Государственный университет

Высшая школа экономики

Нижегородский филиал

### Эссе по эконометрике

**Тема: «Эконометрический анализ влияния экономических показателей на численность пользователей Интернета»**

### Нижний Новгород

### 2008 г.

В наше время Интернет получил большое распространение. Пользователями Интернета в более развитых странах являются почти все слои населения, в менее экономически успешных государствах люди никогда не слышали о компьютерах и Интернете. Цель данной работы – показать зависимость численности пользователей Интернет в конкретной стране от экономических показателей, таких как ВВП на душу населения, национальный доход на душу населения, количество пользовательских компьютеров, а также степень урбанизации населения. Казалось бы, связь ясна: чем больше ВВП и НД, тем больше компьютеров в стране и соответственно больше пользователей всемирной паутины; чем больше городского населения относительно сельского, тем оно образованней и «продвинутей». Однако на практике это оказывается не совсем так. Ряд африканских стран вообще живет по племенным законам.

В этой работе я попытаюсь доказать существование прямой взаимосвязи между численностью пользователей Интернет и ВВП, НД и др. Попытаюсь доказать, что именно эти факторы влияют на количество пользователей ПК и Интернете в большей степени, попробую объяснить полученные результаты теоретически и подведу итог исследованию, сделав собственные выводы на основе проведенных исследований.

Сбор данных осуществлялся при использовании сайта www.geohive.comGeoHive: Global Statistics. В работу включена информация о выборке из 172 стран нашей планеты. Чтобы сделать моё исследование наиболее эффективным, я постараюсь следовать плану:

1. определить зависимую переменную и выбор регрессоров
2. построить регрессию (модель)
3. протестировать модель, оценить её «качество»
4. проанализировать результаты
5. сделать собственные выводы

В качестве метода исследования я использую эконометрический анализ, который буду осуществлять с помощью эконометрического пакета EViews 3.1, разработанного специально для этих целей.

Для описания зависимости я выбрала 6 переменных:

1. intusers – количество пользователей Интернет в стране
2. pc – численность пользовательских компьютеров в стране
3. gdp – Gross Domestic Product – ВВП на душу населения
4. gni – Gross National Income – НД на душу населения
5. urban – численность городского населения
6. rural - численность сельского населения

Выбрав 172 страны, я занесла данные в EViews и настало время для анализа данных. В первую очередь проверяем данные на ошибки.

ВВП на душу населения: нет отрицательных величин, но колеблется переменная значительно. Объяснить это легко, так как в выборке присутствуют как беднейшие страны, так и богатейшие.



Остальные переменные также необходимо смотреть на наличие ошибок, однако чтобы не загромождать эссе, графики я приводить не буду.

Далее смотрим взаимную корреляцию переменных:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | URBAN | GDP | GNI | INTUSERS | PC | RURAL |
|  |  |  |  |  |  |  |
| URBAN | 1.000000 | 0.056682 | 0.089996 | 0.736664 | 0.557379 | 0.873801 |
| GDP | 0.056682 | 1.000000 | 0.780379 | 0.302719 | 0.331656 | -0.068260 |
| GNI | 0.089996 | 0.780379 | 1.000000 | 0.400436 | 0.438161 | -0.060708 |
| INTUSERS | 0.736664 | 0.302719 | 0.400436 | 1.000000 | 0.964982 | 0.426228 |
| PC | 0.557379 | 0.331656 | 0.438161 | 0.964982 | 1.000000 | 0.211864 |
| RURAL | 0.873801 | -0.068260 | -0.060708 | 0.426228 | 0.211864 | 1.000000 |

Видим, что на численность пользователей Интернет огромное влияние оказывает число компьютеров в стране. Кроме того, немаловажное значение имеет численность урбанизированного населения.

Численность городского населения сильно зависит от национального дохода на душу населения.

Количество компьютеров в стране также связано с числом пользователей Интернет и степенью урбанизации населения.

А на число сельского населения оказывают влияние ВВП и НД в обратной зависимости, т.е. чем меньше ВВП и НД, тем больше населения занимается ручным трудом и сельским хозяйством. Это является показателем отсталости экономики и подтверждает правильность строящейся регрессии.



