Break Detection Methods Applied for International Industrial Property Time-Series Data

BeomYong Kim[1]
[1]Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, South Korea

Abstract
The US changed its patent policy toward pro-patent-ism in 1980s. Japan discarded one of dual IP (Industrial Property) infringement resolution systems in 1960. Korea adopted material (composition of matter) patent in 1987. These 3 break cases can be detected by iteratively moving QLR (Quandt Likelihood Ratio) tests through using patent or the like time-series data.

1st, using Stata command "varsoc", we can select the lag-orders for the level data themselves or the natural log data thereof. We may have to use the difference data from the above data. The selection criteria are FPE, AIC, HQIC, SBIC.

2nd, using Stata the command "cusum6", we can select the appropriate & final model from the above candidate models. Selection criteria are the cumulative sums(CUSUM) of the recursive residuals and their squares from the above models' regressions.

3rd, in applying the general regression to the time series data we tend to exaggerate the both ends and so should adopt the centered 70% range. The dummy variable "di" indicates the point of the break. The multiplications of independent variables of the above final model by the dummy variable "di" are required for the above QLR test coding contents.

Key Words: time-series, break, detection, policy, justification, verification
Note: Graphic & tabular images herein were produced by Stata except for self-products.
The following coding contents are modified from Torres-Reyna (2014).

1. Introduction

1.1 Statistical Significance: Statistics Informs International IP Policies

Statisticians select variables, ending up to prepare candidate models, and test the final result, that empowers industrial property policy stake-holders to justify decisions about the most effective and efficient international industrial property policies, such as pro-patent-ism adoption, dispute resolution simplification and patent scope enlargement. Statisticians contribute to providing more and better information for an international spectrum of decision makers - the United States, Japan and Republic of Korea.

1.2 Legends of Data Usage Herein

The US patent application data, including utility patent, design patent and plant patent, were used as “uspap”. “p” stands for patent and “ap” application. The Japanese Industrial property (patent, utility model, design and trademark) lawsuit application data were quoted as "jpsap". "s" means lawsuit.

The Korean patent application data were analyzed as "krpap", to which "t" could be added meaning "total". The Korean patent registration data were also utilized as "krpre" , to which "t" could be added meaning "total", either. "re" symbolizes registration. Both of the Korean patent application and registration data were divided into foreign and domestic data, from which we could find more detailed results in chapter 5.

For example, "krpapf" is intended to be Korean patent application foreign data. "ln" could be added for natural log. "d1", "d2", "L1", "L2" might also be combined together with the dot in front of the original data or the natural log data thereof, which are supposed to be difference data or the lag data from them.

Table 1: Data Collections Hereof

<table>
<tr><th colspan="2">Data</th><th colspan="2">Usage</th></tr>
<tr><td colspan="2">US Patent Application Data
(uspap)</td><td colspan="2">US Pro-patent-ism Proof</td></tr>
<tr><td colspan="2">JP IP(Industrial Property) Lawsuit Application Data (jpsap)</td><td colspan="2">Balloon Effect Proof from JP Dispute Resolution Simplification</td></tr>
<tr><td rowspan="2">KR Patent Application Total Data
(krpapt)</td><td>KR Pat App Foreign
(krpapf)</td><td rowspan="4">Patent Scope Enlargement Proof</td><td rowspan="4">Who Wins from KR Patent Scope Enlargement?</td></tr>
<tr><td>KR Pat App Domestic
(krpapd)</td></tr>
<tr><td rowspan="2">KR Patent Registration Total Data
(krpret)</td><td>KR Pat Reg Foreign
(krpref)</td></tr>
<tr><td>KR Pat Reg Domestic
(krpred)</td></tr>
</table>

1.3 Time Series Data Corrections and the CHOW Test Premises

Let's suppose we have an ADL(1,1) model (Autoregressive Distributed Lag) like the following function. Let τ symbolize the hypothesized break time point. Let $D_t(\tau)$ be the dummy variable that equals 0 before the break time point and 1 after. Accordingly, $D_t(\tau) = 0$ if $t \leq \tau$ and $D_t(\tau) = 1$ if $t > \tau$.

$$Y_t = b_0 + b_1 Y_{t-1} + d_1 X_{t-1} + g_0 D_t(\tau) + g_1[D_t(\tau) \times Y_{t-1}] + g_2[D_t(\tau) \times X_{t-1}] + u_t.$$

Under the H0 of no break, $g_0 = g_1 = g_2 = 0$. Under the H1 that there is a break, at least one of g's is nonzero. Thus H0 against H1 could be tested using the F-statistic, which is called a Chow test for a break at a known break date.

Most of the time a break is expected to be located within a range of dates of $\tau 0$ and $\tau 1$, for every τ between of which the Chow test can be applied and the 'largest' of the resulting F-statistics can also be selected. This modified Chow test is named as QLR (Quandt likelihood ratio) statistic and sometimes as the sup-Wald statistic.

For large-sample approximation the endpoints τ0 and τ1 cannot be too close to the beginning or the end of the sample. This is the reason why the common choice is to use 15% trimming, ending up to be centered 70% chosen. Upon both of the conditions of the above 15% trimming premise and the number "q" of dummy-interaction coefficients including the dummy itself, the QLR statistic $F_{q,\infty}$ distribution table is shown in Stock & Watson (2012, pp.599~601). Critical values for other trimming percentages are given in Andrews (2003).

Stock & Watson (2012) showed the following "Critical Values of the QLR Statistic with 15% trimming". The following symbols such as "uspap", "krpapf" and "jpsap" are corresponding to the right side values. For example, 7.12 of 10% is applied for "uspap".

Table 2: Critical Values of the QLR Statistic with 15% Trimming[F(q, ∞)]

Number of Restrictions(q)	10% Significance Level		5% Significance Level		1% Significance Level	
1	"uspap"	7.12		8.68		12.16
2		5.00		5.86		7.78
3		4.09		4.71	"krpapf"	6.02
4		3.59		4.09	"jpsap" "krpapt" "krpret" "krpred"	5.12
5		3.26		3.66	"krpref" "krpred"	4.53

Here we can change the above ADL(1,1) model into being an AR(1) model, which is written as a time series data model as follows. Similarly, we can have AR(2) and AR(3) models.

$$Y_t = b_0+b_1Y_{t-1}+ g_0D_t(\tau)+g_1[D_t(\tau) \times Y_{t-1}] + u_t$$

$$Y_t = b_0+b_1Y_{t-1}+b_2Y_{t-2}+ g_0D_t(\tau)+g_1[D_t(\tau) \times Y_{t-1}]+g_2[D_t(\tau) \times Y_{t-2}] + u_t.$$

$$Y_t=b_0+b_1Y_{t-1}+b_2Y_{t-2}+b3Y_{t-3}+ g_0D_t(\tau)+g_1[D_t(\tau) \times Y_{t-1}]+g_2[D_t(\tau) \times Y_{t-2}]+g_3[D_t(\tau) \times Y_{t-3}] + u_t.$$

Generally speaking, the following F-statistics are used for ANOVA (Analysis of variance). Here we have breaks or changes in the time series data for some reasons. These reasons are represented as the coefficients of g0, g1, g2, g3, where H0: g0=g1=g2=g3=0 and H1: at least one of g0, g1, g2, g3 is not zero.

$$F= \text{(between sum of squares) / (within sum of squares)} \sim F(k-1, N-k)$$

Because the QLR statistic is the 'largest' of many F- statistics, its distribution is not the same as an individual F-statistic. Instead, the critical values for the QLR statistic must be obtained from a special distribution such as the one in the above table. Like the F-statistic, this distribution depends on the number of restrictions being tested, q, that is the number of coefficients (including the intercept) that are being allowed to break, or change, under the alternative hypothesis. (Stock & Watson, 2012, pp.599~601)

Time series data could be here said to have auto-correlations (nearby interferences) and heteroscedastacities (abnormal distributions). Both of auto-correlations and heteroscedastacities could be removed through adding lag data of themselves and selecting normal distributions.

These addition and selection end up determining the final model. Each & all points of the final model are tested for finding the extent of changing. The values of extents of changing could be said to be the slope and intercept differences between left and right sides at those points.

The traditional CHOW test requires the iid conditions as is illustrated in the following diagram. Every point should be independent each other, with no auto-corelation accepted. Every point should have identically distributed error terms, without heteroskedasticity.

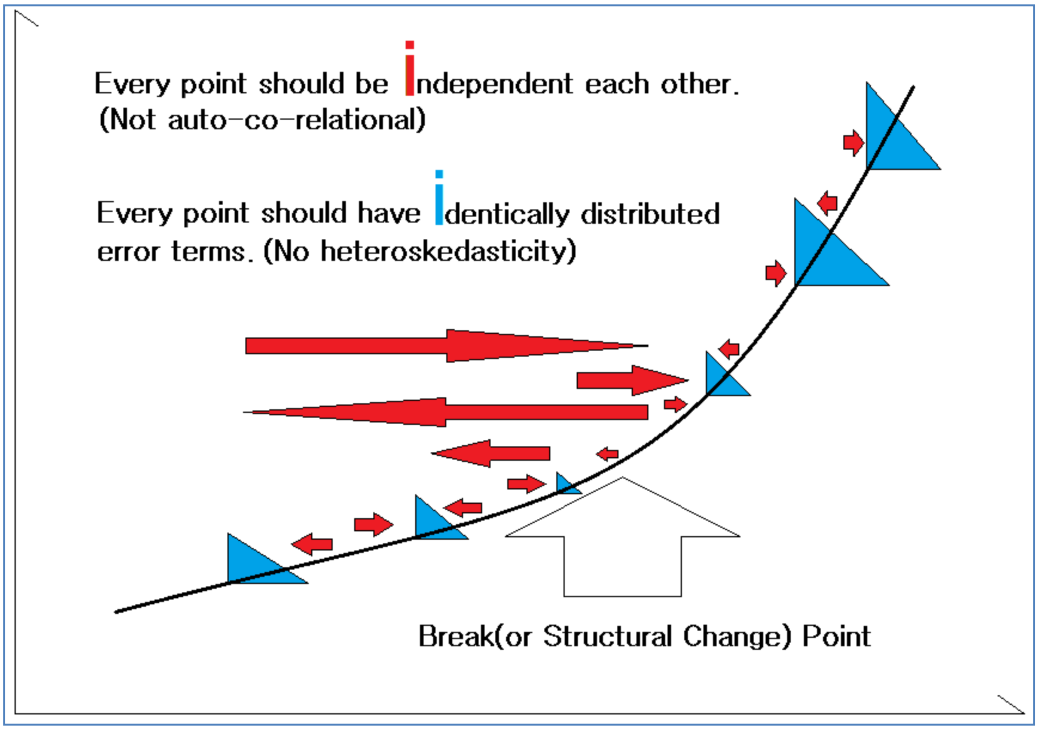


Figure 1: CHOW Test iid Requirements (Source: Kim, 2015)

The above QLR test was implemented to be a computerized coding sample by Torres-Reyna (2014). This thesis shows how to modify the above coding sample in order to adapt to the current international IP (Industrial Property) time series data.

1.4 What This Thesis Shows Compared with the Already Known Theories

This thesis shows how to set up the candidate "p"s and the AR(p) models by using the already known computer software like Stata in order to eliminate auto-correlations. This thesis also shows how to remove heteroscedastacities of the above candidate AR(p) models by selecting the best optimal AR(p) model from the above candidate AR(p) models through using the already known computer software like Stata, either.

Generally speaking, ARIMA (Auto Regressive Integrated Moving-Average) models or the other abnormal time series models like polynomial trend models might be transformed to be normal ones in terms of the iid conditions.

By taking the difference data from the abnormal original time series data, we can get the normal time series data through setting 'd' to be 0, 1, 2, …. in the ARIMA(p,d,q).

Sometimes we use the Box-Cox transform like $(y^{\lambda},-1)/\lambda$, $\lambda \neq 0$, or another transform such as ln(y), $\sqrt{y}$, 1/y. "ln" here stands for natural logarithm.

