Regression Analysis for Global Growth in Skyscrapers

Jason Barr
Rutgers University-Newark
jmbarr@rutgers.edu
December 2017

1. The Model and Data

The goal is to estimate the times-series model:

$$\# \, Skycraper \ Completions_t = \alpha_0 + \boldsymbol{\beta}'\boldsymbol{x_t} + \delta t + \varepsilon_t,$$

where $\# \, Skycrapers \ Completions_t$ is the total number of tall buildings completed each year around the world (here a tall building is 190 meters or taller). The reason for this cut off is that for each year since 1960, except one, at least one 190-meter or taller building was constructed. The CTBUH report focuses on 200-meter or taller buildings. I decided to a slightly lower threshold to increase the number of non-zero observations; though it is very likely the results here are not sensitive to different cut-offs.

α_0 is the constant term, $\boldsymbol{x_t}$ are the exogenous variables, *t* is the trend, measured by the year, and ε_t is the random error term.

Here, the goal is to see how much of skyscraper construction is determined by global gross domestic product (GDP), population growth, and urbanization rates; these variables are hypothesized to increase the demand for tall buildings (Barr, 2016).

Each variable runs from 1960 to 2017. Note that since the independent variables are lagged one or two years (given the lag in construction, and to avoid endogeneity), the right hand side variables are to 2016.

2. Data Sources and Descriptive Statistics

 # Skyscrapers Completed each Year: Skyscraper Center, all completed buildings each year that are 190 meters or taller.

World Gross Domestic Product (Constant 2010 USD): World Bank.

World Population: World Bank.

World Urbanization Rate: World Bank.

Table 1 contains the descriptive statistics.

VARIABLE	OBS.	MEAN	STD. DEV.	MIN.	MAX.
YEAR	58	1988.5	16.9	1960	2017
SKYSCRAPER COMPLETIONS (190M+)	58	27.8	39.5	0	160
WORLD GDP ($USD TRILLIONS)	57	38.34	19.24	11.20	77.53
WORLD POPULATION (BILLIONS)	57	5.14	1.34	3.03	7.44
WORLD URBANIZATION RATE (%)	57	42.81	6.03	33.56	54.30
WORLD POP. GROWTH RATE (%)	56	1.62	0.32	1.18	2.11
WORLD GDP GROWTH RATE (%)	56	3.53	1.63	-1.74	6.66

Table 1: Descriptive statistics for the data set.

3. Unit Root Tests

If the dependent variable contains a unit root then ordinary least squares (OLS) is not an appropriate method of estimation (Wooldridge, 2015). For this reason a series of unit root test were conducted. Based on the tests described below, I do not find evidence for a unit root in the skyscraper completions variable and thus OLS is used for estimation.

Vector autoregression (VAR) and vector error corrections (VEC) models are all provided by below for comparison. All statistical analyses were preformed in Stata 15.1. Note that because of the presence of a zero in the skyscraper completions time series, I work with the variable, *ln(1+#completions).*

The first test was the Augmented-Dickey Fuller (ADF) that included a trend component. The first test excluded lags of the dependent variable, the second had one lag. The BIC suggests that the best specification is without the lag, but results of both tests are given in Table 2. Without the lag, we can reject the null hypothesis of a unit root at greater than 99%; with the lag, the p-value for null rejection is 0.168.

The next test as the Phillips-Perron test with a trend, which produced a p-value of 0.006, also suggesting a rejection of the null hypothesis of a unit root.

