FALL 2015

ECONPress

rethink *your world*

ECONPress
of Northeastern University

ECONPress is a publication for undergraduate compositions in economics. We publish twice a year during each fall and spring semester. ECONPress invites the highest quality submissions from undergraduate students in various economics related disciplines. It provides a forum for the undergraduate economics community to engage in active discussion and debate about the topics, theories, and applications they've learned in the classroom.

Special Thanks

Edward J. Meehan, Jason Fertig, Alexa Nguyen

Edited and published at Northeastern University

Editorial Offices

ECON*Press*

360 Huntington Avenue

301 Lake Hall

Boston, MA 02115-5000

ISSN:

2158-8864 (Print)

2158-8872 (Online)

Electronic Mail:

NUEconPress@gmail.com

Webs Site:

www.ECONPress.org

ECONPress

of Northeastern University

Letter from the Editor

To our readers,

It gives me great pleasure and pride to reveal this new edition of EconPress. Many hands and minds contributed to this journal, and many acknowledgements will be made in this letter. However, I will begin with a quick word on the papers that follow.

We approached our submission reviews this semester with a hope of expanding our target reader audience. Our goal was to show how economic analysis is applicable to a wide range of subjects and is relevant to all walks of life. Within these pages you will find analyses on subjects ranging from minimum wage increases to aggression in professional football. We hope that all of our readers will find topics here that interest and intrigue them. And, while the subjects covered in this journal are diverse, all of the papers are united by their adherence to economic principles and the strong quality of their analysis, as per EconPress' high standards.

This journal could not have come together without the contributions of many individuals. First, I would like to thank all of the authors who submitted their papers for review. We received many high-caliber submissions, and whittling those down to the four we present here was no easy task.

I would also like to thank our advisers, Dr. Nancy J. Kimelman and Dr. Peter Simon, for their guidance throughout the last semester, as well as the entire Northeastern Economics Department for all of their support. Our team is also deeply indebted to Mr. Edward J. Meehan, who has supported EconPress in many ways over the last several years.

I had the privilege of working with an incredibly hard-working and fun team for this journal. Anne-Lise, Ben, Cayla, Celene, Chaitri, Dan, Lauren, Sean and Vinny, all of your contributions were singularly crucial to our successes this semester, and this journal could not have happened without each and every one of you. I also want to thank Alexa Nguyen for her excellent design work, and the entire staff at Smith Print for all of their help.

Lastly, I want to acknowledge you, the readers. All of the work we've put in has culminated with this journal, and we thank you sincerely for taking the time to read it.

Best,

Sohan Shah
Edior-in-Chief

Energy Use in Manufacturing: Evidence from the U.S. Economic Census

Stephen Ngo
Northeastern University

Introduction

Over the past few decades of development, clear examples of advances in energy efficiency can be seen throughout developed economies. In the aviation industry, fuel burn per seat has declined by more than 80% since the late 1950s (World Bank 17). Given that fuel costs form a large proportion of the variable cost of air travel, airlines and other firms in this sector have a strong incentive to invest in more fuel-efficient fleets. The preference of airlines for more fuel-efficient aircraft, in turn, encourages manufacturers like Airbus and Boeing to expend vast resources in the development of more fuel-efficient technologies. Improvements in the availability and utilization of such technologies allow aviation firms to generate higher levels of value for consumers using less fuel.

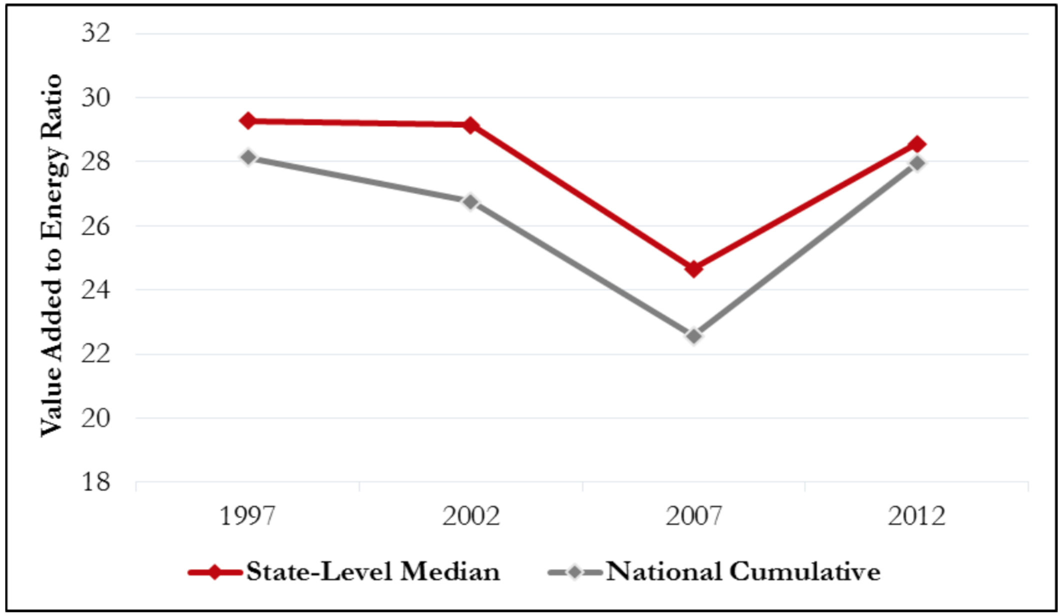
If airlines can be incentivized to adopt more fuel-efficient behaviors through the possibility of lower variable cost, then manufacturing firms ought to be similarly encouraged to make more energy-efficient capital purchases. Improvements in the energy efficiency of aircraft may also be mirrored by increased energy efficiencies in the capital equipment of manufacturers, for whom electricity and fuel costs are often important components.

As with airlines, manufacturers are incentivized to take steps to reduce their marginal fuel costs, most notably by purchasing more energy-efficient production machinery in place of less technologically advanced, more energy-intensive alternatives. These shifts in the deployment of capital assets push manufacturers' marginal energy costs downwards over time. As a result, every unit of energy expended in the creation of economic value should become observably more productive over time (all else equal).

How might one measure this sector-level productivity of energy? The United States Census Bureau's Economic Census gives sector-specific estimates, in dollar terms, of the fuel and energy expenditures of manufacturers within each state, in addition to "value added" as a measure of the cumulative value generation of the manufacturing sector in each state. Assuming (a) that contemporary technology can only either become more productive, or maintain productive parity, and (b) that technological development is the driving force behind the productivity of an input, the aggregate national ratio of value added to total expenditures on energy (electricity + fuel) should rise over time.

Two inferences can be made from Figure 1, which describes the Value Added to Energy Expenditure ratio for the four Economic Censuses conducted since 1997 (the regression model described later was applied to this data set). The first is that the distribution of the ratio is not symmetrical. In all years, the median state was more energy-efficient than the nation as a whole,

Figure 1



implying that more than half the states were more efficient than the 48 contiguous states combined. One might reasonably infer that conditions in some subset of the states with more manufacturing activity are such that these states skew the national ratio below that of the median state.

The second inference is that the ratio is clearly not increasing over this small window of time; rather, it exhibits an overall decline since 1997, with a particularly strong temporary drop seen for the year 2007.

This initial cut of Economic Census data does not lend much credence to technological energy efficiency hypothesis. Yet it does hint at the importance of other factors that may impact the energy efficiency of firms, particularly if this efficiency is measured using metrics that are denoted in financial terms.

Value added, for instance, is "also referred to as gross domestic product (GDP)-by-industry" (BEA). It is a building block of national GDP, which can be interpreted both as a signal of aggregate sector productivity, and of overall macroeconomic conditions. The value added of a single sector, then, can also be swayed by conditions in the wider market (though sector-specific trends may cause deviation from GDP).

Energy expenditures, of course, are immensely affected by shifting commodity and utility prices. Over the span of a single year, they can also be swayed by the arrival of a particularly harsh winter or summer, which drive up heating and cooling costs, respectively.

A more robust method of extrapolating the effective energy efficiency of manufacturing technology should control for such factors.

To calculate this unit productivity of energy – and thus changes in production energy efficiency – this paper examines the validity of an original regression model that seeks to approximate the economic impact of shifting technical knowledge, processes, and more advanced capital goods on the ability of manufacturing firms to extract higher levels of "value added" from energy usage in the production process. The impacts of short-term energy costs, the severity of annual climate patterns, market wages, the domestic macroeconomic environment, and the liberalization of global trade are considered in an observation of energy usage and manufacturing productivity within the United States over time.

Literature Review

A variety of previous studies have attempted to pinpoint factors contributing to changes in industrial energy efficiency, and attempted to measure the effects of these factors on observed energy usage.

Previous researchers have compared energy efficiency regulations enacted by governments, and attempted to measure the effect of these public policies on efficiency outcomes. An analysis of state-level building codes in the United States (Scott et al, 2014) shows that an economic model may make use of regulations specific to energy, as well as regulations specific to pollution (carbon dioxide emissions, for example) as indicators of the zeal of regulators in a given geographic area in enforcing environmental outcomes. While this analysis did not yield robust results from a model predicting the effects of such regulatory activity, the authors constructed their model in order to test the feasibility of their conceptual mode of analysis, and not to prove the statistical significance of the estimated effect of their independent variables on energy consumption and pollution results.

Far more robust models of energy consumption and efficiency in the industrial production process have been constructed. One study attempted to measure the effects of changes in quantity demanded of energy in the short run, and long run changes in firms' capital stocks on total energy consumption in the manufacturing industries of developed nations (Steinbuks and Neuhoff, 2014). Hypothesizing that firms cannot completely adjust their available equipment for energy costs in the short run, and also do so in the long run by gradually shifting their capital structure towards more energy efficient technologies, the study introduced estimates of "capital vintages," or measures of the age of a piece of capital, into their analysis. The authors found evidence of both short run "price-induced and (long run) improvements in the efficiency of capital stock."

While the authors made a great achievement in their estimation of the ability of firm managers to react to changing energy prices through deployment of capital, and is a robust estimator of industrial energy consumption as a result, their model is not effective at analyzing the energy efficiency of their methods in utilizing capital and energy. As the goal of production is to produce a profit stream for the firm, the productivity of the firm's deployment of capital in producing a revenue or profits must also be considered. A capital stock may be considered energy efficient even if it utilizes large volumes of energy, provided that it also outputs large amounts of value for its owner; in contrast, a production process that utilizes a small amount of energy in absolute terms may be energy inefficient if it is not productive. A sufficient model of energy efficiency in the production process, then, must at least contain an indicator of productivity.

Another study, which did not take into account "capital vintage" but did analyze total industrial productivity, made the next great stride towards a comprehensive model of industrial energy efficiency (Zhao et al., 2014). This study attempted to utilize data on industrial production and energy use in the Chinese and Japanese manufacturing industries in order to separate changes in energy efficiency into three categories. The first, the production effect, refers to the tendency of total energy consumption to increase when the overall volume of industrial production increases; the second, the efficiency effect, refers to the share of industrial production value that goes to payments on energy; the last, the structural effect, refers to changes in the mix of industrial sub-sectors within the aggregate industrial sector, which might be an indicator of redeployments of resources, or simply a shift in the mix of production processes observed within the larger industrial sector.

This analysis of industrial energy efficiency is incredibly comprehensive, provided that regulations, macroeconomic conditions independent of the industrial sector, energy prices, weather, and the share of the larger economy involved in import and export activity remain fixed. Absent an accounting of these factors, estimators for the effect of these changes in production techniques and industrial organization may be distorted. As a result, a truly comprehensive model of energy efficiency ought to control for these important conditions.

Model

The regression model described in this paper uses the predicted ratio of "value added" to energy expenditures reported by U.S. manufacturers to represent the production energy efficiency of technology in a given year. Independent of short-run trends that may impact either the numerator (value added) or denominator (energy expenditures) of the ratio, an increase in this measure should reflect higher energy productivity resulting from better availability and utilization of energy-efficient manufacturing technology; a decrease should reflect lower energy productivity, resulting from less utilization of energy-efficient manufacturing technology.

This ratio was calculated using estimates of value added and energy and fuel expenses given by the U.S. Census Bureau's Economic Census. The Census is given every five years, and records various estimates of business activity at both the state and metropolitan levels for various sectors. To ensure the most comprehensive representation of cross-sectional U.S. data, state-level value added to energy expenditure ratios for each year were used; all independent variables used in the regression represent state-level estimates for the given year.

The regression model outlined in this paper takes the form of the below equation and variables:

$$R_{it} = \beta_0 + \beta_1 \ln(V_{it}) + \beta_2 \ln(e_{it}) + \beta_3 \ln(f_{it}) + \beta_4 \ln(L_{it}) + \beta_5 Y_{it} + \beta_6 X_{it} + \beta_7 \ln(H_{it}) + \beta_8 \ln(C_{it}) + \beta_9 T_{2002i} + \beta_{10} T_{2007i} + \beta_{11} T_{2012i}$$

R_{it} represents the ratio of "Value added" to cumulative electricity and fuel expenses, based on estimates of value added and energy expenses given by the Economic Census conducted by the U.S. Census Bureau every five years. This ratio is used as a measure of returns on energy inputs, holding other conditions constant; increases in the value of this value indicate higher productivity within the industrial process per dollar of energy input.

V_{it} represents the total value added of the state's manufacturing sector, using the same estimate from the Economic Census that was used to calculate the dependent variable. It is adjusted for inflation using the annual implicit GDP Deflator provided by the Federal Reserve Bank of St. Louis. States with large manufacturing sectors may enjoy economies of scale in production, independent of energy costs and other variables.

e_{it} represents the inflation-adjusted statewide average price of electricity for industrial production in the given year, based on nominal averages provided by the Energy Information Administration. Prices are expressed in dollars per British Thermal Unit (BTU).

f_{it} represents the inflation-adjusted statewide average price of distillate fuel oil – from which diesel oil is produced – for industrial production in the given year, based on nominal averages provided by the Energy Information Administration and adjusted for inflation using the annual implicit GDP Deflator provided by the Federal Reserve Bank of St. Louis. Prices are expressed in dollars per British Thermal Unit (BTU).

L_{it} represents the inflation-adjusted average annual pay for workers in the state's private manufacturing industry, as estimated by the Bureau of Labor Statistics' Quarterly Census of Employment and Wages. It is denoted in thousands of 2012 U.S. Dollars. Given that labor, like energy, is also a major component of a manufacturing firm's variable cost, it is likely to play a role in the firm's capital asset purchase decisions. However, it is unclear what the effect of higher market wages may have. Higher wages for unskilled workers may incentivize firms to switch to more capital-intensive production processes; on the other hand, higher wages may be an indicator of higher demand for skilled labor, which often results from more usage of capital-intensive production. Since this model utilizes labor cost mostly as a control variable, estimation using an exponential function should account for this phenomenon.

Y_{it} represents the pace of statewide real GDP growth in the given year, as estimated by the latest revised statistic from the Bureau of Economic Analysis. Though trade between states represents the majority of most firms' demand such that strong demand outside of the state is a larger factor than the economic health and growth conditions of a firm's home state itself, the home state's rate of GDP growth is the most accurate indicator of the confidence in market conditions held by firms within the state.

X_{it} represents the share of exports as a percentage of state GDP in the given year, as estimated using export figures from the International Trade Administration and nominal GDP estimates from the Bureau of Economic Analysis. Trade share, the sum of an economy's exports and imports, is an effective indicator of that economy's openness to trade. But it does not communicate the extent to which demand for the industrial base's goods is affected by the severity of exposure to international communication. The similarity of two countries' trade shares may obscure that one economy is running a trade surplus, and the other a trade deficit. In contrast, the export or import share of GDP may indicate both the openness of an economy to international trade, and the competitiveness of its industry. However, a lack of readily accessible historical data on imports (this is also granular at the state level) constrained this model to estimates of nominal export trade share.

H_{it} represents the population-weighted average number of Heating Degree Days (HDD) experienced by the state over the year, denoted in thousands. HDD is a measurement of the energy needed to heat a building to a comfortable temperature during cold weather, as estimated by the National Weather Service.

C_{it} represents the population-weighted average number of Cooling Degree Days (CDD) experienced by the state over the year, denoted in thousands. CDD is a measurement of the energy needed to cool a building to a comfortable temperature during hot weather, as estimated by the National Weather Service.

T_{2002i} represents the difference between the estimated regression intercepts for 1997 and 2002.

T_{2007i} represents the difference between the estimated regression intercepts for 1997 and 2007.

T_{2012i} represents the difference between the estimated regression intercepts for 1997 and 2012.

Estimation

Each of the independent variables represents the impact of a short-run factor on the observed energy efficiency of a state's manufacturing sector. Expression of the impact of factors driving long-run manufacturing energy efficiency, independent of these short-run factors – like technological development – should be contained in each year's unique regression intercept. Evidence of advances in technological efficiency (in addition to other omitted variables) should be contained in the intercept, and the binary variables for the years 2002, 2007, and 2012.

Note that due to a lack of Manufacturing-specific wage data for the state at the

Table 1

Variable	Coefficient	Standard Error
Constant	-112.6981***	32.4198
Natural Log of Total Value Added ($1,000,000s)	1.3654*	0.8009
Natural Log of Price of Electricity ($ per BTU)	5.2041*	2.9228
Natural Log of Price of Distillate Oil ($ per BTU)	17.8995	14.8067
Natural Log of Average Annual Pay ($1,000s)	22.4989***	6.2526
% GDP Change	0.9892**	0.3018
Export Share of GDP	-0.9073***	0.2135
Natural Log of Cooling Degree Days (1,000s)	1.1698	1.6884
Natural Log of Heating Degree Days (1,000s)	1.7664	2.4458
2002 Intercept Differential	-2.2370	2.4071
2007 Intercept Differential	-22.8058	14.3903
2012 Intercept Differential	-21.9853	18.2113
R Square	**0.3314**	

Significance measures:

** Significant at 10% Level*
*** Significant at 5% Level*
**** Significant at 1% Level*

given year, the data point for Wyoming in 2002 was not included in the regression described in this paper. All other figures were formed using calculations that included Wyoming.

Discussion

After controlling for the variables described, an initial examination of the coefficients on the binary variables for the years after 1997 shows that the baseline Value Added to Energy Expenditure ratio for U.S. manufacturing have declined since 1997. Assuming the stability of real distillate fuel prices, real sector wages, GDP growth, and international trade conditions, the regression estimates that states' ratios declined by 2.237 from the year 1997 to 2002, and dropped by an additional 20.5688 by 2007. However, a small recovery of 0.8205 was observed in the ratio predicted by the regression.