Furthermore, by observing the correlograms from the PACF (Partial Auto Correlation Function) and the ACF (Auto Correlation Function) like the following sample table, we can select 'p' for the AR(p) and 'q' for the MA(q). (Johnston & DiNardo, 1997, pp.207~220)

Table 3: Correlograms from the ACF and PACF of the Korean Patent Application Data

LAG	AC	PAC	Q	Prob>Q	Autocorrelation		Partial Autocorr	
					-1 0	1	-1 0	1
1	0.921	1.016	30.6	0		-------		--------
2	0.836	-0.33	56.6	0		------	--	
3	0.755	0.384	78.6	0		------		---
4	0.679	0.145	96.9	0		-----		-
5	0.59	-0.07	111	0		----		
6	0.491	0.104	122	0		---		
7	0.393	0.415	128	0		---		---
8	0.296	2E-04	132	0		--		
9	0.208	0.079	135	0		-		
10	0.131	0.299	135	0		-		--
11	0.055	-0.61	136	0			----	
12	-0.026	-0.36	136	0			--	
13	-0.108	0.104	136	0				
14	-0.171	-0.47	138	0	-		---	
15	-0.224	.	141	0	-			
16	-0.276	.	146	0	--			
17	-0.325	.	154	0	--			
18	-0.368	.	165	0	--			
19	-0.396	.	177	0	---			
20	-0.416	.	193	0	---			

However, the above procedures require repetitive 'trials and errors' from the on-site training experiences in order to get 'rule of thumb' selection wisdom. (Lee, 2013, p.96) For example, 'p' for the AR(p) is selected when ACF diminishes or lapses exponentially in the sine function format and PACF is truncated to be zero after the lag 'p'. And 'q' for the MA(q) is chosen if ACF and PACF have the results on the contrary.

For the principle of parsimony and for the efficiency and effectiveness by utilizing fully the already developed computerized software, this thesis simplifies the above steps as the following chapters. In ARIMA(p,d,q) this thesis only takes "p" and "d" without "q" considered. "q" should be left behind and still alive for showing breaks in this thesis.

2. How to Prove the US Pro-patent-ism

2.1 Research Background

As the following table, the US has taken pro-patent-ism since the 1980s. After the 1980s the US had experienced the increasing patent application numbers. Hall (2004) tried to analyze the US patent data by using the unit root tests and the growth rate thereof.

Table 4: Several Changes in the US Patent System around the years of 1980s

Changes	Results
1980 Diamond v Chakrabarty	Patents for Artificial Generic Organisms
1980 Bayh-Dole Legislation	University Patents Increased
1981 Diamond v Diehr case	Computer Software Patented
1982 CAFC Established	Patent Validity Sustained More
1984 Hatch-Waxman Act	Patent Term Restoration & Generic Drugs
1985/6 TI sues JP semiconductor firms	After winning, funding R&D from royalties
1986 Kodak v. Polaroid	1B$ judgement against Kodak
1994 TRIPS agreement	International harmonization begins
1998 State Street & ATT v. Excel	Patentability of business methods

(Source: Hall, 2004)

2.2 Lag-Order Selection for Level Data or Log Data Thereof

Using Stata command "varsoc", we can select the lag-orders for the level data themselves or the natural log data thereof. We may have to use the difference data from the above data. Balcombe et al (2011) accept the lag order of up to 3, with which we could not agree due to the afterwards seemingly abnormal data, in the following chapter 5, where the trend fluctuation varied extremely. Anyway here in chapters 2, 3 and 4 we maintained the lag of up to 3.

The following table was produced by using the Stata command "varsoc". At the leftmost candidate, "uspap" means the US patent application data and has the lag of 3. In the middle, "d1.uspap" shows that 1st level difference data of "uspap" has the same lag 3, either. In the rightmost, "d1.lnuspap", the natural log data of the middle, has the lag of 0. All of these 3 candidates are still useful.

Table 5: Lag Candidates from the US Patent Application Data

. varsoc uspap /*candidate*/

Selection-order criteria
Sample: 1956 - 1998 Number of obs = 43

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-522.779				2.2e+09	24.3618	24.3769	24.4028
1	-442.155	161.25	1	0.000	5.5e+07	20.6584	20.6886	20.7403
2	-441.265	1.7806	1	0.182	5.5e+07	20.6635	20.7088	20.7863
3	-436.685	9.1593*	1	0.002	4.7e+07*	20.497*	20.5574*	20.6608*
4	-436.313	.74396	1	0.388	4.8e+07	20.5262	20.6017	20.731

Endogenous: uspap
Exogenous: _cons

. varsoc d1.uspap /*candidate*/

Selection-order criteria
Sample: 1957 - 1998 Number of obs = 42

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-437.234				6.8e+07	20.8683	20.8834	20.9097
1	-437.119	.22967	1	0.632	7.1e+07	20.9104	20.9408	20.9932
2	-436.94	.35772	1	0.550	7.3e+07	20.9495	20.995	21.0737
3	-430.297	13.286*	1	0.000	5.6e+07*	20.6808*	20.7415*	20.8463*
4	-430.216	.16203	1	0.687	5.9e+07	20.7246	20.8004	20.9315

Endogenous: D.uspap
Exogenous: _cons

. varsoc d1.lnuspap /*best*/

Selection-order criteria
Sample: 1957 - 1998 Number of obs = 42

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	70.0585				.002184*	-3.2885*	-3.27334*	-3.24713*
1	70.0661	.01506	1	0.902	.00229	-3.24124	-3.21091	-3.1585
2	70.432	.73194	1	0.392	.002361	-3.21105	-3.16555	-3.08693
3	71.7193	2.5746	1	0.109	.00233	-3.22473	-3.16407	-3.05924
4	71.721	.00339	1	0.954	.002444	-3.17719	-3.10137	-2.97033

Endogenous: D.lnuspap
Exogenous: _cons

On the above table, “LL” results from the log likelihood test, “LR” the likelihood ratio test and “df” Augmented Dickey–Fuller unit-root test. FPE (the final prediction error), AIC (Akaike’s information criterion), SBIC (Schwarz’s Bayesian information criterion) and HQIC (Hannan and Quinn information criterion) are also shown.

2.3 Heteroskedasticity Tests for Models

Using the Stata command "cusum6", we can select the appropriate & final model from the above candidate models. Selection criteria are the cumulative sums (CUSUM) of the recursive residuals and their squares from the above models' regressions. (StataCorp, 2009)

The following diagram, both of the leftmost “L(1/3).uspap” and the middle “L(1/3).d1.uspap” have the out of range problem, which is the reason why these candidates should be discarded. The rightmost “d1.lnuspap”, the 1st level difference data from the natural log of the US patent application data, is the only one and best candidate.

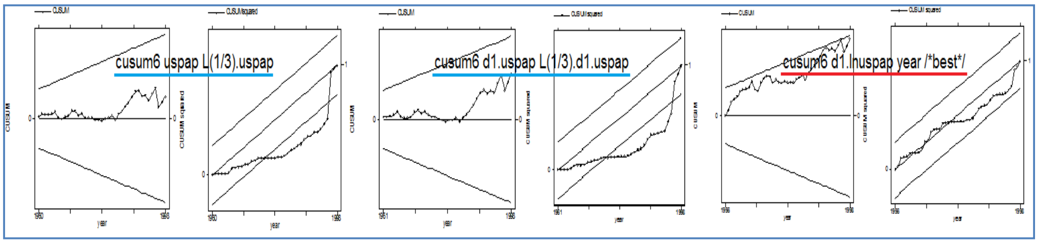


Figure 2: Heteroskedasticity Tests about the US Patent Application Data

2.4 QLR Test Coding for Finding Breaks of the US patent application data

In the following coding contents, both of “local i = round('time'*.15)” and “local f = round('time'*.85)” realize the centered 70% range, that, Mitchell(2014) insists, is necessary in applying the general regression to the time series data.

The following coding contents, both of “local var = "lnuspap"” and “gen diff'var' = d1.'var'” are combined with each other, ending up to become the 1st level difference data from the natural log of the US patent application data. The following coding content, “qui reg diff'var' di,r” means that there are only one independent variable “di” and only one dependent variable “diff'var'”, corresponding to “d1.lnuspap”

The dummy variable "di" indicates the point of the break. The dummy variable "di" and the multiplications of the independent variables thereby of the above final are required for the above QLR test coding contents. This case has the 0(zero) lag, and so, no independent variable multiplied by the dummy variable "di". “local critical=7.12” and “Critical value 10% ('critical')” are required for adaption to data characteristics.

```
use 52-98-jp-us.dta,clear
log using kby15-d1lnuspap-QLR-year.log,replace
tset year
sum year

local time=r(max)-r(min)+1
local i = round('time'*.15)
local f = round('time'*.85)
```

```
local var = "lnuspap"

gen diff'var' = d1.'var'
gen qlr'var' = .
set more off
while 'i'<=('f') {
gen di = (_n >='i')
cap gen d_year = di*year
list year 'var' diff'var' di d_year
qui reg diff'var' di,r
qui test di
cap replace qlr'var' = r(F) in 'i'
dis "'i', QLR of the year " %ty year['i'] " =" %6.2f qlr'var'['i'] " [see above table]"
drop di d_year
local i = 'i' + 1
}
sum qlr'var'
local maxvalue=r(max)
gen maxdate=year if round(qlr'var',0.01)==round('maxvalue',0.01)
local maxvalue1=round('maxvalue',0.01)
local critical=7.12 /*Replace with the appropriate critical value (see Stock & Watson)*/
sum year

local mindate=r(min)
sum maxdate
local maxdate=r(max)
gen break=year if qlr'var'>='critical' & qlr'var'!=.
dis "Below are the break dates..."
list year qlr'var' if break!=.
levelsof break, local(break1)

set more on
twoway connected qlr'var' year,title(breaks in US patent apply#(1952-98)) ///
xlabel('break1', angle(90) labsize(2.3) alternate) ///
yline('critical') ytitle(Quandt-LR(QLR) statistic) xtitle(Time) ///
ttext('critical' 'mindate' "Critical value 10% ('critical')", placement(ne)) ///
ttext('maxvalue' 'maxdate' "Max QLR = 'maxvalue1'", placement(e)) saving(ts3,replace)
scheme(s1manual)
more
twoway (tsline diff'var') lfit diff'var' year,saving(ts4,replace) scheme(s1manual)
more
twoway (tsline 'var') lfit 'var' year,saving(ts5,replace) scheme(s1manual)
more
graph combine ts3.gph ts4.gph ts5.gph,col(1) xsize(9) ysize(14)
```

2.5 The Source Data for Finding Breaks

The following left tables contain the US utility patent application data, the US design patent data, the US plant patent application data, the US total patent application data and the natural log data thereof. Right table has Japanese industrial property lawsuit application data and the natural log data thereof.

Table 6: US Data (Source: USPTO, 2016) and JP IP Lawsuit Data (Source: MIC, 2016)

	Utility Patent Applications (Inventions)	Design Patent Applications	Plant Patent Applications	US Patent Application Total Number	lnuspap		Japan, Industrial property, Lawsuit, Newly received	lnjpsap
1952	64554	4993	84	69631	11.15096515	1952	32	3.4657359
1953	72284	5450	99	77833	11.26232078	1953	42	3.7376696
1954	77185	5465	95	82745	11.32351887	1954	60	4.0943446
1955	77188	5764	118	83070	11.3274389	1955	51	3.9318256
1956	74906	4824	104	79834	11.28770476	1956	56	4.0253517
1957	74197	4714	101	79012	11.27735502	1957	64	4.1588831
1958	77495	4923	134	82552	11.32118368	1958	90	4.4998097
1959	78594	4879	114	83587	11.33364328	1959	68	4.2195077
1960	79590	4525	131	84246	11.34149637	1960	163	5.0937502
1961	83100	4714	107	87921	11.38419396	1961	193	5.2626902
1962	85029	4897	151	90077	11.40842014	1962	220	5.3936275
1963	85869	4968	145	90982	11.41841696	1963	194	5.2678582
1964	87592	5259	120	92971	11.4400429	1964	180	5.1929569
1965	94629	5413	108	100150	11.51442434	1965	146	4.9836066
1966	88525	4853	104	93482	11.44552418	1966	174	5.1590553
1967	85697	4744	103	90544	11.4135912	1967	144	4.9698133
1968	93471	5171	95	98737	11.50021503	1968	169	5.1298987
1969	98750	5496	111	104357	11.55557299	1969	165	5.1059455
1970	103175	5996	188	109359	11.60239133	1970	117	4.7621739
1971	104729	6211	155	111095	11.61814097	1971	135	4.9052748
1972	99298	5867	135	105300	11.5645687	1972	135	4.9052748
1973	104079	5425	118	109622	11.60479336	1973	160	5.0751738
1974	102538	5318	155	108011	11.58998835	1974	159	5.0689042
1975	101014	6292	150	107456	11.58483674	1975	136	4.9126549
1976	102344	7061	175	109580	11.60441016	1976	151	5.0172798
1977	100931	7258	188	108377	11.59337117	1977	192	5.2574954
1978	100916	7538	194	108648	11.59586858	1978	217	5.3798974
1979	100494	7519	196	108209	11.59181982	1979	203	5.313206
1980	104329	7830	220	112379	11.62963237	1980	386	5.9558374
1981	106413	7375	178	113966	11.64365544	1981	305	5.7203118
1982	109625	8174	188	117987	11.67832973	1982	274	5.6131281
1983	103703	8082	255	112040	11.62661123	1983	252	5.5294291
1984	111284	8739	253	120276	11.69754438	1984	284	5.6489742

Table 7: US Patent Data-2 (Source: USPTO, 2016)

1985	117006	9551	231	126788	11.75027168
1986	122433	9912	320	132665	11.79558243
1987	127917	11153	385	139455	11.84549725
1988	139825	11289	377	151491	11.9282815
1989	152750	12615	383	165748	12.01822384
1990	164558	11288	418	176264	12.07973815
1991	164306	13061	463	177830	12.08858332
1992	173075	13078	354	186507	12.13622405
1993	174743	13635	361	188739	12.14812039
1994	189857	15774	459	206090	12.23606825
1995	212377	15409	452	228238	12.33814422
1996	195187	15161	665	211013	12.25967502
1997	215257	16546	621	232424	12.35631857
1998	243062	17107	720	260889	12.47185031
1999	270187	17761	863	288811	12.57352777
2000	295926	18292	797	315015	12.66037554
2001	326508	18280	944	345732	12.75341919
2002	334445	20904	1144	356493	12.78406988
2003	342441	22602	1000	366043	12.81050609
2004	356943	23975	1221	382139	12.8535397
2005	390733	25553	1222	417508	12.94205898
2006	425967	25515	1151	452633	13.02283692
2007	456154	27752	1049	484955	13.09181138
2008	456321	27782	1209	485312	13.09254726
2009	456106	25806	959	482871	13.08750482
2010	490226	29059	992	520277	13.16211664
2011	503582	30467	1139	535188	13.19037337
2012	542815	32799	1149	576763	13.26518672

2.6 QLR Test Result for Finding Breaks of the US Patent Application Data

The US changed its patent policy toward pro-patent-ism in 1980s. Our Humans' viewpoints (left) could be statistically corrected to be gods' (right). This correction washes our illusions away and shows us the way. In the following diagram, we can assume that the US might change its mind to have pro-patent-ism from 1984 on, without otherwise proving data.

