ECOM90003 Applied Microeconometrics Modelling: Paper Overview & Replication

Zelin Chen - 797036

September 18, 2017

Introduction

The phenomenon of rapid growth of Diasability Insurance (DI) payments and Supplemental Securities Income (SSI) over the last 30 years has attracted Dan, Kermit and Seth's attention.[1] They considered its reason as an interesting public policy issue. In labour economics and public policy analysis, many existing literatures have found evidence of correlation between the reduction in labour force participation of less educated men and the reduction in their wages. Authors contributed to these fields from another aspect that whether the reduction of low-skilled men's earnings is correlated with increase in disability payments. There has been some studies on the causation of growth of disability payments, although it has been in huge debate on the validity of the result. Donald O.Parsons (1980, 1982) found that the increase of disability program payments had a large impact on the reduction of labour-force participation. On the other hand, several scholars, although did not deny the correlation between these two factors, but argued that the effect is not that significant due to potential endogeneity issue in the estimations. The nature of this debate is about the elasticity to which disability participation and labour-force participation substitute. Authors' provided more insights into this discussion. Empirically, they found significant impact of permanent job creation and destruction on disability program payments, while transitory changes had little effect. This article will provide an overview of the paper to illustrate how logically authors come up to the certain conclusion, and what econometric techniques were involved. Some replications of statistical results and their commands are also provided if the data is available. [1]

I. Program Description

This section in the paper provides detailed description and economic insights on the two dependent variables that will be analysed later. From the comparison on nature, eligibility and payment structure of DI and SSI programs, they found a key fact that low-wage workers face relatively smaller opportunity cost on their labour income when they decide to withdraw from the labour market.

II. The Impact of the Coal Boom and Bust on Local Economies

This section introduces impact of coal boom and bust on local economy, but, deeply, it was to explain that, due to the exogenous shock of coal price, county coal reserves may be a valid instrumental variable

[1] There might exist a formatting issue when copy stata code back to do-file. In case that copy&paste does not work, please sent email to zelinc@student.unimelb.edu.au, do files for each table is available up on request.

for change of county earnings for regression analysis in the next section. They showed that coal price is determined exogenously due to several economic shocks, which is considered to be not correlated county level variables. So, these IVs should meet the exogeneity condition.

For the relevance, they took several steps to reveal this relationship. Firstly, they compared men and women's employment conditions between coal and non-coal areas in both boom and bust period. They found that, basically, the increase of earnings in coal area during boom period, was mainly due to the increase in hourly wages, not labour supply. On the other hand, both factors contribute to the decline in real earnings of coal areas during bust period. Secondly, instead of gender, they compared changes in labour market conditions for coal and non-coal male workers in coal areas relative to those in non-coal areas. They found that the coal boom had dramatic impact on coal workers' real annual earnings, and also affected noncoal workers' earnings positively for coal areas relative to non-coal areas.

After obtaining the fact that coal boom is positively correlated with earnings in coal-producing areas. In order to make coal area a valid instrument, it also needs to prove that the economic benefit received by coal-producing areas were not spread to other non-coal regions. It is to ensure that a certain county's instrumental variable is only correlated with its real earnings, not other counties'. Authors, firstly, depicted the distribution of three sizes of coal reserves on a map of four states (Figure 2), and another map that demonstrates counties in different quantiles of mean growth in earnings during the bust (Figure 3). Geographically, the comparison of two graphs can approve the negative correlation between the size of coal reserves and its growth in earnings during the bust period, and the effect seemed to be not spreading across counties. Secondly, they demonstrated counties' quantile of mean growth in SSI payment (Figure 4). Compared Figure 2 and 4, authors confirmed co-movements between changes in SSI payment in the bust period and the size of coal reserves. Therefore, IVs should satisfy its relevance and exogeneity conditions.

