ISSN : 0974 - 7435 Volume 10 Issue 3

BioTechnology
An Indian Journal

FULL PAPER

BTAIJ, 10(3), 2014 [659-665]

Chinese regional economic research based on panel data model

Zhihao Su[1]*, Xinghua Su[2], Yajuan Chen[1]

[1]Department of Financial Managment, Langfang Teachers College, Langfang, Hebei, (CHINA)

[2]Department of Business Administration, Langfang Teachers College, Langfang, Hebei, (CHINA)

ABSTRACT

Panel Data model occupies in economic research of higher status, the method is also a newly-developed and practical analysis method. The paper makes research on Hebei, Henan and Shandong three regions gross regional domestic product, primary industry, secondary industry and tertiary industry data, and analyzes data from opening-up and reform to 2011; in the paper, it takes Panel Data model as research method, firstly states Panel Data model principle and model construction moment required mathematical tool-least square method, and then applies each region statistical bureau statistics almanac data analyzing each industry and gross regional domestic product trends followed by time changing, it gets four kind of data has higher correlations by figure, finally it gets three regions' gross regional domestic product increment and three industries increment relations by the paper established mathematical model and computing method, obtained all regression equation multiple correlation coefficients are 1, variance is very little, from which largest variance is Shandong that its variance is 2.2214. By the paper described research methods and research results indicating, Panel Data model can be used as regional economic research's important statistical method, and it can make prediction on future several years' data.

KEYWORDS

Panel data model;
Least square method;
Regression analysis;
Difference equation.

INTRODUCTION

There are lots of factors affect one regional economic development, and gross regional domestic product is an important measurement scale of regional economic development, the value includes economic development process static factors and dynamic stators; the paper puts economics microanalysis aside, carries out research on Hebei, Henan and Shandong's gross regional domestic product from the perspective of lots of data statistics. Economic development includes each kind of industries, now it is used to divide it into three main industries, from which tertiary industry's development is the fastest one, in order to ensure data reliability, the paper respectively takes statistics almanac data from Hebei statistical bureau, Henan statistical bureau and Shandong statistical bureau, in the hope of pursuing objective law from data changes, revealing economic development essential factors and providing data basis and theoretical methods for economic planning.

The paper researches on three regions' four kind of data by applying Panel Data model, in order to better and reasonable select research theories and computing methods, the paper references following documents, with more and more people researching on regional economy, let the field development have better environment, and let regional economic development on track as well as avoid lots of unnecessary detours. Among them, Nanjing information engineering university's Shen Xiao-Jing in her master's thesis "Panel data model and its research on regional economic problems" analyzed Chinese South Jiangsu, Central Jiangsu and North Jiangsu three regions GDP and other relative indicators, she proved Panel Data model can overcome time sequence model's research with smaller precise on shorter period and data in cross section and longitudinal section two directions extended data, and verified model reliability and rationality by FDI and GDP relations[1]; Wuhan University of Technology's Gao Qin in her master's thesis "Port development and regional economic relationships research", she made comprehensive sorting on Chinese and foreign scholars current research status and research methods, and provided thesis main research contents, by stating port and regional economy interactive development theory, she made qualitative evaluation on port and regional economic relationships; applied grey relational analysis method discussing port and region economic indicators correlation degrees, and then utilized DEA method quantitative analyzing port contributions to regional economy, revealed different port and city relationships, and provided suggestive measures for port and city, as well as hinterland harmonious development[2]; Joan teachers training college's Chen Fu-Chuan and others in the article "Hainan province regional economic development advantages analysis", they started from strengthen Hainan province regional economic development's "point" and "axis" selection scientificity, utilized fuzzy comprehensive evaluation model analyzing Hainan province regional development status, constructed growth pole and growth axis structure and layout, which provided decision basis for planning Hainan province regional economic development strategies[3].

The paper on the basis of previous research, applies Panel Data model researching on Hebei, Henan and Shandong three regions gross regional domestic product, in the hope of revealing economic development trend and inflection point from data changes.