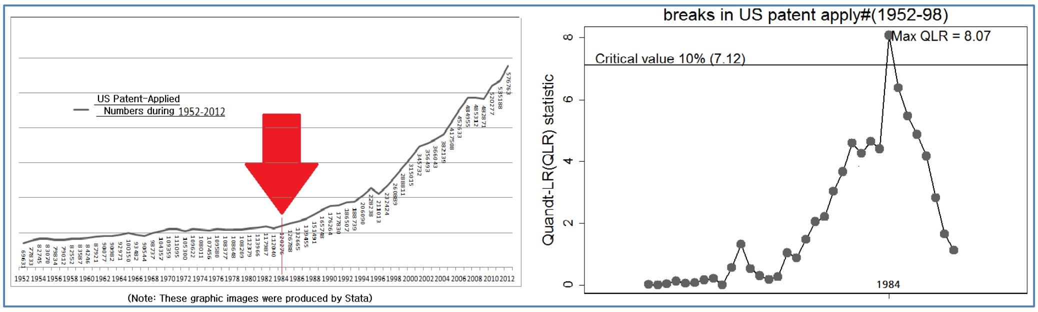


Figure 3: The Break Pont Found from the US Patent Application Data

2.7 QLR Test Result Interpretation

The above left graph hardly shows the exact break point. With the naked eyes, the year of 1996 seems like the break point. Of course, the year of 2009 might also be the break point. On the other hand, the right graph pins the break point of 1984 explicitly, compared to other points. No one can deny the fact that the year of 1984 is just the right break point from the above result.

The following table shows that the dummy variable "di" has the coefficient value 0.031728, 0.0320162, **0.0410059**, 0.038227 and 0.0373452 respectively in the year of 1982~1986. We have also the F-statistic 4.65, 4.40, **8.07**, 6.37 and 5.47 respectively in the year of 1982~1986.

Accordingly, from the year 1984 on, the increasing trend gets to be started, we can infer.

Table 8: The Break Pont Found from the US Patent Application Data

difflnuspap	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]	(1) di = 0
di	.031728	.0147124	2.16	0.037	.002077	.061379	F(1, 44) = 4.65 Prob > F = 0.0365
_cons	.0169893	.0075444	2.25	0.029	.0017845	.0321941	QLR of the year 1982 = 4.65

difflnuspap	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]	(1) di = 0
di	.0320162	.0152573	2.10	0.042	.0012672	.0627653	F(1, 44) = 4.40 Prob > F = 0.0416
_cons	.0175788	.007317	2.40	0.021	.0028324	.0323252	QLR of the year 1983 = 4.40

difflnuspap	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]	(1) di = 0
di	.0410059	.0144314	2.84	0.007	.0119213	.0700905	F(1, 44) = 8.07 Prob > F = 0.0068
_cons	.0153434	.0074294	2.07	0.045	.0003705	.0303163	QLR of the year 1984 = 8.07

difflnuspap	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]	(1) di = 0
di	.038227	.0151446	2.52	0.015	.007705	.0687489	F(1, 44) = 6.37 Prob > F = 0.0153
_cons	.0170806	.0074065	2.31	0.026	.0021538	.0320074	QLR of the year 1985 = 6.37

difflnuspap	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]	(1) di = 0
di	.0373452	.0159719	2.34	0.024	.0051559	.0695345	F(1, 44) = 5.47 Prob > F = 0.0240
_cons	.0181608	.0072639	2.50	0.016	.0035213	.0328003	QLR of the year 1986 = 5.47

3. How to Prove Balloon Effect from JP Dispute Resolution Simplification

3.1 Research Background

Japan discarded one of dual IP (industrial property) infringement resolution systems in 1960. The one was the trials (left) performed by the administrative judges of IP trial & appeal board and the other was the lawsuits (right) by the orthodox judges of district courts. Japan had experienced the radical increase of the other since 1960, just after the above abolishment.

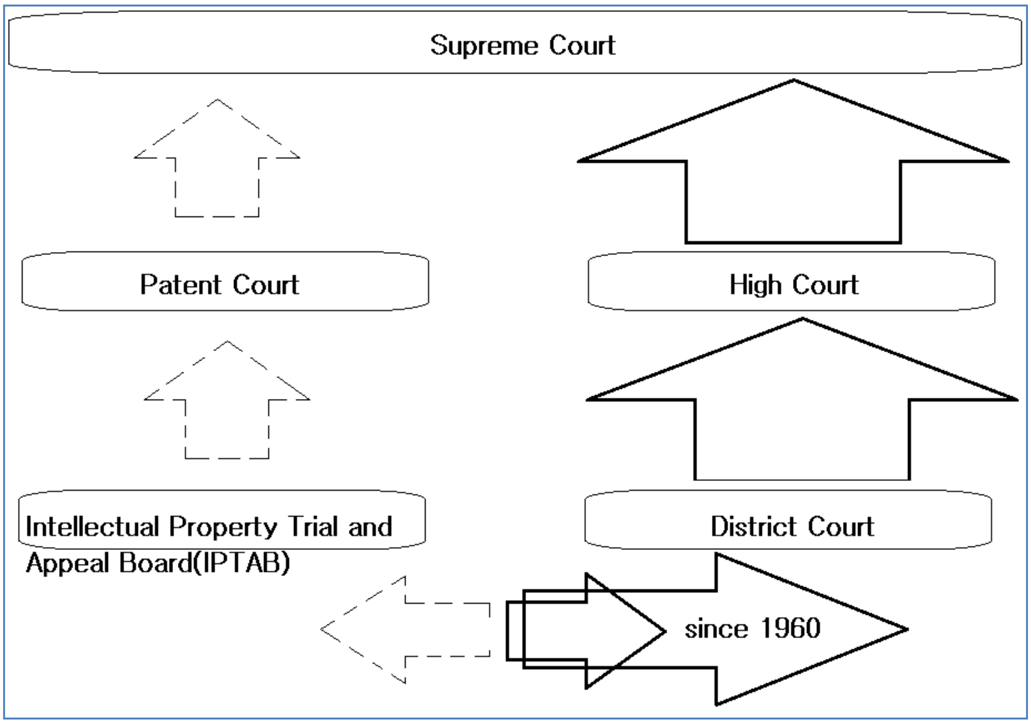


Figure 4: Balloon Effect from JP Dispute Resolution Simplification (Source: Kim, 2015)

3.2 Lag-Order Selection for Level Data or Log Data Thereof

At the left candidate of the following table, “d2.jpsap” means 2nd level difference data of the JP lawsuit application data “jpsap”. “d2.jpsap” has the lag of 3. The right “d2.lnjpsap” shows that 2nd level difference data of “lnjpsap” have the same lag 3, either. “lnjpsap” is the natural log data of “jpsap”.

Table 9: Lag Candidates from the JP Lawsuit Application Data

. varsoc d2.jpsap

Selection-order criteria
Sample: 1958 - 1998 Number of obs = 41

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-236.168				6193.76	11.5692	11.5844	11.611
1	-226.962	18.413	1	0.000	4150.79	11.1689	11.1993	11.2524
2	-222.301	9.3219	1	0.002	3472.6	10.9903	11.0359	11.1157
3	-218.749	7.1023*	1	0.008	3067.35*	10.8658*	10.9267*	11.033*
4	-218.707	.08424	1	0.772	3216	10.9126	10.9886	11.1215

Endogenous: D2.jpsap
Exogenous: _cons

. varsoc d2.lnjpsap (Note:These graphic images were produced by Stata)

Selection-order criteria
Sample: 1958 - 1998 Number of obs = 41

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-18.3692				.150615	.944841	.96006	.986636
1	-7.54052	21.657	1	0.000	.093256	.465391	.49583	.54898
2	-3.95262	7.1758	1	0.007	.082211	.339152	.38481	.464536
3	-.972505	5.9602*	1	0.015	.074669*	.242561*	.303438*	.409739*
4	-.938502	.06801	1	0.794	.078318	.289683	.365779	.498655

Endogenous: D2.lnjpsap
Exogenous: _cons

3.3 Heteroskedasticity Tests for Models

In the following diagram, both of the left “L(1/3).diff2.jpsap” and the right “L(1/3).diff2.lnjpsap” have the normal distribution, which is the reason why we might take either of them. As a final we would rather choose the left “L(1/3).diff2.jpsap” for simplification.

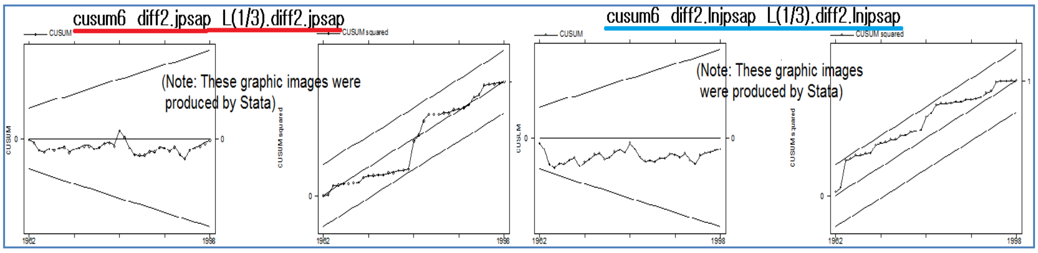


Figure 5: Heteroskedasticity Tests about the JP Lawsuit Application Data

3.4 QLR Test Coding for Finding Breaks of the JP Lawsuit Application Data

The following coding contents, both of “local var = "jpsap"” and “gen diff2`var' = d2.`var'” are combined with each other, ending up to become the 2nd level difference data from JP Lawsuit application data.

The following coding content, “qui reg diff2`var' L(1/3).diff2`var' di d_`var'1 d_`var'2 d_`var'3,r” means that there are 7 independent variables. “L(1/3).diff2`var'” consists of L1.d2.jpsap, L2.d2.jpsap and L3.d2.jpsap due to gen diff2`var' = d2.`var'.

