Information Technology and Entrepreneurship  
John Vandivier  
May 2016

**ABSTRACT**

An analytic model is constructed to deal with the relationship between industrial costs of market entry and the rate of individual-level entrepreneurship. The model is empirically validated against the information technology sector using cross-sectional 2014 PUMS ACS data from the US Census Bureau. Using the IT industry as a case of a low-cost industry, the main hypothesis is that low-cost industries will obtain higher rates of entrepreneurship. A secondary result is that individual wealth will be positively related to the frequency of entrepreneurship. Empirical analysis suggests that individual financial factors have a dominant but complex effect, while the main hypothesis is confounded, laying the groundwork for future research.

1. **Introduction**

This paper abstracts from limited data to address the general relationship between the rate of entrepreneurship and the kind of industry in which such activity is undertaken. It is hypothesized that individuals which produce labor in the information technology industry will obtain a higher rate of entrepreneurship compared to individuals in other industries on average.

The motivation for this paper is threefold. First, the paper adds to the previous literature on the relationship between industry and entrepreneurship by suggesting additional factors to consider and extending the analysis to the individual level. Second, this paper seeks to validate or invalidate a specific model of industry along with certain related assumptions. Third, the health of the information technology industry within a country can be seen as an indicator of innovation and future growth[[1]](#footnote-1). Researchers interested in measuring the health of a country based on an innovation-oriented approach can extend the framework described in this paper.

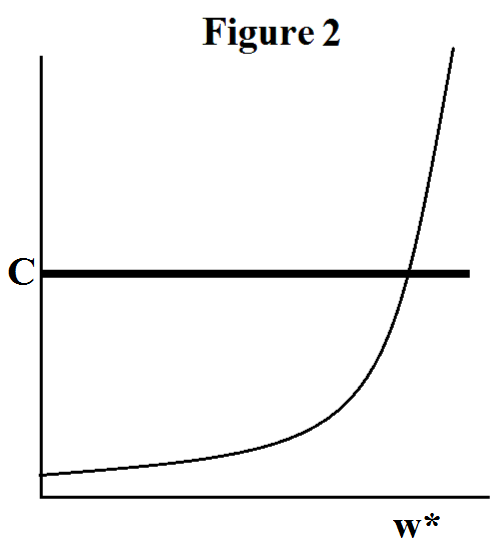
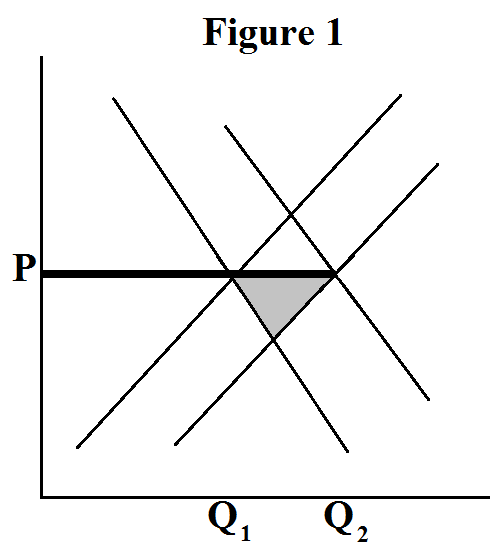
Ceteris paribus, a market with lower entry costs should obtain a larger number of firms under a basic model of market supply and demand. It is taken as a stylized fact that the information technology industry is peculiar due to the low costs of entry which it obtains. A particular model of industry finds that low costs of entry will not only alter the number of firms in a market, but also the organizational distribution of firms in that market. The model suggests that small businesses and entrepreneurs will make up an increasing share of new firms as entry costs decrease. Conditional on that model and the described stylized fact, the hypothesis of interest is directly derived.

The industrial literature on the information technology sector of the economy suggests that submarkets of this industry are observed with medium to high clockspeeds[[2]](#footnote-2) compared to other industries. High clockspeeds and high turnover involve high rates of firm creation, and new firms are generally small. Covin and Douglass observe that entrepreneurship has an importantly different relationship with firm performance in the context of a high technology industry when compared to a low technology industry[[3]](#footnote-3). The present article suggests that the prior observation is in part explained by industry clockspeed.

The model of interest allows firms to vary based only on a single factor, but other papers have established that firms are heterogeneous in other ways which do have important determinant effects in various directions on entrepreneurship. For example, entrepreneurship in simple firms, where simplicity was technically defined, was more potently influenced by the traits of firm leadership in one paper[[4]](#footnote-4).