III. Econometric Specification

Two regression models are specified in this section. The first model is an Ordinary Least Square regression that estimates the correlation between changes in a county's disability payments (dependent variable) and changes in its annual earnings (explanatory variable). It is a first differenced model, which subtracted potential time invariant omitted variables. Some county level time variant variables are included to control for time variant omitted variable bias, for example, the scale of county's population and its growth rate. State-year dummy variables are used to control for any distinctions across states in different years. It also controlled for whether or not the county is in the Metropolitan Statistical Area, because disabled workers may prefer to move into metropolitan areas for better health service and public transportation. Finally, authors considered that counties with different manufacturing structure may have different impact on its disability payment, so fraction of earnings from manufacturing industries in 1969 is controlled for this purpose.

The OLS model may only capture the statistical relationship between changes in a county's earnings and changes in its disability payments, and may not reveal the causal relationship between these two factors, hence further investigation is required. For this purpose, instrumental variables are used to estimate the coefficients. In the two-stage least square model, authors used pre-discussed changes of value of coal reserves and its lag terms to estimate changes of earnings in the first stage equation. The second stage estimates how these coal price-fitted change of earnings influenced disability program payments.

IV. Results

Table 2 in the paper, as a summary statistics of the sample classified by coal price period, provides statistical insights on the relationship between coal value instrument and county earnings. It is shown that, during the coal boom, counties with larger coal value tended to receive higher county earnings. On the other hand, they also tended to suffer greater economy decline during the coal bust period, which had a co-movement with the fall of coal value. This result further enhanced the relevance of coal value as an instrument for county earnings. A replication of table 2 is located in the appendix of this paper. For most of the data, the table produced almost the same results. Some data are deviated within only about 0.02, which I consider it is probably due to rounding error of computers. There is one thing that worth to be noticed that I first stuck at calibrating my fraction of economy in manufacturing (1969) to its correct values, until I filled all missing values to 0 for each county. But I strongly doubt that whether it is valid to simply replace data with value of zero when it is missing.

A. Semiparametric Estimates

Table 3 displays four regressions that estimate the elasticities of changes in disability payments for changes in county's real earnings. From the two OLS regressions on both DI and SSI payments, they found statistically insignificant correlation between these two variables, regardless the inclusion of control variables. In contrast, two-stage method produced statistical significant elasticities on both programs regardless the inclusion of control variables. In the first stage of 2SLS, F-statistic are reported for testing the significance of instruments, and both instruments are found to be jointly significant in each of the four 2SLS's. The 2SLS coefficients from regression with control variables generally show that 10% increase in county's earnings will reduce DI payment by about 3.45%, and reduce SSI payment by 7.13%.

Authors' description of programs in section I provided economic insights on this result. First, surge in earnings increased the opportunity cost for worker to stay in disability program. Second, increase in earnings may cause program participants to become ineligible, because increases of earnings from one of their family members will potential push total family earnings over the program threshold.

After finding the statistical and economic explanation on the correlation between earnings and disability program payments. Authors started to consider whether this finding is applicable to other regions not in this sample? It was worried that, firstly, it may not be a valid representation of urban area, because coal is mined predominately in rural area. Second, due to the nature of coal industry employment structure, the elasticity they found from table 2 may not applicable to the whole economy.

For the first issue, authors used counties with moderate coal reserve and large coal reserve to compare difference of changes in county earnings on disability payments, since counties with different coal reserve size, on average, actually have completely different employment structure, where counties with moderate coal have similar structure as counties in urban areas. The result shows that, although these two types of areas have distinct employment structure, their impact of earnings growth on disability payments does not differ much. However, there is one case that could potential bias the result as physicians may be more likely to certify a patient from rural area than from urban area, because they may think disable workers are more likely to find a job if they live in urban area. The method to investigate this threat was not mentioned.

For the second issue, authors, firstly, made a comparison of employment structure between coal industry and the entire economy. They found that coal industry tended to have much larger proportion of low-skilled worker. They considered that the statistical correlation may be more valid indicator for low-skilled workers, and so they started to conduct another similar estimation on how changes in earnings affect disability payments, with fraction of male employment in primary metals as an instrument. The result from 2SLS produced very similar estimates on the elasticities. It helped to explain that low-skilled workers are more likely to stay in disability program when become unemployed, while, the impact of earnings on other groups of workers (e.g. educated workers) may only have a moderate impact on disability program payments. However, it is worthwhile for authors to conduct an investigation on elasticity for educated workers to enhance their argument.

