Regression Diagnosis: Steps and Considerations in Regional Analysis

***Sung I. Jang*[1]**

[1] Regional Economy and Policy Institute
Email: sijang19@naver.com

Regional Economies and Policies (*REP*)
2017(*2*)*: 33 – 54*

Keywords regression diagnosis, regional analysis, residual analysis

Abstract

This article proposes a systematic procedure of standardized regression diagnosis with respect to regional analysis. The exemplary case is drawn from regional data for demonstrating this procedure. Regression diagnosis is essential for deriving viable implications for regions in terms of both academic and policy discussions, since the application and interpretation of OLS (ordinary least squares) regression analysis are not valid without meeting the statistical requirements for best linear unbiased estimators.

Contents

1. INTRODUCTION

This paper examines the appropriate steps of multiple ordinary least squares (OLS) regression and discusses its implications for regional research and analysis. Regression analysis, together with or without survey, is the most widely and frequently used analytic method across many social sciences including regional research. With cost-benefit analysis, in particular, it has recently been one of the two most preferred methods among policy analysts across disciplines (Morcol and Ivanova, 2010).

This most preferred analytics, however, has been often misused and misinterpreted in regional analysis. Among the three misleading components (i.e., the constant, the coefficient, and the error term) of the basic regression equation, the error term lies at the heart of such misunderstanding. There is no exception in regional research and analysis.

With an emphasis on the error term, this article examines the steps of diagnosing regression with a suggested flow chart for selecting the most appropriate procedures through such steps. For this, it employs practical regional data. In order to enhance the applicability of this diagnostics both in theory and practice, SPSS® (Statistical Package for the Social Sciences) as one of the most widely used statistical packages, was chosen to compare its results of conceptual and theoretical analysis in the form of its practical format.

2. REGRESSION DIAGNOSIS: STEPS AND APPLICATION IN THE STATISTICAL PACKAGE

A multiple regression model was adopted and four independents variables were employed for establishing a more substantial dataset. The exemplary regression analysis and its diagnostics in this study were made following procedures specified in a flow chart (Figure I).

CONSTRUCTING AND RUNNING A REGRESSION MODEL	
Constructing a Model	Reporting and Problem Detection (**esp. Multicollinearity)**
1) to establish a causal relationship based on theory and experience: $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + e$ where $e \sim IID\ N(0, \sigma^2)$	1) to report the model and detect apparent problems 2) to consider transforming, adding, or removing variables; to consider nonlinear models if necessary

↓↑

PRELIMINARY DETECTION	
Drawing the Residual or Leverage Plot	Preliminary Detection (**except Multicollinearity)**
1) plot of the "standardized residuals" (preliminary) 2) considering leverage plots	1) to detect linearity 2) to detect outliers and influence 3) to detect heteroscedasticity

↓↑

ESTIMATING REGRESSION COEFFICIENTS AND ANOVA	
Running a (Stepwise) Regression and Estimation	Test of Significance
1) OLS (Ordinary Least Squares) estimators (BLUE) 2) regression coefficients 3) correlation coefficients 4) coefficients of determination	1) to test the significance of regression coefficients 2) to test the significance of the regression model (*F*) 3) to test the significance of r^2

↓↑

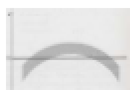
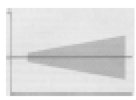
RESIDUAL ANALYSIS			
Diagnostic Methods			Diagnostics
Scatter Plots of Residuals (Y axis) and Predicted Y Values (X axis)		Linearity Problems	Diagnosis: quadratic form, or cubic form, etc.
			Suggestion: to transform, add, or standardize variables (e.g., X^2, X^3, lnX)
		Heteroscedasticity	Diagnosis: The variance is not constant.
			Suggestion: to use WLS (Weighted Linear Squares) if σ^2 is known or to transform the dependent variable
		Specification Errors	Diagnosis: There is a linear relationship.
			Suggestion: New explanatory variables are needed.
Residuals and Influence	Residuals	Outlier	Diagnosis: Outlier if $\lvert$standardized residuals$\rvert$ > 2
		Influence	Diagnosis: Influence if $\lvert$studentized residuals$\rvert$ > 2
	Influence Points	Cook's D	Diagnosis: Too much influence on the overall fit of the model if D > 4/n, where n (96) is the number of observations per variable
		sDFBETA	Diagnosis: Too much influence on specific observations if $\lvert$sDFBETA$\rvert$ > = $2/(n)^{1/2}$, where n (96) is the number of observations per variable
		sDFFITS	Diagnosis: Too much influence on the overall fit of the model if $\lvert$sDFFITS$\rvert$ > $2/(k/n)^{1/2}$, where n (96) is the number of observations and k is the number of parameters
Normality Detection and Tests	Visual Examination	Histogram	Frequency (Y axis) and OLS residuals (X axis)
		Normal Probability Plot	Expected Cumulative Probability (Y axis) and Observed Cumulative Probability (X axis)
	Tests	Skewness	Trouble if the S value is much different from zero
		Kurtosis	Trouble if the K value is much different from three
		Shapiro-Wilk Test	Reject normality if the W value is too small (3<n<2000)
		Kolmogorov-Smirnov	Reject normality if the D value is too small (n>2000)
		Jarque-Bera Test	Reject normality if the JB value is too small (not considered since it is appropriate for a large-sample test).

Figure I. The Flow Chart of the Regression Analysis and Diagnostics in the Study

2.1. The (Unadjusted) Regression Model

The following regression model was assumed. As an "unadjusted" multiple regression analysis, it adopts exemplary four independent variables. The number of observations is 96. The variables themselves are the ones that have been constructed from 96 local-level data of Korean 16 metropolitan areas, regarding the Internet traffic (networks) and Information and Communication Technology (ICT).

$$ICTGRDP = \beta_0 + \beta_1 ICTWORKFORCE + \beta_2 WORKRELATED + \beta_3 BACKBONESUB + \beta_4 X_4 COMNET + e$$

2.2. Preliminary Adjustment of the Model for Finding the Best Fitting Regression Line under the CLRM (Classical Linear Regression Model) Assumptions

2.2.1. Running the "Unadjusted" Regression Model and Detecting Apparent Problems

$$ICTGRDP = 16682700 + 78.889 ICTWORKFORCE + 24.187 WORKRELATED - 1.779 BAKCKBONESUB + 0.003 COMNET + e$$

s.e. =	(16016707)	(17522)	(1467)	(0.679)	(0.001)
t =	(1.042)	(4.502)	(16.483)	(−2.621)	(2.461)
sig. =	(0.300)	(0.000)	(0.008)	(0.010)	(0.016)
VIF =	(1.189)	(6.953)		(6.819)	(1.228)

F = 393.859 (0.000) R^2 = 0.945

Gujarati (2006, 2003: 359-363) proposed six indicators or methods for detecting multicollinearity as follows:

1) high R^2 but few significant t ratios;
2) high zero-order or pair-wise correlations among regressors;
3) examination of partial correlations;
4) auxiliary regressions;
5) eigenvalues and condition index; and
6) tolerance and VIF (Variance Inflation Factor).

In addition to the consequences of multicollinearity above, it is widely accepted that confidence intervals can be artificially wide and OLS estimators can be very sensitive to small changes in the data. Based on the suggestions above, a detection of multicollinearity in the unadjusted model was made. All of the regression coefficients but the constant are greater than two in their t value, and the t statistics were statistically significant because they are less than 0.05. The R^2 value was very high (0.945). As for the pair-wise or zero-order correlations among explanatory variables, the zero-order correlation coefficients between ***WORKRELATED*** and ***BACKBONESUB*** are high. The zero-order correlation coefficients of ***WORKRELATED*** and ***BACKBONESUB*** are 0.956 and 0.857, respectively. Gujarati (2003) discussed that muliticollinearity is a serious problem if the zero-order correlation coefficients between two regressors are higher than 0.8.[1]

[1] Gujarati (2003: 359) specified as follows: "High zero-order correlations are a sufficient but not a necessary condition for the existence of multicollinearity because it can exist even though the zero-order or simple correlations are comparatively low (say, less than 0.50)."

