction pattern.[15]

14 See Appendix C for experimental materials

15 See Hey et al (2008) for a discussion of a "reasonable" benchmark strategy when stock is known but the growth function is unknown.

4. Hypotheses

The current research tests three hypotheses concerning the effects of dispersal on outcomes: (1) subjects in sole ownership without dispersal follow the MAX strategy; (2) subjects in open access follow the MYOPIC strategy, regardless of dispersal condition, and (3) introducing resource dispersal to sole ownership causes subjects to follow the MYOPIC strategy rather than the MAX strategy. These hypotheses rest on the assumption that individual subjects are rational, self-interested, profit-maximizing actors.

Economic theory suggests that the equilibrium extraction level of a sole owner of a renewable resource maximizes the present value of the resource in the absence of price factors (Gordon, 1954; Scott 1955). Past studies have shown that sole owners without full information achieve poor outcomes (Moxnes 1998, 2000, 2004) while subjects not directly given full information but with access to it fall somewhat short of complete maximization (Hey et al, 2008). Subjects in this study are directly given both the growth function as well as a table of every possible integer stock level and its corresponding growth. Thus subjects in SOND are expected to follow the MAX strategy and obtain 100% efficiency.

Rational choice theory, collective action theory, and theory of the commons predict that subjects in the open access treatments will act in their own best interest and not achieve the goals common to each subject. The theory of myopic loss aversion suggests that subjects emphasize present conditions over future conditions and will assign more weight to potential losses than potential gains when making extraction decisions (Benzarti & Thaler, 1993). Results from CPR experiments show that subjects consistently tend to act myopically (Ostrom et al, 1992; Herr et al, 1997; Walker et al, 1990; Walker & Gardner, 1992). From a game-theory perspective, if each subject knows that the other subjects will request all tokens in the final period, it is then strategic to request

all tokens in the penultimate period. Using backwards induction it becomes clear that the dominant strategy of each subject is to extract all tokens in the first period. Because dispersal only affects next period endowments, dispersal should have no effect on open access outcomes. Thus, subjects in OAND and OAD are expected to follow the MYOPIC strategy and achieve 50% efficiency.

It is easy to see how the same concept applies to sole ownership when resource dispersal is present. The benefits of sole ownership imply the owner is in control of not only the resource environment but also the resource itself. While a subject cannot request tokens from another subjects' zone, she can "steal" tokens through the dispersal mechanism. Imagine a game in which subjects A and B play MYOPIC and subject C plays MAX. Subjects A and B each receive 100 tokens in the first period while subject C receives only 50 tokens. In the second period subjects A and B are each endowed with 25 tokens that have dispersed from subjects C's zone, leaving subject C with 25 tokens. All three subjects now play MYOPIC, as subject C can no longer play MAX and realizes her endowment may continue to drop even if she does not extract. All subjects receive 25 tokens and the game ends. Subject C receives a total of 75 tokens compared to the 100 tokens she would have received if she had played MYOPIC from the start. Thus, subjects in SOD are expected to follow the MYOPIC strategy and achieve 50% efficiency.

5. Results

The experimental results are summarized in Table 1. Unless explicitly stated, all reference to statistical significance considers significance at the 5% level. The experiment can generate at most 600 tokens, so individual efficiency is compared to one third of this, or 200 tokens. Individual efficiencies greater than

100% are thus possible in all but the SOND treatment.[16] Subjects in open access make extraction decisions from three zones per period while subjects in sole ownership only make extraction decisions from one zone. Because the three zones cannot be treated as statistically different, I test that the percentage extractions do not differ between zones, and find that only the difference between zones B and C is statistically significant (p=0.0431). With this in mind I construct a single period request that is the individual's average request from the three zones in order to make open access requests comparable to sole ownership requests.

5.1. Open access

Efficiency in open access is negligibly higher without dispersal but this difference was not statistically significant using a two-sided Mann-Whitney test (p=0.6591) (Figure 5). Neither treatment was statistically different from the MYOPIC result of 50% efficiency using a Wilcoxon signed-rank test (OAND p=0.0679; OAD p=0.4004). Three subjects in OAD exceeded the MAX benchmark with a high of 105.2% and one subject in OAND exceeded the benchmark with 101.19%. Open access treatments exceeded the expected experiment length of one period (Figure 6). Still, 11% of groups with dispersal and 44% of groups without dispersal fully exhausted the resource in the first period. Dispersal increased average length by 0.55 periods but this result is not statistically significant (p=0.3091). Only one group in each treatment played the full five periods.

First period extractions are skewed heavily to the left, with the most prominent skew of the four treatments being in OAND (Figure 7). The average per zone request is higher without dispersal but the difference is not statistically

16 In open access this is easy to understand: a subject could obtain up to 600 tokens if the other subjects in her group made no extraction requests. Similarly, subjects in MD can harvest tokens dispersed from other zones, up to a maximum of 335 tokens.

significant (p=0.9035) (Figure 8). Between all three zones, the total individual requests in the first period exceed the initial endowment of 100 tokens. Extraction patterns in the open access treatments clearly reflect the fall-down in resource stock predicted by overharvesting in the first period (Figure 9). First period percentage requests are lower under dispersal but quickly increase to exceed requests in the absence of dispersal (Figure 10). Despite the fact that each of the two groups reaching the final period had an endowment of less than one token, neither group failed to achieve full exhaustion.

5.2. Sole Ownership

Sole ownership efficiency is identical across treatments and far exceeds open access efficiency, but falls short of the MAX prediction of 100% efficiency. Nine subjects in SOD exceeded 100% efficiency with a maximum of 112.43%, the largest of all treatments. Exceeding 100% was not possible in the SOND treatment. All groups in SOD played the full five periods while one subject in SOND exhausted the resource in the second period.

Like in open access treatments, first period requests were right skewed. Sole ownership reduced first period per-zone percentage requests by 6.68 points with dispersal and 5.61 points without dispersal, but this effect was not statistically significant (p=0.0839 and p=0.5757, respectively). Moreover, total individual requests were significantly lower than in open access where subjects withdraw from all three zones. Within sole ownership, dispersal reduced initial requests by 4.75 but was not statistically significant (p=0.1525).

The extraction pattern for SOND pulses around 30 tokens while the pattern for SOD follows a flattened quadratic form. The treatments have an identical final period request that is achieved by a sharp increase in final period requests in SOND. As a percentage of endowment, both treatments maintain their basic form but SOND is now increasing and SOD is increasing at a much faster

rate. First period extraction percentages in SOD are the lowest of all four treatments but steadily increase and triple over the course of the experiment, ending as the highest of the four treatments. Despite the increasing percentage requests, failure to exhaust tokens in the final period was a major problem in both treatments (Figure 11). Only 46.43% of SOND subjects and 36.67% of SOD subjects fully exhausted the resource.[17] The maximum number of tokens remaining was 63.7 tokens with dispersal and 79.6 without. This reduced efficiency by 10.8% in SOND and 6.4% in SOD.