While the coefficients on these binary variables are not statistically significant, it is reasonable to interpret these, as the data are representative of all 48 contiguous U.S. states in every year except 2002, in which only Wyoming was omitted. The regression results, then, can be treated as indicative of correlation within a population, and not a sample of that population.

The finding of lower expected ratios raises questions concerning what may be causing heavier usage of energy in the

domestic manufacturing sector, and whether such omitted variables are impacting the intercepts, or any of the other independent variables.

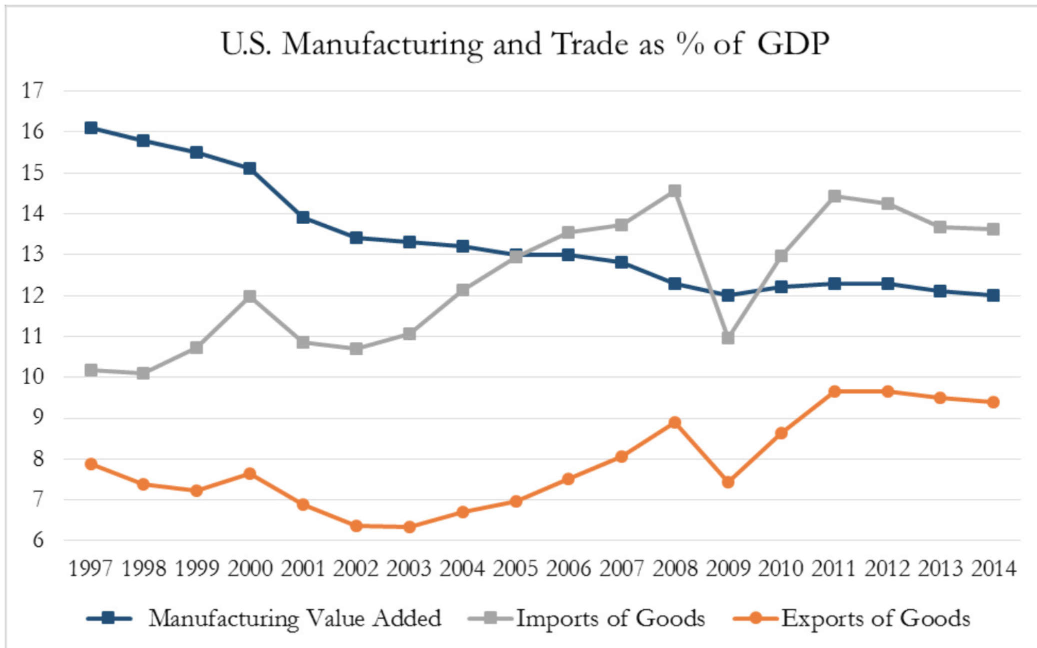
Lagged Effects of Market Indicators

It is no secret that U.S. manufacturing has suffered as a result of increasing global trade. As transportation costs and other barriers to overseas competitors have eroded, domestic manufacturers have responded by offshoring production, improving the efficiency of existing production, or exiting the domestic market. This effect can clearly be seen in the decline of manufacturing as a proportion of U.S. GDP, relative to the momentum of U.S. imports of goods.

The volume of the United States' imports is very well correlated with the volume of its exports. And, the volume of U.S. imports has been much larger than the volume of exports for many years. If a higher volume of imported goods indicates better availability of cheap manufactures for U.S. consumers and lower market prices for domestically made goods, thereby decreasing the value added per unit for domestic manufacturers – then higher volumes of exported goods are indicative of the same effect.

The effect of increasing international trade would be contained in the coefficient for export trade share in any regression model that controlled for exports, but not imports. That is certainly the case for this regression model, which predicts (at a 1% significance) that a 1% increase in state exports as a share of GDP will result in a reduction of the value-energy ratio of 0.9073. This is not an insignificant effect: an increase in the export share of GDP of about 3%, as was seen for the nation from 2002 to 2012, would yield an expected dip in the ratio of 2.7219.

Figure 2



However, the issue with the use of point-in-time export trade share in this regression model is that it assumes such trade activity only affects economic outcomes for the manufacturing sector in the same period. Trade activity in previous years is said to have had no bearing on the financial wellbeing of manufacturing firms, nor on the expectations of future market conditions of firm managers and investors. The extent of international competition in a given year only affects the value generated by production activity in the same year, regardless of whether competition was more or less intense in previous periods.

This is obviously a simplistic assumption. Modifications to stocks of operating capital equipment and similar decisions take place over the span of months or years. The outcomes of production decisions observed now are the result of a series of production choices in that span of time. These production choices were made not with perfect knowledge of market conditions in which their effects would become manifest, but with expectations of market conditions based on the most relevant available information. A perfected regression model, then, would seek to correct for the lagged effects of previous market conditions, in addition to current ones.

Market openness to international trade (export share of GDP) and domestic economic growth (GDP growth) are two such measures for which a lagged regression model might improve upon the current one.

Specialization in the Domestic Labor Market

It is well-known that many, if not most of the domestic manufacturing activity lost to offshoring has been unskilled labor-intensive. Much of the remaining manufacturing sector, then, is more capital and skills-intensive, and may be more difficult to relocate overseas as a result. Such manufacturing activity may also be difficult to relocate overseas due to a lack of expertise in the operation of such capital equipment outside of the United States. Whatever the degree to which this is true for the entire domestic manufacturing sector, it is clear that the model could be improved by distinguishing the costs of unskilled and skilled labor, and the extent to which both are being utilized within each state.

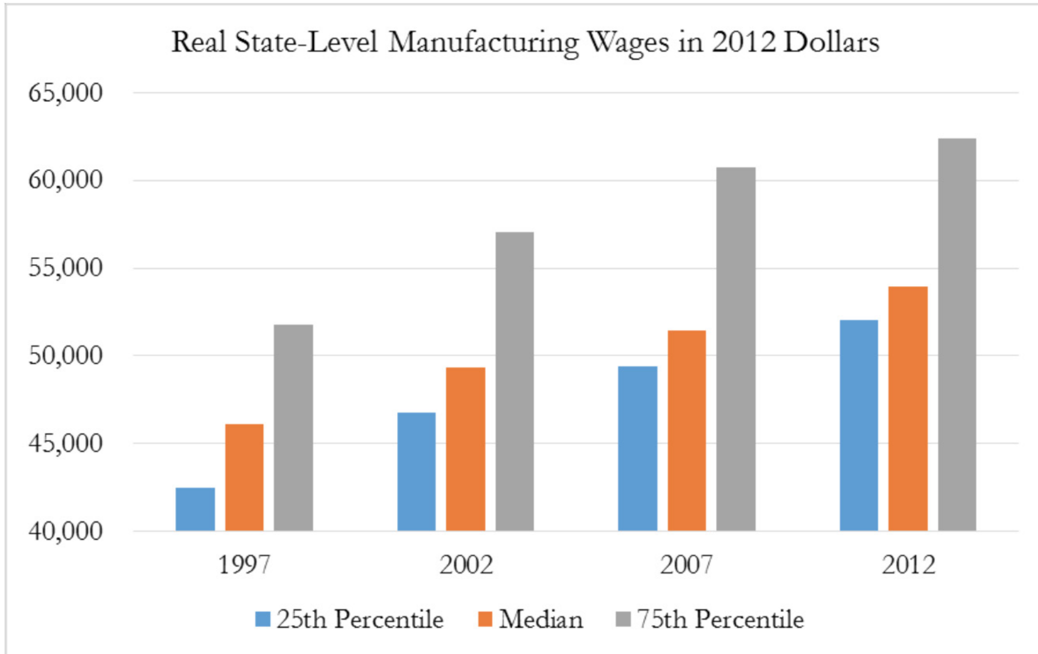
Specialized workers are often paid more than their unskilled counterparts, as they possess not only the given skills, but also education and experience that was accrued in the process of developing such skills. Increasing wages for and the increasing representation of specialized labor in the manufacturing labor force may well be signaled by rising average wages in the sector as a whole.

Indeed, Feenstra and Hanson (1995) “estimate that 15-33% of the increase in the relative wage of (skilled) workers in the U.S. during the 1980’s is explained by rising imports.”

The impact of international trade on U.S. labor specialization may also have worked in parallel with the demands of emerging technologies. Berman, Bound, and Griliches (1994), for example, have found that “biased technological change has (generated) increased demand for skilled labor.”

Whatever the causes of increasing capital-intensity and domestic labor

Figure 3



specialization are, it is possible that their impact on energy efficiency is already exhibited in the regression results. The coefficient on the natural log of real manufacturing wages – statistically significant at a 1% level – estimates that a 1% increase in real wages corresponds to an increase in the value to energy ratio of 0.225. Does this effect result from the energy efficiency of production processes that utilize more specialized labor, or does a generalized increase in market wages regardless of specialization force firms to maintain their productivity while trimming other variable costs? Without a more specific analysis, neither hypothesized effect can be examined.

Capital Vintage

As discussed in Section 2 of this paper, Steinbuks and Neuhoff (2014) found that manufacturers respond to energy price changes by adjusting for current energy prices in the short run, and by making improvements to the energy efficiency of their existing capital stock in the medium to long run.

While the regression examined in this paper does control for the responses of manufacturers to current energy prices, it fails to capture the impact of past energy prices on the decisions manufacturing firms make with regards to the maintenance, replacement, and updating of existing capital equipment. As with the omission of lagged market indicators, such a failure on the part of this analysis leads to ignorance of the impact of all production decisions whose effects are not immediately manifest. Given that this omitted effect is directly related to the energy efficiency of capital, the main vector through which improvements in technology are expressed, this omission represents a major flaw.

This issue highlights an additional dimension of analysis necessary for the quantification of the effect of technological change. While available contemporary production technology can only become more advanced and efficient, firms that might implement such advances do not necessarily do so if such an expense would make them less economically viable. The availability of improved technology is separate from the implementation of these improvements, and ought to be quantified separately. An estimation of the impact of longer run modifications to capital stocks in the regression model would allow the implementation effect to be reflected in the control variables for this impact, and leave the availability effect to be reflected in the moving regression intercept.

How might changes in capital stocks be causing the lower intercept predicted by the regression results in Section 4 of this paper? As previously discussed, in recent years the financial viability of U.S. manufacturing firms has come under pressure from the entry of foreign competitors into the domestic market for manufactured goods. As a result, manufacturing sector output as a percentage of U.S. GDP has continuously declined over the same recent time period.

In response to decreasing demand for U.S. goods and the increasing difficulty of maintaining firm liquidity, manufacturers may be reducing investment in efficient capital stocks. This allows manufacturers to hold more of the cost of such investment as liquid capital, staving off the possibility of bankruptcy. Doing so also allows manufacturers to reduce the scale of production through the retirement of older capital equipment, improving the economic viability of the firm with regards to the quantity of goods supplied, in light of increasing competition.

The aging of capital stocks in the manufacturing sector, however, causes improvements in the efficiency of available production technology to not be utilized in production technology. This aging may also deplete the efficiency of already-operating production equipment, as the engineering issues associated with the upkeep and operation of such equipment mount, potentially requiring the consumption of higher levels of electricity and fuel in performing such activities. These issues would surely impact the aggregate ratio of the manufacturing sector, and must be addressed in a more comprehensive model.

Government Regulations on Energy Usage

One major omitted variable in the proposed model is the effect of government regulations, particularly those on energy and the environment. The Environmental Protection Agency has put in place many federal standards for air particulates, the fuel efficiency of vehicles, and greenhouse gases. Many states have also put similar standards into practice that are even stricter than their federal counterparts. This has resulted in a diversity of both environmental policy and enforcement among the states.

Such regulations can clearly have an effect on energy expenditures. As manufacturers work to ensure that their production assets and practices meet government requirements, their productivity may also suffer, varying based on the severity of these requirements. A more complete model of energy efficiency would attempt to estimate the effects of this regulatory effect.

Conclusion

This paper examined the results of a regression model, which attempted to quantify the impact of various determinants of manufacturing energy efficiency within U.S. states, most notably the impact of advances in efficient production technology. The resulting analysis found clear evidence of a positive correlation of efficient economic value generation with the scale of the manufacturing sector within a state, the current price of electricity, the current level of manufacturing wages, and the GDP growth observed in the state in the same time period. It also found that states with higher levels of exports as a portion of GDP tend to be less energy efficient in economic value generation activity.

However, these results also gave unclear evidence of a general decline in the energy efficiency of manufacturing technology when corrected for these short run factors. The weakness of this evidence, and the ability of the model to explain only 33.14% of the sector energy efficiency variation among the states, implied that there were important variables missing from the analysis performed by the mode.

This author speculates that this was due to the failure of the model to capture the impact of lagged market effects on firm performance, structural changes in the capital intensity and specialization of domestic manufacturing, increases in the vintage of capital equipment within the larger sector, and government environmental regulations. Incorporation of such factors in a future study would provide information not only on the effects of these determinants, but on the impact of the general advancement of production technology. If performed properly, such a model would also provide robust results when applied to other geographic entities or points in time.

Works Cited

Berman, Eli, John Bound, and Zvi Griliches. "Changes in the Demand for Skilled Labor within U.S. Manufacturing: Evidence from the Annual Survey of Manufacturers." *The Quarterly Journal of Economics*, May 1994. Print.

Bureau of Economy Analysis. "Frequently Asked Questions: What is industry value added?" U.S. Department of Commerce, 10 Mar. 2006. Web. 25 Apr. 2015.

Bureau of Economic Analysis. "Interactive Data Application." U.S. Department of Commerce, 11 June 2014. Web. 27 Apr. 2015.

Bureau of Labor Statistics. "Quarterly Census of Employment and Wages." U.S. Department of Labor, 18 Sep. 2014. Web. 1 May 2015.

Economic Census of the United States. "Manufacturing: Geographic Area Series: Detailed Statistics for the State." United States Census Bureau. U.S. Department of Commerce, 27 Jan. 2015. Web. 27 Apr. 2015.

Feenstra, Robert C., and Gordon H. Hanson. "Foreign Investment, Outsourcing, and Relative Wages." The National Bureau of Economic Research, May 1995. Print.

FRED. "Gross Domestic Product: Implicit Price Deflator." *Economic Research*. Federal Reserve Bank of St. Louis, 27 Mar. 2015. Web. 27 Apr. 2015.

National Weather Service. "Degree Days Statistics." National Oceanic and Atmospheric Administration, U.S. Department of Commerce. Web. 25 Apr. 2015.

R Core Team. "R: A language and environment for statistical computing." R Foundation for Statistical Computing.

Schlumberger, Charles E. "Air transport and energy efficiency." The World Bank, 1 Feb. 2012. Web. 1 May 2015.

Scott, Michael J. et al. "Evaluating sub-national building-energy efficiency policy options under uncertainty: Efficient sensitivity testing of alternate climate, technological, and socioeconomic futures in a regional integrated-assessment model." *Energy Economics*. May 2014: 22-33. Print.

State Energy Data System (SEDS). "Energy Prices and Expenditures." Independent Statistics & Analysis. U.S. Energy Information Administration, 27 June 2014. Web. 27 Apr. 2015.

Steinbuks, Jevgenijs and Karsten Neuhoff. "Assessing energy price induced improvements in efficiency of capital in OECD manufacturing industries." *Journal of Environmental Economics and Management*. September 2014: 340-356. Print.

U.S. International Trade Data. "U.S. Exports of Goods by State, Based on Origin of Movement, by NAICS-Based Product." *United States Census Bureau*. U.S. Department of Commerce. Web. 27 Apr. 2015.

Zhao, Yue et al. "A comparative study of energy consumption and efficiency of Japanese and Chinese manufacturing industry." *Energy Policy*. July 2014: 45-56. Print.

The Role of Aggression in the National Football League

Lia Mancuso
Northeastern University

Introduction

There is a subculture of criminal activity, substance abuse, and aggressive behavior within the National Football League. When a player is found guilty of either off-field illegal activity or on-field violence, the player's indiscretion is met with a set standard of sanctions that are meant to discourage the certain behavior, such as a financial fine or multiple game suspension. The NFL does not explicitly condone such violent conduct, but when it comes to winning does having these types of players actually end up hurting the overall performance of the team? This paper explores the specific advantages and disadvantages that come from having a team with more criminal or violent players; whether it ends up stopping the team from winning by having to manage the multiple player fines, suspensions and arrests or whether it actually benefits the team to have more aggressive players.

By looking at the dependent variable (the team's winning percentage of the regular season) and evaluating it as a function of the regressions' independent variables, number of players with suspensions, with fines, and with arrests, and the points and sacks generated by players with with suspensions, with fines, and with arrests, I will infer if a relationship exists and if it is in fact harmful. I will be looking at team-level data for each NFL regular season[1] over a 5-year span, 2007 to 2011. Other variables that may have an influence on the team's winning percentage could include performance factors such as average points scored a game[2], average sacks made a game[3], as well as average age of the players and financial factors such as average salary for each team[4]. I will use these factors as well as the team's fan base and dummy variables, such as prior year winning record variable, to assess what role aggression plays in the NFL and if it is a significant role.

I aim to analyze the effect of fines, suspensions, and arrests on a team's performance as well as the effect of the performance of aggressive, violent and criminal players. Aggressiveness of players on a team is measured as fines (aggressive behavior), suspensions (violent behavior) and arrests (criminal behavior) within a regular season. Team's performance is measured by the team's winning percentage during the regular season. My expectation is that the penalties (fines, suspension, arrests) will have negative effects on the team's winning percentage as they are intended to. Additionally, the performances of the aggressive, violent or criminal players may also result in a harmful effect. This is because most of the players with fines are also penalized yards during the game, which is usually a contributing factor towards why a team loses a game. Players with suspensions and arrests are assumed not to be contributing to the team wins because they are not allowed to play during certain games, so rather than scoring touchdowns they are sitting on the bench. The penalties may also cause motivational problems within the team which could also be a reason for why teams lose games.

1 September 1st through December 31st.
2 By both aggressive and non-aggressive players (combined in regression A, as separate variables in regression B).
3 By both aggressive and non-aggressive players (combined in regression A, as separate variables in regression B).
4 Normalized as a percentage of the total league salaries.

Previous literature has not yet explored the effect of players' aggressive behavior on the outcome of National Football League games, but there has been some insight into whether or not an aggressive player is desired in the NFL draft. The literature suggests that NFL teams avoid drafting prospects with a history of criminal charges or suspensions, but my study would determine if this was a beneficial or detrimental tendency as well as incorporating the players with fines in addition to suspensions and criminal records. The implications of this study could be beneficial for NFL head coaches and general managers as they might have overlooked a player's behavior when assessing the abilities of the player and what effect they have on the team.