Table I. Correlations among Variables

Explanatory Variables	Correlations		
	Zero-order	**Partial**	**Part**
(Constant)	-	-	-
ICTWORKFORCE	*.193*	*.427*	*.113*
WORKRELATED	*.956*	*.865*	*.414*
BACKBONESUB	*.857*	*-.265*	*-.066*
COMNET	*.322*	*.250*	*.062*

As to the VIF values of ***WORKRELATED*** and ***BACKBONESUB***, they were greater than five, and their tolerance values were considered to be problematic because it was closer to 0.1. This also indicates that two explanatory variables, ***WORKRELATED***, and ***BACKBONESUB***, are problematic in multicollinearity, which means their intercorrelations are comparatively high. This multicollinearity problem was assumed to be associated with issues of the sample size because OLS residuals as an estimate of disturbances can be observed better if the sample size is fairly large (Gujarati, 2003: 401). In addition, a small number of samples were more likely to violate the assumption of the normal distribution of residuals because a small number of observations despite required minimum four variables tend to make it hard for the residuals to be scattered randomly in a circular pattern on the scatter plot of the standardized residuals as they can prevent the residual frequency distribution of explanatory variables from being normally distributed in the histogram of residuals and from not being deviated from the diagonal in the normal probability plot of residuals.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Tolerance	VIF
1	(Constant)	16682700	16016707		1.042	.300		
	ICTWORKFORCE	78.889	17.522	.123	4.502	.000	.841	1.189
	WORKRELATED	24.187	1.467	1.092	16.483	.000	.144	6.953
	BACKBONESUB	-1.779	.679	-.172	-2.621	.010	.147	6.819
	COMNET	.003	.001	.069	2.461	.016	.814	1.228

a. Dependent Variable: ICTGRDP

Figure II. A Report of Regression Coefficients of the Model in SPSS

Farrar and Glauber (1967) suggested that the partial correlation coefficients also need to be looked at because high zero-order correlations are a sufficient but not a necessary condition for the presence of multicollinearity. For instance, if $R_{1.234}^2$ is very high but $r_{12.34}^2$, $r_{13.24}^2$ and $r_{14.23}^2$ are comparatively low may imply that the variables X_2, X_3, and X_4 are highly intercorrelated and that at least one of these variables is superfluous (Gujarati, 2003: 360). Despite its usefulness, examination of partial correlation does not guarantee that they will show an infallible guide to multicollinearity, since it may happen that both all the partial correlations and R^2 are sufficiently high. Wichers (1975) discussed that the partial correlation test of Farrar and Glauber was not effective in that a given partial correlation may compatible with different multicollinearity patterns.[2] For example, is high and the partial correlations are high as well, multicollinearity cannot be readily detectable.

[2] Recited from Gujarati (2003: 360). The Farrar-Glauber test has also been criticized by O'Hagan & McCabe (1975) and Kumar (1975).

As a way for detecting which variable is highly correlated with the rest of it in the unadjusted model, each of the explanatory variables was regressed on the remaining explanatory variables in the model so as to look at the corresponding coefficients of determination R^2 in each case. The results showed similar patterns to the results of the examination of zero-order correlations; it also indicated that the multicollinearity problem lies with ***WORKRELATED***, and ***BACKBONESUB***.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.399[a]	.159	.132	753677.42984

a. Predictors: (Constant), COMNET, BACKBONESUB, WORKRELATED

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.925[a]	.856	.851	8999404.2845

a. Predictors: (Constant), COMNET, BACKBONESUB, ICTWORKFORCE

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.924[a]	.853	.849	19460384.179

a. Predictors: (Constant), COMNET, WORKRELATED, ICTWORKFORCE

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.431[a]	.186	.159	11862383088

a. Predictors: (Constant), BACKBONESUB, ICTWORKFORCE, WORKRELATED

Figure III. A Report of Model Summary of the Model in SPSS

To address the multicollinearity problem, several rule-of-thumb methods were considered because there is no sure remedy for it. According to Gujarati (2003: 363-375), such rule-of-thumb methods include:

1) using extraneous or prior information;
2) combining cross-sectional or time-series data;
3) omitting a highly collinear variable'
4) transforming data; and
5) obtaining additional or new data.

In addition to the remedial methods above, a factor analysis of the variables can also help. The method 1) was not readily available because there have been hardly any previous empirical studies in which the collinearity problems happen to less serious, and there were also hardly any available time-series data in respect to the method 2). With regard to the method 3), it was hard to eliminate collinear variable, because they were important in terms of the causal relationship in the model and minimum four explanatory variables were required in this study. Therefore, a natural logarithmic transformation of the ***BACKBONESUB*** variable were made because it had the problematic multicollinearity which has been known from its zero-order correlation, tolerance, VIF value, and so on as we have examined.

2.2.2. Preliminary Visual Examination of Residuals in the “Unadjusted” Model

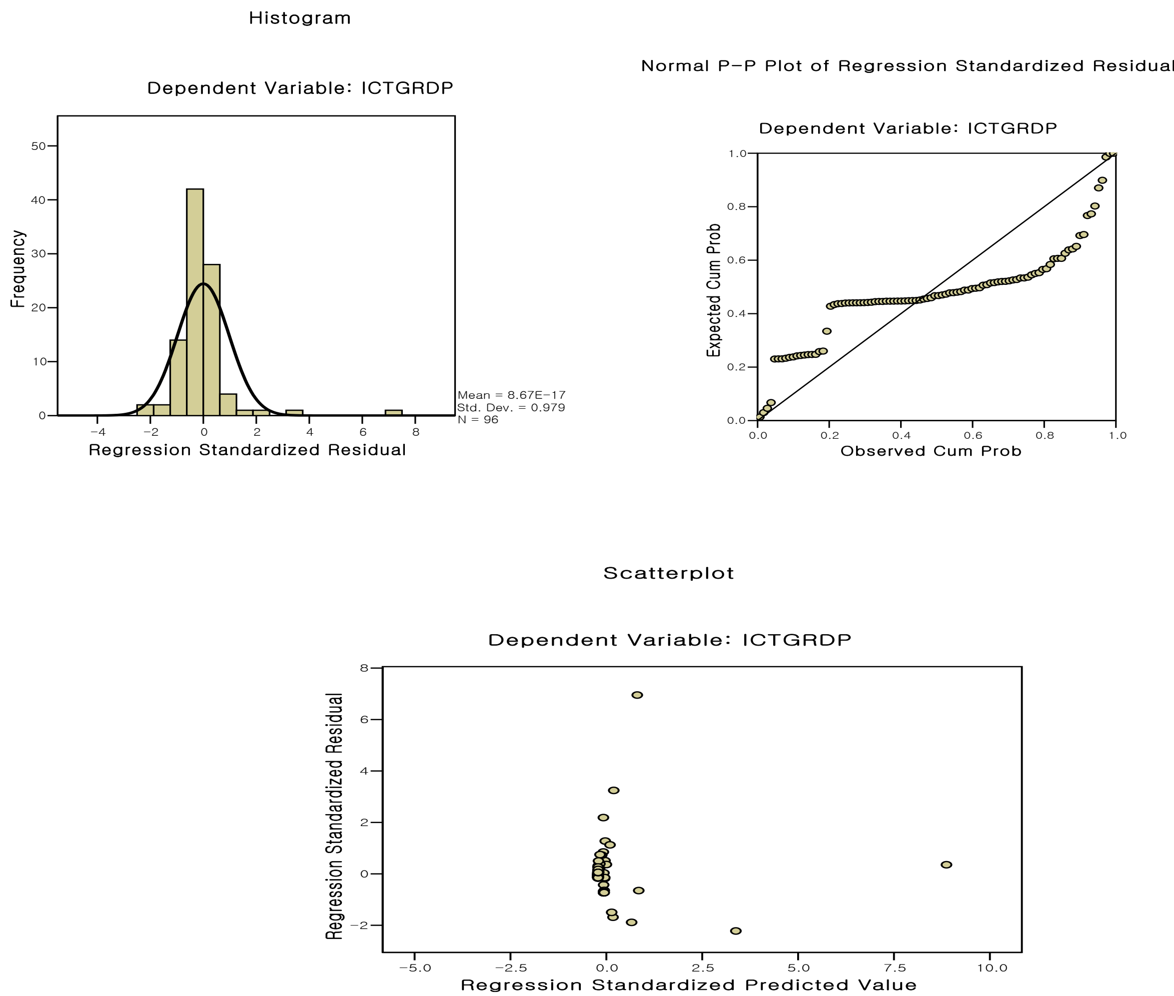


Figure IV. Reports of Histogram and Scatter Plots of the Standardized Residuals in SPSS

Both the histogram of residuals and the normal probability plot of residuals of the unadjusted model indicate the presence of regression problems. The histogram and the normal probability plot above indicate that the distribution of residuals does not follow a normal pattern in a strict sense. In particular, the normal probability plot of residuals shows that residuals are deviated from the diagonal. Many of these problems in the unadjusted model were considered to be also associated with limited (“ninety six”) number of observations per variable. As noted earlier, such a small number of observations despite required minimum four variables can make it hard for the residuals to be scattered randomly in a circular form on the scatter plot of the standardized residuals as they can make the residual frequency distribution of explanatory variables difficult in being normally distributed in the histogram of residuals and in being fitted on the diagonal line in the normal probability plot of residuals.