6. Discussion

6.1 Sole Ownership: Extraction Patterns

The results did not support my primary hypothesis that sole ownership efficiency gains would erode when subjects faced resource dispersal. Rather than follow MYOPIC behavior, subjects in the SOD treatment appeared to have the most cautious behavior in the first period. Considering low first period requests as well as efficiency and game length on par with, and even slightly better than SOND, it would at first seem that adding dispersal did not cause sole ownership subjects to behave more like open access subjects.

On the other hand, a few subjects in SOD appeared to behave according to the MYOPIC strategy. In every period at least one subject requested the full endowment and 30% of subjects achieved over 100% efficiency, an outcome that could only occur if the subject was harvesting tokens that had migrated from other zones. Mean percentage requests also increased over time at an increasing rate. This suggests the possibility that subjects initially playing cautious strategies began to become more aggressive as the experiment progressed. It is impossible

17 In making this calculation, remaining token values of less than one were treated as exhausted. See the next section for a discussion of this and its impact on behavior. The subject in MND who had previously exhausted his resource was not considered.

to predict what would have happened given a longer experiment, but the linear decline in stock hints that early exhaustion may have occurred (Figure 12).

Given that two thirds of SOD subjects failed to exhaust the stock in the final period, it is possible that these subjects were playing as if the game would continue. Extraction in SOND appears to slow in the fourth period after stock finally drops below the optimal level. If the game were longer, this might reflect an adjustment that would bring the stock back towards the optimal level. On the other hand, SOD extraction in the fourth period increases stock deviation from the optimal level at an increasing rate. If this trend continued resource extinction would occur in only a few additional periods. These findings highlight the limitations presented by the length of the experimental design and suggest the effects of dispersal may be witnessed in long-term behavioral adjustment rather than immediate strategy change.

6.2 Sole Ownership No Dispersal: Failure to Maximize

Although these findings show that sole ownership subjects not facing dispersal did not follow the predicted MAX strategy, they still performed better than open access subjects. Subjects in SOND did extend the life of the resource throughout the game despite failing to maximize growth. Only 8 first period requests hit the optimal 50 tokens and 62.07% of requests were at or below 35 tokens. Subjects did not bring tokens to the optimal stock in the second period either, with more than half of subjects having a stock level above the optimal. Clearly, subjects are choosing cautious extraction patterns. Furthermore, the 8 subjects with optimal post-extraction stock in the first period shrank to 3 subjects in the second period and 2 subjects in the third and fourth periods.

The fact that most sole ownership subjects do not reach or even approximate the MAX strategy suggests that subjects do not fully understand the game. This is surprising because subjects are directly given growth information in both equation

and table format. Similar results have been found in another study of sole ownership extraction where subjects were given indirect access to full information (Hey et al (2008). This study found a SOND efficiency of 87.4% compared to 76.4% in the current study. This difference is probably in large part due to three factors. First, subjects in that study tended to over-extract in the first period compared to my research where severe under-extraction was common. Second, failure to extract all tokens in the final period was about 50% more common in my study. Finally, the shorter length of my experiments put greater weight on the first few periods where deviation from optimal behavior was more common. For this reason I expect that SOND efficiency would increase in a longer experiment and more closely resemble the results found in Hey et al (2008).

6.3 Open access: Expected Outcomes, Unexpected Behavior

MYOPIC behavior was expected in both of the open access treatments in addition to SOD. There is evidence to suggest that a small number of subjects followed self-interested strategies although only two subjects in OAND and no subjects in OAD followed the MYOPIC strategy. In both treatments one subject closely approximated this strategy, requesting all but one token. Likewise, one subject in OAND and three subjects in OAD achieved efficiencies greater than 100%. However, most subjects in the open access treatments did not behave aggressively. Over two-thirds of OAND subjects chose to request less than the 'fair' value, which I consider to be one third of the tokens available to the group. In the second and third periods this jumped to 80% and 85%, respectively. The same is true for OAD where half of subjects requested less than the fair value in the first period. This remained true for the second and third periods, but increasing cumulative requests exhausted most of the resource before the fourth period.

Clearly, most subjects were not following the self-interested strategy predicted by rational choice theory. On one hand, the two MYOPIC extraction

requests and the handful of other aggressive requests mixed with many more conservative requests suggests that only a handful. On the other hand, this evidence also supports criticism of the belief that subjects conform to self-interested interpretations of rational choice theory. Regardless of interpretation, the results clearly show that the few subjects acting in self-interest succeeded in ensuring the tragedy of the commons occurred for all members of their groups. Thus, overall efficiency was nearly identical in both treatments and very close to the MYOPIC benchmark. No outcome variables were significantly different at the 5% level and extraction patterns were very similar. The evidence appears to support the hypothesis that dispersal has little effect on behavior and outcomes in open access.

6.4 Experimental Design

When experimental research reveals surprising results, it is necessary to consider whether the results are a consequence of the experimental design. In this experiment the lack of different outcomes between sole ownership treatments, as well as the lack of MYOPIC and MAX strategies in open access and SOND, respectively, are all surprising results. I will examine two factors that could have affected the results: (1) subject understanding (2) parameter values, incentives, & payoff structures.

Because subjects in SOND face no competition, any deviation from the predicted MAX strategy is a result of individual conceptualization and understanding of the experiment. A number of subjects in all treatments asked questions and made statements indicating that they did not fully understand the experiment. The wide range of efficiencies in SOND and the mixed strategies within each treatment backup the anecdotal evidence of misunderstanding (Figure 13). The model of resource growth used is nonlinear in the growth equation, and there is evidence that subjects in experiments consistently misperceive linearity

in non-linear environments. Psychological studies have found that subjects tend to default to linear mental models and struggle to develop a non-linear understanding based upon feedback (Brehmer, 1980, 1992; Sterman, 1994). Even when using perfect property rights to remove the commons problem from resource management, subjects consistently mismanage the resource and achieve sub-optimal outcomes (Hey et al, 2008; Moxnes 1998, 2000, 2004; Sterman, 1994). These studies suggest that mismanagement is worse in experiments with higher degrees of complexity and lower information availability.

Misperceptions of linearity are one optimistic explanation of the mismanagement scene. However, there is evidence that linearity is not the cause. In the study by Hey et al (2008) subjects fell short of maximization despite being given a calculator that they could use to explore post-extraction stock, growth, next period stock, and savings based on possible extraction amounts (Figure 14). Furthermore, subjects in the current study were given a growth table including all integer stock values. Subjects in these experiments can use the resources provided to find the optimal strategy without understanding the non-linear growth function. It was highly unexpected that subjects in the SOND would fail to maximize the resource given the ease of identifying the growth-maximizing stock. The reason for this failure is unknown.