B. Nonparametric Estimates

This subsection provides insights on why the magnitude of 2SLS estimates are much larger than the magnitudes of the OLS estimates. Different instrument variables are proposed to handle this problem. In column (1) and (2) of replicated Table 6 in the appendix, they demonstrated how counties with different coal values differ in their earnings during different periods of time.[2] They found that counties with moderate coal reserves performed relatively better in their economies than counties with little or no coal reserves during boom and peak period, performed relatively worse during the bust period. Counties with large coal reserves have similar patterns, but are doubled in magnitudes except the peak period. So, IVs of coal size and time period interaction are both considered distinct and relevant to growth of earnings.

The rest columns presents how counties with different coal values differ in their DI (column (3) and (4)) and SSI (column (5) and (6)) payment growth during different period of time. They are structural equations that are designed to pre-check statistical relationship between being endowed with large or morderate coal reserves and disability payments. [3] They found that counties with large and moderate coal reserves tended to have less growth in DI payments relative to counties with no or little coal reserves during the coal boom and peak peak, but this advantage shrinked and become almost zero, although still negative, during the bust period. For SSI, counties with large and moderate coal reserve tended to have larger growth in their SSI payments in the coal bust period, and the magnitude for counties with large coal reserves is larger. Overall, it displayed an obvious negative correlation between disability payments and real earnings for counties with large or moderate coal reserves relative to those with little or none. These findings provided relevance condition to treat interaction of regions (as counties with different coal reserves) and periods (as periods with different coal values) as nonparametric instrumental variables to estimate impact of earnings on DI and SSI payments, and to see whether the estimated coefficients will be consistent across periods. [4] The output in Table 7 is replicated in the appendix. My replication matches all the point estimates of the regression, but t-statistics are somehow deviated from original result in range of 0.001 and 0.01. It is probably because of different methods of estimating standard errors.

The key finding is that, the elasticities of changes of real earnings on DI payment growth are not consistent over time, and the volatility is unpredictably large. As it is negative during the coal boom and peak period, but becomes positive during the coal bust. This evidence puts the validity of estimated

[2] The estimated regression is: $\triangle(earnings_{ist}) = \alpha_o + \boldsymbol{year_{st}\alpha_{1st}} + \boldsymbol{C_{ist}}\boldsymbol{\gamma} + u_{ist}$. The coefficients $\boldsymbol{\gamma}$ are reported in replicated Table 6, column (1) and (2). Their coefficients and t-statistics perfectly match the original paper.

[3] The estimated regression is: $\triangle(y_{ist}) = \theta_o + \boldsymbol{year_{st}\theta_{1st}} + \boldsymbol{C_{ist}}\boldsymbol{\tau} + u_{ist}$. The coefficients $\boldsymbol{\tau}$ are reported in column (3), (4), (5) and (6), and they do not reveal causal effects. The point estimates and t-statistics perfectly match the table in the original text.

[4] First stage: $\triangle(earnings_{ist}) = \alpha_o + \boldsymbol{year_{st}\alpha_{1st}} + \boldsymbol{C_{ist}}\gamma + u_{ist}$. Second stage: $\triangle(y_{ist}) = \beta_o + \boldsymbol{year_{st}\beta_{1st}} + \beta_3\triangle(earnings_{ist}) + v_{ist}$. The coefficient β_3 and its t-statistics are reported in replicated Table-7.

elasticities in Table 2 into questions. Fortunately, estimates on SSI payments did not meet this issue, therefore authors continued their detection on only DI payments.

Table 8 displayed several hypothesis regressions that authors proposed to diagnose the coefficient on growth of log earnings. The first step is to include coal size dummies onto the 2SLS in Table 7, as it is considered that counties with large or moderate coal size may have different pattern in their DI payments. The fixed effect is found to be significant for both moderate and large regions, and the coefficient on the growth of earnings increased a little in magnitude. The next three regressions (2-4) were set to analyse whether the coefficients differ across different compositions of sample. [5] No outliers were found across these three estimates, and consistently, fixed effects had similar coefficients. The next three regressions (5-7) carry out two-way comparison by time period. Similar results were found, except for the statistically insignificant coefficient under boom and peak period. Fixed effect for moderate and large coal regions are almost the same across seven regressions. It highlights the necessity to include coal region fixed effects in beginning of semi-parametric estimations. Therefore, authors added fixed effect for large and moderate coal reserves onto regression (4) of Table 3. They also tried to include county dummies to capture each county's fixed effect, which is quite computationally expensive. The comparison between these two regressions can help to identify whether control county fixed effect by classifying their coal values is a right approach. The results show that the coefficient is quite similar, both about -0.27, and more importantly, it is found to be much smaller in magnitude than the point estimate (-0.347) in regression (4) of Table (3). It is also smaller in magnitude than those coefficients estimated under non-parametric estimations.