The scatter plot of the standardized residuals helps to detect outliers and non-linearities because “well behaved” residuals will be spherical (i.e., scattered randomly in an approximate circular pattern). Non-normal distribution of residuals can be a symptom of problems including but not limited to misspecification and heteroscedasticity (Pryce, 2002). Hence, in order to identify what specific problems cause such problems, the scatter plot of the standardized residuals on the predicted values was drawn by SPSS. It can be said that heteroscedasticity is present in the model if the residuals of the plot fans out in or fans in, and non-linearities have not been detected if the residuals of the plot follow a curved pattern. Based on the flow chart of regression analysis and diagnostics of Figure 1, therefore, it is safe to say that the scatter plot of the standardized residuals above indicates some problems of "linearity" and "heteroscedasticity." To solve the linearity problem, in the first hand, logarithmic transformations of each explanatory variable were considered through the procedure as follows.

2.2.3. Preliminary Examination of Heteroscedasticity in the "Unadjusted" Model

In addition to detection through the scatter plot of the standardized residuals, the Levene's test was done for testing which variable causes heteroscedasticity, which is caused by outliers and skewness, and so on. ***BACKBONESUB*** was first tested since a natural logarithmic transformation of it was supposed. In this respect, the Levene's test was made before and after the transformation was made. The results of ***BACKBONESUB*** below show that there was heteroscedasticity before the transformation was made. However, the results of ***BACKBONESUBLOG***, which was transformed, indicate that the null hypothesis of equal variances cannot be rejected. This means that the variance of the residual term does not vary by the ***BACKBONESUBLOG*** variable.

Group Statistics

	BACKBONESUB	N	Mean	Std. Deviation	Std. Error Mean
Unstandardized Residual	>= 9064285.00	6	155700368	400931296.348	163679516
	< 9064285.00	90	-10380025	75038652.0655	7909768.4

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Unstandardized Residual	Equal variances assumed	45.565	.000	3.343	94	.001	166080392	49677364.0	67444822	264715963
	Equal variances not assumed			1.013	5.023	.357	166080392	163870524	-2.55E+08	586734010

Group Statistics

	BACKBONESUBLOG	N	Mean	Std. Deviation	Std. Error Mean
Unstandardized Residual	>= 21.04	52	4592527.0	154792455.373	21465851
	< 21.04	44	-5427532	61717998.8638	9304338.4

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Unstandardized Residual	Equal variances assumed	.777	.380	.403	94	.688	10020058.9	24870930.4	-39361756	59401874
	Equal variances not assumed			.428	69.072	.670	10020058.9	23395586.9	-36651992	56692110

Group Statistics

	WORKRELATED	N	Mean	Std. Deviation	Std. Error Mean
Unstandardized Residual	>= 3608734.00	4	175693918	560386148.528	280193074
	< 3608734.00	92	-7638866	59043742.7516	6155735.7

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Unstandardized Residual	Equal variances assumed	83.445	.000	3.101	94	.003	183332784	59117909.8	65952788	300712779
	Equal variances not assumed			.654	3.003	.560	183332784	280260686	-7.08E+08	1.075E+09

Group Statistics

	WORKRELATEDLOG	N	Mean	Std. Deviation	Std. Error Mean
Unstandardized Residual	>= 8.44	53	5413002.9	311395341.043	42773440
	< 8.44	43	-6671841	109086369.484	16635515

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Unstandardized Residual	Equal variances assumed	.005	.943	.242	94	.809	12084843.8	49835168.2	-86864051	111033738
	Equal variances not assumed			.263	67.022	.793	12084843.8	45894526.0	-79520455	103690142

Figure V. A Report of Group Statistics in SPSS

As a fitted model based on theoretical and empirical reasoning, the regression model with a natural logarithmic transformation of the ***BACKBONESUB*** variable was first transformed because the ***BACKBONESUB*** variable was problematic in the multicollinearity issue which has been detected through the examination of zero-order correlation, its tolerance, VIF value, and so on, in addition to its heteroscedasticity, though they were not obviously identified in the partial regression scatter plots in a matrix form. The heteroscedasticity and linearity problem were improved by transforming the ***BACKBONESUB*** variable into the ***BACKBONESUBLOG*** variable. Many other transformations including quadratic, cubic, and square root were made, but did not improve the fit of the model, heteroscedasticity, multicollinearity, or linearity problems.

2.3. Adjusted model

ICTGRDP = -44410454 + 19.960ICTWORKFORCE + 48.210WORKRELATED + 2782247.2BAKCKBONESUBLOG + 0.004COMNET + e

s.e. =	(22171466)		(0.590)	(17.749)	(305334.667)	(0.001)
t =	(-2.003)	(33.824)	(2.716)	(3.455)		(3.390)
sig. =	(0.48)		(0.000)	(0.008)	(0.001)	(0.002)
VIF =	(1.183)		(1.283)	(1.265)		(1.199)

F = 393.859 (0.000) R^2 = 0.945

The variation in independent variables jointly explains the variation the dependent variable very much because the multiple coefficient of determination in the regression model is very high (0.945). The regression model is significant because the significance of its F statistics (393.589) is 0.000 (Table V). The coefficients of each regression variable indicate how much variability in the dependent variable (***ICTGRDP***) each explanatory variable explains. All the regression coefficients are significant because all the absolute values of their t statistics greater than “two” and their test statistics lies in the critical region of 0.05, which is its significance level (Table VI).

Model Summary [e]

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.956[a]	.914	.913	152189277.63	.914	1003.551	1	94	.000
2	.966[b]	.933	.931	135761593.52	.018	25.125	1	93	.000
3	.969[c]	.938	.936	130360934.98	.006	8.865	1	92	.004
4	.972[d]	.945	.943	123510048.09	.007	11.489	1	91	.001

a. Predictors: (Constant), WORKRELATED
b. Predictors: (Constant), WORKRELATED, ICTWORKFORCE
c. Predictors: (Constant), WORKRELATED, ICTWORKFORCE, BACKBONESUBLOG
d. Predictors: (Constant), WORKRELATED, ICTWORKFORCE, BACKBONESUBLOG, COMNET
e. Dependent Variable: ICTGRDP

Figure VI. A Report of Model Summary of the Adjusted Model in SPSS

ANOVA [e]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2.324E+19	1	2.32438E+19	1003.551	.000[a]
	Residual	2.177E+18	94	2.31616E+16		
	Total	2.542E+19	95			
2	Regression	2.371E+19	2	1.18535E+19	643.119	.000[b]
	Residual	1.714E+18	93	1.84312E+16		
	Total	2.542E+19	95			
3	Regression	2.386E+19	3	7.95252E+18	467.961	.000[c]
	Residual	1.563E+18	92	1.69940E+16		
	Total	2.542E+19	95			
4	Regression	2.403E+19	4	6.00821E+18	393.859	.000[d]
	Residual	1.388E+18	91	1.52547E+16		
	Total	2.542E+19	95			

a. Predictors: (Constant), WORKRELATED

b. Predictors: (Constant), WORKRELATED, ICTWORKFORCE

c. Predictors: (Constant), WORKRELATED, ICTWORKFORCE, BACKBONESUBLOG

d. Predictors: (Constant), WORKRELATED, ICTWORKFORCE, BACKBONESUBLOG, COMNET

e. Dependent Variable: ICTGRDP

Figure VII. A Report of ANOVA of the Adjusted Model in SPSS

As assumed, the Internet utilization through both backbone networks and regional firms' sub-networks (***BACKBONESUBLOG***, coefficient: 2782247.2) shows the highest positive contribution to variability in the GRDP in the ICT sector in the local region (***ICTGRDP***). The work-related Internet utilization of the regional ICT workforce (***WORKRELATED***, coefficient: 48.210) shows second highest in its contribution to ***ICTGRDP***. This means those two are the most important variables contribute to increase in the GRDP in the ICT sector in the local region. The conjoint effects of the regional work related computer utilization of regional ICT workforce and regional network firms' ICT network (***COMNET***, coefficient: .004) shows also a positive contribution to variability in the GRDP in the ICT sector in the local region, but its contribution is not so outstanding. In the case of ICT workforce (***ICTWORKFORCE***, coefficient: 19.960), it impacts the GRDP in the ICT sector in the local region positively. However, the regression model is not so consistent with empirical reasoning because ***ICTGRDP*** has a negative value if ***ICTWORKFORCE***, ***WORKRELATED***, ***BACKBONESUBLOG***, and ***COMNET*** have no value; some measures for fitting a regression without an intercept term can be taken so as to address such a problem. As for multicollinearity, the VIF values in Table VII are much lower, being compared to the VIF values in the unadjusted model.