1. **Theoretical Model**

Suppose a linear competitive aggregate model of supply in a market with a variable rightward shift to demand. This causes some variable amount of market rent and thereby incentivizes a number of firms to enter until the price is restored to the previous price and the rent is dissipated. This case is illustrated in the figure below with new producer surplus in grey.

****

Further suppose that firms can be corporations or entrepreneurs. Corporations generate linearly with the market rent minus the cost of entry. Entrepreneurs are liquidity-constrained individuals who must pay the cost of entry from a stock of wealth. They generate linearly with their individual wealth minus the cost of entry. Finally, suppose individual wealth is nonlinearly skewed right. This situation is illustrated in Figure 2. We can see that as the cost of entry decreases, there is a linear increase in the number of corporations and a nonlinear increase in the number of entrepreneurs. The result is a larger share of entrepreneurs in low-cost industries.

While a dynamic model creates complexity not needed for this paper, a similar theoretical model can be obtained when profits are related to firm generation speed, and corporations are assigned a formation speed which is slower than entrepreneurial firm generation. Another extension would be to examine the case where cost of entry is variable and endogenous to firm size or kind.

A final assumption must be declared before we proceed to the empirical model. The empirical model will analyze cross-sectional data, so there is no mechanism by which we can argue that a shift to demand has occurred. Instead, we must assume that demand is already shifting by some degree in this market and that the market is not already optimized. Fortunately, not only is this assumption plausible in the empirical model, it would be rather strange if that were not the case.

1. **Empirical Model**

The empirical model presented makes use of the 2014 PUMS ACS data set from the US Census Bureau. This is an extensive data set with millions of observations and a number of interesting properties per observation. The data is observed at the individual level. The reason for choosing the 2014 data is simply that it is the most current cross-sectional data available, and this paper does not use a panel design. We will discuss a number of useful extensions to this paper in the conclusions section, and a time-series approach is certainly one of them.

The testable null hypothesis is that IT professionals in the US did not engage in a rate of entrepreneurship which is significantly different compared to other individuals in the year 2014. The dependent variable in the model is whether or not the individual is self-employed, which we will interpret at the model level as the probability that some individual is self-employed given a set of right hand conditions. The variable of interest is whether or not an individual works in the information industry.

The theoretical model is also rich enough to provide us with some additional considerations which the data set is able to support, beyond the question of interest but in a way which reinforces the theoretical framework. One such consideration is that wealth should have a positive effect on the rate of entrepreneurship. Formally, there is a supplemental null hypothesis that wealth has an insignificant coefficient on self-employment status.

The model was estimated using ordinary least squares. Given the dependent variable and right hand variables mentioned above, the short model takes the form below:

X1 is a dummy variable indicating whether or not the individual was employed in the IT industry according to an index constructed from Census Bureau-defined ACS occupation codes[[5]](#footnote-5). The short model was extended to include the various correction variables in the data set, resulting in a long model of the same form with additional factors. These factors were reduced in the model one at a time, until statistically irrelevant and unimportant variables were weeded out.

Variables used in the long model include age, age squared, gender, race, the state of residence, wage, wealth, income, log income, income squared, other income and wealth variables, marital status, years of education, the field of the undergraduate degree if it exists, the field of graduate degree if it exists, a flag for whether the undergraduate degree meets the NSF definition of a science or engineering-related field, citizenship status, hours worked per week, disability status, and body weight.

Given the wide-ranging correction variables, the large sample size, and the stability of the results which will be discussed later on, it’s hard to argue that the results are biased when interpreted in-context. Even still, when we place the results in context we can point to a variety of possible biases and issues. The data is likely to contain a US-level country bias. This bias would tend to inflate income and wealth compared to the rest of the world, and it would also tend to create a variety of other systematic biases related to US institutions including policy, culture, and more. The data will contain any temporal confounders which existed in 2014 as the design limited itself to cross-sectional data in that year. There may be some unobserved bias, but given the wide range of correction data and the invariance of the results it is hard to tell what such bias would include.

In theory there could be a selection bias into the ACS survey. While a response to the survey is required by law and failure to reply may result in thousands of dollars in fines, the US Census Bureau has never prosecuted a case and so it’s possible some individuals may either select not to respond, or they may be structurally inhibited from response. In theory it is possible that this structural inhibition could be causally related to income or wealth issues, and this could have implications for the model of interest.

