CHAPTER 4. METHODOLOGY

Mello and Tiongson (2006) and Huang and Liu (2005) both mentioned that there exists simultaneous problem between income inequality and redistribution. That is, income inequality may have influence on redistribution and redistribution may also have influence on income inequality at the same time. To solve the simultaneity problem, the study adopts the two stage-least squares (2SLS) model. In this chapter, this study will interpret the 2SLS model. Subsequently, this study will show the empirical results of this model in order to understand the impact of the income inequality on the social welfare spending

4.1 Two-Stage Least Squares Model

As we mentioned above, there exists simultaneous problem between income inequality and redistribution. That is, income inequality may have influence on redistribution and redistribution may also have influence on income inequality simultaneously. In order to solve the simultaneity problem, this study adopts the 2LSL model.

The simultaneous equations of this study are as follows:

$$Gini = \alpha_0 + \alpha_1 SW + \alpha_2 X + e \quad (4\text{-}1)$$

$$SW = \beta_0 + \beta_1 Gini + \beta_2 Z + u \quad (4\text{-}2)$$

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. X represents other explanatory variables of Gini index, and Z represents other explanatory variables of the proportion of social welfare spending to total local expenditure (SW). $Gini$ and SW are endogenous variables. In other words, they are jointly determined with each other.

Therefore, the equation we are interested in should be viewed as part of a system of relationships. Familiar examples include market equilibrium, models of macroeconomics and set of factor or commodity demand equations.

If we express endogenous variables, *Gini* and *SW*, in terms of the exogenous variables only, we can get the reduced form of this model. That is:

$$Gini = \frac{(\alpha_0 + \alpha_1\beta_0)}{1-\alpha_1\beta_1} + \frac{\alpha_1\beta_2}{1-\alpha_1\beta_1} Z + \frac{\alpha_2}{1-\alpha_1\beta_1} X + \frac{(\alpha_1 u + e)}{1-\alpha_1\beta_1} \tag{4-3}$$

$$SW = \frac{(\beta_0 + \alpha_0\beta_1)}{1-\alpha_1\beta_1} + \frac{\alpha_2\beta_1}{1-\alpha_1\beta_1} X + \frac{\beta_2}{1-\alpha_1\beta_1} Z + \frac{(\beta_1 e + u)}{1-\alpha_1\beta_1} \tag{4-4}$$

As a result,

$$\begin{aligned} Cov(Gini,,u) &= Cov(\frac{(\alpha_0 + \alpha_1\beta_0)}{1-\alpha_1\beta_1} + \frac{\alpha_1\beta_2}{1-\alpha_1\beta_1} Z + \frac{\alpha_2}{1-\alpha_1\beta_1} X + \frac{(\alpha_1 u + e)}{1-\alpha_1\beta_1}, u) \\ &= \left(\frac{\alpha_1}{1-\alpha_1\beta_1}\right)^2 \sigma_u^2 \neq 0 \end{aligned} \tag{4-5}$$

In (4-5), the covariance between *Gini* and *u* is not equal to zero. That is the disturbance term of *SW* is correlated with the independent variable, *Gini*. And this violates one of the assumptions of ordinary least square model. Consequently, this model cannot be unbiasedly and consistently estimated by ordinary least square model.

The solution of this problem is to adopt 2SLS model. In the first stage, a new endogenous variable, $\hat{G}$, is created by the following regression.

First, we assume that

$$Gini = Gini^m + q \tag{4-6}$$

and

$$Gini^m = a_0 + a_2 X \tag{4-7}$$

where $Gini^m$ has no relationship with the error term *u*. However, *q* is a term that is related to the error term *u*. Then, we can get

$$Gini = (\alpha_0 + \alpha_2 X) + \alpha_1 SW + e = Gini^m + q \tag{4-8}$$

The main model of the first stage is:

$$Gini = \alpha_0 + \alpha_2 X + q \tag{4-9}$$

And $\hat{\alpha}_0$ and $\hat{\alpha}_2$ can be estimated by employing ordinary least square model to (4-9).

Finally,

$$\hat{G} = \hat{\alpha}_0 + \hat{\alpha}_2 X \tag{4-10}$$

By (4-10), we also know that

$$\hat{G} = \hat{G}^m \tag{4-11}$$

Since $\hat{G}$ is uncorrelated with the error term *u*, it does not violate the assumption of ordinary least square. So we can use $\hat{G}$ as the independent variable, and the regression function of second stage is:

$$SW = \beta_0 + \beta_1 \hat{G} + \beta_2 Z + u \tag{4-12}$$

And then $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ can be estimated by ordinary least square fashion.[39]

[39] Theory of 2SLS model is referred to Hill et al. (2001).

