FIN422 CAPSTONE

An Analysis on Gender Disparity in Industry Trades

University of Michigan, Stephen M. Ross School of Business

FIN422

Capstone Research Paper

Ivory Yang

Author Note

The author is a Bachelor of Business Administration candidate at the Stephen M. Ross School of Business of the University of Michigan. This paper was written under the faculty guidance of Tyler Shumway, John C. and Sally S. Morley Professor of Finance and Faculty Director, Tozzi at the University of Michigan.

Author Correspondence

Correspondence regarding this essay should be directed to the author at: Author Name, University of Michigan, Stephen M. Ross School of Business, Ann Arbor, MI 48109

Author Contact Information

Ivory Yang - ivoryang@umich.edu

Introduction

Gender disparity in trading has long been a point of contention in academic research. A six year study by Barber and Odean showed that men on average trade 45% more than women, and that trading reduces men's net returns by 2.65 percentage points a year as opposed to 1.72 percentage points for women (Barber et al., 2001). Similarly, a study conducted at Colorado State University showed that women invest their pensions more conservatively than men, and are more risk averse with their finances in general (Bajtelsmit et al., 1996). In recent years, a new trend has emerged; more and more women are investing in specific value-driven industries. Women have been part of the force responsible for the doubling of ESG (Environmental, Social & Governance) assets to $8.1 trillion worldwide since 2014 and are gradually increasing participation in active investment (InvestmentNews, 2017). Given these points of data, it is logical to consider - do men and women invest in different industries? This paper will seek to examine the effects of gender disparity on industry trades by analyzing a dataset of trades from 1996-1997 to find any significant patterns or correlations.

Hypothesis

The hypothesis of this research paper is that women tend to make more investments in light industries such as Services and Public Administration, while men invest more in heavy industries like Manufacturing and Construction. Light industries are less capital-intensive than heavy industries, and are more consumer-oriented than business-oriented. Given that women tend to traditionally work in the service industry while men take on more labor-intensive careers like construction, the theory is that greater exposure to and knowledge of light industries cause women to invest in businesses operating under such categorizations (Ngai et al., 2017).

Tests

Sorting Trades by Industry

The dataset was first re-sorted with respect to SIC (Standard Industrial Classification) codes in Column Y of the data file, which are four-digit numerical representations of major businesses and industries. A new variable was then created in Column Z titled “Industry”, where the first two digits of each SIC code were matched to the major industry sector to which the business belonged to. The trades were allocated under the following industries - Mining, Construction, Manufacturing, Transportation, Trade, Finance Insurance & Real Estate, Services and Public Administration - based on Table I shown below. This was derived from a SIC Code guideline provided by the McKimmon Center for Extension & Continuing Learning (North Carolina State University, 2020).

Table I: SIC Codes by Industry

SIC Codes (First two digits)	Industry
10, 12, 13, 14	Mining
15, 16, 17	Construction
20-39	Manufacturing
40-49	Transportation
50-59	Trade
60-67	Finance Insurance & Real Estate
70-89	Services
91-98	Public Administration

Exclusion of Selected Trades

After the above steps, trades from Rows 2 - 471 and Rows 207482 - 207498 were excluded from further analysis, as they either did not have four-digit industry codes or had the two-digit code of "99", which represents non-classifiable establishments. Trades from rows 472 - 207481 were retained to run dummy variable regressions.

Dummy Variable Regressions

Based on the variable "Gender" in Column M, a new dummy variable was created in Column AA, "Gender Dummy". Females, presented as "F" were coded with the value of 1, while males presented as "M" were coded with the value of 0. This was replicated another eight times in Rows AB to AI for the different industries, under the names "Mining Dummy", "Construction Dummy" and so on. For example, under Column AB "Mining Dummy", trades with the Mining industry classification were coded with the value of 1, while trades not under the Mining classification were coded with the value of 0.

Using the Data Analysis tool in Excel, a dummy regression test was run using "Gender Dummy" as the y-input range (AA472 - AA207481) and "Mining Dummy" as the x-input range (AB472 - AB207481). This was repeated another seven times with the y-input range staying constant, and the x-input range changing to accommodate the remaining industry-based dummy variables. The regression output results were then recorded and analyzed.

