****

**Operations Expansion Research:**

***Identifying Future Markets Through Internet Growth Factors***

**James Stuart**

**December 17, 2016**

Executive Summary

**Introduction:**

Amazon.com inc., is looking to expand our operations to other countries requires a significant population with access to the internet. Identifying new markets early allows Amazon be first to market in new markets and establish a large foothold before its competitors expand. Therefore, the report observes factors that could possibly influence internet growth. A noted Megatrend of this century is the rise of urbanization, the migration towards cities, with countries in 2014 averaging not immodest 2% growth. Cities are generally have higher density than non-urban areas which means that in cities the infrastructure that supports the internet can cover a smaller area and cover more people. The report also takes into account other variables primarily GDP Growth to determine whether a growing economy leads to more investment in digital infrastructure, General Population Change was included to provide a control for Urban Growth and the Internet Coverage of the previous year was included to capture the phenomena, where the usefulness of connectivity in an area being grown by the number of other users, known as Network Effect.

**Results and Observations:**

There is significant evidence provided by linear regression models that Urban Growth has a relevant positive impact on the percentage change of internet users in a country, the most accurate model predicted that for on average for every percentage point there is .36 increase in the percent of internet growth which is relevant considering that the average percent change is 4.1%. It can also be concluded that GDP Growth has no impact on the change in connectivity what so ever, though it is also positively impacted by Urban Growth. However, the greatest predictor for internet connectivity that was looked at was the previous year's connectivity. Connectivity has a roughly linear relationship with internet growth, which is noteworthy given it should be an exponential decay relationship. Coverage likely picks up on infrastructure development in the countries densest areas happens first so growth slows down due to rural areas taking longer to develop. However, the impact of network economics increasing demand for the internet as the amount of users seems to mitigate the previously discussed effects. It is also interesting to note all explanatory variables tested are correlated to a certain degree

**Conclusion:**

Amazon should look to expand to countries that are demonstrating urban and general population growth. However, further research should be done to determine exactly which factors with cities encourage faster growth than simply adding more users into the market through population growth. The best indicator of a countries future growth is coverage which also is a good measure of the size of a current market, keeping in mind when the network effect and saturation of parts of the population it would a good strategy to target countries that have enough of a market to benefit from the network effect and yet not be saturated, though the impact of each for requires more research.

**Internet Growth Factors and Where to Find Them**

**Amazon Expansion Analysis**

**Introduction:**

As part of Amazon.com inc.’s initiative to expand operations, this report has been constructed to provide a level insight into which countries may be the most relevant targets. Amazon, as a web based service, requires a connected populace to operate. Intuitively the most connected areas provide the best target markets. However, to gain a competitive edge this report aims to look at what factors can point out potential new markets, countries that show high levels of internet growth. Establishing a foothold in a developing market early with allow Amazon establish customer familiarity with its web services before competitors providing a clear advantage in establishing their position in the market. As a result, this report will look at what factors influence internet growth in particular urban growth. As a noted Megatrend this century, the rise of urbanization shows countries in 2014 averaging not immodest 2% growth making this relevant force worth taking that Amazon could potentially advantage of. Cities have a higher density than less urban areas. Therefore, infrastructure that supports the internet can cover a smaller area and cover more people. However, countries are complex entities with numerous forces that impact change within their borders, in order to get a sense of the truer sense of the actual impact of urbanization other variables needed to be considered in order to create a more accurate model for predicting internet growth. The report also looks at GDP growth to determine whether a growing economy leads to more investment in digital infrastructure, general population change was included to provide a control for urban growth and to see if it was significant independent of being an indicator of population change, and the internet coverage of the previous year was included to capture the phenomena of network power and diminishing returns on growth percentages. This reports findings will be based on analysis of linear regression models which will be reinforced by graphs to provide a visual representation of the results.