Строим регрессию, в которую включаем переменные из теоретической модели:

Ls intusers c pc gdp gni urban rural

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dependent Variable: INTUSERS | | | | |
| Method: Least Squares | | | | |
| Date: 02/27/08 Time: 02:03 | | | | |
| Sample(adjusted): 4 172 | | | | |
| Included observations: 132 | | | | |
| Excluded observations: 37 after adjusting endpoints | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | -346430.8 | 250802.2 | -1.381291 | 0.1696 |
| GDP | 10.32608 | 22.95037 | 0.449931 | 0.6535 |
| GNI | 502.9395 | 345.3779 | 1.456201 | 0.1478 |
| PC | 0.719045 | 0.014848 | 48.42816 | 0.0000 |
| URBAN | 0.090404 | 0.010513 | 8.598902 | 0.0000 |
| RURAL | 0.005584 | 0.005304 | 1.052842 | 0.2944 |
| R-squared | 0.989265 | Mean dependent var | | 5812423. |
| Adjusted R-squared | 0.988838 | S.D. dependent var | | 19682933 |
| S.E. of regression | 2079463. | Akaike info criterion | | 31.97751 |
| Sum squared resid | 5.45E+14 | Schwarz criterion | | 32.10854 |
| Log likelihood | -2104.515 | F-statistic | | 2322.154 |
| Durbin-Watson stat | 2.087052 | Prob(F-statistic) | | 0.000000 |

Видим, что незначительной переменной является ВВП, поэтому уберем его из регрессии. Все коэффициенты получились с ожидаемыми знаками, кроме величины сельского населения. Предполагалось, что это отрицательный фактор. Но так как его величина очень близка к 0, не будем обращать на это несовпадение внимания. К тому же его влияние незначительно.

Строим новую регрессию:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dependent Variable: INTUSERS | | | | |
| Method: Least Squares | | | | |
| Date: 02/27/08 Time: 02:09 | | | | |
| Sample(adjusted): 4 172 | | | | |
| Included observations: 132 | | | | |
| Excluded observations: 37 after adjusting endpoints | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | -354918.2 | 249305.2 | -1.423629 | 0.1570 |
| GNI | 618.1578 | 231.0229 | 2.675742 | 0.0084 |
| PC | 0.718812 | 0.014792 | 48.59489 | 0.0000 |
| URBAN | 0.090582 | 0.010473 | 8.649087 | 0.0000 |
| RURAL | 0.005475 | 0.005282 | 1.036557 | 0.3019 |
| R-squared | 0.989247 | Mean dependent var | | 5812423. |
| Adjusted R-squared | 0.988909 | S.D. dependent var | | 19682933 |
| S.E. of regression | 2072923. | Akaike info criterion | | 31.96396 |
| Sum squared resid | 5.46E+14 | Schwarz criterion | | 32.07316 |
| Log likelihood | -2104.621 | F-statistic | | 2920.986 |
| Durbin-Watson stat | 2.087552 | Prob(F-statistic) | | 0.000000 |

Как видно из таблицы, показатель Adjusted R-squared увеличился с 0,988838 до 0,988909. Это значит, что регрессия улучшилась.

Кроме того, регрессор RURAL оказывает незначительное влияние на регрессант, поэтому его можно удалить и построить новую регрессию:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Dependent Variable: INTUSERS | | | | |
| Method: Least Squares | | | | |
| Date: 02/27/08 Time: 02:12 | | | | |
| Sample(adjusted): 4 172 | | | | |
| Included observations: 132 | | | | |
| Excluded observations: 37 after adjusting endpoints | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | -399859.8 | 245577.6 | -1.628242 | 0.1059 |
| GNI | 630.0480 | 230.8051 | 2.729784 | 0.0072 |
| PC | 0.708903 | 0.011291 | 62.78241 | 0.0000 |
| URBAN | 0.100670 | 0.003869 | 26.01779 | 0.0000 |
| R-squared | 0.989156 | Mean dependent var | | 5812423. |
| Adjusted R-squared | 0.988902 | S.D. dependent var | | 19682933 |
| S.E. of regression | 2073526. | Akaike info criterion | | 31.95723 |
| Sum squared resid | 5.50E+14 | Schwarz criterion | | 32.04459 |
| Log likelihood | -2105.177 | F-statistic | | 3892.026 |
| Durbin-Watson stat | 2.066310 | Prob(F-statistic) | | 0.000000 |

Adjusted R-squared незначительно, но уменьшился. А это значит, что модель стала хуже. Поэтому вернемся к предыдущей модели:

Estimation Command:

=====================

LS INTUSERS C GNI PC URBAN RURAL

Estimation Equation:

=====================

INTUSERS = C(1) + C(2)\*GNI + C(3)\*PC + C(4)\*URBAN + C(5)\*RURAL

Substituted Coefficients:

=====================

INTUSERS = -354918.2484 + 618.1577906\*GNI + 0.7188117239\*PC + 0.09058209539\*URBAN + 0.005474726438\*RURAL

Согласно статистике Durbin-Watson stat ( =2.087552, статистика близка к 2) автокорреляция в модели отсутствует.