Results

Mining

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.00139931
R Square	1.9581E-06
Adjusted R S	-2.873E-06
Standard Err	0.27794045
Observations	207009

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.03131269	0.03131269	0.40533755	0.52434622
Residual	207007	15991.4761	0.0772509		
Total	207008	15991.5074			

	Coefficients	*tandard Erro*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.08428326	0.00062532	134.78422	0	0.08305764	0.08550887	0.08305764	0.08550887
1	0.00186343	0.00292687	0.6366611	0.52434632	-0.0038732	0.00760003	-0.0038732	0.00760003

A simple linear regression was calculated to predict trades in Mining based on gender (being female). The regression model produced $R^2 = 0.0000020$, $F(1, 207007) = 0.41$, $p > 0.05$. The results are not statistically significant, meaning that there is **no correlation** between being female and making trades in Mining.

Construction

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.00778107
R Square	6.0545E-05
Adjusted R S	5.5715E-05
Standard Err	0.27793231
Observations	207009

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.96820599	0.96820599	12.534	0.0003997
Residual	207007	15990.5392	0.07724637		
Total	207008	15991.5074			

	Coefficients	*tandard Erro*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.08420193	0.00061267	137.434641	0	0.08300112	0.08540275	0.08300112	0.08540275
0	0.02827754	0.00798724	3.54033884	0.0003997	0.01262274	0.04393234	0.01262274	0.04393234

A simple linear regression was calculated to predict trades in Construction based on gender. The regression model produced $R^2 = 0.000061$, $F(1, 207007) = 12.53$, $p < 0.01$. The results are statistically significant, meaning that there **is correlation** between being female and making trades in Construction.

Manufacturing

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.01032197
R Square	0.00010654
Adjusted R S	0.00010171
Standard Err	0.27792592
Observations	207009

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	1.70378397	1.70378397	22.0575072	2.6477E-06
Residual	207007	15989.8036	0.07724282		
Total	207008	15991.5074			

	Coefficients	*tandard Erro*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.08737421	0.00088474	98.7567068	0	0.08564014	0.08910829	0.08564014	0.08910829
0	-0.005744	0.00122303	-4.6965421	2.6477E-06	-0.0081411	-0.0033469	-0.0081411	-0.0033469

A simple linear regression was calculated to predict trades in Manufacturing based on gender. The regression model produced $R^2 = 0.00011$, $F(1, 207007) = 22.06$, $p < 0.01$. The results are statistically significant, meaning that there **is correlation** between being female and making trades in Manufacturing.

Transportation

SU Enter OUTPUT

Regression Statistics	
Multiple R	0.00235562
R Square	5.5489E-06
Adjusted R S	7.1821E-07
Standard Err	0.27793995
Observations	207009

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.08873594	0.08873594	1.14867612	0.2838277
Residual	207007	15991.4187	0.07725062		
Total	207008	15991.5074			

	Coefficients	*tandard Erro*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.08418815	0.00063359	132.875657	0	0.08294634	0.08542997	0.08294634	0.08542997
0	0.0025595	0.00238812	1.07176321	0.28382765	-0.0021212	0.00724014	-0.0021212	0.00724014

A simple linear regression was calculated to predict trades in Transportation based on gender. The regression model produced $R^2 = 0.0000055$, $F(1, 207007) = 1.15$, $p > 0.05$. The results are not statistically significant, meaning that there is **no correlation** between being female and making trades in Transportation.

Trade

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.0017289
R Square	2.9891E-06
Adjusted R S	-1.842E-06
Standard Err	0.27794031
Observations	207009

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.04780041	0.04780041	0.61876899	0.43150601
Residual	207007	15991.4596	0.07725082		
Total	207008	15991.5074			

	Coefficients	*tandard Erro*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.08453116	0.00064501	131.054332	0	0.08326696	0.08579536
0	-0.0015808	0.00200958	-0.786619	0.43150582	-0.0055195	0.00235795

A simple linear regression was calculated to predict trades in Trade based on gender. The regression model produced $R^2 = 0.0000030$, $F(1, 207007) = 0.62$, $p > 0.05$. The results are not statistically significant, meaning that there is **no correlation** between being female and making trades in Trade.

