CHAPTER 5. EMPIRICAL RESULTS AND MODEL TEST

The purpose of this study is to examine whether the income inequality affect the social welfare spending positively or negatively. The estimating model of this study also includes other variables, such as area variable, time variable, election variable, dependent population variable and so on,[58] to explain social welfare spending. In order to avoid endogeneity, the main methodology this study employs is two-stage least squares (2SLS). Finally, several tests are conducted in this chapter to ascertain whether the empirical result is reliable.

5.1 Empirical Results

This study uses 1994-2006 panel data of the cities and counties of Taiwan to estimate equation (4-13), the impact of income inequality on social welfare spending, and the estimation result is presented in Table 9. According to Table 9, the Hausman test provides strong evidence that the model should be estimated by using 2SLS model at 1% significance level. Moreover, heteroskedasticity problems of the first stage and the second stage models are corrected.[59] In addition, both of the *F*-statistics reject the null hypothesis in α=0.01, which assumes that all coefficients in the models are zero. In the following paragraph, the estimating results of Gini index and its quadratic term will be stated first. Then, the estimating results of the other independent variables will be mentioned.

5.1.1 The Impact of Gini Index and the Square of Gini Index

The results reported in Table 9 show that the impact of Gini index on the

58 Dependent population is measured by the local proportion of population over 65 years of age in this study.

59 χ^2 statistics based on the Breusch-Pangan test of the first stage model and second stage model both reject the critical value at 1% significance level. Furthermore, about the details of Breusch-Pangan test, please refer to Greene (2003).

Table 9: Estimation Result of Two-Stage Least Squares Model (Second Stage)

Variable	Coefficient	t-ratio
B_0	955.36***	3.400
G	-59.64***	-3.268
GSQ	0.91***	3.054
Log(DI)	-1.59	-.997
OLD	1.58***	10.099
Log(EXP)	1.74***	6.925
Log(DEF)	2.47×10^{-2}	0.401
GREEN	1.43**	2.269
ELECT	1.34**	1.978
NORTH	-3.86**	-2.529
MIDDLE	-4.95***	-3.469
SOUTH	-4.37***	-3.402
T	4.66×10^{-2}	0.475
Sample Size	299	
Adjust R^2	0.26	
F-Statistic	9.76***	
RESET Test	1.640	
Hausman Test	4.298***	

Notes: 1. Symbols ***, **and * denote that null hypothesis of estimated parameter equal to zero is rejected at 1%, 5% and 10% significance level.

proportion of local social welfare spending to total local expenditure (SW) is negative and significant in α=0.01. It means that the result supports the theory of imperfect capital market brought up by Benabou (2000). In an imperfect capital market, redistribution, defined as SW in this study, which would increase ex ante welfare gains less support in an unequal society than in a homogenous society.[60] Besides, capital market imperfections make it easier for the rich to become richer and then make the society more unequal. Therefore, the political support on redistribution

[60] It is because the redistribution will bring positive ex ante welfare to those who will get benefit from the redistribution, but negative ex ante welfare to those who will lose because of the redistribution. Therefore, those who can get positive ex ante welfare will support redistribution policies, but those who get negative ex ante welfare will oppose to redistribution policies. However, in a homogenous society, people feel the same way about redistribution since everyone is the same in a homogenous society, so no one will oppose to redistribution policies.

(social welfare spending is one kind of redistribution) will decrease with inequality in the societies with capital market imperfections.

Apart from the negative impact of Gini index on SW, the model also shows a positive and significant impact of the square of Gini index on SW in α=0.01. The result indicates that there is a U-shaped relationship between Gini index and SW, as depicted in Figure 9. In Figure 9, Gini index is set to be 20% to 40%,[61] and other explanatory variables are controlled at their mean values respectively. Furthermore, the constant term and the coefficients of all explanatory variables are the same with the coefficients of our estimating model. The critical value that makes the impact of Gini index on SW turn from negative to positive is about 32.77% in our model. It can be seen in Table 5 that Gini indexes of cities and counties of Taiwan from 1994 to 2006 are all below 32.77%, and that's why our model shows a negative impact of Gini index on SW.