My replication provides exact the same point estimates of Table 8, but the t-statistics are, similar to Table 7, somehow biased within 0.01. One more problem is that coefficients' t-statistic of regression (9) and (10) failed to be calculated under my Stata code. Stata warned that variance matrix is non-symmetric or highly singular. This issue is probably consistent with my deviated t-statistics, as the way of computing variance matrix is different.

Lastly, authors started to explain why the estimates under 2SLS is much larger in magnitude than those under OLS. The key contention is that fitted growth in earnings is much smoother than its actual value, and therefore, it constructs better reflection on the variations in long term patterns, such as changes of disability payments. To prove this argument, authors, firstly, explained why the difference between 2SLS and OLS estimations is due to high volatility of county-level earnings, not endogeneity. Secondly, authors provided some empirical evidence on why they say 2SLS method can reduce transitory shocks.

If there is endogeneity issue, it is believed that the residuals from the first stage regression should be correlated with the explanatory variable, i.e. disability payments. The residuals from the first stage of 2SLS estimation should contain the variation of omitted variables from residuals of OLS, because the exogeneity condition of IVs ensured that the endogeneity effects are not included in the fitted growth of earnings, and instead, they are left in residuals. The experiments showed that neither growth of DI nor SSI payments is correlated with its first stage residuals of real earning growth. It provides evidence that the distortion of OLS estimates are not due to endogeneity.

Next, authors took a contrast situation to prove that why fitted growth of earnings can better reflect long-term variations, and reduce transitory shocks. They took unemployment insurance as the dependent variable and left the rest explanatory variables and IVs same as for the disability payments. They found opposite estimates for unemployment insurance, as OLS estimate is statistically significant,

[5] I first wondered why not only regress it on specific coal size separately, instead of regressing on two coal size combined each time. Then I realised that under the first scenario, coal size fixed effect will have multicollinearity issue.

but 2SLS estimate is not. [6] Economically speaking, unemployment insurance is more sensitive to transitory shocks, therefore, a process that removes transitory shocks (i.e. first stage of 2SLS) will obviously left fitted growth of earnings to have non significant correlation with unemployment insurance payments. Hence, the finding is that OLS estimates of elasticity of changes of real earnings on changes of disability payment is biased due to high volatility, not endogeneity. It further implies that permanent change in labour market had larger impact than transitory shocks.

Conclusion

There are three key findings of this paper. First, empirical evidence showed statistical correlation between growth of real earnings and disability payments, and its causal effect (elasticity) is estimated. Second, it was found that low-skilled workers had higher estimated elasticities than other workers. Third, permanent structural change in labour market had larger impact than transitory shocks.

[6] Replication of Table 9 is displayed in Appendix, all estimates match values from the paper.