If it is determined that having such aggressive players can in fact harm an NFL team's winning percentage, then the implications for the NFL could be quite beneficial. Currently, how a player conducts himself in his daily life, or even how violent the player is on the field, is not of very high concern to NFL teams. Performance and player popularity are the most sought after traits of prospective NFL players, but if my study can sufficiently prove that having violent and criminal players on an NFL team is detrimental to the team's winning percentage, then the NFL should reevaluate each player's significance by factoring in their on and off-field acts of aggression. Many NFL fans place high value on NFL players and as role models their behavior could influence others. This paper is partly motivated on the recent stories about NFL players, such as Ray Rice and Adrian Peterson, displaying violent and abusive behavior in their personal lives. Furthermore, I noticed that Ray Rice and Adrian Peterson are above average players and wanted to further explore if the trait that makes them extraordinary football players is also the reason they are more prone to aggressive and violent behavior in their everyday lives.

Literary Review

The study by Weir and Wu's 2013 paper "Criminal Records and the Labor Market for Professional Athletes: The Case of the National Football League" from the Journal of Sports Economics (Or A study by Weir and Wu first published in the Journal of Sports Economics is highly relevant to this paper. In their article they observe the effect of criminal records on the labor market for professional athletes within the NFL. Weir and Wu (2013) use data from the NFL draft, the order in which prospective college football players are selected to a team, and relate the draft order to each player's "character variable," which is calculated based on the player's criminal record and history of suspensions during college. Their study then examines whether or not a player's history affects his draft position, and also if having a concerning history affects the player's performance. Their results show that NFL prospects with a history of criminal charges or team suspensions fall 16 to 22 spots in the draft order. When the authors separated those players with criminal records from those with team suspensions, they found that players with criminal records dropped roughly 16 spots when compared to players without any record, but players with a history of team suspensions dropped almost 22 spots. The study concluded that players with questionable "character values" did fall to the later rounds of the NFL draft, but players with team problems resulting in suspensions were considered less valuable than players with criminal records, and fell even lower. [Weir and Wu (2013)].

Sanderson explores the natural and unnatural advantages of athletic competition within Major League Baseball (MLB) in his study, "The Many Dimensions of Competitive Balance." While his study focuses mainly on financial benefits, technological enhancements, and natural skills, he also looks at a team's overall player character and integrity. The study discusses how leagues levy punishments on players who display aggression and violence in sports, both on the field and in society at large. He concludes that, the same way natural advantages that are not distributed evenly, "advantages derived from artificial and extraordinary interventions are no less legitimate." [Sanderson (2002)].

A Jones et al. study tests hypotheses relating to violence in the National Hockey League (NHL) and fan attendance. In their 2003 article, Jones et al. remark that in the NHL violence is "inevitable," and note how some players are often overvalued due to their violent tendencies rather than their hockey skills. This is because the fans have a "taste for violence," and the on-ice fights actually attract larger crowds. I found this interesting because, like the NHL fans, NFL fans seem drawn towards the over-confident and aggressive players, which unintentionally can create a league demand for athletes with aggressive or sometimes violent habits. [Jones et al. (1993)]. Since an NFL franchise goal is to make the most profit, and this study shows a correlation between sport violence and fan attendance or game revenue, it is a safe assumption that some NFL teams may be more willing to select more violent players in order to attract a larger fan base and capture a higher profit. Although it seems likely that NFL teams may want more aggressive and violent players on their teams in order to increase profits, Jones et al. do not address whether having these more aggressive and violent players actually helps the team win more of their games – not just put more fans in the seats. The latter idea is what I hope to expand upon in this paper.

Data Sources and Descriptive Statistics

The yearly data for this paper was gathered over a five-year timeframe, from 2007 to 2011, and is characterized as panel data at the team level. Data was collected from each of the 32 National Football League teams, resulting in 160 observations.

The independent variable, winning percentage, and control variables (average points per game scored, average sacks made per game, and team fan base) were collected from ESPN.com, (the website for the Entertainment and Sports Programming Network), the worldwide leader in sports recording and statistics. Each team's winning percentage is calculated by dividing team wins by total games played over the 17 weeks[5] of regular

5 17 weeks with 1 bye week for each team,

season play. Average points per game scored is used as a control variable and proved to highly correlate with winning percentage. Average sacks per game is also included to help capture the defensive performance's influence on the game. In Regression A, which examines the effect of fines, suspensions, and arrests on a team's winning percentage, the variables of points and sacks are calculated with both the aggressive and non-aggressive player performances combined. In Regression B, which examines the effect of the aggressive, violent, and criminal player performance on a team's winning percentage, the points and sacks by aggressive players are separate variables from the points and sacks by non-aggressive players. Other variables include the average age of a team's players, average salary spent on a team's roster[6], a dummy variable for if the team had a winning record the prior year, the size of the team's fan base[7], and the number of a team's players with fines, suspensions, and arrests during the regular season. It is important to note that only acts of aggression resulting in fines, suspensions, and arrests were included. Some examples of acts of aggression include: late hits, helmet-to-helmet hits, spearing, chop blocks, horse collar tackles, and striking/kicking/kneeing.

To find each team's players with fines and suspensions, I used the Spotrac Fines/Suspensions Tracker, affiliated with the USA TODAY Sports Media Group, which provides the largest online sports player system on the internet. The USA Today Salaries Database also provides the average salary data for the financial variable. In order to find the number of players on each team with arrests, I used the Arrest Database for NFL Player Arrests. Arrests and charges on the database are for crimes more serious than common traffic violations, including: theft, drug use, DUIs, gun related crimes, domestic violence, and assault. The data in the Arrest Database comes from media reports and public records, so some player arrests may not be documented for various reasons. This may include lack of media coverage or inaccessible public records. The remaining data was gathered from the Football Database, which compiles NFL team rosters with each player's date of birth, and by calculating each player's age for the given year, I determined the team's average player age.

The average annual number of NFL players arrested (between 2007 and 2011) was 21.6[8], or 1.3% of the total population of 1,696 NFL players[9]. According to the FBI, in 2011, "The arrest rate for violent crime (including murder and nonnegligent manslaughter, forcible rape, robbery, and aggravated assault) was 172.3 per 100,000 inhabitants." That comes out to about 0.17% of the population. The population of the NFL arrested for violent crimes is approximately six times the population of the US arrested. Looking at the descriptive statistics for Regression A in Table 1, the mean number of players on a team with fines is 1.94 with the 75th percentile being three or more players with fines. The standard deviation of players with fines is approximately two players, which

resulting in 16 games played.

6 Normalized as a percentage of the total league salaries.

7 Approximated by the fan attendance at team games both home and away.

8 2.6 avg. arrests for domestic violence, 1.2 avg. arrests for gun related crimes, 7.8 avg. arrests for DUIs, 6 avg. arrests for assault, 0.2 avg. arrests for theft, and 3.8 avg. arrests for drug related crimes.

9 32 teams with 53 players on active roster.

is double the standard deviation for players with suspensions or arrests. This indicates that fines occur more often among players (as arrests have a higher mean and standard deviation than suspensions). The mean for winning percentage is 50 with a standard deviation of almost 20. This is a large standard deviation, indicating a difference of three games won between each deviation.

Scatter plots comparing the main independent and dependent variables, of both Regressions A and B, displayed surprising trend lines. The scatterplot charts[10] suggest a positive correlation between winning percentage and the number of players on a team with fines. An example of this would be the 2007 New England Patriots, who had a perfect winning percentage during the regular season and tied for the third highest number of players with fines. Another instance of this would be in 2009, when every team with four or more aggressive players also had a .500 winning percentage or higher.

The scatter plots for Regression B[11] show that, although there was a positive correlation for both points and sacks made by aggressive, violent, and criminal players and winning percentage, there seems to be a stronger positive correlation with aggressive and violent players compared to criminal players. My goal is to discover what effects players with aggressive, violent, or criminal characteristics have on an NFL team's winning percentage, the scatter plots and Correlation Tables 3 and 4 (found in the Appendix) show significant proof of an existing relationship between aggressive players, number of fines, and winning percentage. Violent player performance was found to have a positive correlation with winning percentage, although not as strong as that of aggressive players, and the number of suspensions a team has possessed a slightly negative correlation. Surprisingly, both criminal player performance and number of arrests were found to a have a marginally positive effect on whether or not a team is more likely to win.

10 Found in Appendix A, section I, (A).

11 Found in Appendix A, section I, (B).

Hypothesis and Econometric Model

Panel data fixed effect regression models were used to test the proposed hypothesis. In addition to the main independent variables, certain variables are used to control for otherwise unobserved effects that may impact a team's winning percentage. Each econometric equation will regress performance of each NFL team with the number of fines, suspensions, or arrests, or the performances of aggressive, violent, or criminal players while controlling for performances of other players, financial, and dummy variables including team fixed effects.

To test my hypothesis, I conducted two separate regression equations in which each of my main independent variables were run with six or seven regressions. I also included a final analysis containing all of the control variables, resulting in a total of 39 regressions. Regression equation A aims to uncover the effect of the penalties (fines, suspensions, or arrests) on the team's winning percentage in order to assess the impact of each "consequence". Regression equation B aims to uncover the effect of the actual players (aggressive, violent, or criminal) on the winning percentage and explore whether players with these characteristics contribute or harm the team. By looking at the results of both regression equations A and B one can estimate the total net effect of both having these aggressive players on the team as well as the penalties these players are inflicted with. A VIF test was also run after running my regression equation in order to detect any multicollinearity.

Empirical Findings

The regression results found in tables five,six and seven proved to be interesting and displayed the effect of fines and arrests on the team's winning percentage. Although fines did not significantly impact statistics regarding the team's winning percentage, there was no statistical significant effect of fines on the team's winning percentage in table five, the coefficient was positive; therefore it implies implying the effects of fines on players may not harm the overall team's winning percentage. The negative coefficient results in table six convey that while suspensions are not significantly proven to harm the team, they certainly do not contribute to the team's winning percentage. Similarly, the effect of arrests on a team's winning percentage was also negative, but was proven significant at a 10% level. The regression results attributed a decrease of 2.4% in a team's winning percentage to every one additional arrest a team has, which indicates that arrests are indeed harmful to a team's winning percentage.

In tables eight, nine and ten, which look at the aggressive, violent and criminal player's performances on the winning percentage, there were some meaningful significances. Table eight proved the effect of both points scored by aggressive players and sacks made by aggressive players are significant at the 1% level. While the effect of points scored by aggressive players (2.941% increase in winning percentage for every one additional point scored by aggressive players) was smaller than the effect of points scored by non-aggressive players (3.011%), the two effects were similar in magnitude. is that a complete sentence? The effect of sacks scored by aggressive players (9.903% increase in winning percentage for every one additional sack made by aggressive players) was found larger than the non-aggressive players effect of 8.145%. These results were similar to the results of violent player's effect in table nine. The points scored by violent players was proven to have significance at the 1% level with an effect of 3.548% increase for every 1 additional point scored by violent players, the violent players effect being larger than the 3.010% effect of the non-violent players. Although the sacks made by violent players was not found to have at any significance, the coefficient (16.03) highly indicates a potentially positive relationship considering the effect of non-violent players sacks is half of the potential effect of violent players sacks. Finally, there was no significance found in table 10 for the performance of criminal players regarding either points or sacks (although sacks were found significant in regression five excluding the effect of a team's fan base). This indicates that criminal players do not possess a competitive advantage compared to their non-criminal counterparts.

The effects of aggressive and violent players on a team's winning percentage, although statistically significant, must be put into perspective to determine the practical magnitude of the relationship. With each win or loss, a team's winning percentage changes by 6.3%.[12] Therefore, in order for the relationship to prove meaningful to a team, the team would need to have aggressive player's' average score greater than three points a game or average sacks of at least 0.6. Table two revealed the mean points scored by aggressive players is 0.82 with standard deviation of 2.49. This means that in order for the aggressive players to improve the team's winning percentage by 1 one win (6.3%) the points scored by aggressive players would need to increase by almost a full standard deviation. The mean of sacks made by aggressive players is 0.26 with a standard deviation of 0.42, which means in order for the aggressive players to improve the team's winning percentage the sacks made by these players would also need to increase by a full standard deviation. This proves that while the majority of aggressive players will not be able to improve their team's winning percentage by 1 win, there are aggressive players performing at or above one standard deviation from the mean. Thus, their effect on the team could result in a substantially improved winning percentage. Although unlikely, it is possible for select players to produce the same results as violent players when violent players are able to score to increase the team's winning percentage by one win and score greater than one standard deviation. Similar to how the violent players points scored could improve the team's winning percentage The fact that the relationship between winning percentage and the performances of aggressive and violent players is statistically significant could mean that there are applicable winning implications to a National Football League team having players with aggressive or even violent tendencies.

Looking at the results of both regression equation A and regression equation B together, one can conclude that aggressive players do have a positive effect on the team, even with the penalty of fines, and can also conclude that criminal players have a negative effect on the team. What is inconclusive is the net effect of violent players, whose performances have been proven to be positive; however, the penalty of suspensions has a negative effect on the team.

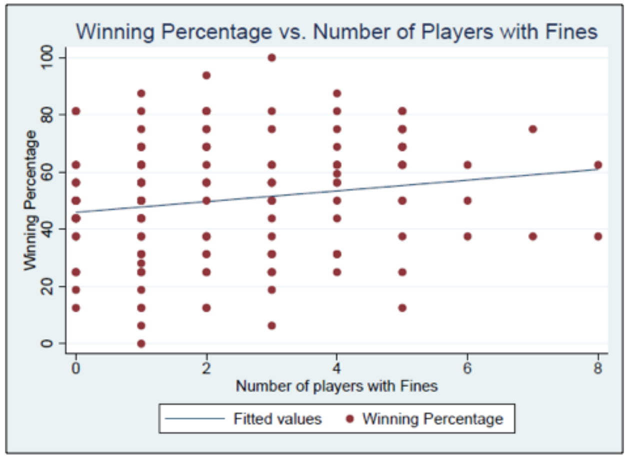
Although the regression in this study has provided ground-breaking new insights into the game of football and how it is played, it is still important to look at the potential flaws in the data. There may be an omitted variable bias due to left out variables (potentially passing and rushing yards per game, yards allowed per game , coaching experience , salary cap efficiency and others) which could overstate or understate the effect that players or penalties have on the team's winning percentage. For future direction of this research, the possibility of studying the data at the player level may help to further assess the effect of specific players on the team and how their distinct behavior may help or harm the team's winning percentage. Regardless, these results have serious implications to the NFL as now coaches and managers will be able to take each player's behavior into consideration, as it has proven likely to have an effect on the overall performance of the team.

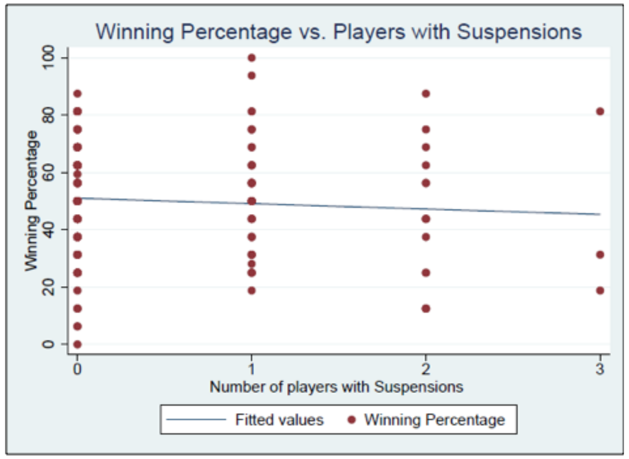
12 One game, divided by 16 total games equals 6.25%.

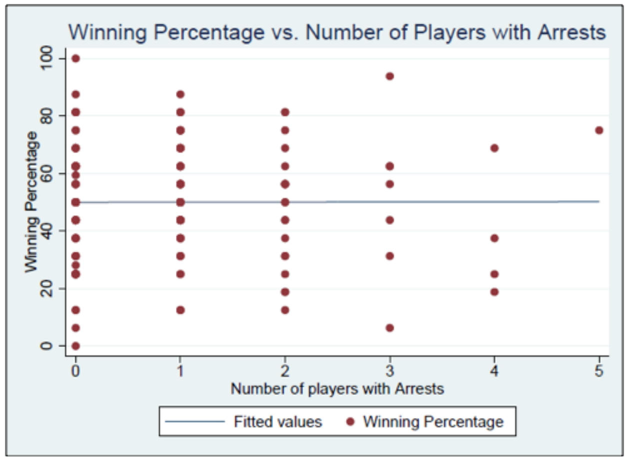
Appendix A

Section I

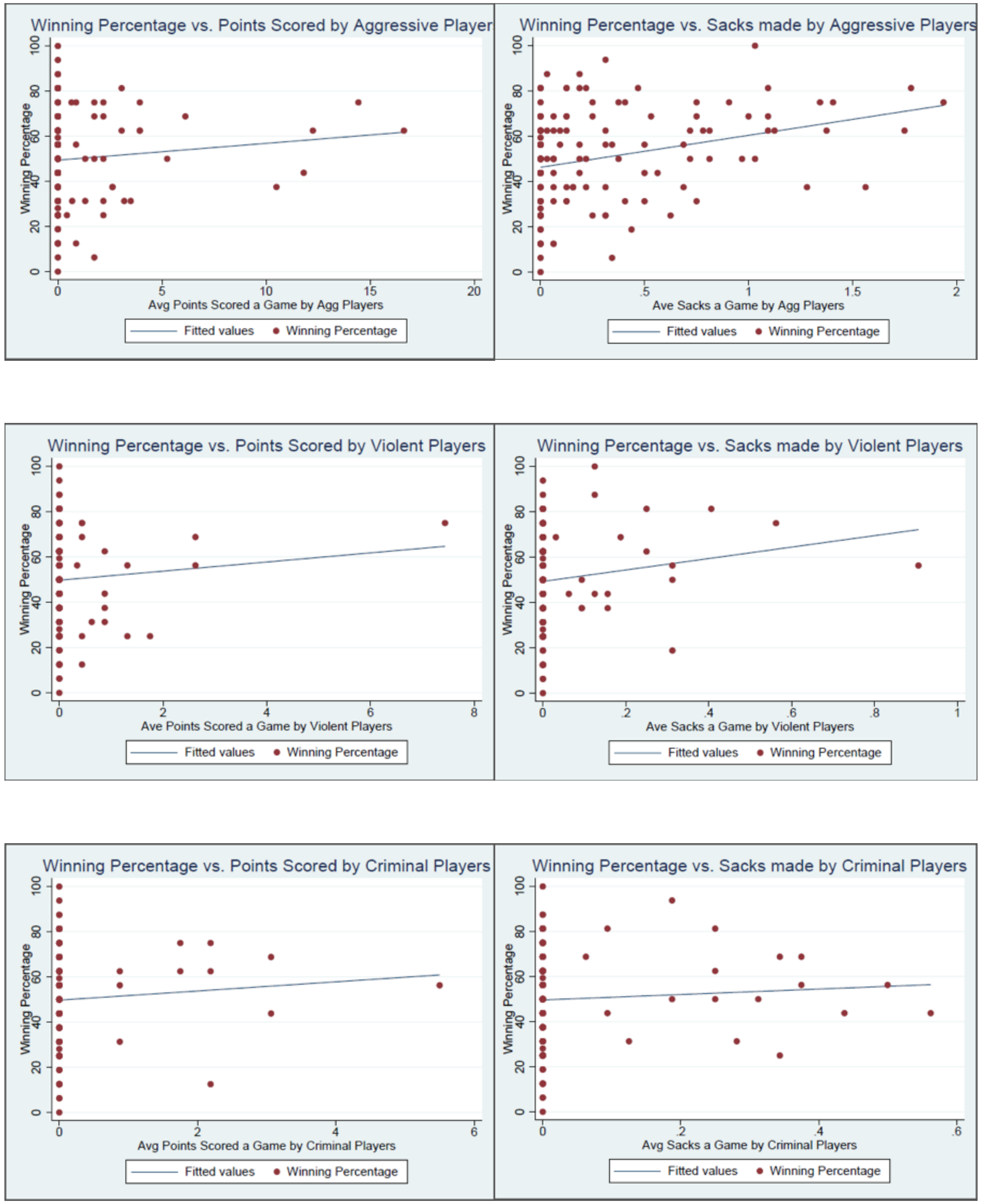
(A)