Retrospectively, I would have also liked to include personality data. It seems that a component of the tendency toward entrepreneurship could be explained by personality traits, and those traits may be endogenous toward many other factors. Unfortunately, this data is not available for the same set, but it presents an opportunity for future work.

1. **Results**

The results of the study appear in part counter to the theory previously proposed as well as expectation. Such results may tend to inflate the value of the paper rather than defeat it. The model demonstrated high significance and explanatory power, but the direction and importance of comparative factor relationships with the dependent variable are in some cases presently absent explanation, as will be further discussed in the conclusions section.

The variables which survived into the final model are summarized in Table 1. Tables 2 and 3 present comparative regression results. The dependent variable is the self-employed flag. The IT flag is the variable of interest. The main explanatory power is derived from the income variables. The primary null hypothesis is rejected, but the direction of the observed effect is opposite the expected result. It turns out that working in an IT field makes an individual less likely, not more likely, to be an entrepreneur.

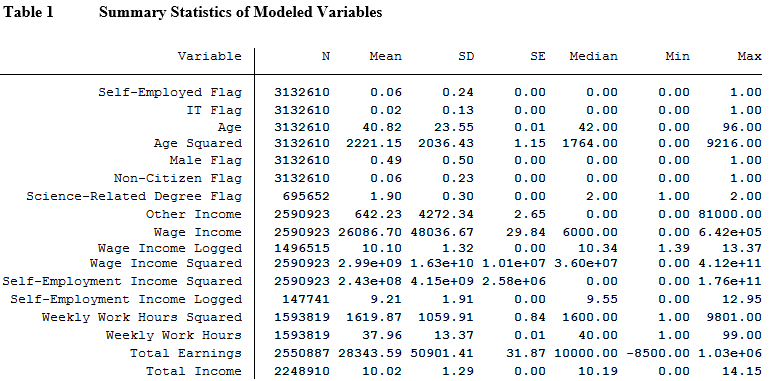
Age and education have an interesting relationship in the model. Age has a positive effect and a negative marginal effect on entrepreneurship, but once the dummy to check whether the field of the undergraduate degree is added to the right hand side, the marginal effect of age disappears. This is observed in the change from model 6 to model 8, for example. Models 6 and 8 are the most predictive models, and model 8 is the same as model 7 after an optional check for robustness is added. Adding the undergraduate degree seems to be an important component of the model, even though it is not statistically significant at the .05 level. It is significant at the .1 level and it also strongly corrects the variable of interest and improves the adjusted explanatory power of the model by a noticeable amount, as is also observed in the move from model 6 to model 8. A final point of interest in the move from model 6 to model 8 is the substantial loss of observations. However, N is still large in the final model.

Income, wages, earnings, and wealth are subtly different concepts with a complex relationship to individual entrepreneurial propensity. As per the theoretical model we should find that higher wealth increases the odds to become an entrepreneur, but separating out a wealth effect from an income effect which may move in the opposite direction is tricky due to the indirect nature of the data on wealth in the ACS data set. Plain wage income has a negative linear relationship with entrepreneurship which may be interpreted as an opportunity cost effect. Wages have a positive marginal effect on entrepreneurship which may be seen as a marginal weakening of the opportunity cost effect as wages rise.

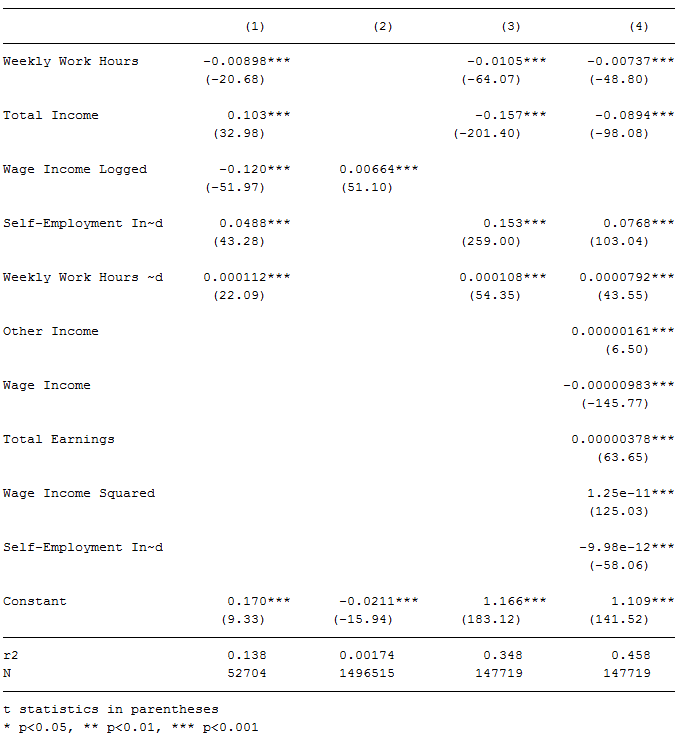
Total income moves in the same direction as wages, but total earnings moves negative with entrepreneurship. The difference between total earnings and total income seems to be personal investment income and financial returns, but if that is the case then we would expect large income with low earnings to represent a proxy of wealth, and we would expect total income to be increasing with entrepreneurship, while the reverse is observed. The presence of the variable for other income may be confounding these relationships as well, because other income may explain some degree of financial investment income.