**Data Sources:**

The regressions in the study contained 196 countries with observations derived from the following sources:

Connectivity data - <http://www.internetlivestats.com/>:

This source provides data pertaining to internet usage, growth, and general population by country. Internet Live Stats is part of the Real Time Statistics Project and provided data for BBC News and the United Nations. This source collection source finalized data for 2014. The data collection 2015 is still being finalized and 2016 is still in progress. This limited our study to a cross-sectional study, a study focused on one year, whereas a study conducted over time would be better suited to answer the posed research question.

Data on GDP and Urban Growth - <http://data.worldbank.org/>:

The World Bank is a global bank with 189 member countries. Their operations aim to provide financial services to developing countries generally at 0% interest, they also provide a platform for knowledge sharing. The World Bank database contains information on a wide variety of queries, including GDP Growth, Urban Growth, and even internet user growth. However, the data here while good for providing the information on GDP Growth and Urban Growth it did not provide the breadth of data on internet connectivity internetlivestats provided so both databases were combined.

Data on Global City: <http://www.lboro.ac.uk/gawc/world2012t.html>:

The Globalization and World Cities group, or the GAWC was created at Loughborough University. GaWC is a research group that specializes in its research in globalization and cities. Part of their research is includes ranking cities based off their international influence.

**Variables and Methods Used:**

|  |  |
| --- | --- |
| Name | Description |
| PUsersChange | Percentage of internet users change over the past total. For instance, if last year the country had 100 users and this year it had 102 it would have a 2% PUsersChange.  Source: <http://www.internetlivestats.com/> |
| PopPChange | Percentage of population change over the past year.  Souce: <http://www.internetlivestats.com/> |
| UrbanPChange | Percentage of the urban population change over the past year  Source: <http://data.worldbank.org/> |
| Coverage | Percentage of the population that had access to the Internet one year ago.  Source: <http://www.internetlivestats.com/> |
| India | A variable that states if the country is India, used as a way to drop or keep it given its extreme outlier status. |
| Alpha | Dummy variable, determines if a country contains an Alpha- or high alpha level city  Source: <http://www.lboro.ac.uk/gawc/world2012t.html> |
| Alpha\_UG | An interaction term computed by Alpha\*UrbanPChange which is meant to evaluate if countries with Alpha cities experienced faster or slower rate of internet growth respect to urban growth than their peers.  Source: Computed |

**What is an Alpha City?**

The GAWC gives cities a categorical ranking based on their global influence. Metrics used to analyze this would be the presence of financial institutions, headquarters of global companies, political decision-making power, education, and infrastructure. Alpha++, Alpha+, Alpha, Alpha- are all counted in the dataset. The Alpha++ cities are London and New York massive cities with global financial services and international business hubs. Cities such as Hong Kong and Paris make up Alpha + which are influential but not to the same degree, Sao Paulo, and Los Angles are considered Alpha cities, and Washington D.C is considered an Alpha– city demonstrating these rankings place finical power over political power. Examples of cities that did not make the cut of alpha are cities Dallas, Bangalore, Rome, Vancouver, and San Jose. Regionally important cities but they lack the same level of international influence. This metric was included as Alpha cities are more globalized and therefore may receive technological advances faster, thus they may be better connected than other cities, therefore, having an alpha city may increase a countries internet growth.

**India: The Extreme Outlier**

**Exhibit 1: Distribution of Internet Growth Percentages**

****

The above graph is the distribution of PUsersChange, the graph shows most of the countries are clustered around the average of 4.138%. However, India is a clear outlier shown in the far right column of the above graph. In 2014 India had a percent change of 30% in the number of internet users in the country, yet after 2014 the country only had 34% of its population. Despite this India is the country with the second largest internet using population by raw numbers. It is unlikely any country will develop in the same way as India already has a massive internet community which according to the network effect makes internet services within extraordinarily desirable for a country that still has a lot of room for development. From a statistical standpoint though India is 6.54 standard deviations from the mean for perspective being 3 standards above the mean would put an observation in the 99% percentile. This means that in many cases dropping India greatly improves the overall accuracy of the model as well as changing the significance of variables involved. In appendix 1 shows that regression model accuracy increases immensely when India is not included. Therefore, for the sake of creating models that are more accurate for general cases, models will be used both with and without India.