3 of “cap gen d_`var'1 = di*l1.diff2`var'”, “cap gen d_`var'2 = di*l2.diff2`var'” and “cap gen d_`var'3 = di*l3.diff2`var'” make multiplication of each of L1.d2.jpsap, L2.d2.jpsap and L3.d2.jpsap by the dummy variable “di”. “local critical=5.12” and “Critical value 1% ('critical')” are required for adaption to data characteristics.

```
use 52-98-jp-us.dta,clear
log using kby14-L3-d2-jpsap-QLR.log,replace

tset year
sum year

local time=r(max)-r(min)+1
local i = round(`time'*.15)
local f = round(`time'*.85)
local var = "jpsap"
gen diff2`var' = d2.`var'
gen qlr`var' = .
set more off
while `i'<=(`f') {
cap gen di = (_n >=`i')

cap gen d_`var'1 = di*l1.diff2`var'
cap gen d_`var'2 = di*l2.diff2`var'
cap gen d_`var'3 = di*l3.diff2`var'

list year `var' d.`var' diff2`var' L(1,3).diff2`var' di d_`var'1
```

```
qui reg diff2`var' L(1/3).diff2`var' di d_`var'1 d_`var'2 d_`var'3,r
/*estat  ovtest*/
qui test di d_`var'1 d_`var'2 d_`var'3

cap replace qlr`var' = r(F) in `i'

dis "`i', QLR of the year " %ty year[`i'] " =" %6.2f qlr`var'[`i'] " [see above table]"
drop di d_`var'1 d_`var'2 d_`var'3
local i = `i' + 1
}
sum qlr`var'
local maxvalue=r(max)
gen maxdate=year if round(qlr`var',0.01)==round(`maxvalue',0.01)
local maxvalue1=round(`maxvalue',0.01)
local critical=5.12 /*Replace with the appropriate critical value (see Stock & Watson)*/
sum year
local mindate=r(min)
sum maxdate
local maxdate=r(max)

gen break=year if qlr`var'>=`critical' & qlr`var'!=.
dis "Below are the break dates..."
list year qlr`var' if break!=.
levelsof break, local(break1)

/*set more on*/
twoway connected qlr`var' year,title(breaks in JP IP lawsuit apply#(1952-98)) ///
xlabel(`break1', angle(90) labsize(3) alternate) ///
yline(`critical') ytitle(Quandt-LR(QLR)statistic) xtitle(Time) ///
ttext(`critical' `mindate' "Critical value 1% (`critical')", placement(ne)) ///
ttext(`maxvalue' `maxdate' "Max QLR = `maxvalue1'", placement(e)) saving(ts3,replace)
scheme(s1manual)
more
twoway (tsline diff2`var' L(1/3).diff2`var') lfit diff2`var' year,saving(ts4,replace)
scheme(s1manual)
more
twoway (tsline `var') lfit `var' year,saving(ts5,replace) scheme(s1manual)
more
graph combine ts3.gph ts4.gph ts5.gph,col(1) xsize(10) ysize(18)
```

3.5 QLR Test Result for Finding Breaks of the JP Lawsuit Application Data

In the following diagram, from the left side we cannot easily recognize the break, in other words, the structural change. However, from the right side we can pin point the year of 1960, which can be detected through the above statistical processes and called as a balloon effect.

Korea has still kept the same dual IP(industrial property) infringement resolution systems running to be maintained to adapt to the increasing dispute cases, as Japan's old fashioned ones. However, there is also a criticism against the overlap problem similar to

Japan's. Accordingly, this balloon effect phenomenon is anticipated to take place likewise in Korea. And so, Korea had better be prepared against the anticipated problems about the capacity or the ability of the related systems.

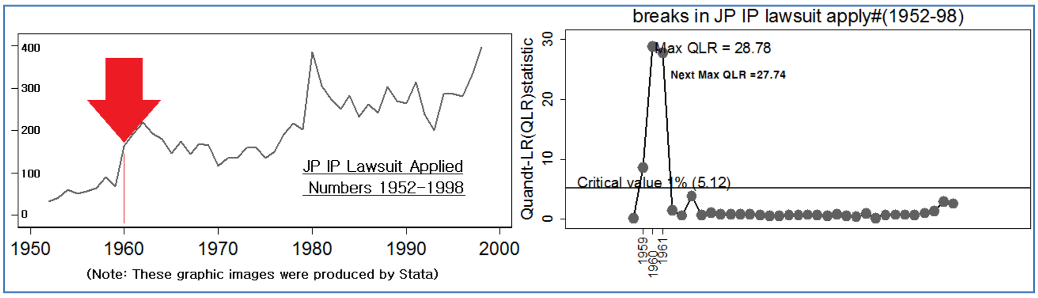


Figure 6: Balloon Effect from the JP Lawsuit Application Data

3.6 QLR Test Result Interpretation

The following two tables show that the dummy variable "di" and the multiplication results of the independent variables by the dummy variable "di" have the coefficient values (0, 0, -0.1287859, 0), (0, 0, 0.1243916, 0.7960382), **(0, 0.4760758, 0.4047516, 0.9737689)**, (-15.42232, 1.562005, 0.351951, 0.570054) and (-27.05154, 0.2839765, 0.1562817, 0.1193771) respectively in the year of 1958~1962.

The coefficient sums of the above variables have the -0.1287859, 0.9204298, **1.8545963**, -12.93831 and -26.4919047 respectively in the year of 1958~1962. We have also the F-statistic 0.12, 8.64, **28.78**, 27.74 and 1.45 respectively in the year of 1958~1962. Accordingly, from the year 1960 on, the increasing trend gets to be started, we can infer.

In the two following tables "di", "d_jpsap1" and "d_jpsap3" were omitted due to collinearities. We can imagine that the omitted variables had also their own roles but due to the collinearities they failed to do so and so had zero coefficient values.

Anyway, the F-statistic of the year 1960 shows that it had the highest break or change causes, compared with other years. And so we can reject the H0: di=d_jpsap1= d_jpsap2= d_jpsap3=0 affirmatively.

Table 10-1: The Break in the JP Lawsuit Application Data - 1

diff2jpsap	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
diff2jpsap						
L1.	-1.066111	.2010208	-5.30	0.000	-1.473418	-.6588038
L2.	-.69304	.2957659	-2.34	0.025	-1.292319	-.0937613
L3.	-.4080003	.1070799	-3.81	0.001	-.6249648	-.1910358
di	(omitted)	because of collinearity				
d_jpsap1	(omitted)					
d_jpsap2	-.1287859	.3699701	-0.35	0.730	-.8784166	.6208448
d_jpsap3	(omitted)					
_cons	2.477473	8.416626	0.29	0.770	-14.57623	19.53118

(1) o.di = 0
(2) o.d_jpsap1 = 0
(3) d_jpsap2 = 0
(4) o.d_jpsap3 = 0
Constraint 1 dropped
Constraint 2 dropped
Constraint 4 dropped

F(1, 37) = 0.12
Prob > F = 0.7297
QLR of the year 1958 = 0.12

diff2jpsap	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
diff2jpsap						
L1.	-1.066431	.2030335	-5.25	0.000	-1.478202	-.6546598
L2.	-.9462012	.4719471	-2.00	0.053	-1.903354	.0109518
L3.	-1.201815	.5587652	-2.15	0.038	-2.335043	-.0685862
di	(omitted)	because of collinearity				
d_jpsap1	(omitted)					
d_jpsap2	.1243916	.5457029	0.23	0.821	-.9823451	1.231128
d_jpsap3	.7960382	.5887287	1.35	0.185	-.3979588	1.990035
_cons	1.997116	8.719631	0.23	0.820	-15.68712	19.68135

(1) o.di = 0
(2) o.d_jpsap1 = 0
(3) d_jpsap2 = 0
(4) d_jpsap3 = 0
Constraint 1 dropped
Constraint 2 dropped

F(2, 36) = 8.64
Prob > F = 0.0009
QLR of the year 1959 = 8.64

Table 10-2: The Break in the JP Lawsuit Application Data - 2

diff2jpsap	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
diff2jpsap						
L1.	-1.537343	.7207316	-2.13	0.040	-3.000506	-.0741803
L2.	-1.22009	.0411529	-29.65	0.000	-1.303635	-1.136545
L3.	-1.374655	.2747233	-5.00	0.000	-1.932373	-.8169366
di	(omitted)	because of collinearity				
d_jpsap1	.4760758	.6925794	0.69	0.496	-.9299351	1.882087
d_jpsap2	.4047516	.2393579	1.69	0.100	-.0811708	.8906739
d_jpsap3	.9737689	.3182484	3.06	0.004	.3276903	1.619848
_cons	2.577615	9.003584	0.29	0.776	-15.70063	20.85586

(1) o.di = 0
(2) d_jpsap1 = 0
(3) d_jpsap2 = 0
(4) d_jpsap3 = 0
Constraint 1 dropped

F(3, 35) = 28.78
Prob > F = 0.0000
QLR of the year 1960 = 28.78

diff2jpsap	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
diff2jpsap						
L1.	-2.602005	3.31e-07	-7.9e+06	0.000	-2.602005	-2.602004
L2.	-1.159299	8.29e-07	-1.4e+06	0.000	-1.159301	-1.159297
L3.	-.9688347	1.17e-07	-8.3e+06	0.000	-.9688349	-.9688344
di	-15.42232	9.101929	-1.69	0.099	-33.91966	3.075027
d_jpsap1	1.562005	.2095348	7.45	0.000	1.136179	1.98783
d_jpsap2	.351951	.22568	1.56	0.128	-.1066859	.8105879
d_jpsap3	.570054	.1103674	5.17	0.000	.3457603	.7943476
_cons	15.87767	.	.	.	.	.

(1) di = 0
(2) d_jpsap1 = 0
(3) d_jpsap2 = 0
(4) d_jpsap3 = 0

F(4, 34) = 27.74
Prob > F = 0.0000
QLR of the year 1961 = 27.74

diff2jpsap	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
diff2jpsap						
L1.	-1.342749	.5961572	-2.25	0.031	-2.554287	-.1312122
L2.	-.9738935	1.476816	-0.66	0.514	-3.975145	2.027358
L3.	-.5250073	1.13624	-0.46	0.647	-2.834124	1.784109
di	-27.05154	20.47928	-1.32	0.195	-68.67044	14.56736
d_jpsap1	.2839765	.6346832	0.45	0.657	-1.005855	1.573808
d_jpsap2	.1562817	1.494099	0.10	0.917	-2.880093	3.192656
d_jpsap3	.1193771	1.141668	0.10	0.917	-2.200772	2.439526
_cons	26.81806	18.28393	1.47	0.152	-10.33936	63.97548

(1) di = 0
(2) d_jpsap1 = 0
(3) d_jpsap2 = 0
(4) d_jpsap3 = 0

F(4, 34) = 1.45
Prob > F = 0.2392
QLR of the year 1962 = 1.45

4. How to Prove the KR Patent Scope Enlargement

4.1 Lag-Order Selection for Level Data or Log Data Thereof

4.1.1 Lag-Order Selection for KR Patent Application Total Data

"kr" is the abbreviation of "Korea". "pap" stands for patent application. "t" equals to total. "krpapt" means KR patent application total data and has the lag 3. "d1.krpapt" is the 1st level difference data of "krpapt" and has the lag 2.

Table 11: Lag Candidates from the KR Patent Application Total Data

. varsoc krpapt /*candidate*/

Selection-order criteria
Sample: 1984 - 2012 Number of obs = 29

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-360.68				4.0e+09	24.9435	24.9582	24.9906
1	-305.8	109.76	1	0.000	9.7e+07	21.2276	21.2571	21.3219
2	-304.31	2.979	1	0.084	9.4e+07	21.1938	21.2381	21.3353
3	-302.04	4.5409*	1	0.033	8.6e+07*	21.1062*	21.1653*	21.2948*
4	-301.74	.60015	1	0.439	9.0e+07	21.1545	21.2283	21.3902

Endogenous: krpapt
Exogenous: _cons

. varsoc d1.krpapt /*candidate*/

Selection-order criteria
Sample: 1985 - 2012 Number of obs = 28

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-295.674				9.3e+07	21.191	21.2056	21.2386
1	-294.279	2.7914	1	0.095	9.1e+07	21.1628	21.1918	21.2579
2	-292.069	4.4186*	1	0.036	8.3e+07*	21.0764*	21.12*	21.2191*
3	-291.809	.52087	1	0.470	8.8e+07	21.1292	21.1874	21.3195
4	-291.697	.22396	1	0.636	9.4e+07	21.1926	21.2654	21.4305

Endogenous: D.krpapt
Exogenous: _cons

4.1.2 Lag-Order Selection for KR Patent Registration Total Data

"kr" is the abbreviation of "Korea". "pre" stands for patent registration. "t" equals to total. "krpret" means KR patent registration total data. "d2.krpret" is the 2nd level difference data of "krpret" and has the lag 2. "lnkrpret" is the natural log data of "krpret" and has the lag 3.

Table 12: Lag Candidates from the KR Patent Registration Total Data

. varsoc krpret /*candidate*/

Selection-order criteria
Sample: 1984 - 2012 Number of obs = 29

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-346.797				1.5e+09	23.986	24.0008	24.0331
1	-322.479	48.636	1	0.000	3.1e+08	22.3779	22.4074	22.4722
2	-319.798	5.3619	1	0.021	2.7e+08	22.2619	22.3062	22.4034
3	-310.733	18.131*	1	0.000	1.6e+08*	21.7057*	21.7648*	21.8943*
4	-310.694	.07651	1	0.782	1.7e+08	21.772	21.8459	22.0078

Endogenous: krpret
Exogenous: _cons

. varsoc d2.krpret /*candidate*/

Selection-order criteria
Sample: 1986 - 2012 Number of obs = 27

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-304.523				3.9e+08	22.6313	22.6456	22.6793
1	-304.283	.4804	1	0.488	4.2e+08	22.6876	22.7162	22.7836
2	-296.157	16.251*	1	0.000	2.5e+08*	22.1598*	22.2026*	22.3038*
3	-296.002	.30959	1	0.578	2.6e+08	22.2224	22.2795	22.4144
4	-294.274	3.458	1	0.063	2.5e+08	22.1684	22.2398	22.4084

Endogenous: D2.krpret
Exogenous: _cons

. varsoc lnkrpret /*candidate*/

Selection-order criteria
Sample: 1984 - 2012 Number of obs = 29

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-49.9431				1.96495	3.51332	3.52808	3.56047
1	-5.84914	88.188	1	0.000	.100627	.54132	.570853	.635616
2	-3.98111	3.7361	1	0.053	.094829	.481456	.525755	.6229
3	-1.00458	5.9531*	1	0.015	.08283*	.345144*	.404209*	.533736*
4	-.98519	.03879	1	0.844	.088777	.412772	.486603	.648512

Endogenous: lnkrpret
Exogenous: _cons

4.2 Heteroskedasticity Tests for Models

4.2.1 Heteroskedasticity Tests for KR Patent Application Total Data

"L(1/3).krpapt" is better than "L(1/2).d1.krpapt" in terms of simplification, even though both of them have the normal distribution.