PANEL DATA PROCESSING MODEL

Panel Data model is the English term of Zhonglie Shuju Muxing. The model was introduced into economic measurement researches by Mundlak, Balestra and Nerlove as earliest, it made up for time sequence model limitation in data processing that can better process cross section and longitude section simultaneous extended panel data, therefore it also called longitude and latitude data as well as panel data. In the following, it makes statements on Panel Data multi-period difference model, balance Panel Data fixed effect model, random effect model and consistency correlation model, in the hope of providing model basis for Chinese regional economic analysis.

Multi-period difference model

Multi-period difference model mathematical expression is as formula (1) show:

$$\mathbf{y}_{\mathbf{it}} = \boldsymbol{\delta}_1 + \sum_{n=2}^{T} \boldsymbol{\delta}_n \mathbf{dn}_t + \sum_{m=1}^{k} \boldsymbol{\beta}_m \mathbf{x}_{\mathbf{itm}} + \mathbf{a}_l + \mathbf{u}_{\mathbf{it}} \tag{1}$$

In formula(1) $i = 1 \cdots N$ represents cross section unit; $t = 1 \cdots T$ represents period; m represents variable mark; dn_t represents annual dummy variable; a_l represents fixed effect, which includes all unobserved and time constant factors that affect y_{it}; u_{it} represents time-varying errors, which stands for unobserved factors change with time and affect y_{it};

When formula(1) T is smaller than N, by first difference, formula(1) can be converted into equation as formula (2) show:

$$\Delta \mathbf{y}_{\mathbf{it}} = \boldsymbol{\alpha}_0 + \sum_{n=3}^{T} \boldsymbol{\alpha}_n \Delta \mathbf{dn}_t + \sum_{m=1}^{k} \boldsymbol{\beta}_m \Delta \mathbf{x}_{\mathbf{itm}} + \Delta \mathbf{u}_{\mathbf{it}}, \mathbf{t} = 2,3,\cdots,\mathbf{T} \tag{2}$$

As formula (2) showed difference equation every unit i is $T-1$ pieces of periods data that totally has $N \times (T-1)$ pieces of observed values, by difference method, it eliminates fixed effect a_l, and can apply least square method to estimate on parameter β.

Balance panel data fixed effect model

If N pieces of cross sections units every unit has same T periods data, and then the kind of data set is called balance Panel Data, its general linear model is as formula (3) show :

$$\mathbf{y_{it} = \beta_0 + \sum_{m=1}^{k} \beta_m x_{itm} + a_l + u_{it}} \tag{3}$$

In formula (3), $i = 1 \cdots N$ represents cross section unit; $t = 1 \cdots T$ represents period; m represents variable mark.

If solve equation average in time on formula(3) every cross section unit i, it can get as formula (4)showed form, and then take formula (3)and formula(4) as differences, it can get formula(5):

$$\mathbf{\bar{y}_l = T^{-1} \sum_{l=1}^{T} y_{lt}} \tag{4}$$

$$\mathbf{y_{it} - \bar{y}_l = \ddot{y}_{lt} = \sum_{m=1}^{k} \beta_m \ddot{x}_{itm} + \ddot{u}_{it}} \tag{5}$$

If formula(5) u_{it} has same variances and no time sequence correlations, and then it can eliminate unobserved effect a_t by formula (5), after that carry out mixed OLS regression with ensemble dividing time average value's variable to do parameters estimation.

Random effect model

For unobserved effect model, it is as formula(6)show:

$$\mathbf{y_{it} = \beta_0 + \sum_{m=1}^{k} \beta_m x_{itm} + a_l + u_{it}} \tag{6}$$

In formula(6) $i = 1 \cdots N$ represents cross section unit; $t = 1 \cdots T$ represents period; m represents variable mark.