**Results:**

**Graph of Percent Change of Internet Users vs. Percent Urban Population Change:**

****

The above graph demonstrates the relationship of internet users change as a result of urban population percent change. From a visual standpoint, the relationship is both strong and linear as urban population change increases so does the percent of internet users change. Appendix 2 regression 1 further states that Urban Growth has a highly significant positive correlation with Internet Growth. However, there is lot of variation as to the exact effect that urban growth has and this is demonstrated as a lot of data falls outside of the gray area representing the 95% confidence interval on the graph, and the aforementioned regression has very little predictive power. Appendix 1, Regression 1 includes an additional variable that was left out of the graph, percent population change. Percent population change is included as a way of attempting to control for the possibility that Urban Growth may only impact internet connectivity because it is an indicator of general population change. Population change picks up on general population change and if urban population change impacts internet growth exclusively for the same reason the variable representing urban population change should become insignificant. However, Urban Growth remains significant with a t-score of 2.18. Furthermore, dropping India as the clear outlier does clearly increases the accuracy of the model (Appendix 1, regression 3). Dropping India also causes population change to become statistically significant. Removing India reduces the significance of percent urban growth slightly but not by a noteworthy amount. These changes are evidence that the outlier India contains a much larger percent urban change in relation to its general population growth and being an immense outlier has a large impression on the regression coefficients. By this evidence, though, both population growth and urban change are significant, and population has the larger impact of the two.

**Considering the impact of GDP or Lack Thereof:**

**Graph of Internet Users Change VS GDP Percent Change**

****

GDP logically seemed like it would perform well as countries with growing economies may also see some wealth invested in internet infrastructure. However, as can been seen from the graph of GDP Percent change and Internet Users Percent the trend is slightly positive but contains a lot of variance. Appendix 2 regression 2, provides further evidence of a significant but highly variable positive trend. Yet when GDP growth is added to the model with urban growth and population growth it is not a significant factor (Appendix 1, Regression 3). However, it also reduced the significance of Urban Growth. Both variables have significant positive relationships with internet growth, however together they both become less significant, these suggests there is multicollinearity. Multicollinearity occurs when two variables are highly correlated. GDP growth is shown to have a strong positive relationship with GDPGrowth, the regression coefficient of GDPChange onto UrbanPChange is 3.16 (Appendix 2, Reference 4), which is statistically significant. This is expected as cities are often more productive than they're rural counterparts thus urban growth is generally accompanied by GDP growth. Given that GDP Growth has t-score of .7 (Appendix 1 Regression 3), when included with Urban Growth this denotes that much of GDP Growth's effect is picked up by Urban Growth. Therefore, GDP Growth was dropped from the model as it is insignificant as a direct indicator of Internet Growth.

**Coverage, The Most Influential Indicator:**

The most accurate model that was created from this data set in regards to predicting the percentage change in internet users is the 4th regression shown in Appendix 1. In this regression, the variable coverage was added to the model. Adding Coverage to the model caused urban growth and population change becomes non-significant. This is again due to multicollinearity. shows Coverage is a good predictor of Urban Growth, with a t-score of -11.10 (Appendix 2, Reference 1) and Population Growth, with a t-score of -11.19 (Appendix 2, Reference 3). Yet there is still value to keeping PopPChange and UrbanPChange in the model as having them both significantly increases the accuracy of the model in predicting internet growth even if only one of them is dropped (Appendix 1 regression 7, Regression 8).