Based on the results and findings by other experimental studies, subject understanding is likely the most serious problem that could be addressed by the experimental design. However, parameter specification also impacts results, specifically experiment length. A length of only five periods highlights initial mismanagement by weighting each period so heavily in overall outcomes. This brevity also may have hidden differences between sole ownership treatments that would have become apparent in a longer experiment. The initial endowment and intrinsic growth rate likely influenced results by limiting absolute change

in marginal token growth to less than 1 token. This specification also created situations where subjects faced fractional endowments. With a payoff ratio of $0.05 per token and a difference between MYOPIC and MAX payouts of $5, subjects had very small real-world incentive to search for the optimal extraction path.[18]

6.5 Rationality

While poor understanding of the experiment was certainly present, the results also suggest that the calculating and self-serving *homo economicus* understanding of rationality may not be the most appropriate model of human behavior. First period withdraws in open access are not significantly different from the 'fair' request of one third of the tokens (p=0.894). Almost a quarter of requests fall between 30-33% of the endowment. This cooperative behavior is better understood through *homo sociologicus*, a definition of rationality focusing on norms and reciprocity rather than solely self-interest (Bruni 2008; Engelen, 2007). Similar results have been found in many experimental studies, some of the most interesting of which are ultimatum games where subjects' modal response is to split the endowment 50-50 (Gintis, 2000).

Homo sociologicus could also explain the increasing requests in SOD. Experimental studies have shown that subjects heavily utilize tools to punish defectors, even when at a cost to themselves (Gintis, 2000;Janssen et al, 2010; Ostrom et al, 1992) or when no direct benefit can be obtained due to group reorganization (Fehr & Gächter,2000). In the absence of punishment tools, subjects may often choose to follow tit-for-tat strategies (Gintis, 2000). In the current study the only way social subjects can punish selfish subjects is by playing more selfish strategies. Thus, the increasing extraction rates in SOD may identify the reciprocal reaction of social subjects to extraction by selfish subjects.

18 It was necessary to limit payoffs in order to achieve a suitable number of subjects.

A final argument to be made against the classical assumptions of rationality is that subjects in SOND that failed to find the optimal growth level of harvest all tokens in the final period clearly were not "weighing off the expected costs and benefits of actions and choosing the action that they consider to be the best" in the manner predicted by *homo economicus* (Engelen, 2007). The same is true for subjects facing fractional endowments because, while possible profit was low, entering a request equal to the endowment was no more costly than entering a request of zero. It appears that subjects were not aggressively pursuing maximization but instead were satisficing with "good enough" (Simon, 1957). *Homo economicus* cannot explain this mismanagement in the presence of full information, regardless of the explanation. Modeling rationality is difficult, and *homo economicus* is a convenient assumption because its simplicity makes analysis and drawing conclusions easier. But humans are intellectually and socially complex actors operating within complex social structures - simplification may not be the best way to examine the results of experimental economics.

7. Conclusion

This paper uses experimental methods to examine the relative performance of open access and sole ownership policies in environments where renewable resources are characterized by equalizing dispersal. Renewable resource policies that assign sole ownership of a resource aim to increase efficiency by removing the perverse incentives subjects face in an open access regime. However, economic theory suggests that sole ownership fails when it provides only exclusive extraction rights and does not give complete control over the physical resource. Such a situation can occur when a mobile resource is capable of crossing ownership boundaries.

As a pilot study, this research has furthered the understanding of how

renewable resource extractors respond to resource dispersal across management zones. As expected, subjects in the open access regime are largely indifferent to resource dispersal because the resource is almost entirely harvested in the first period. The hypothesis that sole owners facing dispersal will demand the entire resource in the first period was not supported by the results. In fact, dispersal did not affect sole ownership efficiency and actually reduced first period requests. However, subject behavior differed in the form of steadily increasing requests not seen by sole owners who did not face dispersal. Likewise, some subjects actually achieved efficiencies of greater than 100% by harvesting tokens dispersed from other subjects' zones. Together, these findings suggest that the effect of dispersal is in long-run adjustments to aggressive behavior rather than in immediate changes to initial strategy. The logical extension of this finding is to ask two questions. First, in a longer experiment will this behavior continue to the point of resource exhaustion? And second, are the increasing requests part of a response by cooperative subjects to the aggressive requests of non-cooperative subjects?

While this research uses an experimental methodology because there are many challenges facing empirical research, it is clear that experimental work is not without its limitations. As the experimenter, my own anecdotal and observational evidence of subject confusion during this research leads me to question the accuracy of the results. The evidence presented here supports this natural skepticism, most notably the failure of most sole owners to fully exhaust the resource in the final period. Given that subjects were informed that the experiment would end, this is in violation of the usual assumptions made when invoking the self-interested rationality defined by *homo economicus*. Similarly, while a small portion of subjects appeared to fit within the *homo economicus* framework, the majority of subjects exhibited cooperative behavior or negative reciprocal response to aggressive subjects. If subjects are neither fully utilizing

the information they are giving nor acting in purely self-interested ways, perhaps another definition of rationality may be more applicable to experimental research.

This research has presented new avenues of inquiry for experimentalists interested in renewable resource extraction. I recommend that future experimental research focus exclusively on sole ownership policy, take measures to address subject understanding, and examine more social definitions of rationality. Key to this will be a design that will examine the long-term effects of dispersal on sole ownership behavior and analyze behavior at the individual level. Based on the evidence presented here, it is my belief that such research will find behavioral response to dispersal and add to the understanding of renewable resource management policy.

Works Cited

Angrist, Joshua, and Jörn-Steffen Pischke. *The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con Out of Econometrics*. National Bureau of Economic Research, 2010.

Benartzi, Shlomo, and Richard H. Thaler. *Myopic Loss Aversion and the Equity Premium Puzzle*. National Bureau of Economic Research, 1993. Web. 20 Feb. 2013.

Beverton, Raymond JH, and Sidney Joseph Holt. *On the Dynamics of Exploited Fish Populations*. Vol. 11. Springer, 1957.

Brehmer, Berndt. "In One Word: Not from Experience." *Acta Psychologica* 45, no. 1 (1980): 223–241.

---. "Dynamic Decision Making: Human Control of Complex Systems." *Acta Psychologica* 81.3 (1992): 211–241. Web. 26 Feb. 2013.

Brundtland, Gro Harlem. *Our Common Future: The World Commission on Environment and Development*. Oxford: Oxford University Press, 1987. Print.

Bruni, Luigino, Mario Gilli, and Vittorio Pelligra. "Reciprocity: Theory and Facts." *International Review of Economics* 55, no. 1–2 (April 1, 2008): 1–11. doi:10.1007/s12232-008-0042-9.

Engelen, Bart. "Rationality and Institutions : an Inquiry into the Normative Implications of Rational Choice Theory," September 28, 2007.

Gintis, Herbert. "Beyond Homo Economicus: Evidence from Experimental Economics." *Ecological Economics* 35, no. 3 (December 2000): 311–322.

Gordon, H. Scott. "The Economic Theory of a Common-Property Resource: The Fishery." *Journal of Political Economy* 62.2 (1954): 124–142. Web. 20 Feb. 2013.

Hardin, Garrett. "The Tragedy of the Commons." *Science* 162.3859 (1968): 1243–1248. Web. 23 Feb. 2013.

Herr, Andrew, Roy Gardner, and James M. Walker. "An Experimental Study of Time-independent and Time-dependent Externalities in the Commons." *Games and Economic Behavior* 19.1 (1997): 77–96. Print.

Hey, John D., Tibor Neugebauer, and Abdolkarim Sadrieh. "An Experimental Analysis of Optimal Renewable Resource Management: The Fishery." Report / Paper. 2008. Web. 2 Feb. 2013.