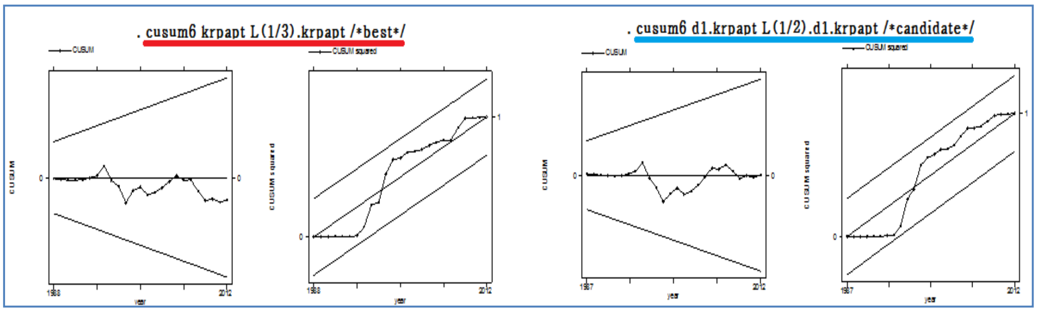


Figure 7: Heteroskedasticity Tests about the KR Patent Application Total Data

4.2.2 Heteroskedasticity Tests for KR Patent Registration Total Data

Both of "L(1/3).krpret" and "L(1/2).d2.krpret" have the out-of-range abnormalities, which are the reason why they should be discarded. The rightmost "L(1/3).lnkrpret", the 1st level difference data from the natural log of the KR patent registration data, is the only one and best candidate.

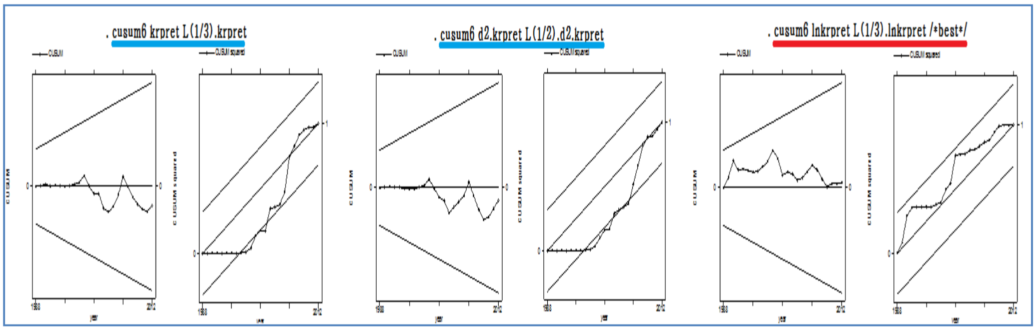
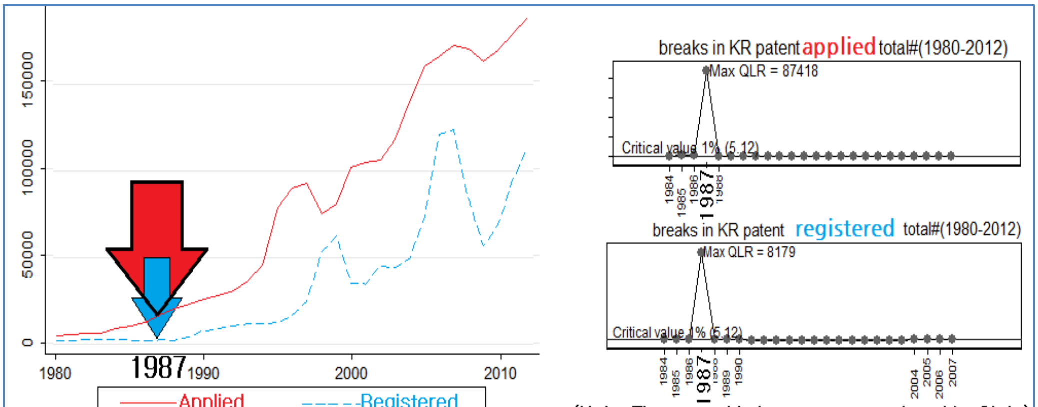


Figure 8: Heteroskedasticity Tests about the KR Patent Registration Total Data

4.3 QLR Test Result for Finding Breaks

Korea adopted material (composition of matter) patent in 1987. The following break value graphs (right side graphs) could be the gods' viewpoints.



(Note: These graphic images were produced by Stata)

Figure 9: Patent Scope Enlargement Proof

4.4 QLR Test Result Interpretation

The following two tables show that the dummy variable "di" and the multiplication results of the independent variables by the dummy variable "di" have the coefficient values (0, 11.62872, -11.99039, 0), (0, -0.4296419, -5.741115, 7.654412), **(55796.99, 4.526771, -1.606583, -13.36916)**, (10422.11, 0.6126575, -1.041757, -0.5218256) and (9095.736, 1.016916, -1.881677, 0.2382782) respectively in the year of 1985~1989.

The coefficient sums of the above variables have the -0.36167, 1.913297, **55786.541028**, -0.9509251, 10421.1590749 and 9095.1095172 respectively in the year of 1985~1989. We have also the F-statistic 381.21, 598.18, **87418.10**, 55.79 and 1.98 respectively in the year of 1985~1989.

In the following table "di" and "d_krpapt3" were omitted due to collinearities. We can imagine that the omitted variables had also their own roles but due to the collinearities they failed to do so and so had zero coefficient values. Anyway, the F-statistic of the year 1987 shows that it had the highest break or change causes, compared with other years. And so we can reject the H0: di=d_jpsap1= d_jpsap2= d_jpsap3=0 affirmatively.

Table 13-1: Extents of Changing in the KR Patent Application Total Data - 1

krpapt	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
krpapt						
L1.	-10.20014	1.300727	-7.84	0.000	-12.88471	-7.515577
L2.	11.1796	.9798191	11.41	0.000	9.157356	13.20185
L3.	.3885592	.138158	2.81	0.010	.1034151	.6737034
di	(omitted)	because of collinearity				
d_krpapt1	11.62872	1.307797	8.89	0.000	8.929561	14.32788
d_krpapt2	-11.99039	1.031695	-11.62	0.000	-14.11971	-9.86108
d_krpapt3	(omitted)					
_cons	5564.225	2472.282	2.25	0.034	461.6853	10666.77

(1) o.di = 0
(2) d_krpapt1 = 0
(3) d_krpapt2 = 0
(4) o.d_krpapt3 = 0
Constraint 1 dropped
Constraint 4 dropped

F(2, 24) = 381.21
Prob > F = 0.0000
QLR of the year 1985 =381.21

krpapt	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
krpapt						
L1.	1.852196	.2486081	7.45	0.000	1.337911	2.366481
L2.	4.928107	.2069104	23.82	0.000	4.50008	5.356134
L3.	-7.261317	1.052949	-6.90	0.000	-9.439508	-5.083125
di	(omitted)	because of collinearity				
d_krpapt1	-.4296419	.3249736	-1.32	0.199	-1.101901	.2426172
d_krpapt2	-5.741115	.3429876	-16.74	0.000	-6.450639	-5.031591
d_krpapt3	7.654412	1.055465	7.25	0.000	5.471016	9.837807
_cons	6102.713	2768.452	2.20	0.038	375.733	11829.69

(1) o.di = 0
(2) d_krpapt1 = 0
(3) d_krpapt2 = 0
(4) d_krpapt3 = 0
Constraint 1 dropped

F(3, 23) = 598.18
Prob > F = 0.0000
QLR of the year 1986 =598.18

Table 13-2: Extents of Changing in the KR Patent Application Total Data - 2

krpapt	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
krpapt						
L1.	-3.111796	.000568	-5478.10	0.000	-3.112974	-3.110618
L2.	.7966992	3.62e-11	2.2e+10	0.000	.7966992	.7966992
L3.	13.76306	.	.	.	.	.
di	55796.99	3145.591	17.74	0.000	49273.44	62320.55
d_krpapt1	4.526771	.1858222	24.36	0.000	4.141399	4.912143
d_krpapt2	-1.606583	.2851746	-5.63	0.000	-2.197999	-1.015167
d_krpapt3	-13.36916	.1403966	-95.22	0.000	-13.66033	-13.078
_cons	-49175.35	6.541878	-7517.01	0.000	-49188.92	-49161.78

(1) di = 0
(2) d_krpapt1 = 0
(3) d_krpapt2 = 0
(4) d_krpapt3 = 0

F(4, 22) =87418.10
Prob > F = 0.0000
QLR of the year 1987 =87418.10

krpapt	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
krpapt						
L1.	.7973684	.4184988	1.91	0.070	-.0705451	1.665282
L2.	.2332223	.4664091	0.50	0.622	-.734051	1.200496
L3.	.9168232	.2334601	3.93	0.001	.4326565	1.40099
di	10422.11	3645.238	2.86	0.009	2862.345	17981.87
d_krpapt1	.6126575	.4578309	1.34	0.195	-.3368256	1.562141
d_krpapt2	-1.041757	.5460316	-1.91	0.070	-2.174157	.0906436
d_krpapt3	-.5218256	.2717118	-1.92	0.068	-1.085321	.0416702
_cons	-3439.348	765.9891	-4.49	0.000	-5027.912	-1850.784

(1) di = 0
(2) d_krpapt1 = 0
(3) d_krpapt2 = 0
(4) d_krpapt3 = 0

F(4, 22) = 55.79
Prob > F = 0.0000
QLR of the year 1988 = 55.79

krpapt	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
krpapt						
L1.	.3872796	.6057773	0.64	0.529	-.8690256	1.643585
L2.	1.0698	.8264422	1.29	0.209	-.6441358	2.783737
L3.	.1604572	.5468334	0.29	0.772	-.9736059	1.29452
di	9095.736	4305.862	2.11	0.046	165.9239	18025.55
d_krpapt1	1.016916	.6336269	1.60	0.123	-.297146	2.330977
d_krpapt2	-1.881677	.8734651	-2.15	0.042	-3.693133	-.0702211
d_krpapt3	.2382782	.563537	0.42	0.677	-.930426	1.406982
_cons	-1344.019	1620.803	-0.83	0.416	-4705.359	2017.321

(1) di = 0
(2) d_krpapt1 = 0
(3) d_krpapt2 = 0
(4) d_krpapt3 = 0

F(4, 22) = 1.98
Prob > F = 0.1333
QLR of the year 1989 = 1.98

4.5 The Source Data for Finding Breaks

"krpapd" and "krpapf" stands for KR patent application domestic and foreign data. "lnkrpapf" is the natural log of "krpapf". "krpref" and "krpred" stands for KR patent registration foreign and domestic data. "lnkrpred" is the natural log of "krpred". "krpapt" and "krpret" stands for KR patent application total and registration total data.