The U-shape result is consistent with the theory of Benabou (2000) and the empirical test of Mello and Tiongson (2006). The U-shaped relationship means that the negative impact of Gini index on SW will weaken with the increase of Gini index. In other words, the negative impact of income inequality on SW will weaken with the deterioration of wealth distribution. Hence, when the Gini index is larger than a critical value, which means inequality problem is serious to a certain extent, more unequal societies will finally turn to spend more on social welfare.

The reason that causes the U-shaped relationship between income inequality and redistribution explained by Benabou (2000) is that as inequality rises, the proportion of those who will lose ex ante welfare from redistribution, usually the rich, increases.

[61] Gini index is set to be 20% to 40% in Figure 9 because the range of Gini index is more reasonable for Taiwan.

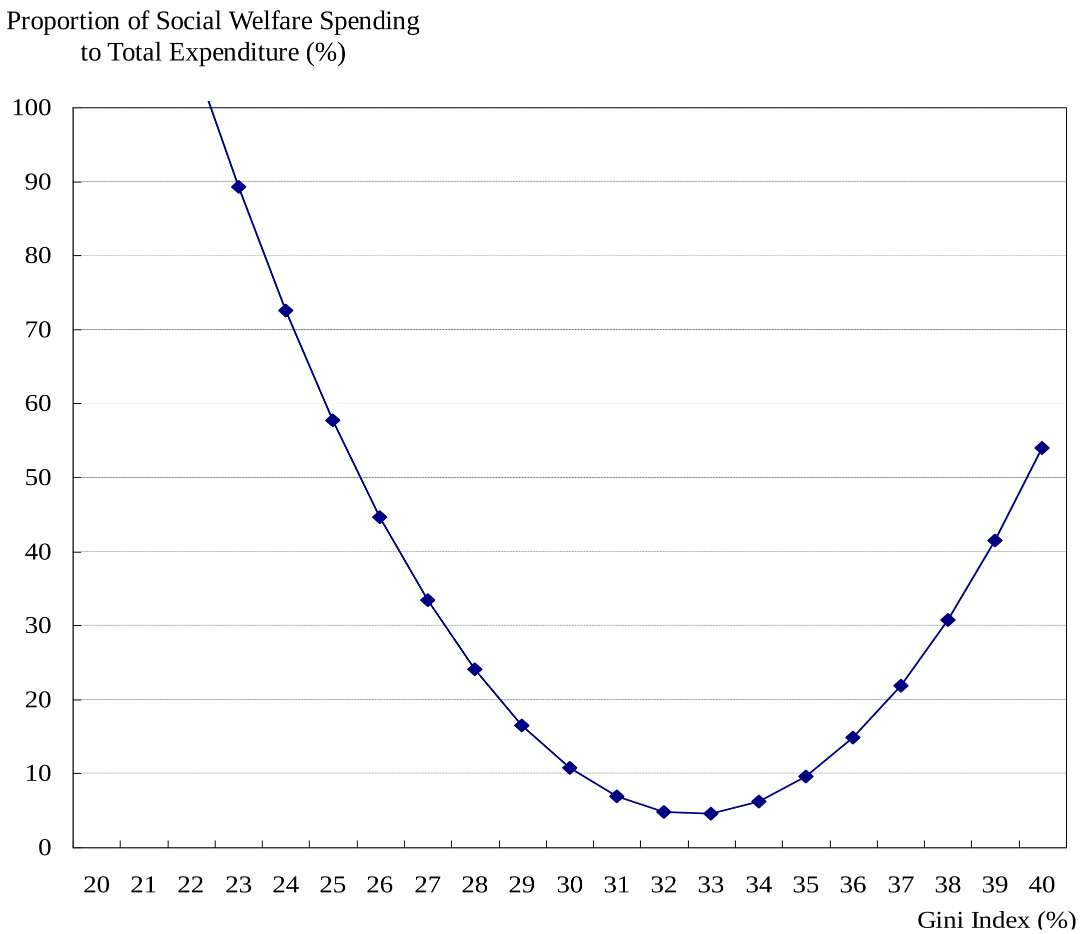

Figure 9: A U-shaped Relationship between Gini Index and Social Welfare Spending

Because the income distribution is right-skewed, the median income is below the mean income. Also, the increase of the proportion of the rich will raises the mean income. As a result, the proportion of those with endowments below the mean increases. And there is no doubt that those with income endowment below the mean income will support redistribution.[62] Thus, redistribution will take place. So when the inequality is large enough, the skewness effect finally dominates. To short, as the inequality rises to some extent, the skewness effect dominates, negative impact of income inequality on redistribution weakens, and finally there will turn to be a positive impact of income inequality on redistribution.