If formula(6) unobserved effect a_l has no correlations with every explanatory variable, then make formula (7) true:

$$\mathbf{Cov(x_{itj}, a_l) = 0, (t = 1, \cdots, T, j = 1, \cdots, k)} \tag{7}$$

When formula (7) is true, it calls formula (6) as random effect model, introduce into composite disturbance formula(8) and input formula(8) into formula(6) that can get another form random effect model, as formula (9)show:

$$\mathbf{v_{it} = a_l + u_{it}} \tag{8}$$

$$\mathbf{y_{it} = \beta_0 + \sum_{m=1}^{k} \beta_m x_{itm} + v_{it}} \tag{9}$$

If $\sigma_a^2 = Var(a_l)$ and $\sigma_{it}^2 = Var(u_{it})$, define a parameter λ, its expression is as formula (10) show, and input formula (10) into formula(9)that can get equation final transform as formula (11)show:

$$\mathbf{\lambda = 1 - \frac{\sigma_{it}}{\sqrt{\sigma_{it}^2 + T\sigma_a^2}}} \tag{10}$$

$$\mathbf{y_{it} - \lambda\bar{y}_l = \beta_0(1-\lambda) + \sum_{m=1}^{k} \beta_m (x_{itm} - \lambda\bar{x}_{im}) + (v_{it} - \lambda\bar{v}_l)} \tag{11}$$

In case specific error has no time sequence correlations and possesses same variance, fixed effect method is more effective; in case it cannot regard observed value as random sampling result from an ensemble, it can regard fixed effect a_l as solve-for parameter, adopt fixed effect method can get better effects;if a_l and all x_{it} are uncorrelated, and then adopt random effect model can get better effects.

Consistency correlation model

If let ρ to represent same section unit two variables positive correlation coefficient, I to represent $n \times n$ orders unit matrix, J to represent all elements as 1 $n \times n$ matrix, and then it has relationships as formula (12) show:

$$\mathbf{v_0 = (1-\rho)I + \rho J} \tag{12}$$

Consistency correlation model is as formula (13) show:

$$\mathbf{Y_{ij} = \mu_{ij} + U_i + Z_{ij}} \tag{13}$$

In formula(13), $\mu_{ij} = E(Y_{ij}), U_i$ and Z_{ij} are all independent same distribution normal random variables, the two are mutual independent, therefore it makes formula(14) true:

$$\begin{cases} \mathbf{U_i \sim N(0, v^2)} \\ \mathbf{Z_{ij} \sim N(0, \tau^2)} \\ \sigma^2 = \mathbf{v}^2 + \tau^2 \\ \rho = \dfrac{\mathbf{v}^2}{\mathbf{v}^2 + \tau^2} \end{cases} \tag{14}$$

For same section observed value Y_{ij}, Y_{ik}, set its correlation coefficient as v_{jk}, as formula(15)show:

$$\mathbf{v_{jk} = Cov(Y_{ij}, Y_{ik})} \quad (15)$$

If for any j, it has relationships as formula (16)show, and then it can get as formula(17) showed v_{jk} expression:

$$\begin{cases} \mathbf{t_{j+1} - t_j = d} \\ \rho = \mathbf{exp}(-\phi \mathbf{d}) \end{cases} \quad (16)$$

$$\mathbf{v_{jk}} = \sigma^2 \mathbf{exp}\left|-\phi(\mathbf{t_j - t_k})\right| = \sigma^2\rho^{|\mathbf{j-k}|} \quad (17)$$

If $Var(Y_{ij}) = Var(W_{ij}) = \sigma^2, Z_{ij} \sim N[0, \sigma^2(1-\rho^2)]$, and then index correlation model is as formula(18)show:

$$\mathbf{Y_{ij} = u_{ij} + W_{ij}, i = 1, \cdots, m; j = 1, \cdots, n;} \quad (18)$$

REGRESSION ANALYSIS MODEL

Multiple linear regression models

Multiple linear regression model general expression is as formula (19) show:

$$\mathbf{y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_{k-1} x_{t(k-1)} + u_t} \quad (19)$$