Coefficients [a]

Model		Unstandardized Coefficients		Standardized Coefficients			95% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	53010415	15719065		3.372	.001	21799844.093	84220986.797					
	WORKRELATED	21.182	.669	.956	31.679	.000	19.855	22.510	.956	.956	.956	1.000	1.000
2	(Constant)	35431296	14454227		2.451	.016	6728066.546	64134525.278					
	WORKRELATED	21.001	.598	.948	35.143	.000	19.814	22.187	.956	.964	.946	.996	1.004
	ICTWORKFORCE	86.486	17.254	.135	5.012	.000	52.223	120.750	.193	.461	.135	.996	1.004
3	(Constant)	-11855734	21091680		-.562	.575	-53745630.196	30034162.075					
	WORKRELATED	20.418	.606	.922	33.676	.000	19.214	21.622	.956	.962	.871	.892	1.121
	ICTWORKFORCE	69.318	17.543	.108	3.951	.000	34.477	104.159	.193	.381	.102	.889	1.125
	BACKBONESUBLOG	2519074.4	846045.651	.086	2.977	.004	838754.720	4199394.102	.421	.296	.077	.798	1.253
4	(Constant)	-44410454	22171466		-2.003	.048	-88451347.216	-369561.034					
	WORKRELATED	19.960	.590	.901	33.824	.000	18.788	21.132	.956	.962	.829	.846	1.183
	ICTWORKFORCE	48.210	17.749	.075	2.716	.008	12.954	83.466	.193	.274	.067	.779	1.283
	BACKBONESUBLOG	2782247.2	805334.667	.095	3.455	.001	1182548.852	4381945.620	.421	.341	.085	.791	1.265
	COMNET	.004	.001	.091	3.390	.001	.002	.006	.322	.335	.083	.835	1.198

a. Dependent Variable: ICTGRDP

Figure VIII. A Report of Regression Coefficients of the Adjusted Model in SPSS

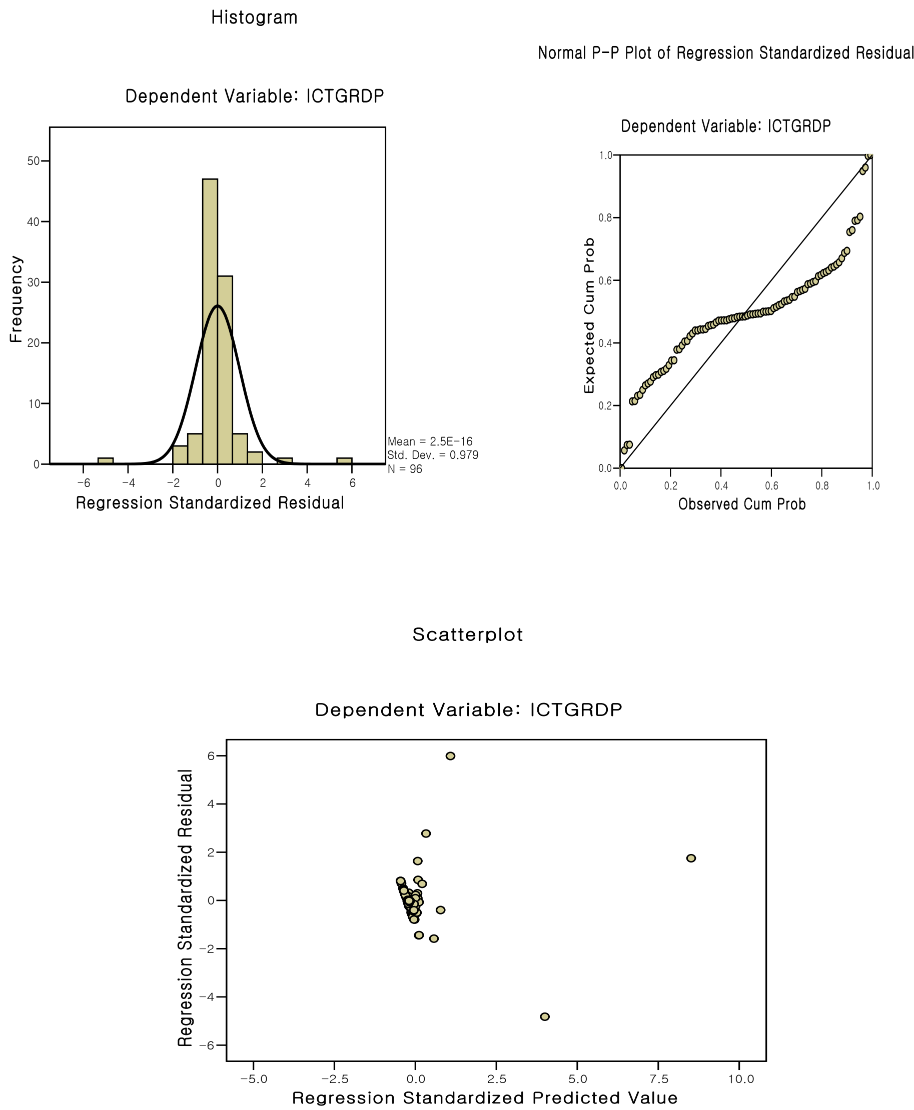
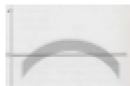
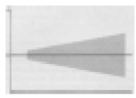


Figure IX. Histogram and Scatter Plots of the Residuals of the Adjusted Model

3. RESIDUAL ANALYSIS AND DIAGNOSTICS IN THE ADJUSTED MODEL

Being compared to the unadjusted model, problems including multicollinearity, heteroscedasiticity, non-normality are detected through many methods which were explained in the previous section and has been improved by transforming the ***BACKBONESUB*** variable. The ***BACKBONESUB*** variable has a problem in the multicollinearity issue which has been detected through the examination of zero-order correlation, its tolerance, VIF value, and so on, in addition to its heteroscedasticity. The heteroscedasticity and linearity problem were also improved by transforming the ***BACKBONESUB*** variable into the ***BACKBONESUBLOG*** variable. However, it is hard to say that the residual distribution is not normal in a strict sense because some causes prevent the residual frequency distribution of explanatory variables from being normally distributed in the histogram of residuals and from not being deviated from the diagonal in the normal probability plot of residuals. *Such a problem in the model is considered to be largely due to limited ("ninety six") number of observations per explanatory variable.*

Table II. Residual Analysis and Diagnostics in the Adjusted Model

RESIDUAL ANALYSIS IN THE ADJUSTED MODEL			
Diagnostic Methods			Diagnostics
Scatter Plots of Residuals (Y axis) and Predicted Y Values (X axis)		Linearity Problems	Diagnosis: Residuals showed a quadratic pattern with some outliers.
			Remedial Measures: A natural logarithmic transformation of the ***BACKBONESUB*** variable was made.
		Heteroscedasticity	Diagnosis: The variance was not constant.
			Remedial Measures: A natural logarithmic transformation of the ***BACKBONESUB*** variable was made.
		Specification Errors	Diagnosis: There was a linear relationship.
			Remedial Measures: New explanatory variables might help, but, instead, the transformation of a variable was made, because of the limitation of the number of observations despite the required minimum number of explanatory variables.
Residuals and Influence	Residuals	Outlier	Diagnosis: Outlier if |standardized residuals| > 2
		Influence	Diagnosis: Influence if |studentized residuals| > 2
	Influence (Points)	Cook's D (COO_1)	Diagnosis: Too much influence on the overall fit of the model if D > 0.0416 (=4/n, where n (96) is the number of observations per variable)
		sDFFITS (SDF_1)	Diagnosis: Too much influence on the overall fit of the model if |sDFFITS| > 0.4083 (=2/(k/n)$^{1/2}$, where n (96) is the number of observations and k is the number of parameters)
		sDFBETA (SDB_1)	Diagnosis: Too much influence on specific observations if |sDFBETA| > 0.2041 (=2/(n)$^{1/2}$, where n (96) is the number of observations per variable)
Normality Detection and Tests	Visual Examination	Histogram	Frequency (Y axis) and OLS residuals (X axis)
		Normal Probability Plot	Expected Cumulative Probability (Y axis) and Observed Cumulative Probability (X axis)
	Tests	Skewness	Trouble if the S value is much different from zero
		Kurtosis	Trouble if the K value is much different from three
		Shapiro-Wilk Test	Reject normality if the W value is too small (3<n<2000)
		Kolmogorov-Smirnov	Reject normality if the D value is too small (n>2000)
		Jarque-Bera Test	Reject normality if the JB value is too small (not considered since it is appropriate for a large-sample test).

It can be said that an observation is an outlier if its absolute value of the standardized residual (SRE_1 in the SPSS data of Appendix II) is greater than 2 because an outlier means "an observation with a large residual." In the same context, it can be said that an observation is an influence point if its absolute value of the studentized residual (SDR_1 in the SPSS data of Appendix II) is greater than 2. It can be said that an observation is an outlier and influence point if its absolute values of both the standardized residual and the studentized residual are greater than 2.[3]

Only "Seoul," "Incheon," and "Suwon," which are three most populated cities in the SMR (Seoul Metropolitan Area), have values greater than two in their standardized residuals. Incheon and Suwon have greater than two in their studentized residuals, as well. In addition, they are only regions with values greater than two both in their standardized residuals and in their studentized residuals (Table IX).