Hilborn, R., C. J. Walters, and D. Ludwig. "Sustainable Exploitation of Renewable Resources." *Annual Review of Ecology and Systematics* 26 (1995): 45–67. Web. 20 Feb. 2013.

Janssen, Marco A. et al. "Lab Experiments for the Study of Social-Ecological Systems." *Science* 328.5978 (2010): 613–617. Web. 20 Feb. 2013.

Johannes, R. E. "Traditional Marine Conservation Methods in Oceania and Their Demise." *Annual Review of Ecology and Systematics* 9 (1978): 349–364. Web. 20 Feb. 2013.

Leamer, Edward E. “Tantalus on the Road to Asymptopia.” *The Journal of Economic Perspectives* 24, no. 2 (April 1, 2010): 31–46. doi:10.2307/25703497.

Levin, Simon A. “Population Dynamic Models in Heterogeneous Environments.” *Annual Review of Ecology and Systematics* 7 (1976): 287–310. Web. 20 Feb. 2013.

Mason, Charles F., and Owen R. Phillips. “Mitigating the Tragedy of the Commons Through Cooperation: An Experimental Evaluation.” *Journal of Environmental Economics and Management* 34.2 (1997): 148–172. Print.

Moxnes, Erling. “Misperceptions of Basic Dynamics: The Case of Renewable Resource Management.” *System Dynamics Review* 20.2 (2004): 139–162. Web. 28 Feb. 2013.

---. “Not Only the Tragedy of the Commons: Misperceptions of Bioeconomics.” *Management Science* 44.9 (1998): 1234–1248. Web. 28 Feb. 2013.

---. “Not Only the Tragedy of the Commons: Misperceptions of Feedback and Policies for Sustainable Development.” *System Dynamics Review* 16.4 (2000): 325–348. Web. 28 Feb. 2013.

---. “Uncertain Measurements of Renewable Resources: Approximations, Harvesting Policies and Value of Accuracy.” *Journal of Environmental Economics and Management* 45, no. 1 (2003): 85–108.

Noailly, Joëlle, Cees A. Withagen, and Jeroen CJM van den Bergh. “Spatial Evolution of Social Norms in a Common-pool Resource Game.” *Environmental and Resource Economics* 36.1 (2007): 113–141. Print.

Olson, Mancur. “The Logic of Collective Action [1965].” *Contemporary Sociological Theory* 2 (2007): 111. Print.

Ostrom, Elinor, James Walker, and Roy Gardner. “Covenants With and Without a Sword: Self-Governance Is Possible.” *The American Political Science Review* 86.2 (1992): 404–417. Web. 20 Feb. 2013.

Rouwette, Etienne AJA, Andreas Größler, and Jac AM Vennix. “Exploring Influencing Factors on Rationality: a Literature Review of Dynamic Decision-making Studies in System Dynamics.” *Systems Research and Behavioral Science* 21.4 (2004): 351–370. Print.

Sanchirico, James N., and James E. Wilen. “Bioeconomics of Spatial Exploitation in a Patchy Environment.” *Journal of Environmental Economics and Management* 37.2 (1999): 129–150. Print.

---. “Optimal Spatial Management of Renewable Resources: Matching Policy Scope to Ecosystem Scale.” *Journal of Environmental Economics and Management* 50.1 (2005): 23–46. Print.

Schaefer, Milner B. “Some Considerations of Population Dynamics and Economics in Relation to the Management of the Commercial Marine Fisheries.” *Journal of the Fisheries Research Board of Canada* 14.5 (1957): 669–681. Web. 20 Feb. 2013.

Schnier, Kurt Erik. “Spatial Externalities and the Common-pool Resource Mechanism.” *Journal of Economic Behavior & Organization* 70.1–2 (2009): 402–415. Web. 23 Feb. 2013.

Scott, Anthony. “The Fishery: The Objectives of Sole Ownership.” *Journal of Political Economy* 63.2 (1955): 116–124. Web. 28 Feb. 2013.

Simon, Herbert A. *Models of Man: Social and Rational : Mathematical Essays on Rational Human Behavior in a Social Setting*. Garland Pub., 1957.

Smith, Martin D., James N. Sanchirico, and James E. Wilen. “The Economics of Spatial-dynamic Processes: Applications to Renewable Resources.” *Journal of Environmental Economics and Management* 57.1 (2009): 104–121. Print.

Sterman, John D. “Learning in and About Complex Systems.” *System Dynamics Review* 10.2-3 (1994): 291–330. Print.

Suter, Jordan F. et al. “Behavior in a Spatially Explicit Groundwater Resource: Evidence from the Lab.” *American Journal of Agricultural Economics* 94.5 (2012): 1094–1112. Web. 23 Feb. 2013.

Walker, James M., and Roy Gardner. “Probabilistic Destruction of Common-pool Resources: Experimental Evidence.” *The Economic Journal* 102.414 (1992): 1149–1161. Web. 20 Feb. 2013.

Walker, James M., Roy Gardner, and Elinor Ostrom. “Rent Dissipation in a Limited-access Common-pool Resource: Experimental Evidence.” *Journal of Environmental Economics and Management* 19.3 (1990): 203–211. Print.

Wiens, John A. “Population Responses to Patchy Environments.” *Annual Review of Ecology and Systematics* 7 (1976): 81–120. Web. 23 Feb. 2013.

Wilen, James E. “Renewable Resource Economists and Policy: What Differences Have We Made?” *Journal of Environmental Economics and Management* 39.3 (2000): 306–327. Web. 20 Feb. 2013.

Zhang, Junjie, and Martin D. Smith. “Estimation of a Generalized Fishery Model: A Two-Stage Approach.” *Review of Economics and Statistics* 93.2 (2010): 690–699. Web. 20 Feb. 2013.

Appendix A: Figures & Tables

Treatment	n	Theoretical Efficiency	Observed Efficiency	Std. Deviation	Mean Length (periods)	Std. Deviation
OAND	27	0.50	0.569	0.218	2.6	1.601
OAD	27	0.50	0.558	0.278	3.1	1.121
SOND	29	1.00	0.764	0.147	4.9	0.557
SOD	30	0.50	0.764	0.283	5.0	0.000

Table 1. Summary of experimental results

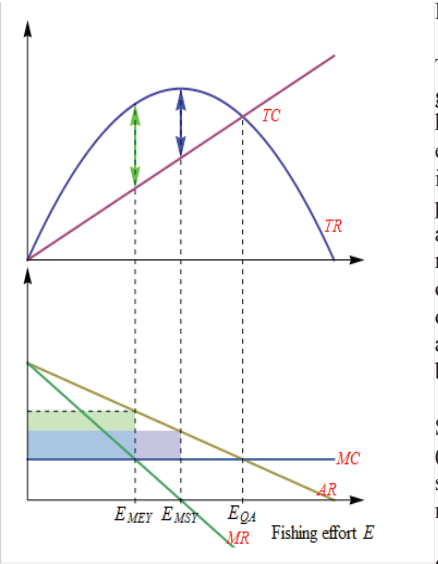


Figure 1: The Gordon Schaefer model.