In formula(19) y_t represents explained variable; x_{tj} represents important explanatory variable; β_i represents regression parameters; u_t represents multiple little factors affect change. If given a sample $(y_t, x_{t1}, x_{t2}, \cdots, x_{t(k-1)})$ and when $t = 1, 2, \cdots, T$ above general model can have as formula (20) showed matrix form:

$$\begin{pmatrix} \mathbf{y_1} \\ \mathbf{y_2} \\ \vdots \\ \mathbf{y_T} \end{pmatrix}_{(\mathbf{T\times1})} = \begin{pmatrix} \mathbf{1} & \mathbf{x_{11}} & \cdots & \mathbf{x_{1j}} & \cdots & \mathbf{x_{1(k-1)}} \\ \mathbf{1} & \mathbf{x_{21}} & \cdots & \mathbf{x_{2j}} & \cdots & \mathbf{x_{2(k-1)}} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ \mathbf{1} & \mathbf{x_{T1}} & \cdots & \mathbf{x_{Tj}} & \cdots & \mathbf{x_{T(k-1)}} \end{pmatrix}_{(\mathbf{T\times k})} \begin{pmatrix} \beta_0 \\ \beta_2 \\ \vdots \\ \beta_{\mathbf{k-1}} \end{pmatrix}_{(\mathbf{k\times1})} + \begin{pmatrix} \mathbf{u_1} \\ \mathbf{u_2} \\ \vdots \\ \mathbf{u_T} \end{pmatrix}_{(\mathbf{T\times1})} \quad (20)$$

In formula(20), if use matrix equation form to express, it as formula(21)show:

$$\mathbf{Y = X\beta + u} \quad (21)$$

In order to get best regression model, it needs to meet following four conditions:

Condition 1 : Random error term is not auto-correlative, it meets every error term average value is zero and same variances are finite values, mathematical expression is as formula (22) show:

$$\mathbf{E(u) = 0; Var(u) = \sigma^2 I} \quad (22)$$

Condition 2 : Explanatory variable and error term are mutual independent, and then it makes $E(X'u) = 0$ true;

Condition 3 : Explanatory variables linear is uncorrelated, and then it makes $\text{rank}(X'X) = \text{rank}(X) = k$ true;

Condition 4 : Explanatory variables have non- randomness, if $T \to \infty$, and then $T^l X'X \to Q$, from which Q represents a finite value non-degenerate matrix.

Least square theory

Least square method principle is solving residual minimum value, when residual arrives at minimum value that error term squares sum is minimum, error minimum value optimization model is as formula (23) show:

$$\begin{aligned} \mathbf{min\,S} &= \mathbf{(Y - X\hat\beta)'(Y - X\hat\beta) = Y'Y \cdot \hat\beta'XY' - Y'X\hat\beta + \hat\beta'X'X\hat\beta} \\ &= \mathbf{Y'Y - 2\hat\beta'X'Y + \hat\beta'X'X\hat\beta} \end{aligned} \quad (23)$$

Due to $Y'X\hat\beta$ is a scalar, so it exists $Y'X\hat\beta = \hat\beta'X'Y$, and formula(23) first order condition meets formula (24), it can get formula(25) by simplifying:

$$\mathbf{\frac{\partial S}{\partial \hat\beta} = -2X'Y + 2X'X\hat\beta = 0} \quad (24)$$

$$\mathbf{X'Y = X'X\hat\beta} \quad (25)$$

Due to $(X'X)$表 represents a non-degenerate matrix, so it makes formula(26) true:

$$\mathbf{\hat\beta = (X'X)^{-1}X'Y} \quad (26)$$

Formula(23) second order condition is as formula (27) show:

$$\mathbf{\frac{\partial^2 S}{\partial \hat\beta' \partial \hat\beta} = 2X'X \geq 0} \quad (27)$$

To sum up:formula(26)is the solution of formula (23).

Due to X elements have non-randomness, and $(X'X)^{-1}$ is a constant matrix, $\hat{\beta}$ is Y linear combination.

RESEARCH RESULT AND ANALYSIS

Three regions economic data

In TABLE 1, A represents Gross Domestic Product; B represents Primary Industry; C represents Secondary Industry ;D represents Tertiary Industry.

By TABLE 1, it can get each region each industry changes trends with years as Figure 1 show.