Descriptive Statistics

	N	Minimum	Maximu	Mean	Std.	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
ICTGRDP	96	32560.63	4.628E+09	129451568	517290600	7.461	.246	62.106	.488
ICTWORKFORCE	96	37.59	5836733.43	210832.55	808759.83	5.625	.246	33.757	.488
WORKRELATED	96	.03	206358732	3608734.3	23351761	7.804	.246	63.950	.488
BACKBONESUBLOG	96	-20.79	59.11	21.0437	17.69492	-.117	.246	-.241	.488
COMNET	96	6151960.90	4.104E+10	9.115E+09	1.294E+10	1.011	.246	-.809	.488
Valid N (listwise)	96								

Figure X. A Report of Descriptive Statistics of the Adjusted Model in SPSS

The normal distribution has the value of "zero" in skewness and has the value of "three" in kurtosis. As for skewness, ***ICTGRDP***, ***ICTWORKFORCE***, ***WORKRELATED***, and ***COMNET*** have values which are much greater than zero. All of their values are positive, and they are positive or right skewed (Table X). ***BACKBONESUBLOG*** has a relatively closer value to zero, and indicates a slight left skewed. As for kurtosis, ***ICTGRDP***, ***ICTWORKFORCE***, and ***WORKRELATED*** have very large values which are far greater than three. This indicates that each distribution of the three variables is very peaked than the normal distribution. In comparison, ***BACKBONESUBLOG*** and ***COMNET*** have smaller values which are closer to three. However, it is hard to say that their distributions follow the normal distribution. Both variables show flatter distributions. *As noted earlier, these problems in the model are considered to be largely due to limited ("ninety six") number of observations per explanatory variable.*

Table III. Skewness, Kurtosis, and JB Statistics of the Adjusted Model

Variable	*Skewness*	*Kurtosis*	*JB Statistic*
ICTGRDP	*7.461*	*62.106*	*11351656931*
ICTWORKFORCE	*5.625*	*33.757*	*3590808.16*
WORKRELATED	*7.804*	*63.95*	*23662816.76*
BACKBONESUBLOG	*-0.117*	*-0.241*	*8660189.257*
COMNET	*1.011*	*-0.809*	*43769883.19*

Considering that, under the null hypothesis that the residuals are normally distributed, the JB (Jarque-Bera) test of normality showed its statistics follows the chi-square distribution with 2 df (degree of freedom) asymptotically, it is safe to say that the residual distributions of all variables are extremely different from the normal distribution (Table X). *The JB test for normality is appropriate for large sample tests, and it is not the best test for normality in this study.*

[3] Outliers, influence points, and leverage points are closely related to one another, and cause certain influences on the OLS assumption that OLS gives equal weight to every observation in a regression model (Gujarati, 2003: 540-541).

Table XI shows that only the ***BACKBONESUBLOG*** variable is normally distributed because its statistics of the Kolmogorov-Smirnov test and the Shapiro-Wilk test are above 0.05 in both tests. This non-normality reported by both tests are not necessarily associated with the issue of limited ("ninety six") number of observations per explanatory variable, because each variable has ninety-six observations. *However, it is safe to assume that more observations will be needed since many values of each variable are too large, being compared to its number of observations* (e.g., 41037563685, the value of "Seoul" observation of the ***COMNET*** variable).

Tests of Normality

	Kolmogorov-Smirnov [a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
ICTGRDP	.401	96	.000	.239	96	.000
ICTWORKFORCE	.409	96	.000	.273	96	.000
WORKRELATED	.464	96	.000	.143	96	.000
BACKBONESUBLOG	.059	96	.200*	.989	96	.629
COMNET	.428	96	.000	.625	96	.000

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

One-Sample Kolmogorov-Smirnov Test

		ICTWORKFORCE
N		96
Normal Parameters [a,b]	Mean	210832.55
	Std. Deviation	808759.83
Most Extreme Differences	Absolute	.409
	Positive	.409
	Negative	-.397
Kolmogorov-Smirnov Z		4.003
Asymp. Sig. (2-tailed)		.000

a. Test distribution is Normal.

b. Calculated from data.

One-Sample Kolmogorov-Smirnov Test

		WORKRELATED
N		96
Normal Parameters [a,b]	Mean	3608734.3
	Std. Deviation	23351761
Most Extreme Differences	Absolute	.464
	Positive	.464
	Negative	-.439
Kolmogorov-Smirnov Z		4.542
Asymp. Sig. (2-tailed)		.000

a. Test distribution is Normal.

b. Calculated from data.

One-Sample Kolmogorov-Smirnov Test

		BACKBONESUBLOG
N		96
Normal Parameters [a,b]	Mean	21.0437
	Std. Deviation	17.69492
Most Extreme Differences	Absolute	.059
	Positive	.043
	Negative	-.059
Kolmogorov-Smirnov Z		.577
Asymp. Sig. (2-tailed)		.893

a. Test distribution is Normal.

b. Calculated from data.

One-Sample Kolmogorov-Smirnov Test

		COMNET
N		96
Normal Parameters [a,b]	Mean	9.115E+09
	Std. Deviation	1.294E+10
Most Extreme Differences	Absolute	.428
	Positive	.428
	Negative	-.241
Kolmogorov-Smirnov Z		4.189
Asymp. Sig. (2-tailed)		.000

a. Test distribution is Normal.

b. Calculated from data.

Figure XI. A Report of Tests of Normality of the Adjusted Model in SPSS

In terms of Cook's D ("COO_1" in Table XII and the SPSS data of Appendix II), the value of the thirtieth observation ("Seoul") has the extremely large value (See Appendix II (COO_1: 14.32358)). "Incheon" has also the second largest value (1.74491). It seems clear that Seoul and Incheon have too much influence on the overall fit of the model, since the overall fit is much influenced if D > 0.0416 (which equals $4/n$, where n (96) is the number of observations per variable) in the model. Guri (0.36613), Ilsan (0.07897) have also values which are greater than 0.0416. Although a region with a larger value in Cook's D, sDFFITS, or sDFBETA is not necessarily superior or more competitive in the hierarchical urban structure, the four regions have much influence on the overall fit of the model, and that is assumed to be largely due to the agglomeration effects of the SMR (Seoul Metropolitan Area).

Table IV. Residual Test Statistics of the Adjusted Model in SPSS and EXCEL

	Seoul	Incheon	Guri	Ilsan	Suwon
Standardized Residual (ZRE_1)	*1.74926*	*-4.82078*	*-1.58308*	*-0.40004*	*5.98852*
Studentized Residual (SRE_1)	*4.05122*	*-5.47711*	*-1.93252*	*-0.58636*	*6.30583*
Cook's D (COO_1) (critic. 0.0416)	*14.32358*	*1.74491*	*0.36613*	*0.07897*	*0.86508*
sDFFITS (SDF_1) (critic. 0.4083)	*9.29605*	*-3.58776*	*-1.37406*	*-0.62609*	*2.75641*
sDFBETA (SDB1_1) (critic. 0.2041)	*-0.13972*	*0.2675*	*-1.24469*	*-0.58462*	*1.53521*
sDFBETA (SDB2_1) (critic. 0.2041)	*8.53756*	*-2.98733*	*0.06332*	*0.00711*	*-0.05629*
sDFBETA (SDB3_1) (critic. 0.2041)	*-0.68516*	*-0.65954*	*0.07569*	*0.05066*	*1.00006*
sDFBETA (SDB4_1) (critic. 0.2041)	*0.65667*	*1.13140*	*0.09404*	*0.08466*	*0.57441*

In terms of sDFFIT ("SDF_1" in Table XII and the SPSS data of Appendix II), all four regions above are also problematic because their absolute values are greater than 0.4083 (Seoul: 9.29605; Incheon: -3.58776; Guri: -1.37406; Ilsan: -0.62609). These four regions have too much influence on the regression coefficients of the model as well because their absolute sDFBETA values are greater than 0.2041 (which equals $2/(n)^{1/2}$, where n (96) is the number of observations per variable). Suwon also has a very large value of sDFFIT (2.75641).

The sDFBETA value of each observation indicates which observation has too much influence on specific regression coefficients. For SDB1_1[4] (sDFBETA for the regression coefficient of the ***ICTWORKFORCE*** variable), only "Suwon" has a value greater than 0.2041 (-0.21927), except "Seoul," "Incheon," "Guri," and "Ilsan." Suwon is located near to the southern part of the city of Seoul. However, in addition to the adjacency to the city of Seoul, Suwon has another important point to be considered. The large value of Suwon in SDB1_1 is consistent with the fact that Suwon has large-scale Samsung Electronics Complexes and is employing much of the ICT workforce in the country. As noted earlier, however, it can be assumed that larger values in the SDF_1 and SDB_1 of Guri and Ilsan. In the case of "Guri" and "Ilsan," they don't have large-scale employers in the ICT sector. Suwon also has the largest value (1.00006) in SDB3_1 (sDFBETA for the regression coefficient of the ***BACKBONESUB*** variable) and Busan is the third in its value (0.30955). Seoul is the second (-0.68516). It is also consistent with the fact that Suwon has large-scale Samsung Electronics Complexes and has competitive network equipment and structure in the country.