Table 14: KR Patent Application and Registration Data (Source: KIPO, 2016)

year	krpapd	krpapf	lnkrpapf		krpref	krpred	lnkrpred		krpapt	krpret	lnkrpret
1980	1241	3829	8.250359		1446	186	5.225747		5070	1632	7.397562
1981	1319	3984	8.290042		1576	232	5.446737		5303	1808	7.499977
1982	1556	4368	8.38206		2335	274	5.613128		5924	2609	7.866722
1983	1599	4795	8.475329		2188	245	5.501258		6394	2433	7.79688
1984	2014	6619	8.7977		2068	297	5.693732		8633	2365	7.768533
1985	2703	7884	8.97259		1919	349	5.855072		10587	2268	7.726654
1986	3641	9118	9.118006		1436	458	6.126869		12759	1894	7.546446
1987	4871	12191	9.408453		1734	596	6.390241		17062	2330	7.753623
1988	5696	14355	9.571854		1599	575	6.35437		20051	2174	7.684324
1989	7021	16294	9.698552		2791	1181	7.074117		23315	3972	8.287025
1990	9082	16738	9.725437		5208	2554	7.845416		25820	7762	8.956995
1991	13253	14879	9.607706		6137	2553	7.845025		28132	8690	9.069928
1992	15952	15121	9.623839		6932	3570	8.180321		31073	10502	9.259321
1993	21459	15032	9.617936		6901	4545	8.421783		36491	11446	9.345396
1994	28564	17148	9.749637		5909	5774	8.66112		45712	11683	9.365891
1995	59236	19263	9.865941		5937	6575	8.79103		78499	12512	9.434443
1996	68413	21913	9.994835		8195	8321	9.026538		90326	16516	9.712085
1997	67346	25388	10.14203		10082	14497	9.581697		92734	24579	10.10965
1998	50596	24592	10.11018		17000	35900	10.48849		75188	52900	10.87616
1999	55970	24672	10.11342		19321	43314	10.67623		80642	62635	11.04508
2000	72831	29179	10.2812		12013	22943	10.04077		102010	34956	10.46185
2001	73714	30898	10.33845		12842	21833	9.991178		104612	34675	10.45377
2002	76570	29566	10.29438		15123	30175	10.31477		106136	45298	10.72102
2003	90313	28339	10.25199		13640	30525	10.3263		118652	44165	10.69569
2004	105250	34865	10.45924		13784	35284	10.47118		140115	49068	10.80096
2005	122188	38733	10.56445		20093	53419	10.88592		160921	73512	11.2052
2006	125476	40713	10.6143		31487	89303	11.39979		166189	120790	11.70181
2007	128701	43768	10.68666		32060	91645	11.42568		172469	123705	11.72565
2008	127114	43518	10.68093		22408	61115	11.02051		170632	83523	11.33288
2009	127316	36207	10.49701		14603	42129	10.64849		163523	56732	10.94609
2010	131805	38296	10.5531		17439	51404	10.84747		170101	68843	11.13958
2011	138034	40890	10.61864		22462	72258	11.188		178924	94720	11.45868
2012	148136	40779	10.61592		29406	84061	11.3393		188915	113467	11.63927

4.6 QLR Test Coding for Finding Breaks

4.6.1 QLR Test Coding Contents for KR Patent Application Total Data

The following coding content, “qui reg 'var' L(1/3).'var' di d_'var'1 d_'var'2 d_'var'3,r” means that there are 7 independent variables. “L(1/3).`var'” consists of L1.krpapt, L2.krpapt and L3.krpapt, because of “local var = "krpapt"”.

“cap gen d_`var'1 = di*l1.`var'”, “cap gen d_`var'2 = di*l2.`var'” and “cap gen d_`var'3 = di*l3.`var'” make multiplication of each of L1.krpapt, L2.krpapt and L3. krpapt by the dummy variable “di”.

```
use 80-12-krp.dta,clear
log using kby47-L3-krpapt-QLR.log,replace
tset year
sum year
local time=r(max)-r(min)+1
local i = round('time'*.15)
local f = round('time'*.85)

local var = "krpapt"

gen qlr'var' = .
set more off
while 'i'<=('f') {
cap gen di = (_n>='i')

cap gen d_'var'1 = di*l1.'var'
cap gen d_'var'2 = di*l2.'var'
cap gen d_'var'3 = di*l3.'var'

list year 'var' L(1,3).'var' di d_'var'3
qui reg 'var' L(1/3).'var' di d_'var'1 d_'var'2 d_'var'3,r

qui test di d_'var'1 d_'var'2 d_'var'3

cap replace qlr'var' = r(F) in 'i'

dis "'i', QLR of the year " %ty year['i'] " =" %6.2f qlr'var'['i'] " [see above table]"
drop di d_'var'1 d_'var'2 d_'var'3
local i = 'i' + 1
}
/* skip due to overlap by the last coding content */
```

4.6.2 QLR Test Coding Contents for KR Patent Registration Total Data

The following coding content, “qui reg 'var' L(1/3).'var' di d_'var'1 d_'var'2 d_'var'3,r” means that there are 7 independent variables. “L(1/3).`var'” consists of L1.lnkrpret, L2.lnkrpret and L3.lnkrpret, because of “local var = "lnkrpret"”.

“cap gen d_`var'1 = di*l1.`var'”, “cap gen d_`var'2 = di*l2.`var'” and “cap gen d_`var'3 = di*l3.`var'” make multiplication of each of L1. lnkrpret, L2. lnkrpret and L3. Lnkrpret by the dummy variable “di”.

```
use 80-12-krp.dta,clear
log using kby57-L3-lnkrpret-QLR.log,replace
tset year
sum year

local time=r(max)-r(min)+1
local i = round('time'*.15)
local f = round('time'*.85)

local var = "lnkrpret"

gen qlr'var' = .
set more off
while 'i'<=('f') {
cap gen di = (_n>='i')

cap gen d_'var'1 = di*l1.'var'
cap gen d_'var'2 = di*l2.'var'
cap gen d_'var'3 = di*l3.'var'

list year 'var' L(1,3).'var' di d_'var'3

qui reg 'var' L(1/3).'var' di d_'var'1 d_'var'2 d_'var'3,r

qui test di d_'var'1 d_'var'2 d_'var'3

cap replace qlr'var' = r(F) in 'i'

dis "'i', QLR of the year " %ty year['i'] " =" %6.2f qlr'var'['i'] " [see above table]"
drop di d_'var'1 d_'var'2 d_'var'3
local i = 'i' + 1
}
/* skip due to overlap by the last coding content */
```

5. Who Wins from KR Patent Scope Enlargement?

5.1 Lag-Order Selection for Level Data or Log Data Thereof

5.1.1 Lag-Order Selection for KR Patent Application Foreign Data

“kr” is the abbreviation of “Korea”. “pap” stands for patent application. “f” equals to foreign. Accordingly, “krpapf” means KR patent application foreign data. “d2.krpapf” is the 2nd level difference data of “krpapf” and has the lag 2. “d4.krpapf” is the 4th level difference data of “krpapf” and has the lag 3. “lnkrpapf” is the natural log data of “krpapf” and has the lag 1.

Table 15: Lag Candidates from the KR Patent Application Foreign Data

. varsoc d2.krpapf /*candidate*/

Selection-order criteria
Sample: 1986 - 2012 Number of obs = 27

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-257.612				1.2e+07	19.1565	19.1707	19.2045
1	-256.848	1.529	1	0.216	1.2e+07	19.1739	19.2025	19.2699
2	-252.377	8.9418*	1	0.003	9.6e+06*	18.9168*	18.9596*	19.0608*
3	-252.111	.53105	1	0.466	1.0e+07	18.9712	19.0283	19.1632
4	-251.847	.52889	1	0.467	1.1e+07	19.0257	19.0971	19.2657

Endogenous: D2.krpapf
Exogenous: _cons

. varsoc d4.krpapf /*candidate*/

Selection-order criteria
Sample: 1988 - 2012 Number of obs = 25

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-263.502				9.1e+07	21.1602	21.1737	21.2089
1	-258.934	9.137	1	0.003	6.8e+07	20.8747	20.9017	20.9722
2	-250.87	16.127	1	0.000	3.9e+07	20.3096	20.3502	20.4559
3	-245.599	10.542*	1	0.001	2.8e+07*	19.9679*	20.022*	20.163*
4	-245.196	.80703	1	0.369	2.9e+07	20.0157	20.0833	20.2594

Endogenous: D4.krpapf
Exogenous: _cons

. varsoc lnkrpapf /*candidate*/

Selection-order criteria
Sample: 1984 - 2012 Number of obs = 29

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-22.539				.296866	1.62338	1.63814	1.67052
1	29.3922	103.86*	1	0.000	.008855*	-1.88912*	-1.85958*	-1.79482*
2	29.7863	.78819	1	0.375	.009238	-1.84733	-1.80303	-1.70588
3	30.6229	1.6734	1	0.196	.009352	-1.83607	-1.777	-1.64747
4	30.6239	.00193	1	0.965	.010036	-1.76717	-1.69334	-1.53143

Endogenous: lnkrpapf
Exogenous: _cons

5.1.2 Lag-Order Selection for KR Patent Registration Foreign Data

"kr" is the abbreviation of "Korea". "pre" stands for patent registration. "f" equals to foreign. Accordingly, "krpref" means KR patent registration foreign data, which has the lag 3. "d4.krpref" is the 4th level difference data of "krpref" has the lag 4.

Table 16: Lag Candidates from the KR Patent Registration Foreign Data

. varsoc krpref /*temporary best*/

Selection-order criteria
Sample: 1984 - 2012 Number of obs = 29

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-305.068				8.6e+07	21.1081	21.1229	21.1553
1	-282.834	44.467	1	0.000	2.0e+07	19.6438	19.6733	19.7381
2	-280.262	5.1455	1	0.023	1.8e+07	19.5353	19.5796	19.6767
3	-273.414	13.696*	1	0.000	1.2e+07*	19.132*	19.191*	19.3206*
4	-273.253	.32051	1	0.571	1.3e+07	19.1899	19.2637	19.4256

Endogenous: krpref
Exogenous: _cons

. varsoc d4.krpref /*best*/

Selection-order criteria
Sample: 1988 - 2012 Number of obs = 25

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-266.412				1.1e+08	21.393	21.4065	21.4417
1	-265.108	2.6093	1	0.106	1.1e+08	21.3686	21.3957	21.4661
2	-258.755	12.706	1	0.000	7.3e+07	20.9404	20.981	21.0867
3	-255.826	5.8576	1	0.016	6.3e+07	20.7861	20.8402	20.9811
4	-251.038	9.5772*	1	0.002	4.6e+07*	20.483*	20.5506*	20.7268*

Endogenous: D4.krpref
Exogenous: _cons

5.1.3 Lag-Order Selection for KR Patent Application Domestic Data

"kr" is the abbreviation of "Korea". "pap" stands for patent application. "d" equals to domestic. Accordingly, "krpapd" means KR patent application domestic data. "d3.krpapd" is the 3rd level difference data of "krpapd" and has the lag 4.

"d2.lnkrpapd" is the 2nd level difference data of "lnkrpapd" and has the lag 3. "lnkrpapd" is the natural log data of "krpapd".

Table 17: Lag Candidates from the KR Patent Application Domestic Data

. varsoc d3.krpapd /*best*/

Selection-order criteria
Sample: 1987 - 2012 Number of obs = 26

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-287.256				2.5e+08	22.1735	22.1875	22.2219
1	-284.397	5.718	1	0.017	2.2e+08	22.0305	22.0584	22.1273
2	-283.272	2.2493	1	0.134	2.1e+08	22.0209	22.0627	22.1661
3	-278.63	9.2846	1	0.002	1.6e+08	21.7408	21.7965	21.9343
4	-275.9	5.4592*	1	0.019	1.4e+08*	21.6077*	21.6774*	21.8497*

Endogenous: D3.krpapd
Exogenous: _cons

. varsoc d2.lnkrpapd /* candidate */

Selection-order criteria
Sample: 1986 - 2012 Number of obs = 27

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	6.20985				.039805	-.385915	-.371644	-.337921
1	7.14315	1.8666	1	0.172	.040012	-.380974	-.352432	-.284986
2	7.43574	.58517	1	0.444	.042192	-.328573	-.28576	-.184591
3	12.7592	10.647*	1	0.001	.030669*	-.648828*	-.591743*	-.456852*
4	12.8364	.15442	1	0.694	.032908	-.580473	-.509117	-.340503

Endogenous: D2.lnkrpapd
Exogenous: _cons

5.1.4 Lag-Order Selection for KR Patent Registration Domestic Data

"kr" is the abbreviation of "Korea". "pre" stands for patent registration. "d" equals to domestic. The following leftmost "d2.krpred" has the lag 2. The middle "d4.krpred" has the lag 4 and the rightmost "lnkpred" the lag 3.