(B)



Section II

(A)

Table 1. Summary statistics:	**NFL Team data**								
	mean	**sd**	**min**	**p5**	**p25**	**p50**	**p75**	**p95**	**max**
Team	16.50	9.26	1.00	2.00	8.50	16.50	24.50	31.00	32.00
Year	2009	1.42	2007	2007	2008	2009	2010	2011	2011
Number of Players with Fines	1.94	1.78	0.00	0.00	1.00	1.00	3.00	5.00	8.00
Number of Players with Suspensions	0.51	0.74	0.00	0.00	0.00	0.00	1.00	2.00	3.00
Number of Players with Arrests	0.85	1.04	0.00	0.00	0.00	1.00	1.00	3.00	5.00
Winning Percentage	50.02	19.92	0.00	12.50	37.50	50.00	62.50	81.30	100.00
Avg Points Scored a Game by Team	21.89	4.72	10.90	13.90	18.30	22.35	24.60	29.15	36.80
Avg Sacks a Game by Team	2.17	0.50	0.63	1.44	1.81	2.06	2.53	3.00	3.69
Average Age of Players	27.02	0.90	25.00	26.00	26.00	27.00	28.00	28.50	29.00
Team Salary % of Avg Salary for League Teams	3.13	0.43	1.84	2.44	2.86	3.13	3.37	3.92	4.20
Prior Year Winning Record	0.44	0.50	0.00	0.00	0.00	0.00	1.00	1.00	1.00
Fan Base (thousands)	1052	112	427	919	1021	1061	1106	1194	1307
Observations	160								

(B)

Table 2. Summary statistics:	**NFL Team data**								
	mean	**sd**	**min**	**p5**	**p25**	**p50**	**p75**	**p95**	**max**
Team	16.5	9.26	1	2	8.5	16.5	24.5	31	32
Year	2009	1.42	2007	2007	2008	2009	2010	2011	2011
Winning Percentage	50.02	19.92	0.00	12.50	37.50	50.00	62.50	81.30	100.00
Avg Points Scored a Game by Agg. Players	0.82	2.49	0.00	0.00	0.00	0.00	0.00	3.94	16.63
Avg Points Scored a Game by Non-Agg. Players	21.06	5.03	8.59	12.57	17.55	21.32	24.32	28.40	36.80
Avg Sacks a Game by Agg. Players	0.26	0.42	0.00	0.00	0.00	0.05	0.34	1.20	1.94
Avg Sacks a Game by Non-Agg. Players	1.91	0.51	0.00	1.06	1.63	1.88	2.13	2.80	3.69
Avg Points Scored a Game by Violent Players	0.15	0.65	0.00	0.00	0.00	0.00	0.00	0.88	7.44
Avg Points Scored a Game by Non-Violent Players	21.74	4.75	10.90	13.65	18.06	22.20	24.60	29.15	36.80
Avg Sacks a Game by Violent Players	0.03	0.11	0.00	0.00	0.00	0.00	0.00	0.22	0.91
Avg Sacks a Game by Non-Violent Players	2.14	0.51	0.63	1.44	1.81	2.06	2.50	3.00	3.69
Avg Points Scored a Game by Criminal Players	0.15	0.65	0.00	0.00	0.00	0.00	0.00	1.31	5.50
Avg Points Scored a Game by Non-Criminal Players	21.74	4.75	10.90	13.90	18.10	22.10	24.60	29.15	36.80
Avg Sacks a Game by Criminal Players	0.03	0.10	0.00	0.00	0.00	0.00	0.00	0.30	0.56
Avg Sacks a Game by Non-Criminal Players	2.14	0.51	0.63	1.38	1.81	2.06	2.50	3.00	3.69
Average Age of Players	27.02	0.90	25.00	26.00	26.00	27.00	28.00	28.50	29.00
Team Salary % of Total League Salary	3.13	0.43	1.84	2.44	2.86	3.13	3.37	3.92	4.20
Prior Year Winning Record	0.44	0.50	0.00	0.00	0.00	0.00	1.00	1.00	1.00
Fan Base (thousands)	1052	112	427	919	1021	1061	1106	1194	1307
Observations	*160*								

Section III

(A)

Table 3 Correlation table for NFL Team data	**Team**	**Year**	**Number of Players with Fines**	**Number of Players with Suspensions**	**Number of Players with Arrests**	**Winning Percentage**	**Avg Points Scored a Game by Team**	**Avg Sacks a Game by Team**	**Avg Age of Players**	**Team Salary % of Avg Salary for Leage Teams**	**Prior Year Winning Record**	**Fan Base**
Team	1											
Year	0	1										
Number of Players with Fines	0.0879	0.2862	1									
Number of Players with Suspensions	0.0155	0.3099	0.0931	1								
Number of Players with Arrests	0.0112	-0.0728	-0.0694	0.1903	1							
Winning Percentage	-0.0212	-0.0002	0.1970	-0.0702	0.0021	1						
Avg Points Scored a Game by Team	-0.0443	0.0300	0.1581	-0.0540	0.0674	0.7851	1					
Avg Sacks a Game by Team	0.1711	0.1469	0.2602	0.0714	-0.0392	0.3951	0.3270	1				
Avg Age of Players	0.0341	0.1087	0.1702	-0.1702	-0.0840	0.1815	0.1422	0.2070	1			
Team Salary % of Avg Salary for Leage Teams	0.0675	0.0000	0.1279	-0.0213	-0.0298	0.1284	0.1311	0.2429	0.1644	1		
Prior Year Winning Record	0.0191	0.0535	0.2266	0.0701	-0.0549	0.2815	0.2804	0.1236	0.1585	0.0883	1	
Fan Base	0.0237	0.2167	0.1183	0.0061	-0.0780	0.1708	0.1271	0.1013	0.0925	0.1346	0.1734	1
Observations	160											

(B)

Table 4 Correlation table for NFL Team data

	Team	Year	Wining Percentage	Avg Points Scored a Game by Agg Players	Avg Points Scored a Game by Non-Agg Players	Avg Sacks a Game by Agg Players	Avg Sacks a Game by Non-Agg Players	Avg Points Scored a Game by Violent Players	Avg Points Scored a Game by Non-Violent Players	Avg Sacks a Game by Violent Players	Avg Sacks a Game by Non-Violent Players
Team	1										
Year	0.00	1									
Wining Percentage	-0.0212	-0.0002	1								
Avg Points Scored a Game by Agg Players	0.0029	0.0766	0.0930	1							
Avg Points Scored a Game by Non-Agg Players	-0.0430	-0.0097	0.6906	-0.3683	1						
Avg Sacks a Game by Agg Players	0.0752	0.1283	0.2992	0.0848	0.1776	1					
Avg Sacks a Game by Non-Agg Players	0.1061	0.0380	0.1405	0.0811	0.0796	-0.4328	1				
Avg Points Scored a Game by Violent Players	0.0678	0.0934	0.0702	0.0085	-0.0240	0.1425	0.0000	1			
Avg Points Scored a Game by Non-Violent Players	-0.0535	0.0161	0.7642	0.1316	0.8632	0.2101	0.1258	-0.1658	1		
Avg Sacks a Game by Violent Players	0.0461	0.0805	0.1328	0.0516	0.1044	-0.0007	0.0272	-0.0359	0.1418	1	
Avg Sacks a Game by Non-Violent Players	0.1588	0.1279	0.3614	0.1407	0.2054	0.3995	0.6323	0.1259	0.2709	-0.1801	1
Avg Points Scored a Game by Criminal Players	0.0987	-0.1215	0.0668	0.0084	0.0153	0.0769	0.0278	0.0614	0.0115	-0.0459	0.1011
Avg Points Scored a Game by Non-Criminal Players	-0.0576	0.0466	0.7707	0.1326	0.8637	0.2219	0.1229	-0.0294	0.9813	0.1439	0.2773
Avg Sacks a Game by Criminal Players	0.0171	0.0735	0.0607	-0.0678	0.0743	-0.0488	0.0509	-0.0676	0.0526	-0.0790	0.0265
Avg Sacks a Game by Non-Criminal Players	0.1647	0.1299	0.3763	0.1645	0.2121	0.4084	0.6270	0.1316	0.2894	0.0420	0.9561
Average Age of Players	0.0341	0.1087	0.1815	0.1031	0.0824	0.1273	0.0981	-0.0325	0.1450	0.1406	0.1746
Team Salary % of Avg Salary for League Teams	0.0675	0.0000	0.1284	0.0531	0.0967	0.1247	0.1356	0.0523	0.1217	0.0930	0.2198
Prior Year Winning Record	0.0191	0.0535	0.2815	0.0348	0.2459	0.1301	0.0135	0.0819	0.2647	0.0945	0.1021
Fan Base	0.0237	0.2167	0.1708	0.0819	0.0788	0.0328	0.0726	-0.0196	0.1283	-0.0684	0.1138
Observations	160										

Appendix B

(A) Fines - Table 5

	(1) Winning Percentage	(2) Winning Percentage	(3) Winning Percentage	(4) Winning Percentage	(5) Winning Percentage	(6) Winning Percentage	(7) Winning Percentage
Number of Players with Fines	0.200	0.145	0.0861	0.0865	0.0875	0.150	0.0829
	(0.839)	(0.678)	(0.693)	(0.683)	(0.682)	(0.689)	(0.722)
Avg Points Scored a Game		3.364^{***}	3.073^{***}	3.059^{***}	3.057^{***}	3.041^{***}	3.041^{***}
		(0.270)	(0.296)	(0.299)	(0.303)	(0.310)	(0.317)
Avg Sacks a Game			8.779^{***}	8.605^{***}	8.597^{***}	8.576^{***}	8.581^{***}
			(2.244)	(2.224)	(2.238)	(2.292)	(2.339)
Average Age of Players				0.850	0.846	0.946	0.853
				(0.894)	(0.917)	(0.896)	(0.901)
Team Salary % of Avg Salary for League Teams					0.205	0.410	0.198
					(3.025)	(3.158)	(3.108)
Prior Year Winning Record						-1.225	-1.258
						(2.654)	(2.664)
Fan Base							0.00000662
							(0.0000107)
Constant	49.63^{***}	-23.90^{***}	-36.46^{***}	-58.77^{**}	-59.23^{**}	-61.75^{***}	-65.41^{***}
	(1.630)	(6.149)	(6.570)	(23.11)	(21.99)	(21.98)	(22.52)
Observations	160	160	160	160	160	160	160
R^2	0.000	0.492	0.538	0.540	0.540	0.541	0.543
Adjusted R^2	-0.006	0.485	0.530	0.528	0.525	0.523	0.522
F	0.0571	78.00	61.14	49.51	41.50	34.83	32.16
rmse	15.23	10.90	10.42	10.43	10.46	10.49	10.51

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

(A) Suspensions - Table 6

	(1) Winning Percentage	(2) Winning Percentage	(3) Winning Percentage	(4) Winning Percentage	(5) Winning Percentage	(6) Winning Percentage	(7) Winning Percentage
Number of Players with Suspensions	-2.540	-1.392	-2.241	-2.135	-2.136	-2.098	-2.123
	(2.474)	(1.687)	(1.838)	(1.863)	(1.868)	(1.886)	(1.881)
Avg Points Scored a Game		3.341^{***}	3.012^{***}	3.007^{***}	3.005^{***}	2.995^{***}	2.993^{***}
		(0.277)	(0.300)	(0.303)	(0.304)	(0.310)	(0.315)
Avg Sacks a Game			9.486^{***}	9.358^{***}	9.348^{***}	9.325^{***}	9.332^{***}
			(2.088)	(2.068)	(2.086)	(2.115)	(2.175)
Average Age of Players				0.466	0.460	0.532	0.430
				(0.872)	(0.895)	(0.883)	(0.872)
Team Salary % of Avg Salary for League Teams					0.250	0.381	0.162
					(3.027)	(3.168)	(3.138)
Prior Year Winning Record						-0.806	-0.892
						(2.512)	(2.546)
Fan Base							0.00000731
							(0.00000967)
Constant	51.32^{***}	-22.40^{***}	-35.35^{***}	-47.60^{**}	-48.15^{**}	-49.93^{**}	-54.08^{**}
	(1.268)	(6.230)	(6.590)	(22.47)	(21.41)	(21.74)	(23.32)
Observations	160	160	160	160	160	160	160
R^2	0.014	0.495	0.549	0.549	0.549	0.550	0.551
Adjusted R^2	0.008	0.489	0.540	0.538	0.535	0.532	0.531
F	1.054	75.99	61.31	49.75	41.46	34.69	32.07
rmse	15.13	10.86	10.30	10.33	10.36	10.39	10.41

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

(A) Arrests - Table 7

	(1) Winning Percentage	(2) Winning Percentage	(3) Winning Percentage	(4) Winning Percentage	(5) Winning Percentage	(6) Winning Percentage	(7) Winning Percentage
Number of Players with Arrests	-0.585	-2.487^{*}	-2.467^{*}	-2.420^{*}	-2.422^{*}	-2.469^{*}	-2.440^{*}
	(1.862)	(1.300)	(1.362)	(1.344)	(1.345)	(1.361)	(1.378)
Avg Points Scored a Game		3.478^{***}	3.187^{***}	3.175^{***}	3.172^{***}	3.154^{***}	3.152^{***}
		(0.276)	(0.304)	(0.308)	(0.310)	(0.316)	(0.324)
Avg Sacks a Game			8.755^{***}	8.635^{***}	8.623^{***}	8.606^{***}	8.603^{***}
			(2.119)	(2.107)	(2.115)	(2.183)	(2.225)
Average Age of Players				0.591	0.585	0.705	0.636
				(0.871)	(0.899)	(0.870)	(0.878)
Team Salary % of Avg Salary for League Teams					0.282	0.535	0.373
					(3.177)	(3.319)	(3.297)
Prior Year Winning Record						-1.542	-1.602
						(2.512)	(2.532)
Fan Base							0.00000536
							(0.00000794)
Constant	50.52^{***}	-23.99^{***}	-36.63^{***}	-52.13^{**}	-52.75^{**}	-55.65^{**}	-58.87^{***}
	(1.583)	(5.780)	(6.140)	(21.98)	(20.65)	(20.58)	(20.84)
Observations	160	160	160	160	160	160	160
R^2	0.001	0.513	0.560	0.561	0.561	0.562	0.563
Adjusted R^2	-0.005	0.507	0.551	0.549	0.546	0.545	0.543
F	0.0985	81.98	66.71	55.47	45.28	38.14	36.91
rmse	15.23	10.67	10.17	10.20	10.23	10.24	10.27

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

(A) Aggressive Players - Table 8

	(1) Winning Percentage	(2) Winning Percentage	(3) Winning Percentage	(4) Winning Percentage	(5) Winning Percentage	(6) Winning Percentage
Avg Points Scored a Game by Agg Players	3.312***	3.001***	2.975***	2.974***	2.971***	2.941***
	(0.365)	(0.364)	(0.354)	(0.356)	(0.355)	(0.390)
Avg Points Scored a Game w/o Agg Players	3.371***	3.045***	3.031***	3.028***	3.005***	3.011***
	(0.277)	(0.312)	(0.317)	(0.321)	(0.332)	(0.336)
Ave Sacks a Game by Agg Players		10.01***	9.877***	9.874***	10.01***	9.903***
		(2.993)	(2.899)	(2.906)	(3.008)	(3.019)
Avg Sacks a Game w/o Agg Players		8.384***	8.190***	8.178***	8.113***	8.145***
		(2.448)	(2.448)	(2.472)	(2.540)	(2.585)
Average Age of Players			0.878	0.873	0.980	0.892
			(0.893)	(0.914)	(0.891)	(0.901)
Team Salary % of Avg Salary for League Teams				0.240	0.467	0.251
				(3.059)	(3.232)	(3.176)
Prior Year Winning Record					-1.320	-1.367
					(2.698)	(2.728)
Fan Base						0.00000657
						(0.0000106)
Constant	-23.71***	-35.22***	-58.23**	-58.76**	-61.21***	-65.18***
	(6.018)	(7.120)	(23.04)	(22.10)	(21.96)	(22.53)
Observations	160	160	160	160	160	160
R^2	0.491	0.540	0.542	0.542	0.543	0.544
Adjusted R^2	0.485	0.528	0.527	0.524	0.522	0.520
F	76.04	47.63	41.11	35.54	30.74	29.58
rmse	10.90	10.44	10.45	10.48	10.50	10.52