For SDB2_1 (sDFBETA for the regression coefficient of the ***WORKRELATED*** variable), there are hardly any observations or regions whose values are greater than 0.2041. For SDB4_1 (sDFBETA for the regression coefficient of the ***COMNET*** variable), only "Suwon" has a value greater than 0.2041 (-0.57441), except "Seoul" (0.65667) and "Incheon" (1.13140). This suggests that further study on why such an "influential" observation have too much influence on the regression model can be needed, given that the regression model of this study has been properly established and improved.

4. CONCLUSION

With the given regional dataset, some of the problems from the unadjusted regression model have been detected and improved. The heteroscedasticity and linearity problem were improved by transforming the ***BACKBONESUB*** variable into the ***BACKBONESUBLOG*** variable. However, the normality examination of skewness and kurtosis as well as the normality tests including the Jarque-Bera test, the Kolmogorov-Smirnov test, and the Shapiro-Wilk test,

[4] In the SPSS data in Appendix II, "SDB1_1" refers to sDFBETA for the regression coefficient of the ***ICTWORKFORCE*** variable. "SDB2_1" refers to sDFBETA for the regression coefficient of the ***WORKRELATED*** variable; likewise, "SDB3_1" for the regression coefficient of the ***BACKBONESUB*** variable, and "SDB4_1" for the regression coefficient of the ***COMNET*** variable.

indicate that the residual distributions are not normal in a strict sense. Such problems in the regression model are considered to be largely due to limited ("ninety six") number of observations per explanatory variable. In addition, it is safe to assume that more observations will be needed since many most of each variable are too large, being compared to its number of observations.

The model can be more improved by acquiring more substantial or additional data as well as finding and including more appropriate explanatory variables. Although some problems remain to be solved, it can be said that the implications of this study are especially the importance of addressing the error term appropriately. In particular, the outliers and influence points can be interpreted as an "influential" region though their influence can be more detected both in the regression model and in the urban hierarchical or the network society in the Information Age context. With those statistical nuisances appropriately controlled in the procedure that has been systemized in this article, regional research and analysis will be enhanced richly and corroborated scientifically.

REFERENCES

Adams, P. 1999. "Bringing Globalization Home: A Homeworker in the Information Age." *Urban Geography* 20: 356-376.

Berry, B. & Pred, A. 1961. *Central Place Studies: A Bibliography of Theory and Applications*. Philadelphia: Regional Science Research Institute.

Castells, M. 2004. *The Network Society: A Cross-Cultural Perspective.* Northamton, MA: Edward Elgar.

Castells, M. 2003. *The Power of Identity,* 2nd ed. Oxford: Blackwell Publishers.

Castells, M. 2000. *The Rise of the Network Society,* 2nd ed. Malden, MA: Blackwell Publishers.

Castells, M. 1989. *The Information City: Information Technology, Economic Restructuring and the Urban-Regional Process*. Cambridge, MA: Basil Blackwell.

Castells, M. & Hall, P. 1994. *Technologies of the World: The Making of Twenty First-Century Industrial Complexes.* London: Routledge.

Cukier, K. 1999. "Bandwidth Colonialism? The Implications of Internet Infrastructure on International E-Commerce." *INET 99 Proceedings*, VA: Internet Society, Available at: www.isoc.org/inet99/proceedings/le/le_2.htm.

Ellison, G. & Glaeser, E. 1994. "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach" (August 1994). *NBER Working Paper*, No. W4840, Available at SSRN: http://ssrn.com/abstract=227956.

Fararr, D. & Glauber, R. 1967. "Multicollinearity in Regression Analysis: The Problem Revisited." *Review of Economics and Statistics* 49: 92-107.

Gaspar, J. & Glaeser, E. 1998. "Information Technologies and the Future of Cities." *Journal of Urban Economics* 43: 136-156.

Gilder, G. 2002. *Telecosm*. NY: Touchstone Books.

Graham, S. & Marvin, S. 1996. *Telecommunications and the City: Electronic Spaces, Urban Places*. London: Routledge.

Gujarati, D. 2006. *Essentials of Econometrics*, 3rd ed. Singapore: McGraw-Hill.

Gujarati, D. 2003. *Basic Econometrics*, 4th ed. Singapore: McGraw-Hill.

Kim, H. & Kwan, M. 2003. "Space-Time Accessibility Measures: A Geocomputational Algorithm with a Focus on the Feasible Opportunity Set and Possible Activity Duration." *Geographical Systems* 5: 71-91.

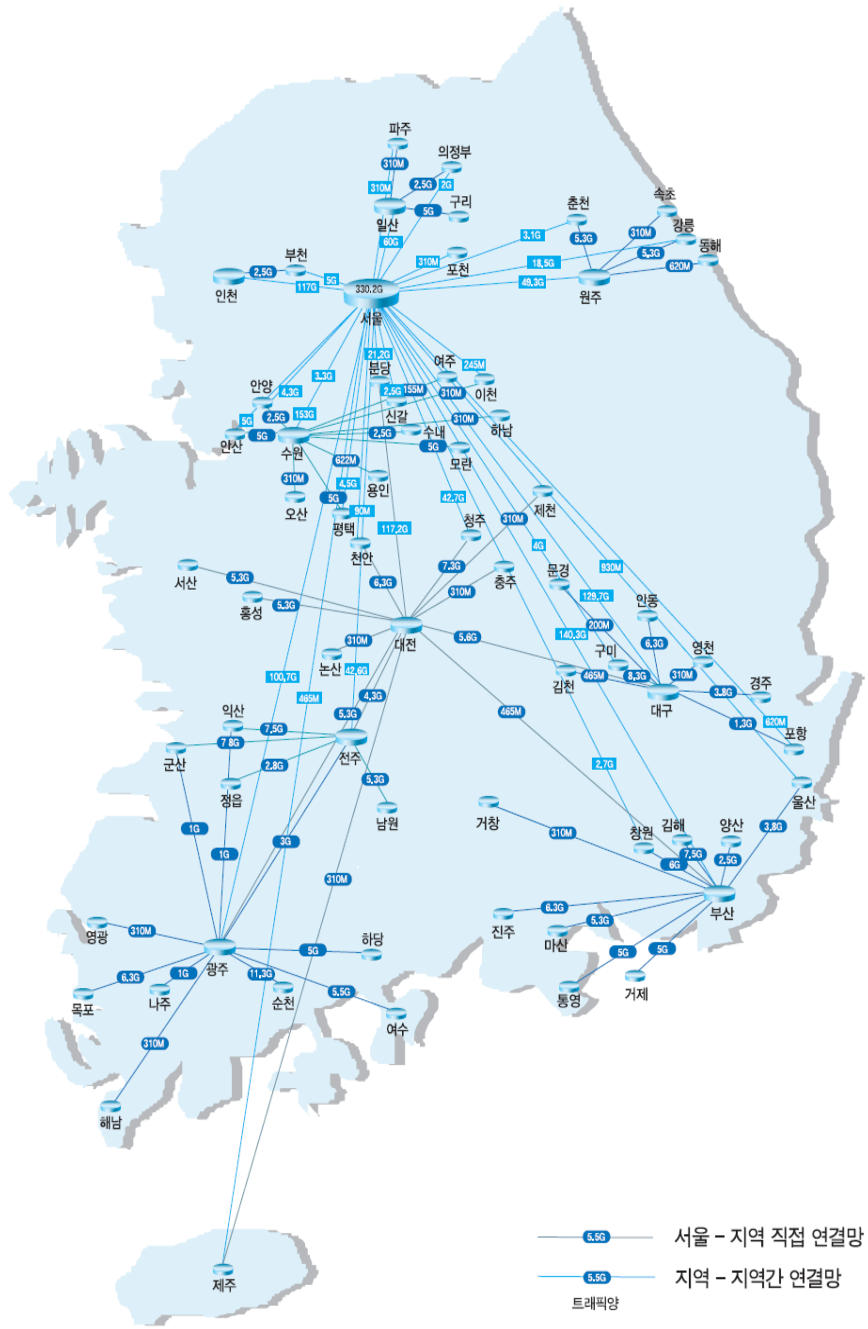
Koo, J. 2006. "The City faces Its Future: Technology and the Future of the City." In R. Steinbacher & V. O. Benson (eds.), *Introduction to Urban Studies*. Dubuque, Iowa: Kendall/Hunt.

Janelle, D. & Hodge, D. 2000. *Information, Place, Cyberspace, Issues in Accessibility.* Berlin: Springer-Verlag.