[62] The redistribution is defined as a complete redistribution. That is, resources are pooled and redistributed to individuals equally.

5.1.2 The Impact of Other Explanatory Variables

Apart from Gini index and its quadratic term, there are also other explanatory variables listed in Table 9. Except for local average disposable income per capita, local public deficit and time trend variable, other variables in the model are all significant.

In terms of the population over 65 years of age, the empirical result presented in Table 9 indicates that local proportion of population over 65 years of age has a significant and positive effect on SW at 1% significance level. As we mentioned in chapter 2, Mello and Tiongson (2006), Milanovic (2000), Perotti (1996) and Bassett, Burkett and Putterman (1999) all found significant and positive impact of population over 65 years of age on social welfare spending in their models. The result makes sense because people over 65 years of age are those who need social welfare service. Therefore, cities and counties with higher proportion of population over 65 years of age will spend more on social welfare.

Furthermore, total local expenditure also has a significant and positive impact on SW at 1% significance level. The total local expenditure represents the scale of local finance in this study and Wang (2003).[63] And it is assumed that larger scale of local finance will cause more social welfare spending because the government will have more financial resources if the scale of local finance is large. The result of our model is consistent with the assumption.

As for the local political party variable, all cities and counties are categorized into two parts, the pan-blue coalition and the pan-green coalition, by the political preference of the county magistrates or city mayors. The estimating result indicates that the SW is larger if the county magistrate/ city mayor is pan-green. Namely,

[63] Total local expenditure is considered as a factor of scale of local finance in Wang (2003).

pan-green governments tend to spend more on social welfare. This result is consistent with that of Wang (2003) and Wu (2007). Both Wang (2003) and Wu (2007) contended that the higher spending on social welfare of pan-green governments, relative to pan-blue governments, is related to the composition of their supporters. Most supporters of the Pan-Green Coalition are blue-collar, and blue-collar people prefer more social welfare spending. Apart form that, most of the people who live in the regions, which mainly focus on conventional industries, are supporters of Pan-Green Coalition, and workers in conventional industries are also those who need social welfare services. However, this study want to point out that the main reason for the higher social welfare spending of Pan-Green Coalition is its own ideology. Pan-Green Coalition is more left-wing. In politics, the left are those who seek to reform or abolish existing social hierarchies and promote a more equal distribution of wealth and privilege.

In terms of the election variable, *ELECT*, which is defined by the election of county magistrates and city mayors. In order to win more votes, politicians usually like to offer more social welfare service before election, so this study expects a positive impact of *ELECT* on *SW*. And the empirical result meets the expectation. That is, election of county magistrates and city mayors affects SW positively and that is because candidates want to please voters. The result might respond to the situation that candidates all took social welfare policy as one of their main policies in the 1993 election for county magistrates and city mayors.

Finally, coefficients of the area dummy variables in the model, *NORTH, MIDDLE* and *SOUTH*, are all negative and significant. This means that north area, middle area and south area all spend less on social welfare relative to the east area. That is to say, the SW is largest in the east area. Though Table 7 seems to show that

SW is higher in the north area than in the east area, the empirical model still tells us that SW is highest in the east area. It is because coefficients of area variables represent the effect of area difference on SW when the other variables are controlled to be invariable. So the information provided by Table 7 might be a mistaken impression. Furthermore, reasons that cause the east area to have the highest SW might be the mountain topography, the lack of resources and the close relationship between people in the east area.

5.2 Specification Test

In order to ascertain whether the empirical results can provide accurate conclusions, Hausman test (correlation between explanatory variables and the error term), Regression Specification Error test (misspecification), and the test for multicollinearity are conducted in the following. [64]

5.2.1 Testing for Correlation between Explanatory variables and the Error Term (Hausman Test)

As we mentioned above, endogeneity problem might cause a non-zero correlation between explanatory variables and the error term. Models cannot be unbiasedly and consistently estimated by using ordinary least square method if there is an endogeineity problem. Therefore, it is necessary to use the Hausman test to examine whether there is a correlation between explanatory variables and the error term.