4.2 Empirical Model and Data Source

This study uses 2SLS model to examine whether more unequal societies spend more on redistribution. In the first stage, we modify the model in Huang and Liu (2005) and use the modified model to predict Gini index, $\hat{G}$. There are 3 differences between the first stage model of this study and that of Huang and Liu (2005). First, social welfare variable is not one of the explanatory variable in this first stage model, but it is in the model of Huang and Liu (2005).[40] It is because the main purpose of the first stage is to create a Gini index which is uncorrelated with the error term of the second stage model. Social welfare variable is the dependent variable of second stage, so it can not be included in the first stage model.

Second, the data of Gini index is different in these two studies. As we mentioned in 2.1, this study takes weight into consideration when computing Gini index,[41] but Huang and Liu (2005) did not. Third, instead of using a dummy variable, Y2001, to control the time trend, this study use T, denote the year of the data of social welfare variable, to control the time trend. It is because the situation that the value of Gini index is extremely high in 2001 does not happen to the Gini index in this paper. However, it is still necessary to consider the time trend, so this study puts T into the model.

The data of the first stage model are collected from Department of Budget, Accounting and Statistics of every county and city government, Ministry of Audit of Republic of China,[42] *Report on the Survey of Family Income & Expenditure* and

[40] Huang and Liu (2005) used social welfare spending as the social welfare variable. This study uses the proportion of social welfare spending to total local expenditure, which is denoted as SW in this study, as the social welfare variable.

[41] The weight means how many households are the same with an observation. Namely, if the weight equals to 10, it means that the observation represents 10 households and the income status of those 10 households are the same.

[42] Settled account of social welfare expenditure from 1994-1999 is collected from Department of

Yearbook of Manpower Survey Statistics.[43] And this study also updates the data to 2006 year.

As to the second stage, it is the key model that this study is interested in, and the dependent variable of it is the social welfare spending in percent of total local expenditure (SW). According to the literature, SW is a function of the Gini index (*G*), the square of Gini index (*GSQ*),[44] average disposable income per capita(*DI*),[45] the proportion of population over 65 years of age (*OLD*), total local expenditure (*EXP*), public deficit of county and city government (*DEF*),[46] political party dummy variable (*GREEN*),[47] local election dummy variable (*ELECT*), area dummy variables (*NORTH*, *MIDDLE* and *SOUTH*) and time variable (*T*). Therefore, the second stage model of this study is established as follows:

$$\begin{aligned} SW = {} & \beta_0 + \beta_1 G + \beta_2 GSQ + \beta_3 \log(DI) + \beta_4 OLD + \beta_5 \log(EXP) + \\ & \beta_6 \log(DEF) + \beta_7 GREEN + \beta_8 ELECT + \beta_9 NORTH + \\ & \beta_{10} MIDDLE + \beta_{11} SOUTH + \beta_{12} T + u \end{aligned} \tag{4-13}$$

In equation (4-13), β_0 is a constant term and *u* represents the error term with zero mean and variance σ^2.[48] Furthermore, apart from *G*, *GSQ*, *GREEN*, *ELECT*, *NORTH*, *MIDDLE*, *SOUTH* and *T*, the other independent variables are lagged one year in order to avoid any potential endogeneity problems between any dependent and

Budget, Accounting and Statistics of every county and city government; settled account of social welfare expenditure from 2000-2006 is collected from Ministry of audit, R.O.C..

[43] Average income of households is collected from *Report on the Survey of Family Income & Expenditure.* Population of employees in good-producing industries, Population of employees in service-producing industries, female labor participation rate and education variation are collected from *Yearbook of Manpower Survey Statistics.*

[44] Both Benahou (2000) and Mello and Tiongson (2006) support that the association between income inequality and redistribution is nonlinear.

[45] For lack of regional GDP data in Taiwan, this study uses the average disposable income per capita instead.

[46] Public deficit is defined as the total expenditure subtracts real revenue. And the real revenue is: total revenue - receipts from public debt - balance of previous years – subsidies.

[47] Kuomintang (KMT), People First Party (PFP) and New Party (NP) are classified as Pan-Blue coalition. Democratic Progressive Party (DPP) is classified as Pan-Green Coalition.

[48] It is normally distributed and also satisfies the general assumption of an independent and identical distribution (iid).

independent variables. In order to eliminate any fluctuations in prices, all value variables in this study are adjusted by the CPI deflator (base year=1994) of Taipei city, Kaohsiung city and Taiwan province respectively.

The data of second stage model is collected from the Department of Budget, Accounting and Statistics of every county and city government,[49] Ministry of Audit of Republic of China,[50] Central Election Commission,[51] *Report on the Survey of Family Income & Expenditure,*[52] and *Urban and Regional Development Statistics Republic of China (Taiwan)*.[53] Because issues about social welfare have been widely talked about since the 1993 election for county magistrates and city mayors, the research period covered by this study is 1994 to 2006.