**Can a meaningful conclusion be drawn?**

For every point of urban growth in a year, internet connectivity will increase by roughly .5697% as show in appendix 1 regression 7, which is the most accurate model that provides a significant coefficient for urban growth. This is a relevant level of impact considering that the average internet growth per year is 4.138%. However, in this model has disadvantages as it picks up on other indicators such as general population growth. Finally, it should also be noted that if India is included in the model Urban Growth has more of an impact than Population Change as India has a higher rate of percent urban growth than urban growth. Therefore, without further research into the factors that separate urban population growth from general population growth, it will be hard to separate the impact of urban growth and population change though both are relevant.

**Global Cities: Further Exploration of Urban Change:**

Alpha cities were put into the data set to attempt to provide another metric exploring Urban Growths impact on internet connectivity. Alpha denotes a City that has large economic and cultural influence on the world and they tend to be well developed with a strong infrastructure. Also given the fact that are better connected to the global economy that other cities may mean they get innovations in technology faster. Therefore, if more people are moving to these Cities it may improve a Country’s connectivity rate. This is why the variable Alpha\_UG was included, if Alpha\_UG was significant it would provide evidence that countries with Alpha cities would experience different levels of internet growth as a result of urban growth than when compared with the rest of the world. However, the variables Alpha and Alpha\_UG can be seen to make little impact when India has been removed from the data set. This may be because most Alpha cities are in countries already developed or because Alpha cities do necessarily have universally strong infrastructure. If India is added back into the data set Alpha\_UG’s significance will increase but not enough for it to be considered. India contains an Alpha level city and has average urban growth, therefore, weights the results heavily given its extraordinary internet growth (Appendix 1 Regression 10).

**Discussion: Coverage**

**Graph 2, 3**



The regressions showed that by far the most powerful indicator of the percentage change in internet users was last year’s Coverage. Coverage’s power as indicator is due to numerous reasons first being mathematical, as when determining percentage change when the numerical growth is linear, the percentage growth represents a smaller piece of the whole each improvement. However, if the mathematical explanation was the only influencing factor the trend line would represent an exponential decay as the reduction would grow proportionally smaller each year, the graphs above show a more linear relationship. It must also be taken into account the possible effect of saturation, it takes more effort to connect more spread out rural areas on than does to connect devices to the internet in the denser city. Take Google's balloons to spread the internet to rural Africa as an example of the types of innovation that need to occur to get everyone connected, where it cities cables can provide the internet to a large concentration of people. Finally, there is the idea of a network effect where the value of the internet to the user’s changes by the number of other users, (Appendix 3). The network effect would create a snowball effect towards growth as demand, or motivation to get connected, increases as more users are added. Many services on the internet have their value is defined by the size of their user base, Facebook's popularity comes from the fact that almost everyone is on it. The Network Effect could be a possible that Coverage is shaped as a more gradual linear decline than steeper exponential decay.

**Conclusion and Recommendation:**

There is significant evidence that increases in urban population have a positive impact on Internet Connectivity, however due to multicollinearity within almost all of the explanatory variables determining an exact relationship is hard without more detailed research. For instance, it is hard to say if Urban Growth is viable primarily because it is an indicator of population growth or because it is an indicator of urban environments being easier to gain access to the internet in. Appendix 1 Regression 1 and 2 suggest that there is some merit to urban growth picking up on factors aside from population change, however when coverage has introduced the significance of both variables drop to the point where one or the other would have to be dropped in order to determine an accurate measure of its relationship to internet growth. The past year’s percentage of population connected is the strongest indicator of what the Percentage Change in Internet Users beyond the level of simply being an indicator of linear growth in internet users. This is because it picks up on factors such as the Network Effect and possibly saturation. As a result, this is a far more important indicator than other factor such as urban growth. In conclusion, as Amazon looks to identify potential markets for expansion, picking countries that are showing urban growth and population growth together have been demonstrated by this research to be good targets despite possible ambiguity of exact effects. In regards to coverage, it is a key metric as it demonstrates how large a market already is, but also is a good indicator of how much a country will grow as discussed in our section on coverage. Future studies should look to see if there a value of coverage where a country is growing fast due to not being saturated while still having enough of a current market to offset start-up costs and benefit from the network effect. Future studies should also take a more detailed look at what makes urban and rural environments different in terms of connectivity is necessary to attempt to unravel the multicollinearity and create more precise models.