Panel data model data analysis

Research on Hebei, Henan and Shandong three regions GDP and three industries increments relationships from 1978 to 2011, applied model is as formula (28)

TABLE 1 : Hebei, Henan, Shandong three gross regional domestic products and each industry data table

Year	Hebei				Henan				Shandong			
	A	B	C	D	A	B	C	D	A	B	C	D
1978	183.06	52.20	92.38	38.48	162.92	64.86	69.45	28.61	225.45	75.06	119.35	31.04
1979	203.22	61.11	101.76	40.35	190.09	77.30	80.52	32.27	251.60	91.12	127.68	32.80
1980	219.24	68.09	105.88	45.27	229.16	93.23	94.44	41.49	292.13	106.43	146.11	39.59
1981	222.54	71.03	103.15	48.36	249.69	106.04	95.79	47.86	346.57	132.21	155.41	58.95
1982	251.45	85.59	107.83	58.03	263.30	108.18	102.76	52.36	395.38	154.07	166.05	75.26
1983	283.21	102.10	114.89	66.22	327.95	143.49	116.36	68.10	459.83	185.57	178.75	95.51
1984	332.22	111.46	145.84	74.92	370.04	155.28	136.29	78.47	581.56	222.13	239.27	120.16
1985	396.75	120.34	184.26	92.15	451.74	173.43	170.07	108.24	680.46	235.96	293.07	151.43
1986	436.65	123.45	207.28	105.92	502.91	179.02	202.15	121.74	742.05	252.73	313.21	176.11
1987	521.92	137.66	255.97	128.29	609.60	220.22	230.25	159.13	892.29	287.31	384.57	220.41
1988	701.33	162.31	323.40	215.62	749.09	240.72	299.83	208.54	1117.6	331.94	497.10	288.62
1989	822.83	196.35	374.92	251.56	850.71	289.95	317.13	243.63	1293.9	359.14	579.65	355.15
1990	896.33	227.89	387.52	280.92	934.65	325.77	331.85	277.03	1511.1	425.29	635.98	449.92
1991	1072.0	236.89	459.91	375.27	1045.7	334.61	388.09	323.03	1810.5	521.85	745.90	542.79
1992	1278.5	257.08	573.15	448.27	1279.7	353.92	545.21	380.62	2196.5	534.62	999.11	662.80
1993	1690.8	301.68	847.92	541.24	1660.1	410.45	764.20	485.53	2770.3	596.63	1355.7	818.03
1994	2187.4	451.91	1053.1	682.46	2216.8	546.68	1058.8	611.26	3844.5	775.03	1891.4	1178.0
1995	2849.5	631.34	1322.7	895.41	2988.3	762.99	1394.9	830.40	4953.3	1010.1	2355.7	1587.4
1996	3452.9	700.94	1664.6	1087.4	3634.6	937.64	1677.6	1019.4	5883.8	1200.1	2784.0	1899.5
1997	3953.7	761.76	1934.3	1257.6	4041.0	1008.5	1861.2	1171.2	6537.0	1195.0	3147.3	2194.7
1998	4256.0	790.60	2084.3	1381.0	4308.2	1071.3	1937.8	1299.0	7021.3	1215.8	3408.0	2397.4
1999	4514.1	805.97	2188.5	1519.6	4517.9	1123.1	1981.0	1413.7	7493.8	1221.0	3644.3	2628.5
2000	5043.9	824.55	2514.9	1704.4	5052.9	1161.5	2294.1	1597.2	8337.4	1268.5	4164.4	2904.4
2001	5516.7	913.82	2696.6	1906.3	5533.0	1234.3	2510.4	1788.2	9195.0	1359.4	4556.0	3279.5
2002	6018.2	956.84	2911.6	2149.7	6035.4	1288.3	2768.7	1978.3	10275	1390.0	5184.9	3700.5
2003	6921.9	1064.0	3417.5	2439.6	6867.7	1198.7	3310.1	2358.8	12078	1480.6	6485.0	4112.4
2004	8477.6	1333.5	4301.7	2842.3	8553.7	1649.2	4182.1	2722.4	15021	1778.4	8478.6	4764.7
2005	10012	1400.0	5271.5	3340.5	10587	1892.0	5514.1	3181.2	18366	1963.5	10478	5924.7
2006	11467	1461.8	6110.4	3895.3	12362	1916.7	6724.6	3721.4	21900	2138.9	12574	7187.2
2007	13607	1804.7	7201.8	4600.7	15012	2217.6	8282.8	4511.9	25776	2509.1	14647	8620.2
2008	16011	2034.5	8701.3	5276.0	18018	2658.7	10259	5099.7	30933	3002.6	17571	10358
2009	17235	2207.3	8959.8	6068.3	19480	2769.0	11010	5700.9	33896	3226.6	18901	11768
2010	20394	2562.8	10707	7123.7	23092	3258.0	13226	6607.8	39169	3588.2	21238	14343
2011	24515	2905.7	13126	8483.1	26931	3512.2	15427	7991.7	45361	3973.8	24017	17370