TEST	P-VALUE	BIC
ADF WITH TREND, NO LAGS	0.004	66.83
ADF WITH TREND, ONE LAG	0.168	68.88
PHILLIPS-PERRON WITH TREND, NO LAGS	0.006	

Table 2: Unit Root Tests for ln(1+# skyscraper completions)

The modified Dickey-Fuller t test ("dfgls" in Stata) also suggests that the optimal lag length for the ADF is 0. As given here:

```
. dfgls ln(1+#completions), maxlag(3)

DF-GLS for ln(1+#Completions)                               Number of obs =    54

               DF-GLS tau      1% Critical       5% Critical      10% Critical
  [lags]     Test Statistic        Value             Value             Value
------------------------------------------------------------------------------
    3           -2.180          -3.740            -3.073            -2.780
    2           -2.568          -3.740            -3.112            -2.816
    1           -2.809          -3.740            -3.145            -2.846
```

```
Opt Lag (Ng-Perron seq t) = 0 [use maxlag(0)]
Min SC   = -1.827771 at lag  1 with RMSE  .3724115
Min MAIC = -1.520928 at lag  1 with RMSE  .3724115

---------------------------------------------------------------------------
DF-GLS for ln(1+#completions)                              Number of obs =    57

               DF-GLS tau      1% Critical       5% Critical      10% Critical
  [lags]     Test Statistic        Value             Value             Value
------------------------------------------------------------------------------
    0            -4.275           -3.740            -3.156            -2.856
```

Table 3: Modified Dickey-Fuller t test results (DF-GLS).

Furthermore, a regression of *ln(1+#completions)* on the lag of *ln(1+#completions)* and the *year*, produces a coefficient estimate of *0.493* (with *p-value=0.00*) and residuals that have no serial correlation (see Table 3). In short, the evidence suggests that the count variable, once detrended, is a stationary AR(1) process, with coefficient estimate well below one (note that the upper 95% confidence interval value is 0.734).

```
. reg ln(1+#completions) l.ln(1+#completions) year

      Source |       SS           df       MS      Number of obs   =        57
-------------+----------------------------------   F(2, 54)        =    248.39
       Model |  80.1599007         2  40.0799503   Prob > F        =    0.0000
    Residual |  8.71355019        54   .16136204   R-squared       =    0.9020
-------------+----------------------------------   Adj R-squared   =    0.8983
       Total |  88.8734509        56  1.58702591   Root MSE        =     .4017

------------------------------------------------------------------------------
 ln(1+#completions)  |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ln(1+#completions).L1|   .4936614   .1201187     4.11   0.000     .2528378     .734485
              year   |   .0364769   .0089662     4.07   0.000     .0185007     .054453
             _cons   |  -71.20107   17.55221    -4.06   0.000    -106.3911     -36.011
------------------------------------------------------------------------------

Breusch-Godfrey LM test for autocorrelation
---------------------------------------------------------------------------
    lags(p)  |          chi2               df                 Prob > chi2
-------------+-------------------------------------------------------------
       1     |          1.956               1                   0.1619
       2     |          1.960               2                   0.3753
---------------------------------------------------------------------------
                        H0: no serial correlation
---------------------------------------------------------------------------
```

Table 4: Regression results which suggest that ln(1+#completions) *is stationary around its trend; the Breusch-Godfrey test suggests no serial correlation of the residuals.*

4. Regression Results

Based on preliminary regressions, it appears that $\Delta\ln(Pop)_t$ is a better measure than $ln(Pop)_t$. That is, what seems to matter is the rate of population growth rather than the size of the population per se. Rapid population growth would suggest rising demand for new structures.

Table 5 presents the results for four specifications for *ln(1+#completions)*. Equation (1) includes only $lnGDP_{t-2}$; Equation (2) includes $lnGDP_{t-2}$ and $ln(1+\#completions)_{t-1}$. Equation (3) includes $lnGDP_{t-2}$, $\Delta\ln(Pop)_{t-1}$, *urbanization rate*, and *year*. Equation (4) is the same as Equation (3), but also includes the *ln(1+#completions)*$_{t-1}$. In short, GDP, population growth and urbanization rate are all positive and statistically significant as expected.