Finance, Insurance, Real Estate

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.0017289
R Square	2.9891E-06
Adjusted R S	-1.842E-06
Standard Err	0.27794031
Observations	207009

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.04780041	0.04780041	0.61876899	0.43150601
Residual	207007	15991.4596	0.07725082		
Total	207008	15991.5074			

	Coefficients	*tandard Erro*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.08453116	0.00064501	131.054332	0	0.08326696	0.08579536	0.08326696	0.08579536
0	-0.0015808	0.00200958	-0.786619	0.43150582	-0.0055195	0.00235795	-0.0055195	0.00235795

A simple linear regression was calculated to predict trades in Finance, Insurance and Real Estate based on gender. The regression model produced R^2 = 0.0000030 , $F(1, 207007) = 0.62$, $p > 0.05$. The results are not statistically significant, meaning that there is **no correlation** between being female and making trades in Finance, Insurance and Real Estate.

Services

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.00322375
R Square	1.0393E-05
Adjusted R S	5.5619E-06
Standard Err	0.27793928
Observations	207009

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.16619267	0.16619267	2.15135466	0.14244568
Residual	207007	15991.3412	0.07725024		
Total	207008	15991.5074			

	Coefficients	*tandard Erro*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.08401332	0.00065708	127.859364	0	0.08272547	0.08530118	0.08272547	0.08530118
0	0.00261655	0.00178391	1.46674942	0.14244575	-0.0008799	0.00611297	-0.0008799	0.00611297

A simple linear regression was calculated to predict trades in Services based on gender. The regression model produced R^2 = 0.000010, $F(1, 207007) = 2.15$, $p > 0.05$. The results are not statistically significant, meaning that there is **no correlation** between being female and making trades in Services.

Public Administration

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.00396128
R Square	1.5692E-05
Adjusted R S	1.0861E-05
Standard Err	0.27793854
Observations	207009

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.25093399	0.25093399	3.24834334	0.07149714
Residual	207007	15991.2565	0.07724983		
Total	207008	15991.5074			

	Coefficients	*tandard Erro*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.08440406	0.0006112	138.09577	0	0.08320612	0.085602	0.08320612	0.085602
0	-0.0339453	0.01883429	-1.8023158	0.07149719	-0.0708601	0.00296941	-0.0708601	0.00296941

A simple linear regression was calculated to predict trades in Public Administration based on gender. The regression model produced $R^2 = 0.000016$, $F(1, 207007) = 3.25$, $p > 0.05$. The results are not statistically significant, meaning that there is **no correlation** between being female and making trades in Public Administration.