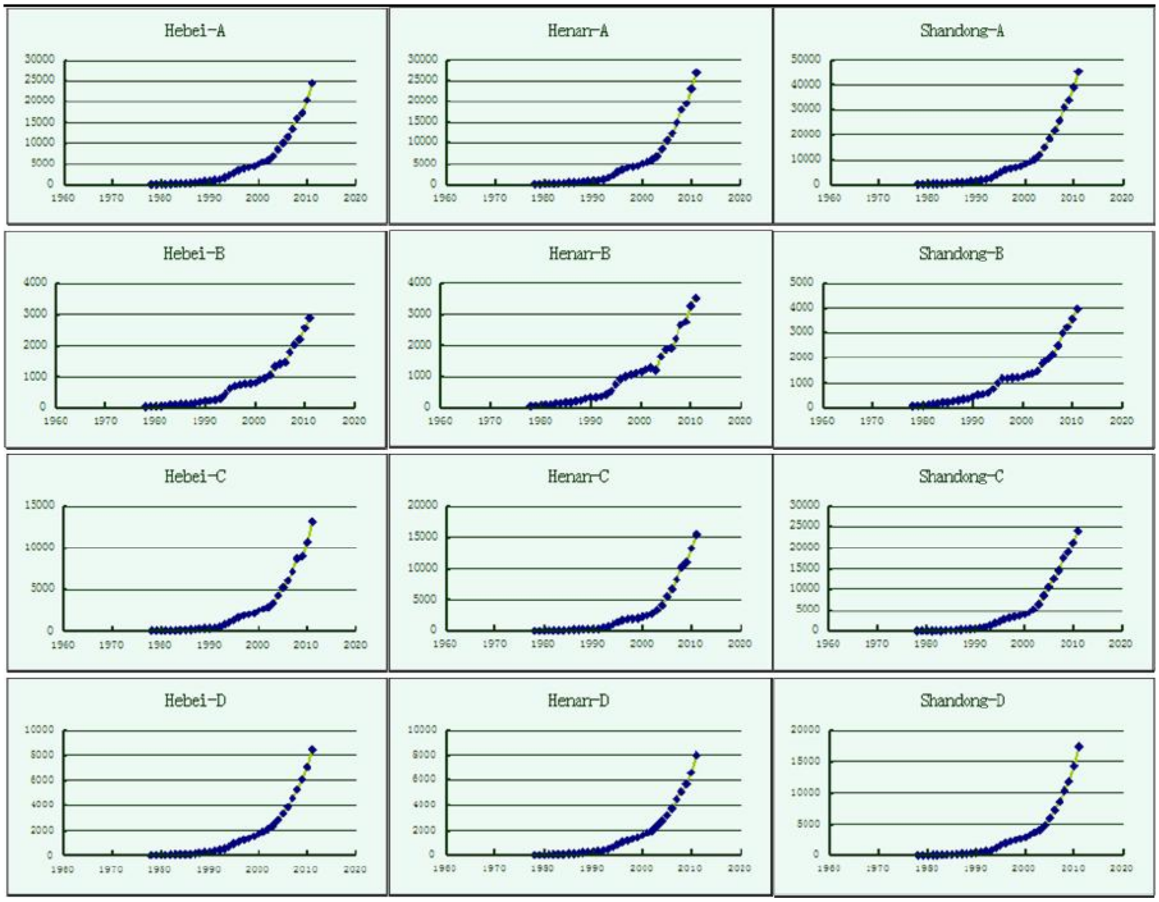

Figure 1 : Each region each industry changes trends with time

show:

$$\begin{cases} y_{1t} = \alpha_0 + \alpha_1 x_{n1} + \alpha_2 x_{n2} + \alpha_3 x_{n3} + a_1 + u_{1t} \\ y_{2t} = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \beta_3 x_{n3} + a_2 + u_{2t} \\ y_{3t} = \gamma_0 + \gamma_1 x_{n1} + \gamma_2 x_{n2} + \gamma_3 x_{n3} + a_3 + u_{3t} \end{cases} \quad (28)$$

In formula(28), $y_{it}\,(i=1,2,3)$ respectively represents Hebei gross regional domestic product, Henan gross regional domestic product and Shandong gross regional domestic product ; α, β, γ respectively represent Hebei each industry regression coefficient, Henan each industry regression coefficient and Shandong each industry regression coefficient, do difference treatment with every period with last period, as formula (29) show:

$$\Delta y_{it} = \mu_1 \Delta x_{it1} + \mu_2 \Delta x_{it2} + \mu_3 \Delta x_{it3} + \xi_{it} \quad (29)$$

By above data, it can get model results as formula(30), (31) and (32)show:

$$\Delta y_{1t} = 1.0009\Delta x_{1t1} + 0.9994\Delta x_{1t2} + 1.0008\Delta x_{1t3} - 0.0358 \quad (30)$$

$$\Delta y_{2t} = 1.0004\Delta x_{2t1} + 1.0004\Delta x_{2t2} + 0.9992\Delta x_{2t3} - 0.0107 \quad (31)$$

$$\Delta y_{13} = 1.0008\Delta x_{3t1} + 1.0003\Delta x_{3t2} + 0.9996\Delta x_{3t3} - 0.1001 \quad (32)$$

Formula(30) obtained regression equation statistical parameters: residue squares sum is 2.3041, multiple correlation coefficient is 1, variance is 1.5179;Formula(31)obtained regression equation statistical parameters:residue squares sum is 1.1053, multiple correlation coefficient is 1, variance is 1.0513;Formula(32)obtained regression equation statistical parameters:residue squares sum is 4.9347, multiple correlation coefficient is 1, variance is 2.2214. From above statistical parameters, it is clear that apply the method can better obtain each gross regional domestic product and each industry gross value transformation relations.

CONCLUSIONS

The paper applies Panel Data model better researching Hebei, Henan and Shandong three regions each industry gross value and gross regional domestic product relations from opening-up and reform to 2011; each industry output value and gross regional domestic product have higher correlations, random factors are in mutual independent normal distribution, which has no big effects on model construction; The paper established model applies least square principle getting very high goodness model result, and variance value is very little, the model can research on future several years' data status in certain ranges.

REFERENCES

[1] Chen Fu-Chuan; Hainan province regional economic development advantages analysis [J]. Journal of Hainan University (Humanities and Social Science), **27(6)**, 601-606 **(2009)**.

[2] Yuan Li-Bing; Discuss on regional economic growth dynamic model[J]. Journal of Xuzhou Normal University (Natural Science), **29(1)**, 117-120 **(2003)**.

[3] Fei Zhong-Hua; Regional economic development strategic mathematical model and empirical analysis[J]. Mathematics in Practice and Theory, **37(15)**, 12-19 **(2007)**.

[4] Li Da-Ming; Regional economic development evaluation indicator system and evaluation method[J]. Journal of Xian Institute of Finance, **21(5)**, 27-31 **(2008)**.

[5] Zheng Shi; Regional economic system ARX parameters model estimation[J]. Mathematics in Practice and Theory, **37(9)**, 6-11 **(2007)**.