Janelle, D. 1997. “Sustainable Transportation and Information Technology: Suggested Research Issues.” *Journal of Transport Geography* 5: 39-40.
Janelle, D. 1995. “Metropolitan Expansion, Telecommuting, and Transportation. In Hanson, S. (ed.), *The Geography of Urban Transportation*, 2nd ed. NY: Guilford Press: 407-434.
Janelle, D. 1974. “Transportation Innovation and the Reinforcement of Urban Hierarchies.” *High Speed Ground Transportation Journal* 8: 261-269.
Janelle, D. 1969. “Spatial Reorganization: A Model and Concept.” *Annals of the Association of American Geographers* 59: 348-364.
Janelle, D. 1968. “Central Place Development in a Time-Space Framework.” *The Professional Geographer* 20: 5-10.
Lenntorp, B. 1976. *Paths in Space-Time Environment: A Time Geography Study of Movement Possibilities of Individuals, Studies in Geography 4.* Lund, Sweden: CWK Gleerup.
Lorenzton, S. 1995. “The Use of ICT in TNCs: A Swedish Perspective on the Location of Corporate Functions.” *Regional Studies* 29: 673-6895.
Malecki Edward J. 2002. “The Economic Geography of the Internet’s Infrastructure.” *Economic Geography* 78 (4): 399-424.
Moss, M. 1998. “Technology and the Cities.” *Cityscape* 3: 107-127.
Moulaert, F. & Gallouj, C. 1993. The Locational Advanced Producer Service Firms: The Limits of Economies of Agglomeration. *Service Industries Journal* 13: 91-106.
Pred, A. 1971. “Large-City Interdependence and the Preelectronic Diffusion of Innovations in the U.S.” *Geographical Analysis* 3: 165-181.
Pred, A. 1972. “Urban Systems Development and the Long-Distance Flow of Information through Preelectronic U.S. Newspapers.” *Economic Geography* 47: 498-524.
Pred, A. 1973a. *Urban Growth and the Circulation of Information: The United States System of Cities, 1790-1840*. Cambridge: Harvard University Press.
Pred, A. 1973b. “The Growth and Development of Systems of Cities in Advanced Economies.” In Pred, A. & Törnqvist, G., *Systems of Cities and Information Flows: Two Essays* (Lund Studies in Geography, Ser. B., Human Geography, No. 38:1-82).
Pred, A. 1974. “Industry, Information, and City-System Interdependencies.” In Hamilton, I. (ed.), *Spatial Perspectives of Industrial Organization and Decision-Making*, NY: John Wiley: 105-139.
Pred, A. 1964. “Toward a Typology of Manufacturing Flows.” *The Geographical Review* 54: 65-84.
Pryce, G. 2002. “Heteroscedasticity: Testing and Correcting in SPSS” (Unpublished Electronic Document).
Rodrigue, J.-P., Comtois, C., and Slack, B. 2006. *The Geography of Transport Systems.* NY: Routledge.
Scott, A. 2000. “Economic Geography: The Great Half Century.” *Cambridge Journal of Economics* 24:483-504.
Taylor, P. 2001. “Visualizing New Metageography: Exploring in the World-City Space.” In Dijkink, G. & Kinippenberg, H, (eds.), *The Territorial Factor: Political Geography in a Globalizing World*: 113-128, Amsterdam: Vossiuspers UvA.
Townsend, A. 2001. “The Internet and the Rise of the New Network Cities, 1969-1999.” *Environment and Planning B: Planning and Design* 28: 39-58.
Zook, M. 2000. “The Web of Production: The Economic Geography of Commercial Internet Content Production in the United States.” *Environment and Planning A* 31: 411-426.

TeleGeography (www.telegeography.com)

Appendix I. Map of the Interregional Internet Traffic



Source: http://isis.nida.or.kr

Appendix II. SPSS Data Editor Results ("Saved" Values)

Region	ZRE_1	SRE_1	COO_1	SDF_1	SDB1_1	SDB2_1	SDB3_1	SDB4_1
(Seoul)*	2.77302	2.94763	0.22573	1.11089	-0.5307	-0.34503	0.50644	0.97233
Gangreung	0.15688	0.15842	0.0001	0.02214	0.00301	0.00338	-0.01091	-0.0102
Gwangju	-0.02157	-0.02174	0	-0.00272	0.00019	0.00015	-0.00039	0.00133
Gumi	0.28648	0.29047	0.00047	0.04835	0.0097	0.01043	-0.03368	-0.01838
Daegu	0.05127	0.05167	0.00001	0.00641	-0.00069	-0.00053	0.00136	-0.00283
Daejeon	-0.03793	-0.03824	0	-0.00489	0.00051	0.00043	-0.00121	0.00227
Busan	0.32258	0.3249	0.00031	0.03889	-0.00371	-0.00245	0.00496	-0.01639
Bundang	-0.47551	-0.48595	0.0021	-0.10195	0.03725	0.02205	0.00496	-0.08631
Yongin	-0.53367	-0.54703	0.00303	-0.12269	0.02869	0.01229	0.04333	-0.09366
Suwon	0.85196	0.8762	0.00886	0.21022	-0.09915	-0.07606	0.10618	0.16482
Pyeongtaek	-0.55188	-0.56514	0.00311	-0.12419	0.03266	0.01536	0.03675	-0.09757
Incheon	0.17434	0.17562	0.00009	0.02119	-0.00162	-0.00105	0.00211	-0.00945
Ansan	-0.61177	-0.62577	0.00363	-0.13419	0.03942	0.02084	0.02916	-0.10883
Paju	-0.23678	-0.24575	0.00093	-0.06793	0.00513	-0.00237	0.04344	-0.04032
Pocheon	-0.19821	-0.20614	0.00069	-0.05861	0.0036	-0.00276	0.0388	-0.03369
Eujungbu	-0.49436	-0.50733	0.00274	-0.1165	0.0245	0.00916	0.04694	-0.08635
Anyang	-0.59366	-0.6075	0.00348	-0.13148	0.03675	0.01863	0.03313	-0.10526
Ulsan	0.50705	0.51888	0.00254	0.11226	0.02981	0.03024	-0.09098	-0.04262
Wonju	-0.02236	-0.02253	0	-0.00273	0.00008	0.00002	0.00005	0.0013
Icheon	0.70609	0.73266	0.00823	0.20234	0.05757	0.05884	-0.18079	-0.05896
Ilsan	-0.24248	-0.24822	0.00059	-0.05405	0.02454	0.01726	-0.01563	-0.04561
Jeonju	0.0594	0.05986	0.00001	0.00744	0.00025	0.00039	-0.00143	-0.00367
Jeju	0.81314	0.84926	0.0131	0.25553	0.07667	0.07698	-0.23028	-0.08003
Changwon	0.31281	0.3175	0.00061	0.05489	0.01161	0.01238	-0.03942	-0.02089
Cheonan	0.80616	0.84252	0.01309	0.25546	0.07569	0.07645	-0.23189	-0.0744
Cheongju	0.11593	0.1169	0.00005	0.01506	0.0011	0.00131	-0.00418	-0.0077
Chuncheon	0.38586	0.39276	0.00111	0.07426	0.01747	0.01823	-0.05649	-0.02832
Pohang	0.48861	0.50019	0.0024	0.1091	0.02795	0.02896	-0.09088	-0.03513
Bucheon	-0.63008	-0.64433	0.0038	-0.13738	0.04126	0.02246	0.02699	-0.11222
Seoul	**1.74926**	**4.05122**	**14.32358**	**9.29605**	**-0.13972**	**8.53756**	**-0.68516**	**0.65667**
Giheung	-0.7951	-0.81186	0.00562	-0.16725	0.06106	0.04527	-0.01705	-0.14331
Hanam	-0.72449	-0.74003	0.00475	-0.15372	0.05036	0.03203	0.01219	-0.12941
Seongnam	0.68733	0.70657	0.00567	0.16789	-0.0151	-0.0634	0.09724	0.10893
Anyang	-0.5281	-0.54077	0.00284	-0.11866	0.04905	0.04129	-0.04511	-0.09763
Ansan	0.08729	0.08964	0.00009	0.02084	-0.00899	-0.00768	0.01021	0.01642
Pyeongtaek	-0.07264	-0.07454	0.00006	-0.01706	0.00531	0.00652	-0.00901	-0.01285
Osan	-0.73704	-0.75276	0.00489	-0.15591	0.05103	0.03317	0.01062	-0.13152
Yeoju	-0.67451	-0.68952	0.00428	-0.14584	0.04037	0.02364	0.02993	-0.11875
Icheon	-0.79087	-0.8073	0.00547	-0.16507	0.05237	0.03779	0.00454	-0.14002
Suwon	**5.98852**	**6.30583**	**0.86508**	**2.75641**	**1.53521**	**-0.05629**	**1.00006**	**0.57441**