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

(A) Violent Players - Table 9

	(1) Winning Percentage	(2) Winning Percentage	(3) Winning Percentage	(4) Winning Percentage	(5) Winning Percentage	(6) Winning Percentage
Ave Points Scored a Game by Violent Players	4.221***	3.360***	3.444***	3.441***	3.489***	3.548***
	(0.834)	(0.966)	(0.979)	(0.982)	(0.996)	(0.976)
Avg Points Scored a Game w/o Violent Players	3.360***	3.046***	3.034***	3.032***	3.015***	3.010***
	(0.272)	(0.309)	(0.309)	(0.313)	(0.323)	(0.333)
Ave Sacks a Game by Violent Players		15.58	15.08	15.08	15.28	16.03
		(11.37)	(11.08)	(11.09)	(11.44)	(11.75)
Avg Sacks a Game w/o Violent Players		8.857***	8.660***	8.651***	8.627***	8.626***
		(2.360)	(2.361)	(2.374)	(2.440)	(2.501)
Average Age of Players			0.838	0.833	0.937	0.833
			(0.905)	(0.928)	(0.899)	(0.911)
Team Salary % of Avg Salary for League Teams				0.220	0.423	0.192
				(3.064)	(3.219)	(3.156)
Prior Year Winning Record					-1.235	-1.346
					(2.637)	(2.677)
Fan Base						0.00000786
						(0.0000105)
Constant	-23.65***	-36.11***	-58.07**	-58.56**	-61.05***	-65.66***
	(5.971)	(6.443)	(23.39)	(22.35)	(22.20)	(22.69)
Observations	160	160	160	160	160	160
R^2	0.493	0.540	0.542	0.542	0.543	0.545
Adjusted R^2	0.486	0.528	0.527	0.524	0.522	0.521
F	76.14	45.35	38.88	34.09	29.15	28.31
rmse	10.89	10.43	10.45	10.48	10.50	10.52

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

(A) Criminal Players - Table 10

	(1) Winning Percentage	(2) Winning Percentage	(3) Winning Percentage	(4) Winning Percentage	(5) Winning Percentage	(6) Winning Percentage
Avg Points Scored a Game by Criminal Players	2.051	1.858	1.867	1.867	1.928	1.926
	(2.223)	(2.159)	(2.236)	(2.243)	(2.346)	(2.356)
Avg Points Scored a Game w/o Criminal Players	3.377***	3.069***	3.054***	3.054***	3.042***	3.043***
	(0.268)	(0.293)	(0.297)	(0.301)	(0.309)	(0.313)
Avg Sacks a Game by Criminal Players		19.76*	19.89*	19.89*	19.78*	18.96
		(10.57)	(10.67)	(10.76)	(10.84)	(11.61)
Avg Sacks a Game w/o Criminal Players		8.780***	8.599***	8.599***	8.590***	8.588***
		(2.191)	(2.172)	(2.191)	(2.232)	(2.268)
Average Age of Players			0.889	0.889	0.957	0.889
			(0.858)	(0.878)	(0.866)	(0.874)
Team Salary % of Avg Salary for League Teams				0.00374	0.142	0.0128
				(2.887)	(3.044)	(3.008)
Prior Year Winning Record					-0.834	-0.898
					(2.738)	(2.789)
Fan Base						0.00000478
						(0.0000106)
Constant	-23.69***	-36.37***	-59.68**	-59.69***	-61.33***	-64.08***
	(5.921)	(6.370)	(21.95)	(21.18)	(21.42)	(22.46)
Observations	160	160	160	160	160	160
R^2	0.494	0.545	0.547	0.547	0.547	0.548
Adjusted R^2	0.487	0.533	0.532	0.529	0.526	0.524
F	81.70	48.02	42.56	37.72	32.40	31.67
rmse	10.88	10.38	10.39	10.42	10.45	10.48

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

	(1) Winning Percentage	(2) Winning Percentage	(3) Winning Percentage	(4) Winning Percentage	(5) Winning Percentage	(6) Winning Percentage	(7) Winning Percentage	(8) Winning Percentage	(9) Winning Percentage
Number of Players with Fines	1.061 (0.779)	0.338 (0.921)	0.318 (0.952)	1.642** (0.822)	2.418** (0.913)	2.384** (0.905)	2.366** (0.920)	2.385** (0.936)	2.308** (0.965)
Number of Players with Suspensions		-2.616 (2.498)	-2.570 (2.528)	-1.844 (1.963)	-2.343 (2.150)	-2.159 (2.167)	-2.163 (2.179)	-2.055 (2.205)	-2.073 (2.197)
Number of Players with Arrests			-0.214 (1.878)	-1.188 (1.577)	-0.839 (1.578)	-0.805 (1.566)	-0.820 (1.544)	-0.895 (1.528)	-0.864 (1.520)
Average Points Scored a Game w/o Agg. Players				2.170*** (0.365)	2.251*** (0.370)	2.233*** (0.372)	2.221*** (0.381)	2.204*** (0.389)	2.221*** (0.388)
Average Sacks a Game w/o Agg. Players					7.163** (2.799)	6.994** (2.794)	6.905** (2.821)	6.698** (2.922)	6.681** (2.977)
Average Age of Players						0.875 (1.290)	0.841 (1.306)	1.015 (1.307)	0.881 (1.308)
Team Salary % of Total League Salary							1.486 (3.995)	1.811 (4.249)	1.509 (4.125)
Prior Year Winning Record								-1.950 (3.099)	-2.014 (3.121)
Fan Base									0.00000954 (0.0000119)
Constant	47.69*** (2.655)	50.62*** (2.198)	50.82*** (2.897)	3.942 (8.396)	-12.89 (10.10)	-35.90 (35.76)	-39.14 (35.73)	-43.33 (36.40)	-48.71 (37.66)
Observations	160	160	160	160	160	160	160	160	160
R-squared		0.015	0.015	0.275	0.314	0.316	0.317	0.319	0.322
Adjusted R-squared		0.002	-0.004	0.256	0.291	0.289	0.285	0.283	0.281
F		0.566	0.385	12.25	11.44	9.384	9.779	8.769	10.99
rmse	17.21	15.17	15.22	13.10	12.79	12.81	12.84	12.86	12.88

Significance levels: * p<0.10, ** p<0.05, *** p<0.01

Works Cited

ESPN. ESPN Internet Ventures, n.d. Web. 23 Oct. 2014.

FBI. FBI, 09 Aug. 2012. Web. 19 Oct. 2014.

Jones, J. C. H., D. G. Ferguson, and K. G. Stewart. "Blood Sports and Cherry Pie: Some Economics of Violence in the National Hockey League." *American Journal of Economics and Sociology* 52.1 (1993): 63-78. JSTOR. Web. 16 Sept. 2014.

Sanderson, Allen R. "The Many Dimensions of Competitive Balance." *Journal of Sports Economics* 3.2 (2002): 204-28. *Elsevier*. Web. 16 Sept. 2014.

The Football Database. N.p., 2014. Web. 19 Oct. 2014.

"The Latest Sports Contract News, Figures, Projections and Breakdowns, including Salaries, Cap Hits, and Bonuses." Spotrac: *Sports Contracts, Salaries, Caps, Bonuses, & Transactions*. N.p., n.d. Web. 19 Oct. 2014.

"USA TODAY: Latest World and US News." *USA Today*. Gannett, n.d. Web. 19 Oct. 2014.

Weir, Kendall, and Stephen Wu. "Criminal Records and the Labor Market for Professional Athletes: The Case of the National Football League." *Journal of Sports Economics* (2013): 1527002513487739. *Elsevier*. Web. 16 Sept. 2014.

Effect of Increases in Minimum Wage on State Employment Rate

Steven Siano
Northeastern University

Introduction

This paper explores the connection between increases in the minimum wage and unemployment levels in states. In the past, economists have claimed that raising the minimum wage will increase unemployment because businesses will hire fewer workers due to the larger wage. However, recent studies have called this claim into question and argue that an increase in the minimum wage in a state does not cause an increase in the unemployment rate for that state. In this paper I will analyze a panel set of data from the 50 states and Washington, D.C. from 2012-2014. The results of the model used in this study demonstrate that increases in the minimum wage have no statistically significant effect on the unemployment level for a state.

Literary Review

Over the past few years, studies have cast doubt on the notion that higher minimum wages cause higher unemployment in that affected area. Two recent studies in particular shed light on different aspects of this debate. In January 2015, Jonathon Meer and Jeremy West released a study titled, "Effects of the Minimum Wage on Employment Dynamics." This study explores how the minimum wage affects employment over time. Another recent study by Joseph Sabia, titled "The Effects of the Minimum Wage over the Business Cycle" and published in the May 2014 edition of the Journal of Labor Research, examines the effects of a minimum wage increase. Sabia explores whether the effects of a minimum wage increase on low-skilled employment differ over the course of a business cycle.

In their study "Effects of the Minimum Wage on Employment Dynamics," Meer and West seek to prove that minimum wage increases cause changes in employment over time through the growth rate of new employment. They illustrate that traditional approaches measuring the effect of a minimum wage increase are misstating the true effects of the increase. By using a difference-in-difference identification strategy exercising state panel data sets, they estimate these effects. The data they use include the Business Dynamics Statistics (BDS), the Quarterly Census of Employment and Wages (QCEW), and the Quarterly Workforce Indicators (QWI). The panel measures data from all 50 states from 1975 to 2012.

Meer and West use a distributed lag model to best capture the dynamics of a minimum wage increase's effect on the growth of new employment. The estimated model from their data is as follows:

$$\textit{Predicted Employment} = -0.0825(\log MinWage) - 0.0524(1^{st}\text{ lag of }\log MinWage) - 0.0503(2^{nd}\text{ lag of }\log MinWage) - 0.0552(3^{rd}\text{ lag of }\log MinWage)$$

The coefficients for log*MinWage*, the 1st lag, and the 2nd lag are all statistically significantly different from zero at the 99% level. The coefficient for the 3rd lag is not statistically significantly different from zero at any level.

This study concludes that previous methods of estimating the effect an increase in the minimum wage has on employment levels systemically err if the true effects of the increase are on growth rates. They also find that the practice of including jurisdiction-specific time trends in the analysis bias the estimates towards zero. The main finding, though, establishes that the effects of an increase in the minimum wage reduce employment over a longer period than previously examined in literature.

In his study, "The Effects of the Minimum Wage over the Business Cycle," Joseph Sabia examines whether the impact of an increase in the minimum wage on low-skilled employment differ over the state of the business cycle. He reveals that, when controlling for spatial heterogeneity and state-specific non-linear trends, minimum wage increases from 1989 to 2012 reduced low-skilled employment more during recessions than expansions. He studies a set of panel data from the CPS Merged Outgoing Rotation Group (MORG) from 1989 to 2012 and generates state-by-year aggregate variables from the cross sections.

Sabia utilizes a weighted OLS model to estimate the effects of changes in the minimum wage and the employment effect of low-skilled workers across the state business cycle. He adds state effects, year effects, and state-specific controls into his model to best isolate the effect of the change in the minimum wage. He estimates these effects on two panels: first by real state GDP and by state unemployment rates. The controls he uses include: the natural log of the prime-age male unemployment rate, the average adult wage rate, real per capita retail GDP, real per capita manufacturing GDP, percent of the population ages 16-19 and 55-64. His estimates are represented in Table 5.

The most significant estimate is the (4) outlined by the box. In this regression that the effect of the log(*MW*) is significant at the 1% level in both panels.

In this study, Joseph Sabia concludes that increases in the minimum wage reduce employment of teenagers and younger workers without a high school degree regardless of the state of the business cycle. He also finds that increases in the minimum wage made during troughs likely have greater adverse effects on employment levels in states than they do during peaks of the business cycle. This is consistent with economic theory because firms are likely to have more elastic demand for labor during slack labor market conditions than they would in a tight labor market. Sabia advises states that wish to increase their minimum wage to enact the change during tight labor markets and at peaks of the business cycle to minimize the adverse effects of an increase. He also finds inconclusive evidence that minimum wages indexed to measures such as inflation do not have different effects than increases made by law.

Theoretical Model

The model I used for this analysis

Table 5

Heterogeneity in the low-skilled employment effects of the minimum wage across the state business cycle, 1989–2012

	(1) Teenagers ages 16 to 19	(2)	(3)	(4) Ages 16 to 24 without HS degree	(5)	(6)
	Panel I: Employment Effects by Real State GDP Growth					
Log (minimum wage)	−0.072 (0.074)	−0.120 (0.124)	0.066 (0.095)	−0.213*** (0.069)	−0.125 (0.180)	0.018 (0.142)
Real finance GDP growth of 0 to 2.5 %*log (minimum wage)	−0.089* (0.050)	−0.082** (0.037)	−0.109** (0.032)	−0.113*** (0.033)	−0.113** (0.033)	−0.135** (0.036)
Real finance gdp growth of<0 %*log (minimum wage)	−0.071+ (0.045)	−0.064 (0.045)	−0.091* (0.053)	−0.113*** (0.034)	−0.087* (0.052)	−0.125* (0.064)
	Panel II: Employment Effects by State Unemployment Rate					
Log (minimum wage)	−0.080 (0.078)	−0.126 (0.129)	0.055 (0.101)	−0.253*** (0.076)	−0.164 (0.201)	0.016 (0.147)
Prime age UR of 4.5 to 7.9 %*log (minimum wage)	−0.061* (0.033)	−0.071* (0.038)	−0.079** (0.037)	−0.047 (0.041)	−0.047 (0.056)	−0.124** (0.054)
Prime-age UR≥8 %*log (minimum wage)	−0.260*** (0.126)	−0.170* (0.101)	−0.083 (0.109)	−0.268** (0.121)	−0.124 (0.115)	−0.074 (0.126)
State effects?	Yes	Yes	Yes	Yes	Yes	Yes
Year effects?	Yes	Yes	Yes	Yes	Yes	Yes
State-specific controls?	Yes	Yes	Yes	Yes	Yes	Yes
State 3rd order polynomial trends?	No	Yes	Yes	No	Yes	Yes
Census division year effects?	No	No	Yes	No	No	Yes
N	1,224	1,224	1,224	1,224	1,224	1,224

Notes: Weighted OLS estimates obtained using CPS Merged Outgoing Rotation Group files from 1989 to 2012. State-specific time-varying controls include the natural log of the prime-age male unemployment rate, the average adult wage rate, real per capita retail GDP, real per capita manufacturing GDP, percent of population ages 16 to 19, and the percent of population ages 55 to 64. Standard errors corrected for clustering on the state are in parentheses

***Statistically significant at 1 % level **at 5 % level *at 10 % level + at 15 % level

draws from a set of panel data for the years 2009 to 2014 covering all 50 states and Washington, D.C. It is a time fixed-effects panel regression model. The dependent variable for this analysis is the unemployment level for each state over the period 2009 to 2014. This variable will be estimated from the independent variable of the change in the state minimum wage, the percentage change in national GDP from the previous year, and the binary variables for each of the years from 2009 to 2013. The binary variables of the years included in the period studied are included in order to account for any omitted variables that vary over time but not across the states. The variable of the change in the minimum wage was used to best capture the effect caused by a change in the minimum wage and not the nominal wage itself. The variable of the percentage change in national GDP was included to differentiate the effect of the growth of the US economy from other time variable effects across the panel. Originally, the variables per capita income and the change in the labor force were included in the model to attempt to isolate the effect of the minimum wage from other macroeconomic factors. However, during the estimation of the econometric model both were excluded for reasons that will be discussed in later sections. The theoretical model for this analysis is:

$$\textit{Predicted Unemployment Level} = \beta_0 + \beta_1 \text{Log}(Min\_Wage) + \beta_2(\%change\_in\_GDP) + \delta_2(2009) + \delta_3(2010) + \delta_3(2011) + \delta_4(2012) + \delta_5(2013)$$

Where β_0 is the intercept for the regression, β_1 is the effect of a change in the minimum wage on the overall unemployment level, β_2 is the effect of a percent change in national GDP on the unemployment level, and δ_1-δ_5 represent the intercepts for each year.

Data

The data for this analysis was acquired from the Bureau of Labor Statistics (BLS) and the Labor Department. The unemployment levels for each state represent the average annual levels of unemployment recorded by the BLS in the Local Area Unemployment Statistics program. The data gathered in the program was recorded from 1976 to 2014; however, only the period from 2009 to 2014 is analyzed here. There are no gaps in the data and the levels were calculated using the same methods for each year. The levels represent the number-of-persons population estimates utilized by the labor force to stay consistent with the latest Census numbers.

The data for the minimum wages for each state from 2009-2014 was obtained from the Labor Department. The Labor Department keeps records of all nominal minimum wage levels in each state, including Washington DC and the US territories. Some states incorporate ranges for their minimum wages depending on the type of work that is done. For these states, the highest minimum wage level was chosen. In addition, some states do not have a minimum wage set by state law. For these states, I used the federal minimum wage level that was set at $7.25 per hour in 2009. Some states have set minimum wages below the federal minimum wage level. Those levels were included in this analysis because they affect some jobs within the state.

The time period from 2009 to 2014 is of importance. The last time the federal minimum wage was raised was in 2009, to $7.25 per hour. In order to isolate the effects of a state level increase in the minimum wage on the unemployment level, this period was chosen. An increase in the federal minimum wage would dilute the effects of an individual state raising its minimum wage on its own due to the mobility in labor. If the minimum wage for the country increases, all states would be affected equally. The objective of this analysis is to study how changes at the individual state level affect the unemployment level in that state.

The percentage change in the national GDP was obtained from the Bureau of Economic Analysis. The data was taken from the National Income and Product Accounts tables. The levels represent the annual percentage change from the preceding year in real Gross Domestic Product. This data was included to separate the effects of the growth rate on unemployment from the other omitted variables captured by the binary time variables. As GDP increases, the unemployment level should decrease to keep pace with the growing economy.

Estimated Model

The estimated model used in this analysis is a time fixed effects model for panel data. The panel data are strongly balanced because there are no missing observations from any state in any year. The total number of observation is 306 from 51 different groups in the data. The dependent variable in this regression is the unemployment level. The independent variables used are the log of the minimum wage, the percentage change in national real GDP, and the binary variables for the years 2009-2013. The binary variable for 2014 was not included in the estimated regression to avoid perfect multicollinearity from the dummy variable trap. During the estimation, the binary variable for 2009 was omitted by STATA because of collinearity with the percentage change in real GDP. The estimated result is detailed below[1]:

Predicted Unemployment Level =
325716.2* – *48208.05*(Log(*Min_wage*))**
– 16374.66*(*%_change_real_GDP*)**
+ 102325.9*(*y2010*) + 67752.82***(*y2011*)**
+ 53730.66*(*y2012*) + 31730.77*** (*y2013*)**

The standard errors used in this regression were clustered robust standard errors and the attached values of the standard errors for the coefficients are: 122000.7, 62646.99, 18915.44, 14071.9, 10953.17, 5978.515, and 2844.45 respectively. The only variable that was not statistically different from 0 at any level was the coefficient for log(*min_wage*), while the rest of the coefficients were significantly different from 0 at the 99% level. The F statistic for the overall regression is F(6,50) = 5.9 which is significant at the 99% level. The R^2 value for within the panel data is 0.339, the R^2 between years is 0.0238 and the R^2 overall is 0.0079.