Self-employment income predictably has a positive logarithmic relationship with entrepreneurship, but it does have a negative marginal relationship which doesn’t seem to have an obvious explanation. It may indicate that wealthier individuals face a smaller incentive to engage in entrepreneurship. Weekly hours worked are strongly negatively correlated with entrepreneurship, but they have a positive marginal relationship. This doesn’t seem to be open for simple causal interpretation.

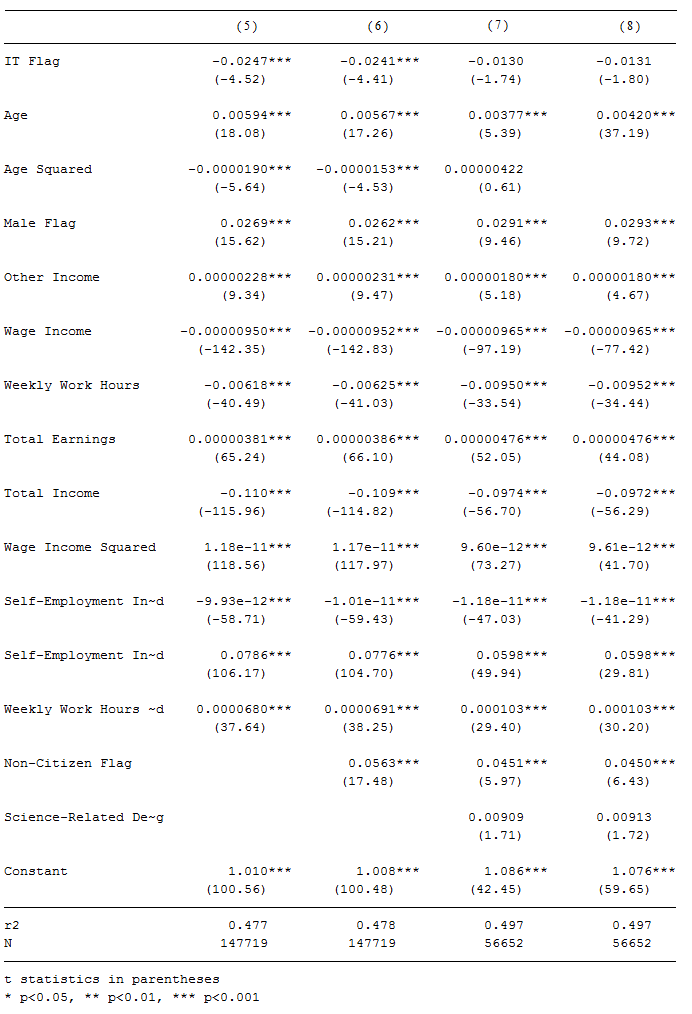
The log of wages is an interesting variable which was included in the model early on but it was found to have a confounding effect. Looking at models 1, 2, and 3, we see that the direction of the relationship between log income and entrepreneurship flipped in a long and short model. When the variable was omitted from model 3, the explanatory power of that model more than doubled compared to the original. It is interesting to note that wage, income, and earnings alone are regressed in model 4 which obtains an R2 of about .46, while the final model obtains an R2 of about .5. This indicates that the financial factors dominate the other factors in explaining the left hand side.

****

**Table 2 Regressions 1 - 4**



**Table 3 Regressions 5 - 8**



1. **Conclusions**

As previously mentioned, the data suggested a highly significant relationship in the direction opposite the expected result. As per the introduction, the hypothesis follows directly from the assumption of a particular stylized fact in the context of a particular model. I logical conclusion would be that there are one of three explanations for the mismatch between the observation and the expected result. First, the stylized fact may not be so. Second, the theoretical model may be decoupled from reality. Finally, the observed data may contain some sort of confounding information which caused the expected relationship to be lost in this particular cross-section.