Total Revenue (also representing population growth) increases when effort allows larger harvests, but decreases when too much effort reduces the resource stock beyond its maximum growth level. The model predicts over-allocation of effort in open access because individual firms receive the market's average revenue rather than their own marginal revenue (E_{QA}). This essentially occurs because part of the harvest earned by a new firm entering the industry would have been harvested by other firms.

Schaefer's Maximum Sustainable Yield (E_{MSY}) creates profit and maintains resource stock by harvesting such that the growth-maximizing stock level is maintained.

Gordon's Maximum Economic Yield is the optimal outcome (E_{MEY}) that maximizes profit by producing where MR=MC. This corresponds with the lowest level of effort and highest resource stock.

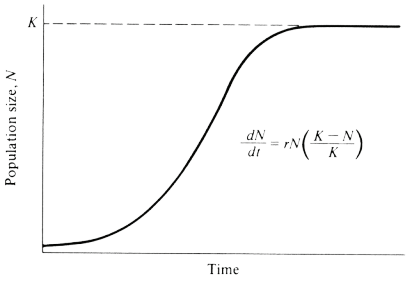
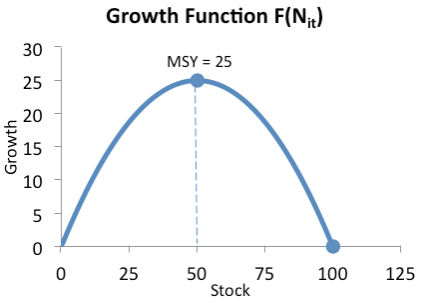


Figure 2: Standard population model depicting logistic growth. The population grows very slowly when the population density is very low or very high, and much faster when the population density is near half the carrying capacity.



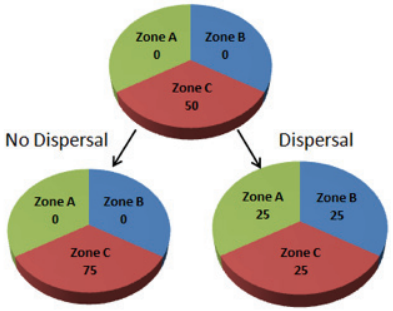


Figure 3: Logistic growth function used in this experiment with K=100 and r=1(left). Example of equalizing dispersal where Zones A and B grow no tokens while Zone C grows 25 tokens (right).

Figure 4: The maximization strategy (MAX) brings stock to MSY in the first period, harvests only growth in intermediate periods, and harvests all tokens in the final period.

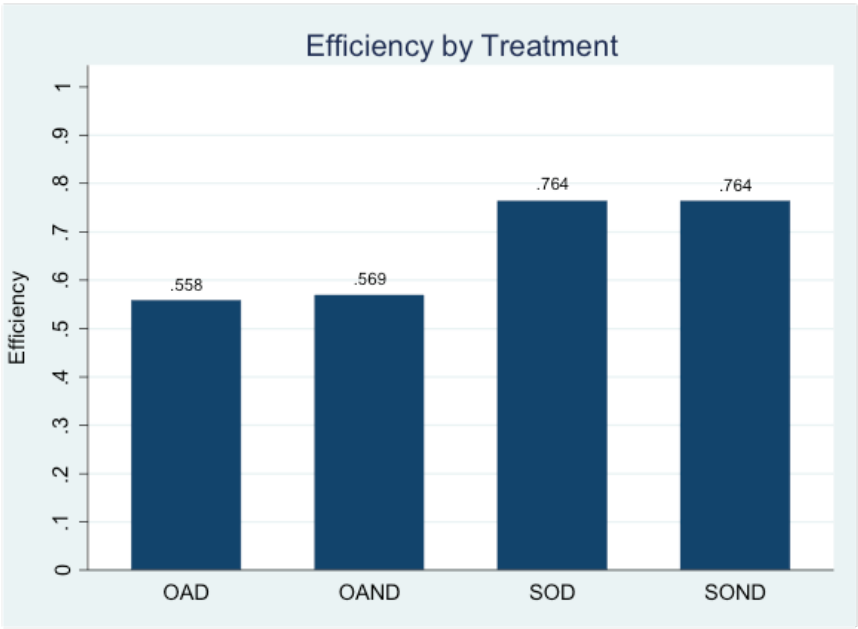


Figure 5: Mean efficiency by treatment. Differences between policies are significant at the 5% level while differences between dispersal conditions are not.

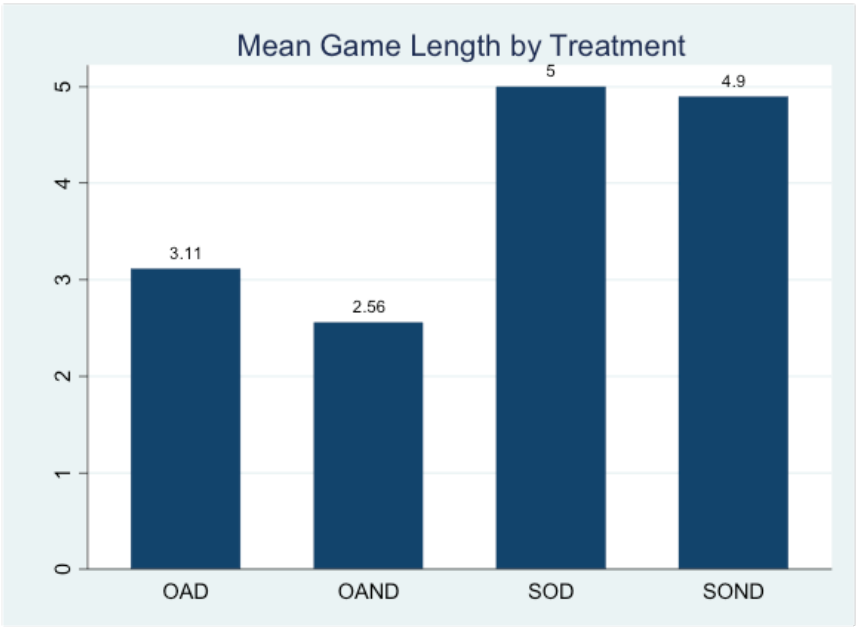


Figure 6: Mean game length by treatment. Differences between policies are significant at the 5% level while differences between dispersal conditions are not.