VARIABLE	(1)	(2)	(3)	(4)
LN(GDP)$_{T-2}$	2.11	0.91	3.26	2.54
	0.00	0.00	0.05	0.07
ΔLN(POP)$_{T-1}$			61.8	60.3
			0.10	0.14
URBANIZATION RATE$_{T-2}$			0.59	0.49
			0.00	0.00
YEAR			-0.24	-0.19
			0.03	0.03
LN(1+COMPLETIONS) $_{T-1}$		0.58		0.15
		0.00		0.32
CONSTANT	-63.0	-27.1	350.4	279.6
	0.00	0.00	0.03	0.04
N	56	56	56	56
R^2	0.92	0.89	0.92	0.92
BGODFREY P-VALUE	0.00	0.13	0.13	0.28
HETTEST P-VALUE	0.93	0.01	0.10	0.04

Table 5: Regression Results. Dependent variable is ln(1+#completions)$_t$. *Note equations (1) was estimated via Newey-West regression (with two lags). Equations (2), (3) and (4) were estimated via OLS with robust standard errors. p-values below coefficient estimates. p-values less than 0.1 suggest statistical significance.*

Table 6 presents results from vector autoregression (VAR) and vector error correction (VEC) models, using specification (4), assuming an exogenous trend. Only presented are the skyscraper completions equations. The VAR results are similar to the OLS results. The coefficient for the co-integrating equation is not statistically significant, suggesting VEC estimation is not necessary.

	VAR		VEC
Variable	**Ln(1+#completions)**	**Variable**	**Δln(1+#completions)**
		C.E.$_{t-1}$	-0.024
			0.65
Ln(GDP)$_{t-2}$	2.58	**ΔlnGDP$_{t-1}$**	-0.85
	0.14		0.84
Δln(Pop)$_{t-1}$	60.6	**Δ^{2}lnPop$_{t-1}$**	132.6
	0.19		0.24
Urbanization rate$_{t-2}$	0.50	**Δurbanization rate$_{t-1}$**	0.65
	0.00		0.52
year	-0.19	**ΔlnCount$_{t-1}$**	-0.41
	0.08		0.00
Ln(1+#completions)$_{t-1}$	0.15	**Constant**	0.03
	0.24		0.93
Constant	286.7	**Cointegrating Equation**	
	0.08	**Ln(1+#completions)$_t$**	1.00
			na
		Ln(GDP)$_t$	10.2
			0.00
		Δln(Pop)$_t$	-543.3
			0.00
		Urbanization rate$_t$	-1.03
			0.00
		Constant	-262.3, na
N	55		54
R^2	0.92		0.29
		C.E. χ^2 p-value	0.00

Table 6: VAR and VEC models for ln(1+#completions). Note p-values below coefficient estimates. Note only the skyscraper completions equations are given.

Finally, Table 7 gives Poisson regressions for the skyscraper completions count (i.e., in levels). The results are qualitatively similar to those in Table 5.

VARIABLE	(1)	(2)	(3)
$LN(GDP)_{T-2}$	1.22	3.81	3.85
	0.33	0.01	0.01
$\Delta LN(POP)_{T-1}$		62.5	42.8
		0.31	0.44
URBANIZATION $RATE_{T-2}$		0.57	0.50
		0.00	0.00
YEAR	0.03	-0.24	-0.22
	0.45	0.01	0.02
# $COMPLETIONS_{T-1}$	0.01	0.00	
	0.00	0.38	
CONSTANT	-94.8	334.4	300.5
	0.02	0.01	0.03
N	56	56	56
LL	-172.3	-160.3	-160.7
AIC	352.6	332.6	331.4
BIC	360.7	344.8	341.5

Table 7: Poisson regressions with dependent variable of # skyscraper completions. p-values below estimates, from robust standard errors.

References

Barr, J. M. (2016). *Building the Skyline: The Birth and Growth of Manhattan's Skyscrapers*. Oxford University Press.

Wooldridge, J. M. (2015). *Introductory Econometrics: A Modern Approach*. Nelson Education.