To test for heteroskedasticity within the regression, a modified Wald test for group wise heteroskedasticity in fixed effect regression models was used. The null hypothesis of the test is that there is homoskedasticity within the model. The chi^2 value returned by the test was 84441.94, which corresponds to a 0.000 p-value for homoskedasticity being present in the regression. This results in rejection of the null hypothesis meaning that there is a heavy presence of heteroskedasticity present in the regression. To correct for this in my model I used robust standard errors.

1 *** Statistically significant at the 99% level,
** Statistically significant at the 95% level,
*Statistically significant at the 90% level

To test whether there was time fixed effects in the model, I tested the parameters of the binary variables. The F-test for this has a null hypothesis of the time binary variables jointly equaling zero. The test returned an F-statistic of F (4, 50) = 8.89. This corresponds to a p-value of 0.000, leading to a reject null hypothesis. This means that the binary variables for the time fixed effects should be included in the model.

In order to test for multicollinearity within the analysis, I compared the R^2 values of the estimated regressions of the minimum wage and the time fixed variables on the change in the unemployment level and of the time fixed variables on the minimum wage. The R^2 overall for the latter regression was 0.073 is greater than the 0.008 value for the model. Next, I analyzed the correlations of the independent variables on themselves and on the dependent variable with the purpose of finding strong correlations between variables. The highest correlation was *y2009* to *%_change_real_GDP* at -0.989, which is big enough to cause multicollinearity but was dropped from the regression to correct for this. I concluded that there was not a significant presence of multicollinearity within the model used.

To confirm that the explanatory variables should be included in the regression, I tested the joint hypothesis that they were jointly equal to zero. This test returned an F-statistic of F (2, 50) = 16.91 that corresponds to a p-value of 0.000. This results in the rejection of the null hypothesis that the effects of the variables are jointly equal to zero. This is likely due to the significance of the percentage change in real GDP on the unemployment level.

The major finding from this model is that increases in the minimum wage do not significantly affect the unemployment level in that state. In fact, the negative coefficient on the log(*min_wage*) suggested that a 1% increase in the wage would translate to a decrease in the unemployment level of 48,208 workers. This is contrary to economic theory that would have predicted that an increase in the minimum wage would cause an increase in the unemployment level. The t-statistic for the change in the minimum wage equals -0.77 with a corresponding p-value of 0.445. This high p-value means that the effect of the change in the minimum wage is not statistically significant from zero at any confidence level below 55%. The 95% confidence interval for the effect of a 1% change in the minimum wage is -174038 – 77622, meaning that the estimated effect is extremely variable.

The time binary variables are all statistically significant and should be included in the regression. The t-statistics are: 5.41 in 2010, 4.81 in 2011, 4.91 in 2012, and 5.31 in 2013. The coefficients of these variables decrease in value throughout the period included from 102325.9 in 2010 to 31730.77 in 2013. This suggests that the variables affecting the unemployment rate that are consistent across states but vary over time, are improving over the period. This is consistent with economic conditions seen in this period. By most standards the national economy has improved since 2009 that would correspond to declining intercepts for each year included.

The large t-statistic and low p-value on the variable of *%_change_real_GDP* suggests that the improvements in the national economy, seen in the time binary variables, have the largest effect on state unemployment levels. According to the model, each percent increase in the national real GDP translates into a decrease of 16,375 workers in the unemployment level. This variable carries a t-statistic of -5.76 and a corresponding p-value of 0.000. According to this model, the most significant effect on unemployment levels in states during the examined period is the growth of the national economy. This is consistent with economic theory, which says that as the economy improves and grows, employment should grow with it and inversely unemployment levels will decline over time.

Conclusion

The overall result of this analysis is that the minimum wage does not influence the unemployment level. Instead, effects in the national economy that change over time play a more important role in determining the unemployment levels. This result breaks from traditional economic thought on the effects of increasing minimum wages but is more consistent with recent studies done on these effects.

Traditional economic thought argues that an increase in the minimum wage would increase the unemployment level in two ways. The first being that employers simply cut back on the number of employees they have because of the rise in wages. If the minimum wage increases, the rest of wages will likely follow over time. This causes the cost of employing workers to increase and some employers would reduce their number of employees in response to the rising cost. The second way an increase in the minimum wage would increase the unemployment level is that it would attract more people to

the labor force because of the increase in the opportunity cost of not working. The potential to make higher wages would likely cause people to enter the labor force and look for jobs but were not actively searching for jobs at the original minimum wage.

Traditional theory overstates the magnitude of both these effects. The argument of employers cutting jobs in response to a minimum wage increase over simplifies the rationale of hiring additional workers. Firms will hire employees to the point that the additional worker costs the firm more than that worker adds in productivity. A small increase in the minimum wage will raise the cost of hiring that worker, but the increase in consumption by workers associated with that increase might account for the increased cost or be even greater than the cost. Unless the increase in the minimum wage is substantially large, the effect of rising employer costs is likely overstated. In addition, a state level increase that is large enough to cause employers to cut jobs is unlikely because too high a minimum wage could cause some employers to move their production to another state with a lower minimum wage over the long run.

Traditional theory also overstates the additions to the labor force caused by an increased minimum wage. There is not a significant portion of those out of the labor force that choose to be out of the labor force because the minimum wage is not high enough currently for them to want to work. Most people that are not in the labor force are that way because of retirement, disability, or educational purposes. The segment of people who do not work because they would not make enough to cause them to give up their leisure is very small if at all existent. Most people who choose not to work because they would not make enough for them to sacrifice their leisure likely already have a level of income that can sustain their lifestyle without working. These people would not be affected at all by an increase in the minimum wage.

Recent studies have suggested what this analysis has concluded, that the effect of an increase in the minimum wage is not as big as once thought. This is especially true for increases at the state level because of the mobility of labor. The mobility of labor in today's economy creates incentives for states to keep their minimum wage at levels that do not seriously affect hiring decisions by employers in the state and cause them to relocate somewhere else. This forces states to adjust the increases to their minimum wages to keep pace with the overall growth in the economy and the cost of living in that state.

An increase in the federal minimum wage, though, would likely have different effects than one at the state level. An increase in the national minimum wage level may cause greater effects on employment because employers would not be able to move production to somewhere that can support them. This constraint to the mobility of labor could cause employers to either raise prices to the appropriate level or decrease employment to necessary levels; but, this effect should also be small due to the relatively small effect minimum wages play on overall unemployment levels.

The change in our economy towards a service based economy can be a possible explanation for the effects of changes in the minimum wage on in-state unemployment levels being so insignificant. Over the course of the last few decades, globalization has caused many manufacturing and production industries to move to countries with cheaper labor. This has been accompanied by increases of employment in service industries. These industries have much more inelastic demand for labor curves because of the localized nature of these jobs. These companies are more likely to substitute capital for labor over the long run if the wage rate increases too much.

The more influential effects on unemployment are the changes to national economic conditions. The model reflects this by the significance of both the time variables and the change in real GDP variable. This suggests that unemployment levels depend on factors in the overall economy that change over time and not on minimum wage levels. Economic theory agrees with this conclusion. As consumption and aggregate demand increase, firms will hire more workers to keep pace with the growing product demand. This will increase the employment levels at all wage rates through the scale effect.

In conclusion, the effect of a change in the minimum wage rate for a state does not significantly affect the unemployment level in that state. The model estimated in this analysis refutes economic theory that an increase in the minimum wage level translates into an increase in the unemployment level. Instead, the model supports the claim that conditions in the national economy have bigger effects on state unemployment levels than the minimum wage. These findings are consistent with recent studies that have demonstrated that there is a small effect on unemployment, if any at all, by increases in state minimum wages.

Works Cited

"Changes in Basic Minimum Wages in Non-farm Employment under State Law: Selected Years 1968-2014." US Department of Labor, Dec. 2014. Web. 25 Apr. 2015.

"Employment Status of the Civilian Noninstitutional Population, Annual Averages." *Civilian Noninstitutional Population and Associated Rate and Ratio Measures for Model-Based Areas*. Bureau of Labor Statistics, 27 Mar. 2015. Web.

Meer, Jonathon, and Jeremy West. "Effects of Minimum Wage on Employment Dynamics." *SSRN Working Paper Series* (2015): n. pag. Web.

"Percent Change From Preceding Period in Real Gross Domestic Product." *National Income and Product Accounts Tables*. US Bureau of Economic Analysis, 27 Mar. 2015. Web. 25 Apr. 2015.

Sabia, Joseph J. "The Effects of Minimum Wages over Business Cycle." *Journal of Labor Research* 35.3 (2014): 227-45. *Springer Link*. Web. 20 Mar. 2015.

Income Inequality and Financial Market Participation: Rural and Urban China

Yidan Jin
Smith College

Introduction

China's economy has grown significantly during the past decades since the economic reform in 1978, driven by increasing exports and investment. At the same time, dramatic growth had also increased structural and trade imbalances, which are related to income inequality (Zhu and Wan, 2012). As lower income groups cannot afford consumption which affects exports and investments, the engines of economic growth, the issue of income inequality has been widely discussed to explain the imbalance. Among various dimensions of income inequality, such as race and gender, the rural-urban gap in China is one of the largest in the world and would be even greater if differences in standard of living, welfare benefits, and infrastructure were taken into consideration (Wang and Piesse, 2010).

China's financial market has been developing since the economic reform and liberalization. A variety of financial products are available in China's market, including stocks, securities investment funds, bonds, and commodity futures (China's Financial Markets: An Insider's Guide to How the Markets Work, 2006, p.2). Household investment outcomes are playing an increasingly critical role in household wealth accumulation (Zhen, 2013) and households' behaviors in the financial market affect asset pricing and consequently determine market efficiency (DeLong et. al, 1990; Dumas, Kurshev, & Uppal, 2009). In this context, participation of households in financial markets has implications in maintaining households' wealth accumulation and in decreasing rural-urban wealth inequality.

The main purpose of this paper is to examine two topics: 1) rural-urban inequality in individual labor income, and 2) determinants of household financial market participation in investment products and loans. First, the paper focuses on highlighting the determinants of individual labor income by analyzing demographic factors and human capital. The rural-urban inequality in human capital explains most of the rural-urban labor income difference. This study also shows that there is a rural-urban gap in financial market participation and concludes that household income and availability of financial resources positively influence financial market participation. Then this study justifies the political recommendation to decrease rural-urban income inequality and wealth inequality.

Literary Review

Individual Labor Income Study

In 2009, urban residents earned 2.33 times more than those in rural areas, while the income of rural residents in coastal provinces tripled from 1989 to 2004. Since the 1980s, income inequality in China has risen at a faster pace than in the United States.

From 1980 to 2012, China's Gini coefficient increased from 0.30 to 0.55, surpassing the U.S. coefficient of 0.45. (Xie and Zhou, 2014)

Compared with other occupations, the overall level of farmers' income in China is low. Rural individuals have lower labor income because farming is more concentrated in rural areas. Yusuf and Saich (2008) explain that the size of the rural-urban income gap is influenced by the integration of rural-urban labor markets. They suggest that rural industrialization and rural enterprises have important roles in increasing rural labor income and minimizing the gap (p50). Lee (2013) also points out that the income inequality for urban households in China is mainly related to the coastal provinces with relatively higher return to capital, capital intensity, and thus capital income in the state sector. Similarly, Xia et al. (2013) demonstrate that urban wage inequality is affected by the changes in wage structure and employment share of the state sector. Besides the effect of the labor market's structure and employment share, Sicular et al. (2005) demonstrate that differences in educational characteristics between rural and urban areas contribute substantially to the gap. Zhu and Wan (2012) also confirm the rural-urban income inequality and suggest that government interventions can target rural-urban disparity through rapid urbanization, and tackle regional inequality by developing financial markets and ensuring progressive allocation of fiscal resources.

An important contribution of this study is that it draws from a well-censored sample of widely distributed respondents in China. It uses a decomposition method that is often used in gender income inequality studies to quantitatively analyze determinants of rural-urban income inequality.

Household Financial Market Participation Study

The rural-urban inequality in financial market participation is an indicator of rural-urban wealth inequality. According to the 2013 China Household Finance Survey Report, the Chinese household financial market participation rate is low and informal financial sectors are very active. Rural households are more active in participating in informal financial sectors (China Household Finance Survey Report, 2013). While the formal (bank) financing is often claimed to be the main engine for economic growth (Ayyagari, Demirgüç-Kunt and Maksimovic 2010), informal financing accounts for about 28% of the total borrowing in China (Li and Hsu 2009). Formal financial services such as loans and insurance are absent in rural areas (Wang & Moll, 2010) and the demand for insurance in rural areas is constrained by the lack of insurance knowledge, compared with urban China (Cai, de Janvry, & Sadoulet, 2013). As a means of wealth accumulation, rural-urban inequality in household financial market participation would exacerbate the imbalance of rural-urban economic development. The absence of financial and insurance markets can also lead to highly variable household income and persistent poverty (Dercon and Christiaensen, 2011; Jensen, 2000; Rosenzweig and Wolpin, 1993).

Many studies have discussed the determinants of financial market participation such as the ownership of stocks and bonds.

Household income, gender, marital status, education, financial literacy, and culture all influence the participation rate. Income is crucial to the financial market participation. Grinblatt, Keloharju, and Linnainmaa (2011) suggest that household income and education are all key contributors to financial market participation. Education is found to have a strong positive effect on households' stockownership (Haliassos & Bertaut, 1995). Cole and Shastry (2009) report a remarkable 7% to 8% increase in the probability of financial market participation with only one additional year of schooling. Van Rooij, Lusardi, and Alessie (2011) find that those with low literacy are much less likely to invest in stocks. Nguyen (2006) finds household financial activity in Vietnam is determined by household size and agricultural work rather than distance to the nearest bank branch.

This paper contributes to the analysis of Chinese financial market participation and aims to determine how to increase wealth accumulation through participating in formal financial markets. It also gives particular predicted probabilities of households holding investment products and loans.

Data

We obtained the data from the China Household Finance Survey (hereafter CHFS), a nationally representative survey in China conducted by the Survey and Research Center for China Household Finance[1] from 2011 to 2013. It examines detailed information about household finances and assets including non-financial assets, financial assets and other household assets. It collects demographic data and labor income on an individual basis as well as financial market participation of households. The non-response rate was 11.6% (16.5% in cities and 3.2% in rural areas relatively), which is lower than that of other finance surveys conducted in China in the past, such as the Survey of Consumer Finance in 2010. It was conducted by face-to-face interviews with 29,324 individuals in 8438 households[2] covering 29 provinces and 1,048 communities.

Individual Labor Income Study

This study uses individual data set from the CHFS, excludes unemployed and retired individuals, and uses a subset of respondents who reported being employed at the time of the survey, reported their annual labor income (or being imputed[3] by CHFS), and were no younger than 16 years old. Self-employed workers, freelancers, and farmers are included in the analysis as the salary gap of these occupations is an essential determinant of the urban-rural gap in annual labor income. The analytic sample of N=7,074 is derived from list-wide deletion of respondents who had missing values on individual labor income or any one of the demographic variables of age, marital status, gender, education levels, occupation types and living in an urban area. Table 1 shows the

1 The Survey and Research Center for China Household Finance is based at Southwestern University of Finance and Economics.

2 Respondents of individual and household data sets are same but the head of family answers household survey.

3 In order to solve the problem of missing data, some important variables are imputed by the CHFS, this paper will use actual individual income as well as imputed individual income as dependent variables. The imputed variable of individual income is 34.10% in a sample size of 7,079 in this study.

Table 1

Individual labor income demographic statistics (N=7,074)[1]

		Total	Urban	Rural
Average annual individual labor income (RMB)		28,343.38	32,110.89	20,073.24
Medium annual individual labor income (RMB)		19,572.42	22,180.00	14,400.00
Average logarithm of annual individual labor income		9.81	9.95	9.49
Age	Average age	38.47	39.32	36.58
	Medium age	38.00	39.00	36.00
	Average age square	1,602.32	1,655.03	1,486.61
			Percentage(%)	
Gender	Male	61.04	58.62	66.35
	Female	38.96	41.38	33.65
Education	Below or primary school	14.05	9.28	24.53
	Junior high education	33.83	27.82	47.02
	High school education	37.26	43.64	23.26
	Four-year college degree or above	14.86	19.26	5.19
Marital Status	Single	17.50	13.44	26.42
	Married/living with a partner	80.11	83.87	71.86
	Seperated/divorced/widowed	2.39	2.70	1.72
Occupation	Farming	1.94	0.51	5.06
	Self-employed/freelance	7.85	7.37	8.90
	Employed by others	90.22	92.12	86.04

1 For categorical factors, index of dissimilarity is calculated in the way in the footnote as indicatd in the paper; it is the mean difference for continuous factors.

basic demographic data for relevant variables.

Respondents' annual labor income is in RMB. The individual labor income contains income they gain from their first job and second job if applicable. Marital status has three levels: single, married or living with a partner, separated or divorced or widowed. We examine human capital, measured by education level and occupation types. The CHFS uses nine rank ordered degrees to represent education levels of respondents: never attended school, primary school, junior high, high school, secondary/vocationalschool[4], college/vocational[5], undergraduate degree, Master's degree, and PhD degree. According to the average education level in mainland China and the nine year compulsory education system which requires people to finish their junior high school education, this study uses simplified variables to represent the educational achievement of respondents: below primary school or primary school education, junior high school education, senior high education, and four-year college degree or above. We also use a simplified classification of occupation: farmers, self-employed or freelance workers, and employed by other parties, including government

4 Secondary/vocational schools refers to the same level of education as high school but graduates will go to work directly rather than going to universities in China.

5 College/vocational refers to two or three years of college education but does not offer bachelor's degrees to graduates.

agencies, public institutions, military, NGOs, private enterprises, and others. The annual salary difference between occupations, especially between farmers and employees of formal enterprises, has an impact on the urban-rural labor income inequality since farmers earn far less than employees in China.

In general, the average annual individual labor income of urban residents is much higher than that of rural residents. The average annual labor income in rural areas is only 62.51% of that in cities. Table 1 shows us the general level of wage inequality in rural and urban areas in China. Rural and urban residents do not have a large age gap, but rural residents tend to be younger than urban residents. Because farming requires heavy physical labor and because of the tradition of a preference for sons in rural areas, men constitute a larger percentage of the total population when compared with cities. Married people and people living with a partner make up the major part of the respondents of our sample, but the rural area has a larger single percentage.

It is interesting to note that there is a big difference in human capital, measured by education and occupation. As expected, urban residents have higher education in general compared with rural residents. Especially the percentage of lowest education level in rural areas is more than twice of that in urban area. Moreover, the percentage of respondents with a four-year college degree or above is much lower in rural areas than in cities, which implies that higher education is not balanced. It is worth noting that higher education or even the nine year compulsory education system is not universal and promoted all around China. Since the percentage of farmers is not significant in our sample, farming has the smallest percentage in both cities and rural areas. However, farming is more common in rural areas.