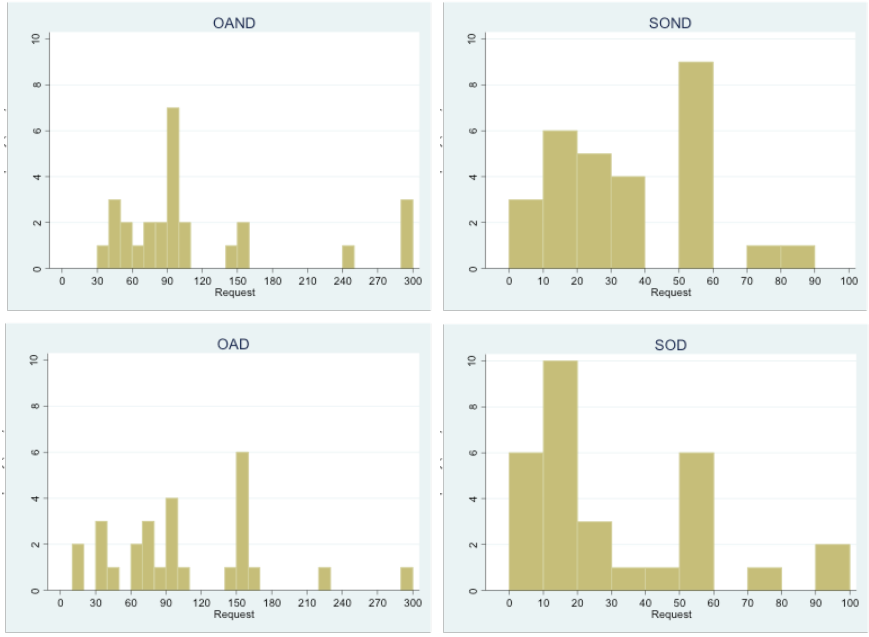


Figure 7: Histograms displaying frequency distribution of first period requests.

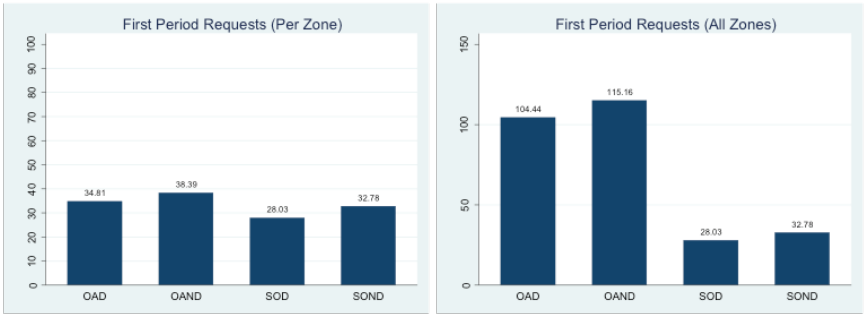


Figure 8: First period individual requests by treatment. Requests are shown per zone (left) and combined from all zones (right) to highlight the impact of multiple zones in open access treatments.

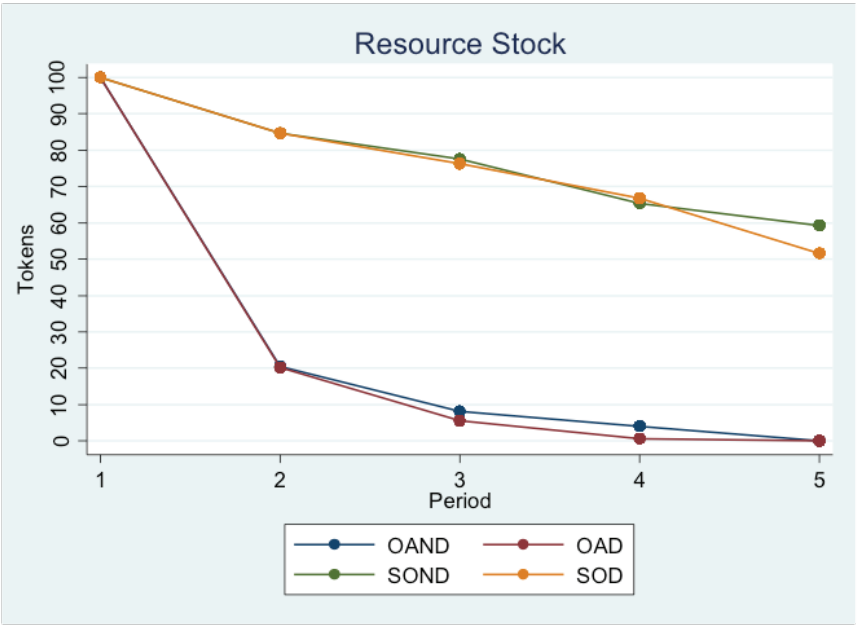


Figure 9: Extraction pattern by period: absolute token requests.

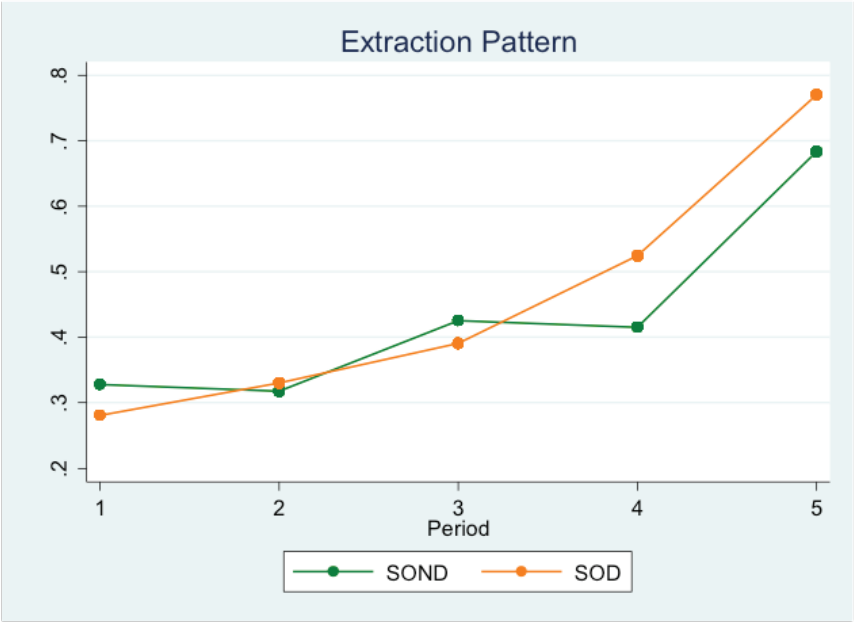


Figure 10: Extraction pattern by period: token requests as a percentage of available endowment.

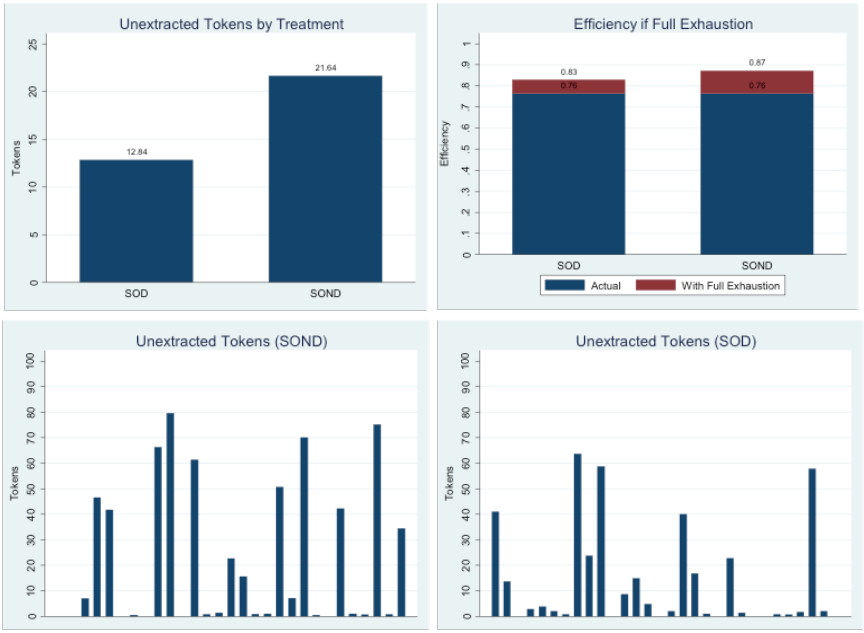


Figure 11: Tokens not extracted during final period (top left) and the resulting efficiency loss (top right). Graph depicting number of tokens remaining in final period with individual subjects on x axis (bottom).