Table 18: Lag Candidates from the KR Patent Registration domestic Data

. varsoc d2.krpred /*candidate*/

Selection-order criteria
Sample: 1986 - 2012 Number of obs = 27

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-296.864				2.2e+08	22.064	22.0783	22.112
1	-296.62	.48795	1	0.485	2.4e+08	22.12	22.1485	22.216
2	-287.772	17.696*	1	0.000	1.3e+08*	21.5387*	21.5815*	21.6826*
3	-287.593	.35813	1	0.550	1.4e+08	21.5995	21.6565	21.7914
4	-286.004	3.1777	1	0.075	1.4e+08	21.5558	21.6272	21.7958

Endogenous: D2.krpred
Exogenous: _cons

. varsoc d4.krpred /*temporary best*/

Selection-order criteria
Sample: 1988 - 2012 Number of obs = 25

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-291.378				8.4e+08	23.3902	23.4038	23.439
1	-290.876	1.003	1	0.317	8.8e+08	23.4301	23.4572	23.5276
2	-284.3	13.153	1	0.000	5.6e+08	22.984	23.0246	23.1303
3	-280.999	6.6014	1	0.010	4.7e+08	22.7999	22.854	22.995
4	-275.086	11.826*	1	0.001	3.2e+08*	22.4069*	22.4745*	22.6507*

Endogenous: D4.krpred
Exogenous: _cons

. varsoc lnkrpred /*best*/

Selection-order criteria
Sample: 1984 - 2012 Number of obs = 29

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	-58.8021				3.61985	4.12428	4.13905	4.17143
1	-7.11698	103.37	1	0.000	.109821	.628757	.65829	.723054
2	-5.90466	2.4246	1	0.119	.108282	.614114	.658413	.755559
3	-1.86984	8.0696*	1	0.005	.087923*	.404816*	.463881*	.593409*
4	-1.70422	.33123	1	0.565	.09329	.46236	.536191	.698101

Endogenous: lnkrpred
Exogenous: _cons

5.2 Heteroskedasticity Tests for Models

5.2.1 Heteroskedasticity Tests for KR Patent Application Foreign Data

The following rightmost "L1.lnkrpapf" is the best in terms of simplification, even though all of the following 3 have the normal distribution.

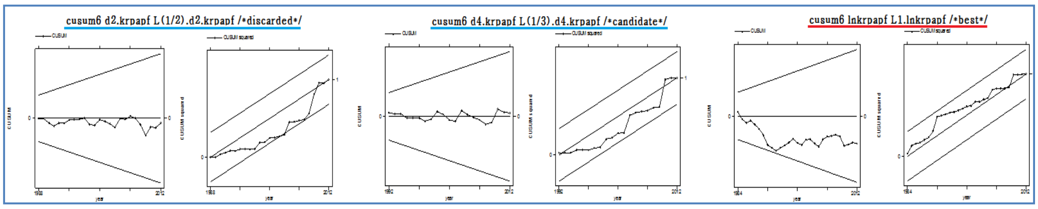


Figure 10: Heteroskedasticity Tests about the KR Patent Application Foreign Data

5.2.2 Heteroskedasticity Tests for KR Patent Registration Foreign Data

The following left "L(1/3).krpref" has the out-of-range abnormality, which is the reason why it should be discarded. The right "L(1/4).krpref", the lags 1, 2, 3, 4 of the 4th level difference data from the KR patent registration foreign data, is the only one and best candidate.

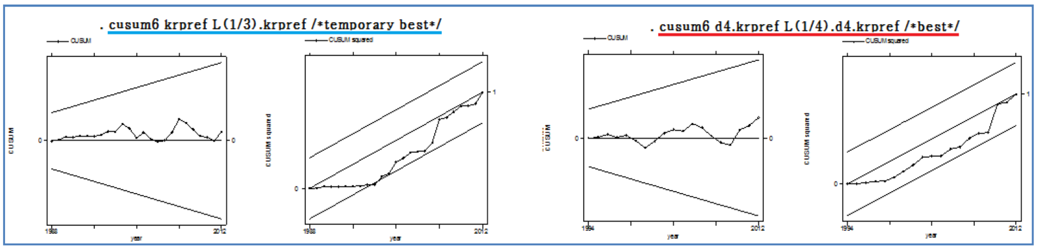


Figure 11: Heteroskedasticity Tests about the KR Patent Registration Foreign Data

5.2.3 Heteroskedasticity Tests for KR Patent Application Domestic Data

The left "L(1/4).d3. krpapd" is the only one and best candidate in the following diagram.

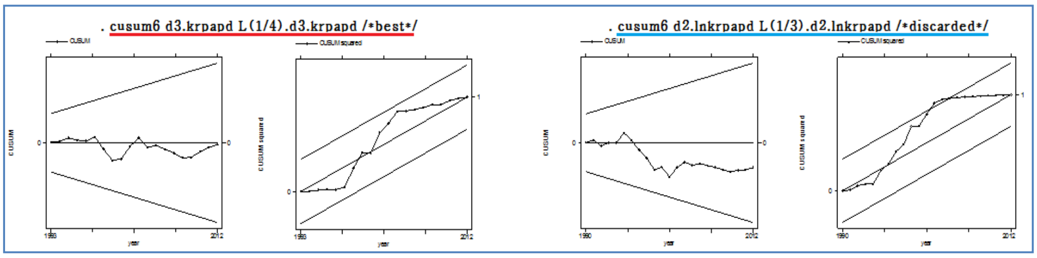


Figure 12: Heteroskedasticity Tests about the KR Patent Application Domestic Data

5.2.4 Heteroskedasticity Tests for KR Patent Registration Domestic Data

The following leftmost “L(1/2).d2.krpred” has the out of range problem. Both of the middle “L(1/4).d4.krpred” and the rightmost “L(1/3).lnkpred” have the normal distribution. The rightmost “L(1/3).lnkpred” is the best in terms of simplification.

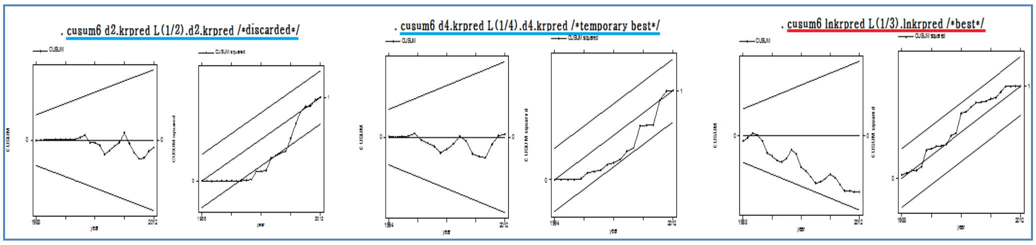


Figure 13: Heteroskedasticity Tests about the KR Patent Application Domestic Data

5.3 QLR Test Result for Finding Breaks

We divided both of the application data and registration data of the KR patent into foreign and domestic data, from which we can find the following secrets.

Foreigners applied earlier but took results later, and the domestic people did contrariwise. Domestic people had already enjoyed the above adoption since 1987 and didn’t need to hurry.

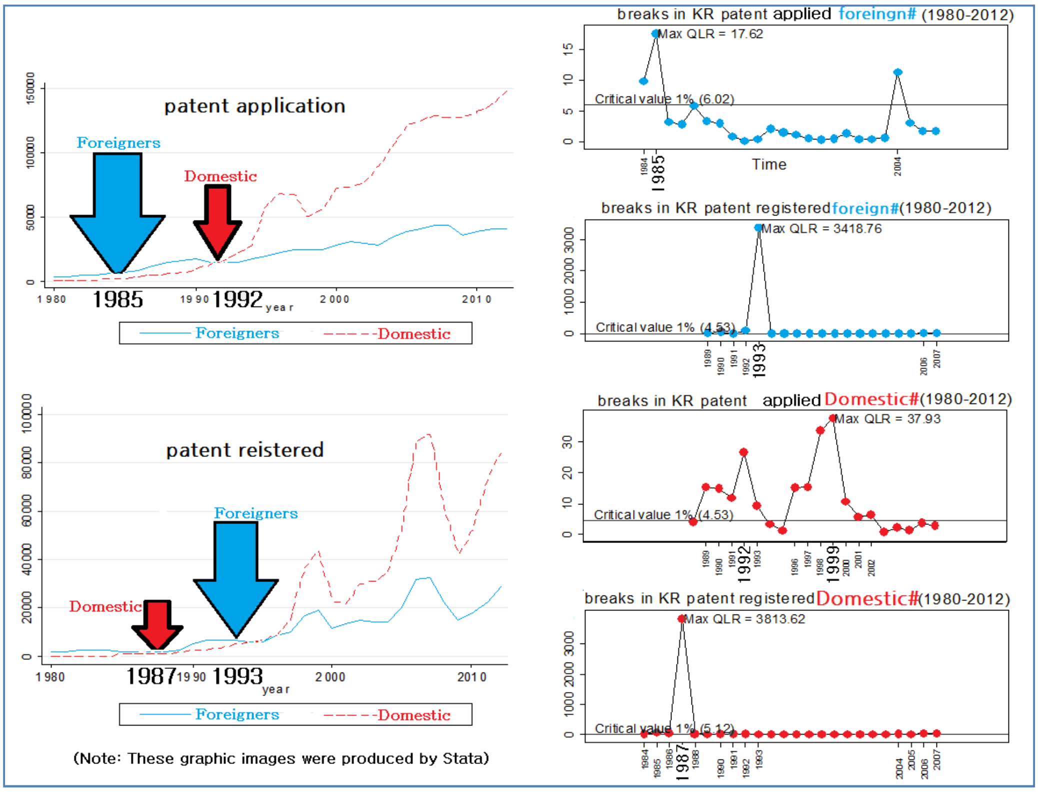


Figure 14: Who Wins from KR Patent Scope Enlargement

5.4 QLR Test Coding Contents for Finding Breaks

5.4.1 QLR Test Coding for KR Patent Application Foreign Data

The following coding content, “qui reg 'var' L1.'var' di d_'var',r” means that there are 4 independent variables. “L1.`var'” is L1.lnkrpapf, because of “local var = "lnkrpapf"”.

“cap gen d_'var' = di*l1.'var'” makes multiplication of L1.lnkrpapf by the dummy variable “di”. “local critical=6.02” and “Critical value 1% ('critical')” are required for adaption to data characteristics.

```
use 80-12-krp.dta,clear
log using kby24-L1-lnkrpapf-QLR.log,replace
tset year
sum year
local time=r(max)-r(min)+1
local i = round('time'*.15)
local f = round('time'*.85)

local var = "lnkrpapf"

gen qlr'var' = .
set more off
while 'i'<=('f') {
cap gen di = (_n >='i')

cap gen d_'var' = di*l1.'var'

list year 'var' l1.'var' di d_'var'
qui reg 'var' L1.'var' di d_'var',r

qui test di d_'var'

cap replace qlr'var' = r(F) in 'i'

dis "'i', QLR of the year " %ty year['i'] " =" %6.2f qlr'var'['i'] " [see above table]"
drop di d_'var'
local i = 'i' + 1
}
sum qlr'var'
local maxvalue=r(max)
gen maxdate=year if round(qlr'var',0.01)==round('maxvalue',0.01)
local maxvalue1=round('maxvalue',0.01)
local critical=6.02 /*Replace with the appropriate critical value (see Stock & Watson)*/
sum year
local mindate=r(min)
sum maxdate
local maxdate=r(max)

gen break=year if qlr'var'>='critical' & qlr'var'!=.
dis "Below are the break dates..."
```

```
list year qlr'var' if break!=.
levelsof break, local(break1)
set more on
twoway connected qlr'var' year,title(breaks in KR patent apply foreign#(1980-2012)) ///
xlabel('break1', angle(90) labsize(2.3) alternate) ///
yline('critical') ytitle(Quandt-LR(QLR) statistic) xtitle(Time) ///
ttext('critical' 'mindate' "Critical value 1% ('critical')", placement(ne)) ///
ttext('maxvalue' 'maxdate' "Max QLR = 'maxvalue1'", placement(e)) saving(ts3,replace)
scheme(s1manual)
more
twoway (tsline 'var' L1.'var') lfit 'var' year,saving(ts4,replace) scheme(s1manual)
more
twoway (tsline 'var') lfit 'var' year,saving(ts5,replace) scheme(s1manual)
more
graph combine ts3.gph ts4.gph ts5.gph,col(1) xsize(9) ysize(14)
```

5.4.2 QLR Test Coding for KR Patent Registration Foreign Data

The following coding content, “qui reg diff4'var' L(1/4).diff4'var' di d_'var'1 d_'var'2 d_'var'3 d_'var'4,r” means that there are 10 independent variables. “L(1/4).diff4'var'” consists of L1.d4.krpref, L2.d4.krpref, L3.d4.krpref and L4.d4.krpref, because of “local var = "krpref"” and “gen diff4'var' = d4.'var'”.