In the regression model of this study

$$SW = \beta_0 + \beta_1 Gini + \beta_2 Z + u \qquad (5\text{-}1)$$

[64] Details about Hausman test and RESET test, please refer to Hill et al. (2001).

where *Gini* is the Gini index of the cities and counties of Taiwan, and *SW* is the social welfare spending in percent of total local expenditure. *Z* represents other explanatory variables of the proportion of social welfare spending to total local expenditure (*SW*). In equation (5-1), we wish to know that whether *Gini* is correlated to *u*.

Let *X* be other explanatory variables of Gini index. That is, *X* are instrumental variables of *Gini*. Then, estimate the model

$$Gini = \alpha_0 + \alpha_1 X + e \tag{5-2}$$

And obtain the residual

$$\hat{e} = Gini - \hat{\alpha}_0 - \hat{\alpha}_1 X \tag{5-3}$$

Then, include the residual computed above as an explanatory variable in (5-1).

$$SW = \beta_0 + \beta_1 Gini + \beta_2 Z + \beta_3 \hat{e} + u \tag{5-4}$$

Estimate (5-4) by ordinary least square, and employ t-test for the null hypothesis H_0: $\beta_3 = 0$ against the alternative hypothesis, If H_1: $\beta_3 \neq 0$. If H_0 is rejected, it means that there is no correlation between *Gini* and *u* and the model can be estimated by using ordinary least square model. However, if the test fails to reject H_0, it means that there is not enough evidence to infer that the model has no correlation between *Gini* and *u*. Then, the model should be estimated by using a 2SLS model.

The result of Hausman test of the model indicates that the corresponding *p*-values of the coefficient of $\hat{e}$, 4.298, is significant in α=0.01, so there are enough evidences to reject H_0. Therefore, the Hausman test suggests that the model should be estimated by using 2SLS model.

5.2.2 Testing for Model Misspecification: The Regression Specification Error test (RESET Test)

This study adopts the RESET test to test model misspecification. RESET test is used to detect omitted variables and incorrect functional forms. It can examine whether the functional form of models suffered from misspecification errors and omitted variable bias. Furthermore, it is also a general test for the correlation between independent variables and the error terms caused by measurement error in the independent variables, simultaneity considerations, or the combination of a lagged dependent variable with autocorrelated error terms.

RESET is calculated by regressing the residuals on the independent variables and the square of the fitted dependent variable. It proceeds as follows. First, let the predicted values of the *SW* be

$$S\hat{W} = \beta_{0i} + \sum_{k=1}^{K} \beta_k X_{kit} \qquad (5\text{-}5)$$

where *k* represents the *k*'th explanatory variables, *i* and *t* represent respectively *i*'th city/county and *t*'th year. Then add the predicted values of *SW* to the regression model:

$$S\hat{W} = \beta_{0i} + \sum_{k=1}^{K} \beta_k X_{kit} + \gamma_1 S\hat{W}_{it}^2 + \varepsilon_{it} \qquad (5\text{-}6)$$

In equation (5-2), a test for misspecification is a test of the null hypothesis

H_0:$\gamma_1 = 0$ against the alternative hypothesis, H_1:$\gamma_1 \neq 0$.

If H_0 is rejected, it means that the original model is inadequate and can be improved. However, if the test fails to reject H_0, it means that there is not enough evidence to infer that the model has any misspecification problems. That is, if this study can significantly improve the model artificially, including powers of the predictions of the model, then the original model must have been inadequate. Adding the square of *SW* prediction to estimate the model, the result of RESET test indicates that the corresponding *p*-values of the coefficient of $S\hat{W}^2$, 1.64, is above the conventional significance level of 0.1. Therefore, there is insufficient evidence from the RESET test to suggest that the model is inadequate. The lack of significance of RESET test implies that the model of this study is without misspecification problems.

5.2.3 Testing for Model Multicollinearity

It is common that variables may move together in systematic ways. Such variables are known as collinear variables. The collinearity problem may lead to incorrect conclusions and which is called collinearity or multicollinearity. Multicollinearity is a common problem of regression model analyses. A simple test of multicollinearity is to look at the correlation coefficients among explanatory variables. Generally, the correlation coefficients among explanatory variables can not exceed 0.8. And all correlation coefficients among explanatory variables of this study are not over 0.8. Consequently, there is no multicollinearity problem occurring in this study.