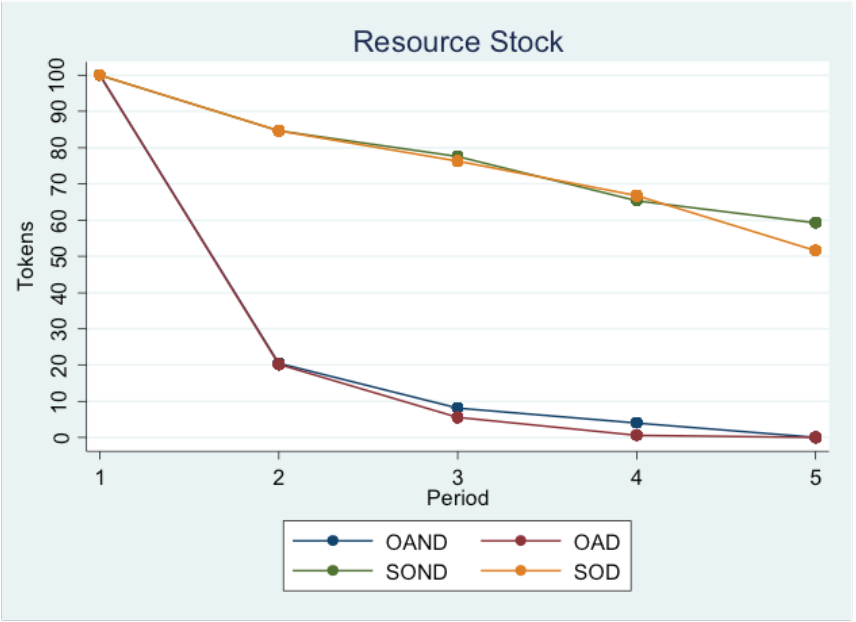


Figure 12: Deviation from the optimal stock of 50 tokens in periods 1 through 4 and 0 tokens in period 5.

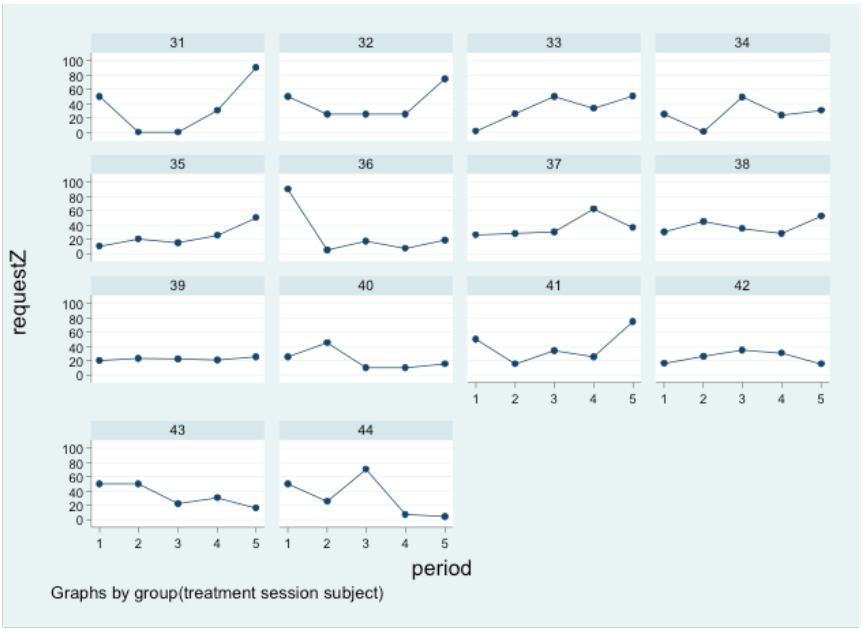


Figure 13: Extraction patterns from first session of SOND that show no general strategy.

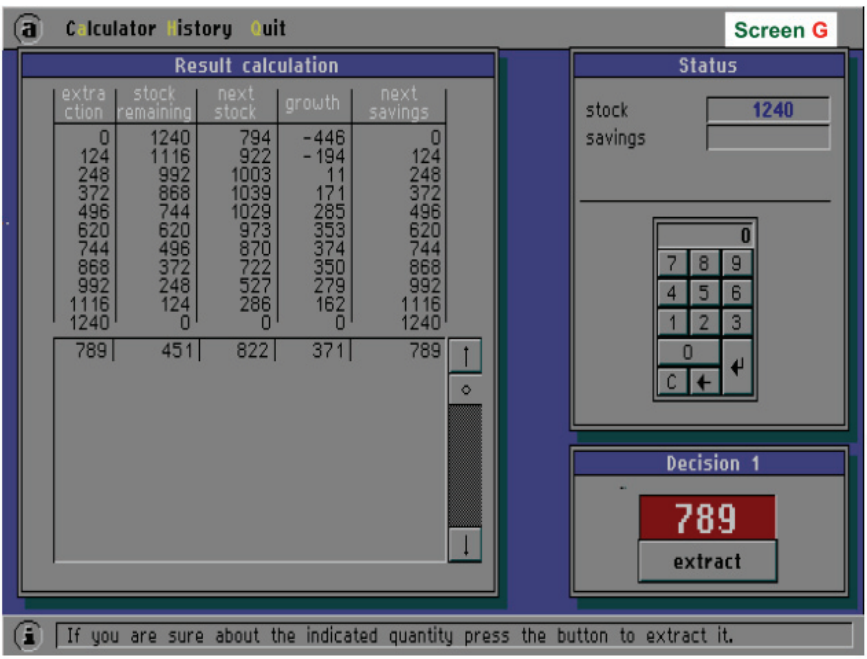


Figure 14: Calculation tool used in Hey et al (2008) to give subjects stock & growth information.

Appendix B: Recruitment Email

Subjects' names were gathered using the freely available information in the college's email system. The order of the subjects' names were randomized and any ineligible students were removed (those who assisted the experimenter, had prior knowledge of the experiment, or had previously participated). Recruitment emails were then sent to these students. An announcement was also placed in the "Student Digest," a collection of announcements sent to students in email form. The general wording was as follows:

> You are invited to participate in an experiment on [date]. **By participating in the experiment you will earn $5. You can also earn additional money based on your performance, up to $20**. The experiment should take at most 45 minutes. The sessions are as follows:
>
> - [Date, time, location]
>
> **To sign-up, please email [author email]** which session you would like to participate in.
>
> Spots will be filled on a first-come, first-served basis. The experiment is being conducted by Kevin Lugo, a senior Economics major, in conjunction with Professor John Cadigan. If you have already participated in one of my experiments last semester you may not participate again.
>
> Questions should be addressed to Kevin Lugo [author email].