**Appendix 1:**

Regression Table for PUsersChange: All Regressions had 196 Observations.

t=8.13 and t=2.14

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Explanatory**  **Variable** | Regression  1 | Regression  2 | Regression 3 | Regression  4 | Regression 5 | Regresion 6 | Regression 7 | Regression 8 | Regression 9 | Regression  10 |
| **Intercept** | 1.697  (4.45) | 5.830  (7.33) | 1.581  (4.91) | 5.394  (8.16) | 5.342 (8.03 ) | 5.412247  (8.10) | 5.643  (8.90) | 5.5179  (8.67) | 1.528  (4.62) | 5.815  (7.42) |
| **UrbanPChange** | .689  (2.18) | .375  (1.26) | .573  (2.15) | .287  (1.17) | .243 (0.96) | .289  (1.14) | .563  (3.90) |  | .524  (1.90) | .22  (0.73) |
| **PopPChange** | .8954  (1.84) | .297  (.64) | 1.059 (2.58) | .502  (1.31) | .573 (1.48) | .525  (1.35) |  | .919  (4.16 ) | 1.11  (2.66) | .522  (1.13) |
| **Coverage** |  | -.059  (-5.80) |  | -.054  (-6.44) | -.054  (-6.42) | -.054  (-6.08) | -.056  (-6.88) | -.055  (-6.72) |  | -.062  (-5.82) |
| **GDPChange** |  |  |  |  | .038 (0.73) |  |  |  | .041  (0.70) |  |
| **India** | yes | yes | no | no | no | no | no | no | no |  |
| **Adj R2** | 0.2585  (3.474) | 0.365  (3.216) | 0.329  (2.929) | 0.4448  (2.664) | 0.443  (2.667) | 0.439  (2.677) | 0.4421  (2.671) | 0.4426  (2.6595) | 0.326  (2.933) |  |
| **Alpha** |  |  |  |  |  | .224  (0.23) |  |  |  | -.521  (-0.45) |
| **Alpha\_UG** |  |  |  |  |  | -.093  (-0.15 ) |  |  |  | 1.35  (1.82) |

(denotes t-statistic, for Adj R2 denotes mean squared error)

**Appendix 2:**

Additional Regressions Referenced:

Dependent Variable: PUsersChange

|  |  |  |
| --- | --- | --- |
| Explanatory Variable | Regression 1 | Regression 2 |
| **Intercept** | 1.834  (4.87) | 3.756783  (11.61) |
| **UrbanPChange** | 1.204  (8.13) |  |
| **GDPGRowth** |  | .183  (2.41) |
| **Adj R2** | 0.2495  (3.496) | 0.0239  (3.9868) |

(denotes t-statistic, for Adj R2 denotes mean squared error)

Appendix 3: Correlations in Explanatory Variables

|  |  |  |  |
| --- | --- | --- | --- |
| Reference | Dependent Variable | Explanatory Variable | T-Statistic |
| 1 | PUsersChange | Coverage | -10.05 |
| 2 | UrbanPChange | Coverage | -11.10 |
| 3 | PopPChange | Coverage | -11.19 |
| 4 | GDPChange | UrbanPChange | 3.16 |

**Appendix 3**

**Excerpt from Investopedia:**  
What is the 'Network Effect':

The network effect is a phenomenon whereby a good or service becomes more valuable when more people use it. The internet is a good example. Initially, there were few users of the internet, and it was of relatively little value to anyone outside of the military and a few research scientists. As more users gained access to the internet, however, there were more and more websites to visit and more people to communicate with. The internet became extremely valuable to its users.

Found at: <http://www.investopedia.com/terms/n/network-effect.asp>,

Accessed December 15, 2016, No author listed.