Household Financial Market Participation Study

The CHFS provide information on household characteristics including household sizes, subjective attitude toward finance, non-financial assets, financial assets, income, and expenditures. Because this paper studies the difference between household financial market participation in rural and urban areas, it will focus on three main financial activities: investment products holdings, formal loans holdings, and informal loans holdings. After excluding respondents who do not report having any of the variables: holding investment products, holding formal, informal loans and responding to questions about demographic categories, the sample size for investment products analysis is 7,343 households and 8,050 households for having loans.

This study also examines holding investment products, having formal loans, informal loans, annual household income (RMB), household size, using credit cards, using any formal sources of information from media (newspapers, magazines, television, radio, and Internet[6]), owning non-financial assets (land, real estate, and vehicles), and interest in economics, politics, and social topics. We also include a variable that measures the patience of respondents and financial knowledge about interest rates and returns to make long-term financial plans.

6 SMS is counted as an informal source since people in china often use SMS as interpersonal communication.

This paper studies investment products holdings including owing stocks, bonds, mutual funds, derivatives, or wealth management products[7] by analyzing categories such as no available or convenient financial service[8], thinking the market is bad[9], insufficient knowledge of investment products[10], and respondents' investment attitudes. The above variables are represented as dummies: they will be counted as "1" if respondents answered "yes" and "o" otherwise, except investment attitude which has four levels: above average risk and return, average risk and return, below average risk and return, and not willing to answer or do not know[11].

To study Chinese households' loan holdings, this paper counts having loans from a formal bank as formal loans and borrowing from relatives, friends and colleagues, informal financial organization, and others as informal loans. Insufficient knowledge of loans or inconvenience of application is counted as 1 if respondents choose "do not know how to apply", "do not have confidence the loan would be grated at all", or "the application process is too troublesome".

7 Financial products do not include deposits, funds bonds, equities, derivatives, business assets, real estate, and personal property. It includes those offered by banks, brokers, or trust.

8 Including too far away from the security company, do not know where to open an account, cumbersome procedures, and limited financial resource.

9 Including too risky, returns are too slow, lost money previously, and term is too long.

10 Including do not know how to open an account, lack relevant knowledge, never heard of them, and afraid of being cheated.

11 The precise wording of the question can be found in the appendix: CHFS survey Part 1 A4012.

Table 2 displays demographic statistics of household financial market participation, showing the large rural-urban difference in the average and medium annual household income. Indeed, the urban average annual household income is more than twice of that of rural households. Rural households have higher average household size than urban households, which could be explained by higher demand of agricultural labor and relatively flexible one child policy in rural areas. For women who had a second child, those whose first child had been a daughter were often officially permitted to have a second child under the reformed family planning policy in rural China (Hesketh, Li, and Zhu 2005). Moreover, rural households have much higher agricultural work participation than expected. Respondents to this survey have a high percentage of owning non-financial assets including land, real estate, and vehicles, but the rural-urban gap is not remarkable. Rural households usually own land as non-financial assets, having a higher percentage than urban households in this category.

The difference of subjective attitude toward finance is not significant between rural and urban areas. Urban respondents have a higher rate of using any formal sources of information from media, but both groups of respondents have a low percentage (less than 10%) in only using informal sources. Compared with rural respondents, urban households are more impatient in financial investment, whereas rural respondents are more willing to wait for higher return.

Urban residents have an overwhelmingly higher rate of having investment products (14.03% compared with 1.87% of rural households) and express that the stocks and

Table 2

Household financial market participation demographic statistics (N=8,050 unless otherwise indicated)

	Total	Urban	Rural
Average annual household income (RMB)	55,379.77	69,000.13	33,801.43
Medium annual household income (RMB)	30,000.00	38,029.75	18,200.00
Average logarithm of annual household income	10.12	10.40	9.67
Average household size	3.49	3.24	3.89
		Percentage(%)	
Having non-financial assets[1]	93.86	90.98	98.43
Having a credit card	5.54	7.29	2.76
Using any formal sources of information	94.62	96.78	91.20
Interested in economics, politics, and social topics	77.43	81.24	71.40
Financial knowledge and patience of respondents			
Impatient	69.63	72.36	65.30
Patient	29.33	26.69	33.52
Do not know and not willing to answer	1.04	0.95	1.19
Loans			
Having fomal loans[2]	15.20	16.11	13.77
Having informal loans[3]	33.12	27.52	41.99
Insufficient knowledge of loans or inconvenience of application process[4]	18.29	12.79	27.00
Investment products (N=7,343)			
Having investment products[5]	8.98	14.03	1.87
No available or convenient services[6]	52.44	52.56	52.28
Respondents think the market is bad[7]	22.15	29.60	11.64
Insuffienct knowledge of investment products[8]	66.03	57.77	77.66
Investment attitude			
Above average risk and return	11.79	12.40	10.92
Average risk and return	24.32	25.66	22.40
Below average risk and return	62.43	60.68	64.91
Not willing to answer/ don't know	1.47	1.26	1.77

1 Including land, real estate, and vehicles

2 Having loans from a formal bank

3 Borrowing from relatives, friends and colleague, informal financial organization, and others

4 Including do not know how to apply, do not have confidence the loan would be grated at all, or the application process is too troublesome

5 Including owing a stock account, bonds, mutual funds, derivatives, and wealth management products

6 Including too far away from the security company, do not know where to open an account, cumbersome procedures, and limited financial resource

7 Including too risky, returns are tow slow, lost money previously, and term is too long

8 Including don't know how to open an account, lack relevant knowledge, never heard of them, and afraid of being cheated

bonds market is bad for investment. Rural households show a large percentage in reflecting insufficient knowledge of investment products. However, the availability and convenience of investment services are not significantly different between rural and urban areas as well as investment attitude, but most respondents accept below- average risk and return.

The rate of having formal loans is 15.2% and the rate of having informal loans is 33.12%. These numbers are not high in general, probably because not all respondents are in need of loans. Ma and Yi (2010) states that the average saving rate has been rising over time, so that the aggregate marginal propensity to save exceeds 50% in the 2000s. High savings imply that people have sufficient funds and do not often need loans. However, the informal loan rate has a large gap between rural and urban regions: the percentage of having informal loans in rural areas is 1.5 times that of cities.

Methodology

Individual Labor Income Study

In this section, we employ 1) Duncan's D-index of dissimilarity, 2) OLS regression, and 3) regression decomposition methods to examine determinants of rural-urban labor income inequality.

We first use Duncan's D-index of dissimilarity (Duncan and Duncan, 1955) to measure the compositional differences among categorical factors (marital status, gender, education levels, and occupation types) and the mean differences between rural and urban areas among continuous factors (labor income and age). This index is a measurement of social segregation, sensitive to changes in population distribution (Social research update, 2000). It is calculated as

where $D = \frac{1}{2}\sum_j \left|\frac{u_j}{U} - \frac{r_j}{R}\right|$ **U** is the total number of the urban residents, u_j is the number of urban residents in the j-th group, **R** is the total number of the rural residents, and r_j is the number of rural residents in the j-th group. The D-index can be interpreted as the percentage of urban (or rural) who need to switch groups before urban and rural distributions become equal.

Secondly, as described below, an OLS regression model as shown in equation (1) is developed to examine the associations of the logarithm of annual individual labor income (Y) with age, gender, education levels, marital status, occupation types and urban residency.

(1) $$Ln(Y) = \beta_0 + \beta_1 age + \beta_2 age^2 + \beta_3 male + \beta_4 education + \beta_5 marital + \beta_6 occupation + \beta_7 urban$$

To examine how much rural-urban differences in the means of each independent variable explain the average labor income gap, we apply an OLS regression model which eliminates "$\beta_7 urban$" separately on urban and rural laborers as shown in equation (2).

(2) $$Ln(Y) = \beta_0 + \beta_1 age + \beta_2 age^2 + \beta_3 male + \beta_4 education + \beta_5 marital + \beta_6 occupation$$

The decomposition method is similar

to those of Chang and England (2011), who show a precise amount of gender wage gap that is explained by discrimination in industrialized East Asia. It was developed from an extended model of the Oaxaca decomposition method (1973). Kim and Shirahase (2014) use the same method in testing cross-Hnational differences in income distribution between males and females.

As Chang and England (2011) have pointed out, coefficients of regression for separated groups tell us the rate of return to a unit change in the variable. Oaxaca (1973) and Jones and Kelley (1984) argue that coefficients of separated groups (urban and rural) equally evaluate how much differences in the means of each independent variable explain the average labor income gap. Therefore, we present the results of the percentage explained by independent variables by using both urban and rural coefficients. Using rural coefficients, we examine how different the average ln labor income of rural respondents would be if rural people retained their rate of return to the factor but moved to the urban group. The equation is given as follows:

$$\%\ of\ gap\ explained\ using\ urban\ slope = \frac{(urban\ mean - rural\ mean) * urban\ coefficient}{difference\ between\ urban\ and\ rural\ mean\ of\ (\ln income)}$$

$$\%\ of\ gap\ explained\ using\ rural\ slope = \frac{(urban\ mean - rural\ mean) * rural\ coefficient}{difference\ between\ urban\ and\ rural\ mean\ of\ (\ln income)}$$

The percent of ln income gap explained by the mean difference in each independent variable is achieve by taking the product of the independent variable's mean difference and its coefficient, which is then divided by the mean difference between urban and rural in ln labor income (Dummy factors such as education are explained by the sum of mean differences on all the dummies). They show the percent of the ln labor income gap that is explained by mean differences in each explanatory variable, with two estimates provided- one using urban and one using rural coefficients.

Household Financial Market Participation Study

We employ 1) Duncan's D-index of dissimilarity and 2) logistic regression models to examine determinants of household financial market participation. Using Duncan's D-index of dissimilarity, we measure the mean difference between rural and urban areas to measure continuous factors (income and household size) and the compositional difference among categorical factors (all other variables).

The logistic regression model–equation (3)–assesses the associations of holding investment products with annual household income, household size, financial knowledge and patience of respondents, investment attitude, urban residency, having a credit card, being interested in economics, politics, and social topics, having non-financial assets, having no available or convenient financial service, thinking the market is bad, and having insufficient knowledge of investment products as described in the data section. p represents the probability of having investment products:

$$(3)\quad Ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 HHI + \beta_2 urban + \beta_3 formal + \beta_4 cc + \beta_5 patience + \beta_6 nfa + \beta_7 size + \beta_8 serv + \beta_9 topics + \beta_{10} bad\ market + \beta_{11} insufficient\ knlg + \beta_{12} investment\ attitude$$

Another logistic regression model–equation (4)–estimates the association of having formal loans from banks with all independent variables, including annual household income, household size, financial knowledge and patience of respondents, urban residency, having a credit card, being interested in economics, politics, and social topics, having non-financial assets, having insufficient knowledge of loans or inconvenient application process. The CHFS combines questions of knowledge of loans and the convenience of loan services together so these two aspects are considered as one dummy variable reflecting the level of understanding and the service of loans. p is the probability of having formal loans:

$$(4)\quad Ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 HHI + \beta_2 urban + \beta_3 formal + \beta_4 cc + \beta_5 patience + \beta_6 nfa + \beta_7 size + \beta_8 topics + \beta_9 serv\&knlg$$

The above expression is also applied to informal loans, where p represents the probability of having informal loans.

Results

Results of Individual Labor Income Regression Model and Decomposition

The difference in the distributional disparity between urban and rural laborers varies in education levels and marital status as shown in Table 3.

The D-index for level of education is the highest among all variables, meaning that as many as 34.45% of rural laborers would need to change their educational degrees to achieve balance with the educational distribution of laborers in urban areas, or vice versa. Besides this highest compositional difference in education due to the unbalance of educational resources, there is also a relatively high level of rural-urban segregation in marital status when compared with gender and occupation. The D-index for marital status is 12.99%, showing that 12.99% of rural laborers would need to change the distribution of marital status to have the same distribution as urban laborers. The demographic data of marital status in Table 1 shows that rural areas have a higher percentage of single people, but marital status is partially influenced by age. Older people would have higher percentage of being married or living with a partner. It can be shown that the mean difference of age is 2.74: rural residents are 2.74 years younger than urban residents on average. Therefore, the D-index of marital status is not influential or significant.

With this understanding of the differences in urban and rural attributes, we now consider the analysis of earnings. The urban-rural earning gap, measured by the urban-rural difference in the average logarithm of annual individual labor income, is 0.46. Urban residents earn RMB 12,037.65 ($1,965.97) more than rural residents in China on average.

Table 4 presents the results of OLS regression–equation (1). The joint significance of age, gender, education and occupation were statistically significant with the overall F-test (p<0.0001). However, marital status was not statistically significant. Holding other independent variables constant, older people had a higher annual labor income, which could be explained by increased working experience. Holding other factors

Table 3

Index of dissimilarity--comparing the conpositional difference by urban and rural areas[1]

Individual laborers (N=7,074)	
Age	2.74
Age square	168.43
Annual individual labor income	12,037.65
Logarithm of annual individual labor income	0.46
Education	34.45%
Marital status	12.99%
Gender	7.73%
Occupation	6.08%
Households (N=8,050 unless otherwise indicated)	
Annual household income	35,198.70
Logarithm of annual household income	0.72
Household size	0.65
Formal sources of information	5.57%
Having a credit card	4.53%
Interested in economics, politics, and social topics	9.84%
Financial knowledge and patience of respondents	7.06%
Having non-financial assets	7.44%
Loans	
Having fomal loans	2.02%
Having informal loans	14.47%
Insufficient knowledge of loans or inconvenience of application process	14.21%
Investment products (N=7,343)	
Having investment products	12.16%
No available or convenient services	0.28%
Respondents think the market is bad	17.95%
Insuffienct knowledge of investment products	19.89%
Investment attitude	2.62%

1 For categorical factors, index of dissimilarity is calculated in the way in the footnote as indicatd in the paper; it is the mean difference for continuous factors.

Table 4

OLS regression results of annual individual labor income-equation (1) (N=7,074)

		Coefficient	Std. Error	Overall Significance Test
Age		0.051***	0.006	F(2, 7062) = 45.64 Prob > F = 0.0000
Age square		-0.000632***	0.000	
Urban	Rural (reference)			F(1, 7062) = 43.64 Prob > F = 0.0000
	Urban	0.162***	0.025	
Gender	Female (reference)			F(1, 7062) = 173.72 Prob > F = 0.0000
	Male	0.285***	0.022	
Education	Below or primary school (reference)			
	Junior high education	0.139***	0.034	F(3, 7062) = 323.77 Prob > F = 0.0000
	High school education	0.468***	0.035	
	Four-year college degree or above	1.105***	0.042	
Marital Status	Single (reference)			
	Married/living with a partner	0.049	0.037	F(2, 7062) = 2.09 Prob > F = 0.1243
	Seperated/divorced/widowed	-0.058	0.077	
Occupation	Farming (reference)			
	Self-employed/freelance	1.263***	0.084	F(2, 7062) = 150.75 Prob > F = 0.0000
	Employed by others	1.334***	0.077	
Intercept		6.859***	0.138	

F- test for joint significance

Adjusted R-squared = 0.2339

***p<0.0001

**p<0.01

*p<0.05

constant, men earn 28.48% more than women in China, demonstrating a large gender gap in annual labor income.

Education levels have a significant effect on annual labor income: the annual labor income increases with higher education levels achieved. When comparing respondents who received a four-year college degree or above with people who only attended primary school or below, the OLS results show a college degree or above is associated with 110.55% higher annual income than primary school, holding other factors constant. Occupation also illustrates the income difference between farmers, the self-employed and those employed by others. Among respondents, both the self-employed and those employed by others earn over 100% more than farmers in China, controlling for other variables.

Table 5

Detailed decomposition of rural-urban inequality in annual individual labor income-equation (2) (N=7,074)

Independent variables	Mean (Urban)	Mean (Rural)	Rural-urban gap[1]	Urban Coefficients	% of gap explained using urban coefficients[2]	Rural Coefficients	% of gap explained using rural coefficients[3]
Total % explained by mean differences					**0.7896**		**0.7206**
Age					**0.0780**		**0.0523**
Age	39.33	36.59	2.74	0.055***	0.3243	0.039***	0.2300
Age square	1655.03	1486.61	168.42	-0.0006***	-0.2464	-0.0005***	-0.1777
Gender					**-0.0429**		**-0.0600**
Male	0.59	0.66	-0.08	0.255***	-0.0429	0.357***	-0.0600
Education					**0.5849**		**0.5811**
Junior high education	0.28	0.25	0.03	0.195***	0.0139	0.08+	0.0057
High school education	0.44	0.23	0.20	0.539***	0.2358	0.315***	0.1378
Four-year college degree or above	0.19	0.05	0.14	1.103***	0.3352	1.44***	0.4376
Marital Status					**0.0058**		**0.0189**
Married/living with a partner	0.84	0.72	0.12	0.027	0.0070	0.088	0.0228
Seperated/divorced/widowed	0.03	0.02	0.01	-0.055	-0.0012	-0.179	-0.0039
Occupation					**0.1639**		**0.1283**
Self-employed/freelance	0.07	0.09	-0.02	1.555***	-0.0503	1.208***	-0.0391
Employed by others	0.92	0.86	0.06	1.629***	0.2142	1.273***	0.1674

Urban adjusted R-squared = 0.1784; rural adustedj R-squared = 0.2528

+ P<0.1

***p<0.0001

1 Urban-rural gap= urban mean-rural mean

2 Calculated by dividing [urban-rural gap*urban slope] by total difference between urban and rural mean of (ln labor income).

3 Calculated by dividing [urban-rural gap*rural slope] by total difference between urban and rural mean of (ln labor income).