Each treatment was designed to have 30 participants for a total of 120 subjects. Extra participants were recruited in the event subjects did not show up and, if sent home, were given the show-up payment of $5. Despite extra recruitment some sessions were short of the desired 15 subjects. For this reason 27 subjects participated in open access no dispersal, 27 in open access dispersal, 29 in sole ownership no dispersal, and 30 in sole ownership dispersal.

Appendix C: Experimental Materials

Instructions OAND

Introduction

Thank you for agreeing to participate in this experiment. At this time please turn off any cell phones or other electronic devices.

You will earn $5 by participating in the experiment and may earn additional cash based on your performance. You get to keep any money that you earn over the course of the experiment. The experiment may take as long as 45 minutes.

When you are done with each screen, press "OK" to continue.

In this experiment you will be randomly and anonymously placed in a group with two other subjects.

You and the two other subjects will play through a series of periods in which you will make decisions that will earn you experimental tokens. These tokens will remain in your account for the duration of the experiment. At the end of the experiment these tokens will be exchanged for dollars at a rate of 1 token to $0.05.

Experiment Design

There will be five periods. At the beginning of each period, subjects will choose a number of tokens to withdraw from three 'zones' labeled Zone A, Zone B, and Zone C. Your individual requests may not exceed the amount available in each zone. Each zone will begin the first period with 100 tokens.

In any period, if the group requests a total number of tokens that is less than the amount available in that zone, each player will receive the amount she requested from that zone.

If the total number of tokens requested from a zone by the group exceeds the amount available in that zone, each player will receive a number of tokens proportional to her share of the total group request. In other words, if you request R tokens, the group requests a total of X tokens, and T tokens are available, then if X > T you will receive a number of tokens from that zone according to:

$$Your\ Tokens = \left(\frac{R}{X}\right) * T$$

At the end of each period, the number of tokens in a zone will grow based on the number of tokens remaining in that zone (T). Growth will follow the growth function:

$$Growth = T * \left[1 - \left(\frac{T}{100}\right)\right]$$

The number of tokens that grow will be added to the number of tokens remaining in the zone. This new total will be the number of tokens available in the next period. Each of the 5 periods will proceed in the same manner. Note that if, in any period, all remaining tokens are taken then there no growth occurs and there are 0 tokens available from that zone in all subsequent periods.
You have been given a reference sheet with a table displaying the growth associated with every integer token amount. This table also shows the number of tokens that will be available in the next period in each of those circumstances.

After each period, subjects will view a summary of the results from that period. This will include your personal performance as well as the group request from each zone, the actual group withdraw from each zone, and the total group withdraw that period. You will also see how many tokens remain in each zone at the end of the period. Finally, you will see the cash value of the tokens you earned that period.

You may find it helpful to record the results of each period. You have been provided with a paper record sheet to assist you in this process.

Summary

- In each period you will choose how many tokens you wish to withdraw from zones A, B, and C.
- If the group request is **less than** the tokens available, you will receive your request.
- If the group request is **greater than** the tokens available, you will receive tokens proportional to your share of the group request.
- Based on the number of tokens available in each zone, the growth function described above will be used to determine the number of tokens available in the next period.
- The game will last 5 periods.

After the final period, please record your earnings on you receipt form and wait for further instructions.

Throughout this experiment you are not to communicate with other players in any way. You must keep your eyes on your own screens at all times and may not use any electronic devices. Breeching these rules will result in a forfeit of all compensation.
If you have any questions, please ask them now.

Reference Sheet

Growth Table $Growth = T * \left[1 - \left(\frac{T}{100}\right)\right]$

Tokens	Growth	New Tokens	Tokens	Growth	New Tokens	Tokens	Growth	New Tokens	Tokens	Growth	New Tokens
0.00	0.00	0.00	25.00	18.75	43.75	50.00	25.00	75.00	75.00	18.75	93.75
1.00	0.99	1.99	26.00	19.24	45.24	51.00	24.99	75.99	76.00	18.24	94.24
2.00	1.96	3.96	27.00	19.71	46.71	52.00	24.96	76.96	77.00	17.71	94.71
3.00	2.91	5.91	28.00	20.16	48.16	53.00	24.91	77.91	78.00	17.16	95.16
4.00	3.84	7.84	29.00	20.59	49.59	54.00	24.84	78.84	79.00	16.59	95.59
5.00	4.75	9.75	30.00	21.00	51.00	55.00	24.75	79.75	80.00	16.00	96.00
6.00	5.64	11.64	31.00	21.39	52.39	56.00	24.64	80.64	81.00	15.39	96.39
7.00	6.51	13.51	32.00	21.76	53.76	57.00	24.51	81.51	82.00	14.76	96.76
8.00	7.36	15.36	33.00	22.11	55.11	58.00	24.36	82.36	83.00	14.11	97.11
9.00	8.19	17.19	34.00	22.44	56.44	59.00	24.19	83.19	84.00	13.44	97.44
10.00	9.00	19.00	35.00	22.75	57.75	60.00	24.00	84.00	85.00	12.75	97.75
11.00	9.79	20.79	36.00	23.04	59.04	61.00	23.79	84.79	86.00	12.04	98.04
12.00	10.56	22.56	37.00	23.31	60.31	62.00	23.56	85.56	87.00	11.31	98.31
13.00	11.31	24.31	38.00	23.56	61.56	63.00	23.31	86.31	88.00	10.56	98.56
14.00	12.04	26.04	39.00	23.79	62.79	64.00	23.04	87.04	89.00	9.79	98.79
15.00	12.75	27.75	40.00	24.00	64.00	65.00	22.75	87.75	90.00	9.00	99.00
16.00	13.44	29.44	41.00	24.19	65.19	66.00	22.44	88.44	91.00	8.19	99.19
17.00	14.11	31.11	42.00	24.36	66.36	67.00	22.11	89.11	92.00	7.36	99.36
18.00	14.76	32.76	43.00	24.51	67.51	68.00	21.76	89.76	93.00	6.51	99.51
19.00	15.39	34.39	44.00	24.64	68.64	69.00	21.39	90.39	94.00	5.64	99.64
20.00	16.00	36.00	45.00	24.75	69.75	70.00	21.00	91.00	95.00	4.75	99.75
21.00	16.59	37.59	46.00	24.84	70.84	71.00	20.59	91.59	96.00	3.84	99.84
22.00	17.16	39.16	47.00	24.91	71.91	72.00	20.16	92.16	97.00	2.91	99.91
23.00	17.71	40.71	48.00	24.96	72.96	73.00	19.71	92.71	98.00	1.96	99.96
24.00	18.24	42.24	49.00	24.99	73.99	74.00	19.24	93.24	99.00	0.99	99.99
									100.00	0.00	100.00