Выполним тест на гетероскедастичность:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| White Heteroskedasticity Test: | | | | |
| F-statistic | 7.466570 | Probability | | 0.000000 |
| Obs\*R-squared | 43.14884 | Probability | | 0.000001 |
|  |  |  |  |  |
| Test Equation: | | | | |
| Dependent Variable: RESID^2 | | | | |
| Method: Least Squares | | | | |
| Date: 02/27/08 Time: 02:18 | | | | |
| Sample: 4 172 | | | | |
| Included observations: 132 | | | | |
| Excluded observations: 37 | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | -1.19E+12 | 1.87E+12 | -0.639594 | 0.5236 |
| GNI | 2.10E+09 | 3.83E+09 | 0.548274 | 0.5845 |
| GNI^2 | -403321.1 | 1132324. | -0.356189 | 0.7223 |
| PC | 445574.9 | 228912.2 | 1.946488 | 0.0539 |
| PC^2 | -0.002393 | 0.000829 | -2.885490 | 0.0046 |
| RURAL | 74276.26 | 84150.65 | 0.882658 | 0.3791 |
| RURAL^2 | -9.97E-05 | 0.000103 | -0.965351 | 0.3363 |
| URBAN | 163878.1 | 87839.00 | 1.865665 | 0.0645 |
| URBAN^2 | -0.000216 | 0.000157 | -1.372084 | 0.1725 |
| R-squared | 0.326885 | Mean dependent var | | 4.13E+12 |
| Adjusted R-squared | 0.283105 | S.D. dependent var | | 1.35E+13 |
| S.E. of regression | 1.14E+13 | Akaike info criterion | | 63.03441 |
| Sum squared resid | 1.60E+28 | Schwarz criterion | | 63.23096 |
| Log likelihood | -4151.271 | F-statistic | | 7.466570 |
| Durbin-Watson stat | 1.436753 | Prob(F-statistic) | | 0.000000 |

В описываемой модели присутствует гетероскедастичность, т.к. вероятность ошибиться, отвергая гипотезу об отсутствии гетероскедастичности, практически ровна нулю. Но это можно объяснить тем, что выборка большая по размеру и неоднородна по значениям. Если сократить объем данных, то получится избавиться от гетероскедастичности.

Проведем тест Вальда:

Здесь, в данном тесте нам нужно определить, объясняют ли выбранные нами регрессоры регрессант лучше, чем константа. В тесте Вальда предположим все коэффициенты равными 0, т.е. C(1)=0, C(2)=0, C(3)=0, C(4)=0, C(5)=0. Получим, что:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Wald Test: | | | | |
| Equation: Untitled | | | | |
| Null Hypothesis: | C(1)=0 | | | |
|  | C(2)=0 | | | |
|  | C(3)=0 | | | |
|  | C(4)=0 | | | |
|  | C(5)=0 | | | |
| F-statistic | 2544.353 |  | Probability | 0.000000 |
| Chi-square | 12721.76 |  | Probability | 0.000000 |

В результате данного теста, мы получили, что Probability равный 0.000000, т.е. вероятность ошибиться, отклонив гипотезу, что все коэффициенты объясняют регрессию хуже, чем константа ровна нулю, значит, объясняющие переменные хорошо объясняют зависимую.

Выводы:

1. Полученная модель позволяет дать ответ на вопрос о зависимости численности Интернет пользователей от экономических показателей;
2. Согласно этой модели, наибольшее влияние на число пользователей оказывают национальный доход на душу населения, степень урбанизации населения и количество персональных компьютеров;
3. Хотя выявлены определенные закономерности, определяющие численность пользователей Интернет, разброс значений достаточно большой. На это указывает достаточно большая величина стандартного отклонения. Это обусловлено тем, что каждая страна уникальна. Население одной может при высоком доходе совсем не тратить средства на электронные устройства и, соответственно, общение через Интернет, а люди другой – наоборот, покупают всевозможные новинки и жить не могут, если они не на пике популярности технологий. Всех этих факторов учесть невозможно, но это и не было моей задачей. Я искала общие закономерности, и мне их удалось найти. Это главное достижение моей работы.