Conclusion

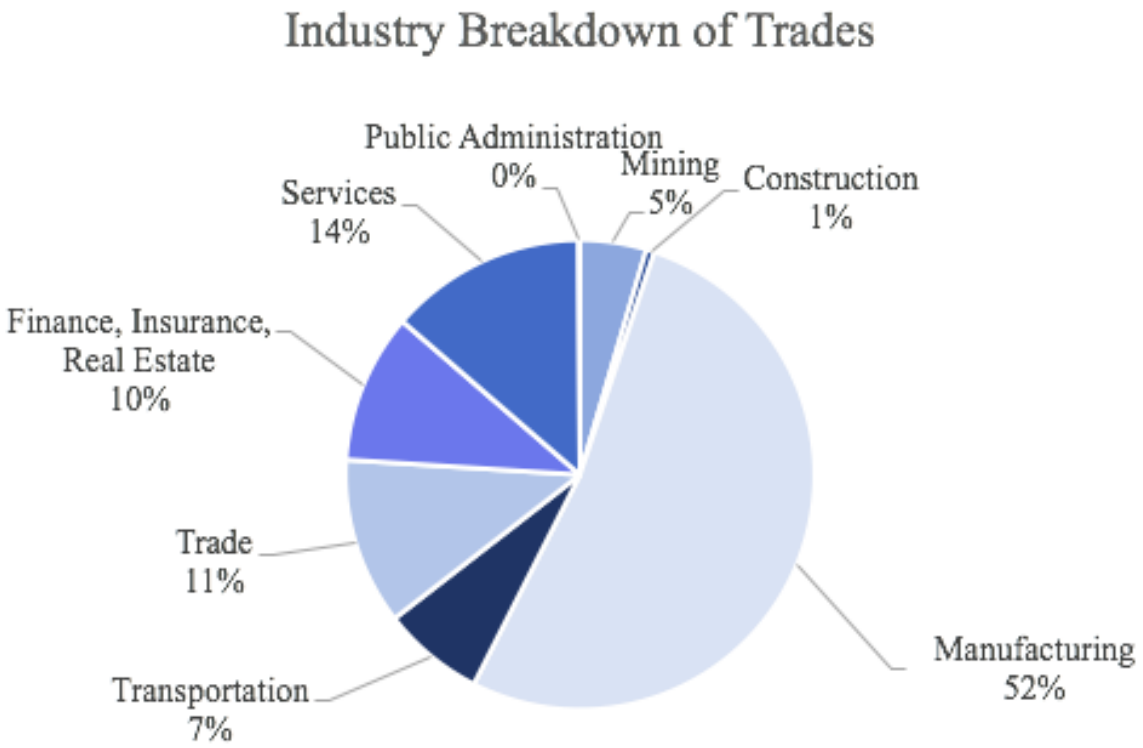
Overall, the dummy variable regression tests showed that there is correlation between being female and trading in Construction and Manufacturing, but no correlation between being female and trading in the remaining six industries. This proves the hypothesis to be **false**, showing that women tend to trade more in Construction and Manufacturing which are considered heavy industries. However, although results were statistically significant ($p < 0.01$) for the Construction and Manufacturing regressions, the R^2 coefficient was very low. This means that the independent variable, while significant, does not explain for any *variation* in the dependent variable. Being female does not account for the variability around the mean for the number of

trades in a particular industry. As such, it might be more effective to run multilinear regression tests in the future that incorporate additional independent variables.

Limitations

The trades in the dataset from Rows 472 - 207481 were broken down by industry. As shown from Table II below, over half of the trades were made on businesses covered within the Manufacturing Industry (52%), while trades in industries like Construction and Public Administration only amounted to about <1% each. Given the disparity in volume of trades across industry, the data analysis conducted may not be the most accurate.

Table II: Industry Breakdown of Trades in Dataset



In addition, keeping in mind how large the dataset was, the trades could only be loosely grouped under eight umbrella industries by their two-digit codes, and not further broken down by sub-industries and operations. It would have been interesting to see if there were very specific

niche industries in which women traded more in. Furthermore, some of the trades could not be used to run the dummy variables regressions as they either did not have four-digit SIC codes or were non-classifiable establishments. The dataset used in this research paper is also very old; the amount of women making investments has increased significantly since the data was collected and general investment behaviors may also have changed over the years based on accessibility of information, volatility and market conditions.

Future Ideas for Research

Given the growing trend of women investors actively making investments in ESG companies, it would be interesting to analyze more recent datasets to test if there exists any correlation between women and investment in sustainability-driven companies. Likewise, focusing on gender breakdown of investments in stereotypically feminine companies under the Retail sector would be an intriguing approach to consider. Beyond the academia component, research in such a direction would be an excellent way to better understand the gender disparity in trading, and how to narrow the gap.

References

Barber, B. M., & Odean, T. (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics*, *116*(1), 261-292.

Bajtelsmit, V. L., & Bernasek, A. (1996). Why do women invest differently than men?. *Financial Counseling and Planning*, *7*.

InvestmentNews (2017). Millennials, women drive assets to ESG strategies. https://www.investmentnews.com/millennials-women-drive-assets-to-esg-strategies-72635

McKimmon Center for Extension & Continuing Education (2020). 2-Digit SIC Codes. *University of North Carolina*. https://mckimmoncenter.ncsu.edu/2digitsiccodes/

Ngai, R. L., & Petrongolo, B (2017). Gender gaps and the rise of the service economy. *American Economic Journal: Macroeconomics, 9*(4), 1-44.