The data set used in this study is a panel data set of Taiwan's 23 counties and cities for 1994-2006. The advantage of panel data is that, compared with time series or cross-sectional data, such data contains more information and observations. Due to the larger sample size, the use of panel data can increase the number of degrees of freedom and make the estimation more accurate. To keep consistency of the data, this study excludes receipt from debt and balance of previous years from all total revenues of local governments. And it also excludes expenditure for sinking fund from all total expenditures of local governments. What is more, this study also adjusts the data to

[49] Settled account of social welfare expenditure, settled account of total expenditure, settled account of total revenue, receipt from debt, the balance of previous years, expenditure for sinking fund and subsidies of 1994-1999 are all collected from Department of Budget, Accounting and Statistics of every county and city government.

[50] Settled account of social welfare expenditure, settled account of total expenditure and settled account of total revenue of 2000-2006 are collected from Ministry of audit, R.O.C..

[51] Regional Election dummy variable and the political party dummy variable are collected from Central Election Commission.

[52] Gini coefficient is calculated by using the data of *Report on the Survey of Family Income & Expenditure* . And average disposable income per capita is collected from *Report on the Survey of Family Income & Expenditure*.

[53] Population over 65 years of age is collected from *Urban and regional development statistics Republic of China (Taiwan).*

meet the current accounting year of Taiwan.[54] The definitions, descriptive statistics, and expected signs of all of the variables are listed and described in Table 8 and analyzed in the paragraphs below.

G represents local Gini index and it's the variable that this study concerns most. Theoretically, Meltzer and Richard (1981) addressed that Gini index has a positive impact on social welfare spending, but Benabou (2000) and Rodriguez (2004) declared that Gini index has a negative impact on social welfare spending. Theoretical debates over the issue also brought about a lot of empirical studies on the issue, such as Meltzer and Richard (1983), Gouveia and Masia (1998), Moffitt, Riber and Putterman (1999), Milanovic (2000), and Mello and Tiongson (2006).[55] This study expects the impact of Gini index on SW to be negative because the mechanism that the rich is easier to become richer through investment in the imperfect market happens in Taiwan. In fact, almost no capital market is completely perfect in the real world. Therefore, this study expects a negative impact of income inequality on social welfare spending, and which is based on the imperfect capital market of Taiwan.[56]

GSQ represents the square of local Gini index. As we known, Benahou (2000) addressed a negative impact of income inequality on redistribution. However, he also addressed that, as the inequality rises to some extent, the skewness effect dominates,[57] negative impact of income inequality on redistribution weakens, and finally there will

[54] The accounting year of Taiwan was changed from July/1-June/30 to January/1-December/31 in 2000, so the official data of 2000 included 1.5 year value, not 1 year.
[55] Details about these literatures are in the 2.3 section of this study.
[56] Wang and Hsu (2002) addressed that as for the derivatives, some mature markets, like S&P 500 index future market, are more perfect. However, some emerging markets, like SGX-DT Taiwan stock index and TAIFEX Taiwan index futures, are less perfect.
[57] Benabou (2000) said that, as inequality rises, the proportion of those who will lose ex ante welfare from redistribution, usually the rich, increases. Because the income distribution is right-skewed, the median income is below the mean income, and 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 those with income endowment below the mean income will support redistribution. Thus, redistribution will take place. So when the inequality is large enough, the skewness effect finally dominates.

be a positive impact of income inequality on redistribution. That is, redistribution is U-shaped with respect to inequality. Furthermore, the empirical study of Mello and Tiongson (2006) also supported a nonlinear relationship between income inequality and redistribution. That's why this study includes the square of Gini index in our model.

DI represents local average disposable income per capita. Local average disposable income per capita is used because of the lack of local GDP data in Taiwan. Mello and Tiongson (2006) and Bassett, Burkett and Putterman, (1999) considered GDP an important variable that influences social welfare spending. Both of their studies support the positive influence of GDP on social welfare spending. It is because high GDP implies high ability of the government to provide social welfare service. Therefore, the influence of GDP is positive. In addition, income level is one of the factors of the degree of industrialization in Wang's (2003) article. Wang (2003) addressed that the higher the degree of industrialization, the more the demand of social welfare since higher degree of industrialization will bring social problems. Therefore, this study expects GDP to positively affect social welfare spending.