On reflection of the assertion of the stylized fact of low costs in the information technology sector, it is plausible that this assertion is incorrect. Information technology is a wide field. The introduction recognized that the previous literature has given a variety of unique values for the clockspeed of certain submarkets within the information technology industry. Heterogeneous clockspeeds indicate that these submarkets may be importantly different from one another in ways which are related to the rate of entrepreneurship.

There are particular submarkets in the IT industry which seem to have low barriers to entry. The markets for web development or graphic design services can be undertaken by an individual for virtually no marginal cost, given that an individual with such skill has likely already invested in a personal computer and that a personal computer is essentially the only capital requirement for such a production function. On the other hand, most of information technology is not concentrated in these two submarkets. Future research could improve on the current paper by narrowing the scope of study.

On reflection of the theoretical model, there may be some confounding simplifications made or important factors left out of consideration. First, risk aversion may play a more serious role than allowed for in the model. Individuals employed in the IT sector often enjoy low-risk employment. Entrepreneurship is generally regarded as a relatively high-risk and high-reward endeavor. It could be that the opportunity cost of an individual in this position is very high when considering self-employment in risk-adjusted terms. Secondly, this theoretical difference may be exacerbated by personality traits if it is found, for example, that introverts tend toward programming, for example, while more extroverted personalities are associated with more entrepreneurship. Finally, it may simply be that the linearization of the theoretical model was a poor representation of the complex nonlinear relationships observed in the ACS data between financial variables and entrepreneurship, or that the choice of a cross-sectional design ignored important dynamic features of the industry which might interact with entrepreneurship. Future work could alter these models appropriately and wrangle time series and personality data.

**References**

Covin, Jeffrey G., and Dennis P. Slevin. "Corporate Entrepeneurship in High and Low Technology Industries." Journal of Euromarketing (1994): 99-127. Web.

Miller, Danny. “The Correlates of Entrepreneurship in Three Types of Firms.” Management Science (1983): 770-791. Web.

Souza, Gilvan C., Barry L. Bayus, and Harvey M. Wagner. "New-Product Strategy and Industry Clockspeed." Management Science 50.4 (2004): 537-49. Web.

Thurik, Roy, and Wennekers, Sander. “Linking Entrepreneurship and Economic Growth.” Small Business Economics (1999): 27-56. Web.

**A Appendix: OCC Codes Included in the Index**

The index dummy variable took a value of 1 if the individual was coded with any of the 14 OCC codes identified in the table below.

|  |  |
| --- | --- |
| **OCC Code** | **Data Dictionary Label** |
| 0110 | MGR-COMPUTER AND INFORMATION SYSTEMS MANAGERS |
| 1005 | CMM-COMPUTER AND INFORMATION RESEARCH SCIENTISTS |
| 1006 | MM-COMPUTER SYSTEMS ANALYSTS |
| 1007 | CMM-INFORMATION SECURITY ANALYSTS |
| 1010 | CMM-COMPUTER PROGRAMMERS |
| 1020 | CMM-SOFTWARE DEVELOPERS, APPLICATIONS AND SYSTEMS SOFTWARE |
| 1030 | CMM-WEB DEVELOPERS |
| 1050 | CMM-COMPUTER SUPPORT SPECIALISTS |
| 1060 | CMM-DATABASE ADMINISTRATORS |
| 1105 | CMM-NETWORK AND COMPUTER SYSTEMS ADMINISTRATORS |
| 1106 | CMM-COMPUTER NETWORK ARCHITECTS |
| 1107 | CMM-COMPUTER OCCUPATIONS, ALL OTHER |
| 1400 | ENG-COMPUTER HARDWARE ENGINEERS |
| 7010 | RPR-COMPUTER, AUTOMATED TELLER, AND OFFICE MACHINE REPAIRERS |

For more information on available codes, view the entire 2014 PUMS ACS data dictionary at: <https://www2.census.gov/programs-surveys/acs/tech_docs/pums/data_dict/PUMSDataDict14.pdf>

1. See Thurik and Wennekers (1999) [↑](#footnote-ref-1)
2. See Gilvan et al. (2004) [↑](#footnote-ref-2)
3. See Covin and Dennis (1994) [↑](#footnote-ref-3)
4. See Miller (1983) [↑](#footnote-ref-4)
5. See Appendix A for a table of the 14 OCC codes considered member to the IT industry, along with Census Bureau-defined labels for each code. [↑](#footnote-ref-5)