“cap gen d_'var'1 = di*l1.diff4'var'”, “cap gen d_'var'2 = di*l2.diff4'var'”, “cap gen d_'var'3 = di*l3.diff4'var'” and “cap gen d_'var'4 = di*l4.diff4'var'” make multiplication of each of L1.d4.krpref, L2.d4.krpref, L3.d4.krpref and L4.d4.krpref by the dummy variable “di”. “local critical=4.53” and “Critical value 1% ('critical')” are required for adaption to data characteristics.

```
use 80-12-krp.dta,clear
log using kby35-L4-d4-krpref-QLR.log,replace
tset year
sum year
local time=r(max)-r(min)+1
local i = round('time'*.15)
local f = round('time'*.85)

local var = "krpref"
gen diff4'var' = d4.'var'
gen qlr'var' = .
set more off
while 'i'<=('f') {
cap gen di = (_n>='i')

cap gen d_'var'1 = di*l1.diff4'var'
cap gen d_'var'2 = di*l2.diff4'var'
cap gen d_'var'3 = di*l3.diff4'var'
cap gen d_'var'4 = di*l4.diff4'var'

list year 'var' d.'var' diff4'var' L(1,4).diff4'var' di d_'var'4
qui reg diff4'var' L(1/4).diff4'var' di d_'var'1 d_'var'2 d_'var'3 d_'var'4,r
```

```
qui test di d_'var'1 d_'var'2 d_'var'3 d_'var'4

cap replace qlr'var' = r(F) in 'i'

dis "'i', QLR of the year " %ty year['i'] " =" %6.2f qlr'var'['i'] " [see above table]"
drop di d_'var'1 d_'var'2 d_'var'3 d_'var'4
local i = 'i' + 1
}
sum qlr'var'
local maxvalue=r(max)
gen maxdate=year if round(qlr'var',0.01)==round('maxvalue',0.01)
local maxvalue1=round('maxvalue',0.01)
local critical=4.53 /*Replace with the appropriate critical value (see Stock & Watson)*/
sum year
local mindate=r(min)
sum maxdate
local maxdate=r(max)

gen break=year if qlr'var'>='critical' & qlr'var'!=.
dis "Below are the break dates..."
list year qlr'var' if break!=.
levelsof break, local(break1)
set more on
twoway connected qlr'var' year,title(breaks in KR patent registered foreign#(1980-2012))
///
xlabel('break1', angle(90) labsize(2.3) alternate) ///
yline('critical') ytitle(Quandt-LR(QLR) statistic) xtitle(Time) ///
ttext('critical' 'mindate' "Critical value 1% ('critical')", placement(ne)) ///
ttext('maxvalue' 'maxdate' "Max QLR = 'maxvalue1'", placement(e)) saving(ts3,replace)
scheme(s1manual)
more
twoway (tsline diff4'var' L(1/4).diff4'var') lfit diff4'var' year,saving(ts4,replace)
scheme(s2manual)
more
twoway (tsline 'var') lfit 'var' year,saving(ts5,replace) scheme(s1manual)
more
graph combine ts3.gph ts4.gph ts5.gph,col(1) xsize(9) ysize(14)
```

5.4.3 QLR Test Coding for KR Patent Application Domestic Data

The following coding content, “qui reg diff3'var' L(1/4).diff3'var' di d_'var'1 d_'var'2 d_'var'3 d_'var'4,r” means that there are 10 independent variables. “L(1/4).diff3'var'” consists of L1.d3.krpapd, L2.d3.krpapd, L3.d3.krpapd and L4.d3.krpapd, because of “local var = "krpapd"” and “gen diff3'var' = d3.'var'”.

“cap gen d_'var'1 = di*l1.diff3'var'”, “cap gen d_'var'2 = di*l2.diff3'var'”, “cap gen d_'var'3 = di*l3.diff3'var'” and “cap gen d_'var'4 = di*l4.diff3'var'” make multiplication of each of L1.d3. krpapd, L2.d3. krpapd, L3.d3. krpapd and L4.d3. krpapd by the dummy variable “di”. “local critical=4.53” and “Critical value 1% ('critical')” are required for adaption to data characteristics.

```
use 80-12-krp.dta,clear
log using kby25-L4-d3-krpapd-QLR.log,replace
tset year
sum year
local time=r(max)-r(min)+1
local i = round('time'*.15)
local f = round('time'*.85)
local var = "krpapd"
gen diff3'var' = d3.'var'
gen qlr'var' = .
set more off
while 'i'<=('f') {
cap gen di = (_n >='i')

cap gen d_'var'1 = di*l1.diff3'var'
cap gen d_'var'2 = di*l2.diff3'var'
cap gen d_'var'3 = di*l3.diff3'var'
cap gen d_'var'4 = di*l4.diff3'var'

list year 'var' diff3'var' L(1,4).diff3'var' di d_'var'4
qui reg diff3'var' L(1/4).diff3'var' di d_'var'1 d_'var'2 d_'var'3 d_'var'4,r

qui test di d_'var'1 d_'var'2 d_'var'3 d_'var'4

cap replace qlr'var' = r(F) in 'i'

dis "'i', QLR of the year " %ty year['i'] " =" %6.2f qlr'var'['i'] " [see above table]"
drop di d_'var'1 d_'var'2 d_'var'3 d_'var'4
local i = 'i' + 1
}
sum qlr'var'
local maxvalue=r(max)
gen maxdate=year if round(qlr'var',0.01)==round('maxvalue',0.01)
local maxvalue1=round('maxvalue',0.01)
local critical=4.53 /*Replace with the appropriate critical value (see Stock & Watson)*/
sum year
local mindate=r(min)
sum maxdate
local maxdate=r(max)

gen break=year if qlr'var'>='critical' & qlr'var'!=.
dis "Below are the break dates..."
list year qlr'var' if break!=.
levelsof break, local(break1)
set more on
twoway connected qlr'var' year,title(breaks in KR patent apply domestic#(1980-2012)) ///
xlabel('break1', angle(90) labsize(2.3) alternate) ///
yline('critical') ytitle(Quandt-LR(QLR) statistic) xtitle(Time) ///
ttext('critical' 'mindate' "Critical value 1% ('critical')", placement(ne)) ///
ttext('maxvalue' 'maxdate' "Max QLR = 'maxvalue1'", placement(e)) saving(ts3,replace)
```

```
scheme(s1manual)
more
twoway (tsline diff3'var' L(1/4).diff3'var') lfit diff3'var' year,saving(ts4,replace)
scheme(s1manual)
more
twoway (tsline 'var') lfit 'var' year,saving(ts5,replace) scheme(s1manual)
more
graph combine ts3.gph ts4.gph ts5.gph,col(1) xsize(9) ysize(14)
```

5.4.4 QLR Test Coding for KR Patent Registration Domestic Data

The following coding content, “qui reg 'var' L(1/3).'var' di d_'var'1 d_'var'2 d_'var'3,r” means that there are 8 independent variables. “L(1/3).'var'” consists of L1. lnkrpred, L2. lnkrpred and L3.lnkrpred, because of “local var = "lnkrpred"”.

“cap gen d_'var'1 = di*l1.'var'”, “cap gen d_'var'2 = di*l2.'var'” and “cap gen d_'var'3 = di*l3.'var'” make multiplication of each of L1. lnkrpred, L2. lnkrpred and L3. lnkrpred by the dummy variable “di”.

```
use 80-12-krp.dta,clear
log using kby37-d3-lnkrpred-QLR.log,replace
tset year
sum year
local time=r(max)-r(min)+1
local i = round('time'*.15)
local f = round('time'*.85)

local var = "lnkrpred"

gen qlr'var' = .
set more off
while 'i'<=('f') {
cap gen di = (_n>='i')

cap gen d_'var'1 = di*l1.'var'
cap gen d_'var'2 = di*l2.'var'
cap gen d_'var'3 = di*l3.'var'

list year 'var' L(1,3).'var' di d_'var'3
qui reg 'var' L(1/3).'var' di d_'var'1 d_'var'2 d_'var'3,r

qui test di d_'var'1 d_'var'2 d_'var'3

cap replace qlr'var' = r(F) in 'i'

dis "'i', QLR of the year " %ty year['i'] " =" %6.2f qlr'var'['i'] " [see above table]"
drop di d_'var'1 d_'var'2 d_'var'3
local i = 'i' + 1
}
/* skip due to overlap by the last coding content */
```

6. Constraints of This Thesis

Balcombe et al (2011) insisted that we should only accept the lag order of up to 3 for autocorrelation elimination. However, in the chapter 5, at the paragraphs of 5.2.2, 5.2.3, 5.4.2 and 5.4.3, we could not help but choose the lag 4 due to the source data fluctuations. The other restriction is the fact that we adopted the significance level 10% for the US patent application data analysis in chapter 2.

Afterwards, somebody could try to use the Box-Cox transform or another transform such as √y and 1/y. Transforming methods herein were the natural logarithm from the level data, the difference data from the level data or the natural logarithm data. The lags were chosen by the criteria of FPE, AIC, HQIC, SBIC, which were given by the already developed computerized software.

7. Further Application Possibilities

There are so many time series data such as the GDP per capita data, the stock price data, the seismic intensity data and etc.

This author tried to apply this method for the stock price data and failed to get satisfactory results due to extreme fluctuations, against which the peculiar transformation like the Box-Cox's, √y or 1/y might have to be used or for which polynomial trend models should be considered.

Generally speaking, if they find the long-run moving average line at the top, the mid-run one at the middle and short-run one at the bottom or vice versa, they call it “golden cross” or “dead cross”. After “cross” the stock price will show breaks upwards or downwards. In ARIMA(p,d,q) they only take “q” without “p” and “d” considered, on the contrary to this thesis.

Accordingly, in ARIMA(p,d,q) both of the one option of “p” & “d” and the other option of “q” are alternatives to each other. We cannot use simultaneously both of the one option of “p” & “d” and the other option of “q”, which are logically exclusive “or”, this author insists.

As for GDP, the implicit price deflator, the PPP(purchasing-power-parity) conversion factor or the other arbitration factors like 1985 Plaza agreement for the dollar depreciation might to have to be applied or considered before this method's being applied.

Once we experienced some odors just before the seismic intensity data's breaks, we faced earthquake breaks, we could suppose. If that was the case, we might be able to apply this thesis’s method or the above “golden cross” or “dead cross” method, which could be the cause of this author’s naive imaginations for the future researches & developments.

References

Andrews (2003), Andrews, Donald W. K., Tests for Parameter Instability and Structural Change with Unknown Change Point: A Corrigendum, *Econometrica, Vol. 71, No.1,* January 2003, 395-397

Balcombe et al (2011), Balcombe, Kelvin G., Fraser, Iain and Sharma, Abhijit, "Bayesian model averaging and identification of structural breaks in time series", *Applied Economics, 43,* 2011, 3805–3818

Chow (1960), Chow, Gregory C, "Tests of Equality Between Sets of Coefficients in Two Linear Regressions". *Econometrica 28 (3):* 591–605

Hall (2004), Hall, Bronwyn H., "Exploring the Patent Explosion", *Journal of Technology Transfer, Volume 30, Numbers 1-2 35-48 0892-9912, Elsevier*, 2004.

Johnston & DiNardo (1997), Johnston, Jack & DiNardo, John, 『Econometrics Methods』, 4th edition, McGraw-Hill, international editions, 1997, pp.207~220

Kim (2015), Kim, BeomYong, Illustrated Intellectual Property Guide, Laon book-publishing Corp, 2015.12.16.

Kim et al (2003), Kim, Woocheol, Kim, Jaejoo, Park, Byeongwook, 『General Statistics』, Youngji Culture Corp, 2003.7.15.

KIPO (2016), Korean IP Office, http://www.kipo.go.kr, 2016

Lee (2013), Lee, Sarngyeol, 『Time Series Analysis - Theory and SAS Practises』, Gyoengki-do, Liberty-Academy corp., pp.88~90, 95~96, 2013.08.20.

MIC (2016), Created by editing this thesis, (Statistics Bureau, Ministry of Internal Affairs and Communications website), (http://www.stat.go.jp/english/data/chouki/28.htm), 2016

Min & Choi (2012), Min, Insick & Choi, Pilseon, 『STATA Basic Statistics & Regression』, Seoul, Jipil-midi corp., pp.28, 50, 214~218, 221~224, 2012.12.15.

Min & Choi (2014), Min, Insick & Choi, Pilseon, 『 STATA Time Series Data Analysis』, Seoul, Jipil-midi corp., pp.13, 14, 17, 169, 206, 2014.6.20

Mitchell (2014), Mitchell, Sara McLaughlin, "Time Series Analysis: Method and Substance, Introductory Workshop on Time Series Analysis", Department of Political Science University of Iowa, p.16, 2014

Quandt (1960), Quandt, Richard, “Tests of the Hypothesis that a Linear Regression Obeys Two Separate Regimes.” *Journal of the American Statistical Association. 55,* pp. 324–30, 1960

StataCorp (2009), Stata: Release 11. Statistical Software. College Station, TX: StataCorp LP., 2009

Stock & Watson (2012), Stock, James H and Watson, Mark M., 『Introduction to Econometrics』, 3rd edition, Global Edition, Boston: Pearson Addison Wesley, pp. 598~602, 2012

Torres-Reyna (2014), Torres-Reyna, Oscar, “Time Series(ver. 1.5)”, Princeton University, Data Consultant, 2014.8.

USPTO (2016), http://www.uspto.gov/web/offices/ac/ido/oeip/taf/h_counts.pdf, 2016

Yang (2013), Yang, Ohseok, 『Impressive Stata At first sight, Stata Application for Management』, 2013.10.25

Woo (2013), Woo, Seokjin, 『STATA for Economic Analysis』, Seoul, Jipil-midi corp., pp.13~18,28,45~47,73,89,93,203,204,219,220,254,271~277, 2013.10.25