Bucheon	*-1.44726*	*-1.48952*	*0.02629*	*-0.36506*	*-0.21927*	*0.03504*	*0.10238*	*-0.15098*
Incheon	***-4.82078***	***-5.47711***	***1.74491***	***-3.58776***	***0.02675***	***-2.98733***	***-0.65954***	***1.1314***
Eujungbu	*-0.50379*	*-0.51598*	*0.00261*	*-0.11373*	*0.04623*	*0.04018*	*-0.04551*	*-0.09288*
Guri	***-1.58308***	***-1.93252***	***0.36613***	***-1.37406***	***-1.24469***	***0.06332***	***0.07569***	***0.09404***
Paju	*-1.4396*	*-1.47313*	*0.02046*	*-0.32191*	*-0.14669*	*0.05996*	*0.02648*	*-0.17127*
Ilsan	***-0.40004***	***-0.58636***	***0.07897***	***-0.62609***	***-0.58462***	***0.00711***	***0.05066***	***0.08446***
Gwangju	*0.0049*	*0.00494*	*0*	*0.00061*	*-0.00001*	*0*	*-0.00001*	*-0.00031*
Daegu	*-0.11575*	*-0.11671*	*0.00005*	*-0.01499*	*0.00216*	*0.00186*	*-0.00496*	*0.00622*
Busan	*-0.13943*	*-0.14065*	*0.00007*	*-0.0186*	*0.00339*	*0.00313*	*-0.00818*	*0.00661*
Jeonju	*-0.04378*	*-0.04411*	*0.00001*	*-0.00541*	*0.00037*	*0.00026*	*-0.00058*	*0.00255*
Cheonan	*0.0376*	*0.03791*	*0*	*0.00491*	*-0.00064*	*-0.00059*	*0.00163*	*-0.00211*
Cheongju	*-0.05318*	*-0.05364*	*0.00001*	*-0.00705*	*0.00087*	*0.00083*	*-0.00236*	*0.00315*
Hongseong	*-0.00167*	*-0.00168*	*0*	*-0.00022*	*0.00003*	*0.00002*	*-0.00007*	*0.00009*
Chungju	*0.33572*	*0.34092*	*0.00073*	*0.05994*	*0.01345*	*0.01388*	*-0.04225*	*-0.02551*
Seosan	*0.00239*	*0.00241*	*0*	*0.00031*	*-0.00004*	*-0.00003*	*0.00009*	*-0.00014*
Jecheon	*0.35966*	*0.36554*	*0.00088*	*0.06602*	*0.01502*	*0.01564*	*-0.04824*	*-0.02652*
Nonsan	*0.36861*	*0.37476*	*0.00095*	*0.06843*	*0.01573*	*0.01637*	*-0.05045*	*-0.0273*
Jeju	*0.44023*	*0.44898*	*0.00162*	*0.08957*	*0.02265*	*0.02306*	*-0.06908*	*-0.03643*
Daejeon	*0.29706*	*0.30419*	*0.0009*	*0.06672*	*-0.01785*	*-0.01391*	*0.05446*	*-0.01559*
Gumi	*0.08098*	*0.0821*	*0.00004*	*0.01362*	*-0.00078*	*-0.00321*	*0.00915*	*-0.00442*
Andong	*-0.05469*	*-0.05535*	*0.00001*	*-0.00861*	*0.00053*	*0.00192*	*-0.00537*	*0.00295*
Pohang	*-0.08684*	*-0.08746*	*0.00002*	*-0.01038*	*0.00025*	*0.0005*	*-0.00075*	*0.00464*
Gimcheon	*0.11939*	*0.12047*	*0.00005*	*0.01621*	*0.00198*	*0.00221*	*-0.00763*	*-0.00702*
Moongyeong	*0.23837*	*0.24135*	*0.00029*	*0.03811*	*0.00718*	*0.00767*	*-0.02499*	*-0.01502*
Gyeongju	*-0.10469*	*-0.10565*	*0.00004*	*-0.01431*	*0.00159*	*0.00257*	*-0.00685*	*0.00531*
Youngcheon	*0.16517*	*0.16688*	*0.00012*	*0.02392*	*0.00368*	*0.00399*	*-0.01336*	*-0.01*
Daegu	*0.22596*	*0.23087*	*0.00047*	*0.04815*	*-0.01326*	*-0.0112*	*0.03887*	*-0.01049*
Gunsan	*-0.01848*	*-0.01862*	*0*	*-0.0023*	*-0.00005*	*-0.00009*	*0.00035*	*0.00113*
Iksan	*-0.04081*	*-0.04112*	*0.00001*	*-0.00504*	*-0.00004*	*-0.00011*	*0.00052*	*0.00248*
Naju	*0.18341*	*0.1853*	*0.00014*	*0.02653*	*0.00395*	*0.00433*	*-0.01367*	*-0.01239*
Jeonju	*-0.17609*	*-0.17755*	*0.00011*	*-0.02283*	*0.00297*	*0.00265*	*-0.00716*	*0.00985*
Younggwang	*0.24558*	*0.24865*	*0.00031*	*0.03922*	*0.00742*	*0.0079*	*-0.02457*	*-0.01726*
Mokpo	*-0.07096*	*-0.07197*	*0.00003*	*-0.01214*	*0.00303*	*0.00285*	*-0.00829*	*0.00354*
Haenam	*0.22045*	*0.22304*	*0.00024*	*0.03416*	*0.0061*	*0.00653*	*-0.02039*	*-0.01532*
Sooncheon	*-0.06953*	*-0.07052*	*0.00003*	*-0.01187*	*0.00296*	*0.00278*	*-0.0081*	*0.00347*
Yeosoo	*-0.15124*	*-0.15276*	*0.00009*	*-0.02155*	*0.00414*	*0.00376*	*-0.01066*	*0.00812*
Gwangju	*0.09461*	*0.09699*	*0.0001*	*0.02179*	*-0.00582*	*-0.00534*	*0.01802*	*-0.00482*
Gunsan	*-0.30963*	*-0.31337*	*0.00048*	*-0.04866*	*0.01006*	*0.01042*	*-0.02988*	*0.01618*
Namwon	*-0.30552*	*-0.30877*	*0.00041*	*-0.04496*	*0.00808*	*0.00858*	*-0.02437*	*0.01643*
Iksan	*-0.27478*	*-0.27785*	*0.00035*	*-0.04145*	*0.00839*	*0.00831*	*-0.02369*	*0.01444*
Jeongeup	*-0.14315*	*-0.14426*	*0.00006*	*-0.01791*	*0.00155*	*0.00132*	*-0.00332*	*0.00824*
Jeonju	*-0.15157*	*-0.15522*	*0.00023*	*-0.03409*	*0.00772*	*0.00923*	*-0.02807*	*0.00756*
Sokcho	*-0.10902*	*-0.10983*	*0.00004*	*-0.01332*	*0.00026*	*0.0002*	*-0.00004*	*0.00641*
Gangreung	*-0.44294*	*-0.45255*	*0.0018*	*-0.09437*	*0.02075*	*0.02577*	*-0.07598*	*0.02177*
Donghae	*-0.14307*	*-0.14414*	*0.00006*	*-0.01763*	*0.00122*	*0.00098*	*-0.00219*	*0.00815*
Wonju	*-0.40269*	*-0.41238*	*0.00166*	*-0.09064*	*0.02008*	*0.02506*	*-0.0748*	*0.01967*

Ulsan	*0.02733*	*0.02755*	*0*	*0.00343*	*-0.00007*	*-0.00004*	*0.00007*	*-0.00176*
Jinju	*-0.03952*	*-0.03987*	*0.00001*	*-0.0053*	*0.00093*	*0.00088*	*-0.00234*	*0.00194*
Masan	*-0.03146*	*-0.03171*	*0*	*-0.00402*	*0.0006*	*0.00053*	*-0.00136*	*0.00158*
Gimhae	*-0.01462*	*-0.01476*	*0*	*-0.00203*	*0.00039*	*0.00037*	*-0.00101*	*0.00071*
Yangsan	*-0.00012*	*-0.00013*	*0*	*-0.00002*	*0*	*0*	*0*	*0.00001*
Gurchang	*0.4052*	*0.41278*	*0.00129*	*0.07981*	*0.0188*	*0.0198*	*-0.06231*	*-0.0283*
Tongyoung	*-0.06998*	*-0.07056*	*0.00002*	*-0.0091*	*0.0014*	*0.00133*	*-0.00344*	*0.0035*
Gurje	*-0.06179*	*-0.06231*	*0.00001*	*-0.00799*	*0.00121*	*0.00113*	*-0.00291*	*0.0031*
Changwon	*-0.0155*	*-0.01562*	*0*	*-0.00194*	*0.00026*	*0.00022*	*-0.00054*	*0.00079*
Busan	*1.63341*	*1.67247*	*0.02707*	*0.37164*	*-0.10508*	*-0.08341*	*0.30955*	*-0.06898*