Table 6

Logistic regression results of having investment products-equation (3) (N=7,347)

		Coefficient	Std. Error	Odds Ratio	Overall Significance Test
Logarithm of annual household income		0.540***	0.046	1.005***	
Urban		1.567***	0.149	4.791***	
Household size		-0.132***	0.034	0.876***	
Products or assets	Having a credit card	0.805***	0.128	2.237***	
	Having non-financial assets	0.499	0.201	1.646	
Financial service factor	No available or convenient services	-0.750***	0.099	0.473***	
Financial knowledge factor	Using formal sources of information	0.543	0.339	1.721	
	Interested in economics, politics, and social topics	0.356**	0.129	1.428**	
	Financial knowledge and patience of respondents				
	Impatient (reference)				chi2(2) = 2.43
	Patient	-0.159	0.102	0.853	Prob > chi2 = 0.2973
	Do not know and not willing to answer	-0.027	0.542	0.973	
	Respondents think the market is bad	0.062	0.098	1.064	
	Insuffienct knowledge of investment products	-0.713***	0.096	0.490***	
	Investment attitude				
	Above average risk and return	0.243	0.132	1.276	
	Average risk and return (reference)				chi2(3) = 36.49
	Below average risk and return	-0.407***	0.102	0.666	Prob > chi2 = 0.0000
	Not willing to answer/ Ddon't know	-1.971	1.020	0.139	
Intercept		-9.109***	0.633	0.00011***	

Chi-square tests for joint significance

Pseudo R^2=0.2061

***p<0.0001

**p<0.01

*p<0.05

Note: Reference groups:Living in rural areas; Not using formal sources of information; Having no credit card; Not interested in Economics, politics, and social topics; Impatient; Having no non-financial assets; Having available or convenient service; Having sufficient knowledge of stocks or bonds; Thinking the market is not bad; Average risk and return

Table 5 refers to the amount of the urban-rural gap caused by the urban-rural difference in each independent variable. The results of decomposition show that overall, 78.96% of the annual labor income gap can be explained by using urban coefficients; 72.06% of the gap can be explained using rural coefficients. The difference of using urban and rural coefficients is small. Among all independent variables, education contributes more than 50% of the annual labor income gap, followed by occupation and age. The main difference in education is that 19% of rural respondents are graduates of four-year college or above, while only 5% of rural respondents are. Meanwhile, the coefficient of this factor is the highest, representing strong influence on annual labor income and then on rural-urban income gap.

Results of Household Financial Market Participation

The index of dissimilarity is reported in Table 3. Among all categories, the difference in household subjective attitude toward finance (such as sources of information, patience, and interests) is not significant from the result of Duncan's D-index of dissimilarity. Together with having credit cards and non-financial assets, they are all below 10%. The high index of dissimilarity appears in the knowledge of investment products, indicating that 19.89% of rural workers would need to change to "have sufficient knowledge of \ investment products" in order to achieve balance with the distribution of household workers, and vice versa. Besides their prominent compositional difference, lack of knowledge of loans and inconvenience of loan application also stands out among all categories. And 17.95% of rural workers would need to change their opinions toward the market in order to be balanced with the distribution of urban households (vice versa). There are rural urban differences in owning investment products and informal loans. Rural has higher household sizes and lower household income, representing a higher financial pressure to maintain daily expenditure.

Table 6 reports the results of the logistic model—equation (3)—of holding investment products. Annual household income, living in urban areas, household size, having a credit card, being interested in economic topics, having no available service, and having insufficient knowledge of investment products are statistically significant. The Chi-square test of investment attitude also shows investment attitude's statistical significance.

In general, the odds ratio of holding investment products increases with household income, living in urban areas, having a credit card, and being interested in economics, politics and social topics. Conversely, larger household size, not having available or convenient financial service, having insufficient knowledge of investment products decreases the log of odds. Meanwhile, using formal sources of information, having non-financial assets, and investing above average risk and return is positively associated with the log of odds, but they are not statistically significant.

Indeed, holding other variables constant, living in urban areas results in 379.1% increase in the odds of having investment products. A 1% increase in household income is associated with a 0.5% increase in the odds of having investment products, controlling for others. This confirms that urban households with higher household income tend to have a larger probability in owning investment products than rural households with lower household income. We can also conclude that households responding that they have no available or convenient financial service are associated with a 52.7% decrease in the odds of having investment products and insufficient knowledge with a 51.0% decrease in the odds, controlling for other variables. These results verify that financial knowledge and available financial service influences the financial market participation rate in terms of investment products.

The following hypothetical cases show how our model predicts the probability of having investment products taking into account living in urban areas, annual household income, insufficient knowledge and the availability of service:

Case 1: *Rural and urban households:*

The predicted probability of having investment products for urban households with medium household income RMB 26,900 ($4,393.27), average household size 3.52 people (N=7,347) and all other reference variables is 5.04%. With other variables constant, a rural household with average income only has a 1.52% probability of having investment products.

Case 2: *Different levels of annual income for urban households:*

The predicted probability of having investment products for urban households with 75% percentile annual income (RMB 50,200($8,198.60)), average household size 3.52 people, and all other reference variables is 10.32% (higher than 5.04%).

Case 3: *Comparison of having and not having available and convenient financial services for urban households:*

The predicted probability of having investment products for urban households with medium household income, average household size 3.52 people, reflecting no available and convenient financial services, and all other reference variables is 3.38% (lower than 5.04%).

Case 4: *Comparison of sufficient and insufficient knowledge for urban households:*

The predicted probability of having investment products for urban households with medium household income, average household size 3.52 people, reflecting insufficient knowledge of investment products, and all other reference variables is 3.51% (lower than 5.04%).

Overall, the probability of holding investment products is low. Household income, available and convenient financial services, and financial knowledge are positively related to the probability. As a way to generate wealth, holding investment products demonstrates a large gap between urban and rural households.

Table 7

' Logistic regression results of holdings of formal loans from banks-equation (4) (N=8,050)

		Coefficient	Std. Error	Odds Ratio	Overall Significance Test
Logarithm of annual household income		0.306***	0.029	1.003***	
Urban		-0.037	0.073	0.964	
Household size		0.110***	0.021	1.117***	
Products or assets	Having a credit card	0.573***	0.115	1.773***	
	Having non-financial assets	2.067***	0.309	7.890***	
Financial knowledge and service factor	Using formal sources of information	0.005	0.163	1.004	
	Interested in economics, politics, and social topics	0.282**	0.085	1.326**	
	Financial knowledge and patience of respondents				
	Impatient (reference)				chi2(2) = 15.69
	Patient	-0.257***	0.073	0.775***	Prob>chi2 = 0.0004
	Do not know and not willing to answer	-0.922*	0.468	0.398*	
	Insufficient knowledge of loans or inconvenience of application process	-0.695***	0.099	0.4995***	
Intercept		-7.346***	0.442	0.000651***	

Chi-square tests for joint significance

Pseudo R^2=0.0623

***p<0.0001

**p<0.01

*p<0.05

Note: Reference groups: Living in rural areas; Not using any formal sources of information; Having no credit card; Not interested in Economics, politics, and social topics; Impatient;

Table 8

Logistic regression results of holdings of informal loans-equation (4) (N=8,050)

		Coefficient	Std. Error	Odds Ratio	Overall Significance Test
Logarithm of annual household income		-0.082***	0.020	0.9992***	
Urban		-0.220***	0.056	0.803***	
Household size		0.260***	0.017	1.297***	
Products or assets	Having a credit card	0.008	0.115	1.008	
	Having non-financial assets	0.594***	0.137	1.811***	
Financial knowledge and service factor	Using formal sources of information	-0.201	0.113	0.818	
	Interested in economics, politics, and social topics	0.049	0.063	1.050	
	Financial knowledge and patience of respondents				
	Impatient (reference)				chi2(2) = 5.72
	Patient	-0.116	0.057	0.891	Prob > chi2 = 0.0573
	Do not know and not willing to answer	-0.361	0.263	0.697	
	Insufficient knowledge of loans or inconvenience of application process	1.498***	0.063	4.471***	
Intercept		-1.365***	0.246	0.255***	

Chi-square tests for joint significance

Pseudo R^2=0.1111

***p<0.0001

**p<0.01

*p<0.05

Note: Reference groups: Living in rural areas; Not using any formal sources of information; Having no credit card; Not interested in Economics, politics, and social topics; impatient;

Table 7 presents the results of logistic model of holding formal loans–equation (4). All independent variables are statistically significant except living in urban areas and formal sources of information. The overall test for financial knowledge and patience of respondents is also statistically significant at the 99.9% level of significance. It is surprising that living in urban areas is not statistically significant, while it suggests no obvious difference in urban and rural households in having formal loans from banks. As expected, the household income is influential in determining the probability of having formal loans. Holding other variables constant, a 1% increase in annual household income increases the odds of having formal loans from banks by 3%. Increasing household sizes increases the potential of having formal loans from banks: one unit increase in household size is associated with an 11.7% increase in the odds of having formal loans, controlling for other variables. Meanwhile, having non-financial assets increases the odds ratio by 689%, which is extremely high. However, it can be argued that households apply for loan mainly in order to purchase non-financial assets such as real equity, land, and vehicles.

Table 8 presents the results of the logistic model of having informal loans–equation (4). From Table 8, annual household income, living in urban areas, household size, having non-financial assets, and insufficient knowledge of loans or inconvenience of application processes are statistically significant. Especially, living in urban areas reduces the odds ratio by 19.74%, controlling for other variables. Increasing household income by 1% will reduce the odds ratio of having informal loans by 0.082%, holding other variables unchanged. These results confirm the rural-urban gap in holding informal loans and demonstrate that higher household income reduces rate of informal loans. The results are compatible with Nguyen (2007) who argues household size determines financial activities in rural Vietnam. However all three cases of financial activities shows financial services influence financial activities, indicating a difference between China and Vietnam.

Here are comparisons of the probability of different hypothetical cases for holdings of informal loans:

__Case 1:__ *Rural and urban households:*

The predicted probability of having informal loans for rural households with medium household income RMB 30,000($4,899.56), average household size 3.49 people and all other reference variables is 21.37%. With other variables constant, an urban household with medium income only has a probability of having informal loans 17.90%. (less than 21.37%)

__Case 2:__ *Different levels of annual income for urban households:*

The predicted probability of having informal loans for urban households with 75% percentile annual income (RMB 55884.18 ($9,126.93)), average household size 3.49 people, and all other reference variables is 17.17%. The difference in household income is not as obvious as expected, but increasing household income is negatively related to informal loans.

__Case 3:__ *Comparison of having insufficient knowledge of loans and inconvenience of application process for urban households:*

The predicted probability of having informal loans for urban households with medium household income, average household size, reflecting insufficient knowledge of loans and inconvenience of application process for urban households, and all other reference variables is 49.38%. (more than 17.90%) Compared with households with knowledge of loans and convenience of application process, these households usually turn to informal loans as alternatives.

In general, households have a high rate of participating in informal loan markets. Rural households more likely have informal loans than urban households. Improving financial knowledge and convenience of formal loan's service can largely reduce the probability and alleviate the situation.

Conclusions and Policy Recommendations

Our models and decomposition are able to explain portions of income and wealth inequality between rural and urban China. The D-index of individual laborers shows that the largest gap in rural and urban areas is education distribution. Higher education highly increases individual labor income while participating in agricultural works reduces labor income when compared with other occupations. The study confirms on the basis of individual data about labor income that education and occupation contribute more to the individual labor income gap between rural and urban areas, revealing that human capital should be a crucial target in policy making.

The household data reveals a rural-urban difference in holding investment products and informal loans but the difference in formal loans participation rate is not significant for rural and urban households. The overall participation rate in the formal financial market is low but informal loans markets are active in China. Household income is positively and significantly related to the participation rate, confirming a previous study that income is an indicator of financial market participation. The availability and convenience of financial services and sufficiency of knowledge also improve the participation rate.

Turning to the policy implications of these findings, we concentrate on those areas where human capital would be a central consideration, such as education and occupation. Improving education in rural areas to raise the rate of higher education would minimize the education gap and then target the individual labor gap. Because income inequality between agricultural workers and other laborers is still remarkable, subsidizing and improving farmers' income can be highly effective in dealing with income inequality.

The high difference between rural and urban financial market participation rates implies that developing rural financial services, simplifying registration and application processes, and promoting financial knowledge about investment products and loans could increase financial market participation, help households accumulate wealth, and alleviate rural-urban wealth inequality.

Our result directs more attention to the importance of human capital to explain the rural-urban variation. However, for individual income study, further research is necessary to identify other potential reasons in determining the individual labor income gap or to use more detailed categories than those used here to interpret the remaining unexplained portion of the income gap. Since the decomposition method is based on OLS regression, another future improvement that could be done is to design a decomposition method for logistic regression in order to explain the wealth inequality.

Appendix

1. The precise wording of the survey question can be found on the website of China Finance Household Survey: http://www.chfsdata.org/intro814.html

References

Ayyagari, M., Demirguc-Kunt, A., & Maksimovic, V. (2010). Formal versus Informal Finance: Evidence from China. *Review of Financial Studies*, 304883097.

Chang, C., & England, P. (2011). Gender Inequality In Earnings In Industrialized East Asia. *Social Science Research*, 1814.

Cole, S., & Shastry, G. (2008). INDUSTRY GEOGRAPHY Browse By Faculty About Us Newsletter Sign-Up 14 JAN 2009 WORKING PAPERS Smart Money: The Effect of Education, Cognitive Ability, and Financial Literacy on Financial Market Participation. *Harvard Business School, 09-071*. Retrieved November 3, 2014, from http://hbswk.hbs.edu/item/6093.html

Dercon, S., & Christiaensen, L. (2011). Consumption risk, technology adoption and poverty traps: Evidence from Ethiopia. *Journal of Development Economics*, 1598173.

Dumas, B., & Kurshev, A. (2007). *Equilibrium portfolio strategies in the presence of sentiment risk and excess volatility.* Cambridge, Mass.: National Bureau of Economic Research.

Duncan, O., & Duncan, B. (1955). A Methodological Analysis of Segregation Indexes. *American Sociological Review*, 2108210.

Grinblatt, M., Keloharju, M., & Linnainmaa, J. (2011). IQ and Stock Market Participation. *The Journal of Finance*, 212182164.

Haliassos, M., & Bertaut, C. (1995). Why do so Few Hold Stocks? *The Economic Journal*, 111081110.

Huang, Z. (2013). A study of household finance in China. *Theses & Dissertations of Lingnan University*, (25).

Jensen, R. (2000). Agricultural Volatility and Investments in Children. *American Economic Review*, 3998404.

Jones, F., & Kelley, J. (1984). Decomposing Differences between Groups: A Cautionary Note on Measuring Discrimination. *Sociological Methods & Research*, 323-343.

Kim, Y., & Shirahase, S. (2014). Understanding intra-regional variation in gender inequality in East Asia: Decomposition of cross-national differences in the gender earnings gap. *International Sociology*, 29(3), 229-248.

Lee, J. (2013). Income Inequality in Urban China and the Role of State Sector. *Hitotsubashi Journal of Economics*, 54(2), 159-76.

Li, J. (2009). *Informal finance in China: American and Chinese perspectives*. Oxford: Oxford University Press.

Neftci, S. (2007). *China's financial markets: An insider's guide to how the markets work* (p. 2). Amsterdam: Elsevier.

Nguyen, C. (2007). Determinants of Credit Participation and Its Impact on Household Consumption: Evidence From Rural Vietnam. *Center for Economic Reform and Transformation, Discussion paper*(703). Retrieved October 30, 2014, from http://econpapers.repec.org/paper/hwecertdp/0703.htm

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 693-693.

Rooij, M., Lusardi, A., & Alessie, R. (2011). Financial literacy and stock market participation. *Journal of Financial Economics*, 449-472.

Rosenzweig, M., & Wolpin, K. (1993). Credit Market Constraints, Consumption Smoothing, and the Accumulation of Durable Production Assets in Low-Income Countries: Investments in Bullocks in India. *Journal of Political Economy*, 223-223.

Sicular, T., Ximing, Y., Gustafsson, B., & Shi, L. (2005). The Urban-Rural Income Gap And Inequality In China. Review of Income and Wealth, 93-126.

Social Networks and the Decision to Insure. (2013). *MPRA Paper 46861, University Library of Munich, Germany*. Retrieved November 29, 2014, from http://mpra.ub.uni-muenchen.de/46861/1/MPRA_paper_46861.pdf

Taylor, C., Gorard, S., & Fitz, J. (2000). A re-examination of segregation indices in terms of compositional invariance. *Social Research Update*.

Therese, H., Li, L., & Zhu, W. (2005). The Effect of China's One-Child Family Policy after 25 Years. *The New England Journal of Medicine*.

Tiwari, A., Shahbaz, M., & Islam, F. (2013). Does financial development increase rural-urban income inequality?: Cointegration analysis in the case of Indian economy. *International Journal of Social Economics*, 151-168.

Wang, H., & Moll, H. (2010). Education Financing of Rural Households in China. *Journal of Family and Economic Issues*, 353-360.

Wang, X., & Piesse, J. (2010). Inequality and the Urban-rural Divide in China: Effects of Regressive Taxation. *China & World Economy*, 36-55.

Xia, Q., Song, L., Li, S., & Appleton, S. (2013). The Effects of the State Sector on Wage Inequality in Urban China: 1988–2007. *IZA Discussion Paper, No. 7142.*

Xie, Y., & Zhou, X. (2014). Income Inequality in Today's China. *Proceedings of the National Academy of Sciences, 111*(19), 6928-2933.

Yusuf, S. (2007). *China urbanizes consequences, strategies, and policies* (p. 50). Washington, D.C.: World Bank.

Zhu, C., & Wan, G. (2012). Rising Inequality in China and the Move to a Balanced Economy. *China & World Economy*, 83-104.

Submissions Process and Guidelines

ECONPress is a peer reviewed undergraduate academic journal where anonymized papers may undergo several rounds of revisions before the paper is accepted publication. All potential authors should email their submissions to *NUEconPress@gmail.com.*

Peer Review Process

When an Author submits a paper to ECONPress, it is reviewed by the entire selection team. The managing editor then leads a discussion on the paper's merits, during which the team makes recommendations to the submissions director who then communicates a decision to the author:

- *Accept Submission*
- *Revisions Required*
- *Resubmit for Review*
- *Decline Submission*

Accept Submission- The paper will be reviewed by a Copy Editor and after minor errors are corrected, the paper will undergo a final proofread before it is put in a queue for publication.

Revisions Required- The paper is suitable for the journal but some revisions are necessary to become ready for publication. The author will receive feedback related to the paper and will be required to submit a revised version. The paper will again be reviewed by at least two reviewers. This process may continue for several rounds and if it is ultimately deemed fit for publication, the paper will be marked Accept Submission.

Resubmit for Review- The paper is suitable for the journal but requires significant revisions. The author is suggested to make extensive changes to the paper and complete a new submission.

Decline Submission- This paper is not suitable for this journal. This may be for many reasons, including the paper not relating to econoimcs, the paper is discovered to have been plagiarized, etc.

Submission Information

Authors should submit their papers via an email to *NUEconPress@gmail.com.*

We welcome all papers related to the study of Economics from students enrolled as an undergraduate at an accredited university at the time of writing. This includes (but is not limited to) interviews with economists, research papers, and opinion papers.

Submissions are accepted on an ongoing processes, with the selection process occuring twice per year: once in the Spring and once in the Fall.