OLD represents local proportion of population over 65 years of age. Intuitively, the higher the proportion over 65 years of age, the more the social welfare spending is, because the old are those who need social welfare spending. As we know, Mello and Tiongson (2006) found that population over 65 years of age is a significant and positive determinant of social welfare spending. Moreover, Milanovic (2000) found that population over 65 years of age is positive and significant in a few models, but insignificant in others. What is more, Perotti (1996) and Bassett, Burkett and Putterman (1999) both found significantly positive effect of population over 65 years of age on social welfare spending. Proportion over 65 years of age is also considered

in dependent population factor in Wang's (2003) study. Hence, this study expects a positive effect of population over 65 years of age on social welfare spending.

EXP represents total local expenditure and it is the measurement of fiscal scale of local government in this study. Wang (2003) addressed that if the fiscal scale of the local government is high, then the financial resources they can utilize is also relatively high. What is more, the enactment and enforcement ability of public policy and the fiscal autonomy will also be high in the cities and counties with high fiscal scale. Therefore, if the demand for social welfare is high, then the city/ county with high total local expenditure will also spend more on social welfare. Wang (2003) also found a positive influence of total local expenditure on social welfare spending empirically. However, the total local expenditure is also the denominator of the dependent variable in this study, so the impact of *EXP* on the dependent variable, SW, is also likely to be negative.

DEF represents local public deficit. Intuitively, the local public deficit is negatively affects the social welfare expenditure. If the local fiscal deficit is high, the government's ability of offering social welfare service is limited. Therefore, this study expect a negative impact of local public deficit on social welfare spending. The model of this study also includes local political party dummy variable (*GREEN*) and local election dummy variable (*ELECT*). As for the political variable, Wang's (2003) study found that the social welfare expenditure is more when the ruling party is pan-green than pan-blue. Wu (2005) and Wu (2007) also agreed with that.

As for the election variable, *ELECT* is defined by the election of county magistrates and city mayors because this study is analyzing the data of counties and cities of Taiwan. In order to win more votes, politicians usually like to offer more social welfare service before election, so this study also takes election into

consideration and expects a positive impact of *ELECT* on *SW*. Finally, this study also includes area dummy variables and time variable (*T*) to control area and time differences. As for the area dummy variables, this study takes *EAST* as reference group and put *NORTH*, *MIDDLE* and *SOUTH* dummy variables into the empirical model.

Table 8: Descriptive Statistics and Definitions of Variables

Variables	Definitions	Mean (S.D.)	Expected Sign
SW	The proportion of local social welfare spending to total local expenditure. (%)	10.89 (4.15)	
G	Local Gini index. (%)	30.06 (0.68)	—
GSQ	The square of local Gini index. (%)	904.16 (41.08)	+
DI	Local average disposable income per capita. (Thousand NTD)	189544.41 (33178.63)	+
OLD	Local proportion of population over 65 years of age. (%)	9.24 (2.04)	+
EXP	Total local expenditure. (Thousand NTD)	23971974 (25916686)	?
DEF	Local public deficit. (Thousand NTD)	8468860 (5491950)	—
GREEN	Local political party dummy variable. The dummy variable equals to 1 if the county magistrate/ city mayor is pan-green. If not, then the dummy variable equals to 0.	0.47 (0.5)	+
ELECT	Local election dummy variable. The dummy variable equals to 1 if there is an election of county magistrates or city mayors. If not, then the dummy variable equals to 0.	0.24 (0.43)	+
EAST	Area dummy variable. East is the reference group of area dummy variables.	0.09 (0.28)	
NORTH	Area dummy variable. The dummy variable equals to 1 if the county/city is in the north area of Taiwan. If not, then the dummy variable equals to 0.	0.30 (0.46)	—
MIDDLE	Area dummy variable. The dummy variable equals to 1 if the county/city is in the middle area of Taiwan. If not, then the dummy variable equals to 0.	0.26 (0.44)	—
SOUTH	Area dummy variable. The dummy variable equals to 1 if the county/city is in the south area of Taiwan. If not, then the dummy variable equals to 0.	0.35 (0.48)	—
T	Time variable represents the year of the data of SW (the dependent variable).	7 (3.75)	+

Source: Department of Budget, Accounting and Statistics of every county and city government, Ministry of Audit of Republic of China, Central Election Commission, *Report on the Survey of Family Income & Expenditure*, and *Urban and regional development statistics Republic of China (Taiwan).*

Note: 1. Except for *G*, *GSQ*, *GREEN*, *ELECT*, area variables and *T*, the other independent variables are lagged one year.

2. Keelung city, Taipei county, Taipei city, Yilan county, Taoyuan county, Hsinchu county and Hsinchu city are included in the north area. Miaoli county, Taichung county, Taichung city, Changhua county, Nantou county and Yunlin county are included in the middle area. Chiayi county, Chiayi city, Tainan county, Tainan city, Kaohsiung county, Kaohsiung city, Pintung county and Penhu county are included in the south area. Taitung county and Hualien county are included in the east area.