Replication of Tables

Replication of TABLE 2

Variables	All counties	Large Coal Counties	Moderate Coal Counties	No Coal Counties
Coal Boom (1970-1977)				
Logarithmic difference in SSI payments	0.063	0.061	0.071	0.060
	(0.217)	(0.189)	(0.212)	(0.225)
Logarithmic difference in DI payments	0.127	0.102	0.119	0.136
	(0.087)	(0.080)	(0.080)	(0.090)
Logarithmic difference in county earnings	0.030	0.058	0.034	0.022
	(0.081)	(0.077)	(0.079)	(0.081)
Logarithmic difference in population	0.013	0.015	0.012	0.012
	(0.018)	(0.018)	(0.021)	(0.018)
Logarithmic difference in real price of coal	0.094	-	-	-
	(0.140)			
Logarithmic difference in coal value instrument	0.252	0.721	0.547	0.049
	(0.648)	(1.080)	(0.823)	(0.223)*
Mean coal reserves	457*	2,563	412	6.43*
	(1108)	(1779)	(257)	(19.4)
Fraction of economy in manufacturing (1969)	0.268	0.158	0.284	0.287
	(0.161)	(0.147)	(0.168)	(0.153)
Fraction of counties with an MSA	0.261	0.191	0.296	0.264*
	(0.439)	(0.394)	(0.457)	(0.441)
Population	84.4	59.1	80.4	91.3*
	(194)	(71.6)	(191)	(212)
Coal Bust (1983-1993)				
Logarithmic difference in SSI payments	0.062	0.067	0.063	0.061
	(0.073)	(0.070)	(0.068)	(0.075)
Logarithmic difference in DI payments	0.033	0.030	0.028	0.035
	(0.091)	(0.090)	(0.099)	(0.089)*
Logarithmic difference in county earnings	0.017	-0.009	0.008	0.026
	(0.077)	(0.086)	(0.058)	(0.079)
Logarithmic difference in population	0.002	-0.007	-0.002	0.005
	(0.013)	(0.012)	(0.012)	(0.013)
Logarithmic difference in coal value instrument	-0.111	-0.318	-0.241	-0.022
	(0.149)	(0.137)	(0.107)	(0.057)
Logarithmic difference in real price of coal	-0.041			
	(0.018)			
Population	85.7	58.5	78.3	94.2
	(179)	(68.6)	(170)	(197)
Number of counties	330	47	71	212

Notes: Standard errors are in parentheses. * if data differs from the original table in the last digit (difference within 0.002 is not labelled), ** if data differs in two last digits, *** if data differs completely.

Replication of TABLE 3

Controls:	(1) OLS	(2) 2SLS	(3) OLS	(4) 2SLS
State-year dummies	Yes	Yes	Yes	Yes
County is in MSA (1990)	Yes	Yes	No	No
County's population	Yes	Yes	No	No
Change in county's population	Yes	Yes	No	No
Fraction of earnings from manufacturing, 1969	Yes	Yes	No	No
Instruments: Change in value of coal reserves and two lagged values	No	Yes	No	Yes
Panel A: Disability Insurance Payments				
Change in county's earnings	-0.002	-0.345	0.002	-0.347
	(0.11)	(3.98)*	(0.10)	(4.40)*
First-stage F-statistic on excluded instruments	-	26.7	-	28.1
N	7260	7260	7260	7260
Panel B: Supplemental Security Income Payments				
Change in county's earnings	-0.023	-0.711	-0.020	-0.636
	(1.57)	(5.36)	(1.40)	(5.51)
First-stage F-statistic on excluded instruments	-	26.7	-	27.9
N	7904	7904	7904	7904

Notes: Absolute values of t-statistics are in parentheses. * if data differs from the original table in the last digit.

Replication of TABLE 6

	Difference in the logarithm of real earnings		Difference in the logarithm of real Disability Insurance payments		Difference in the logarithm of real Supplemental Security Income Payments	
Period	Counties with moderate coal reserves compared to counties with little or no reserves (1)	Counties with large coal reserves compared to counties with little or no reserves (2)	Counties with moderate coal reserves compared to counties with little or no reserves (3)	Counties with large coal reserves compared to counties with little or no reserves (4)	Counties with moderate coal reserves compared to counties with little or no reserves (5)	Counties with large coal reserves compared to counties with little or no reserves (6)
1970-1977	0.014	0.035	-0.016	-0.030	-0.004	-0.017
	(3.96)	(6.48)	(5.47)	(7.67)	(0.85)	(3.56)
1978-1982	0.004	0.002	-0.010	-0.024	0.007	0.001
	(1.00)	(0.30)	(3.36)	(6.56)	(1.93)	(0.18)
1983-1993	-0.018	-0.034	-0.003	-0.004	0.004	0.014
	(6.58)	(8.20)	(1.60)	(1.31)	(1.69)	(4.28)

Notes: Absolute values of t-statistics are in parentheses.

Replication of TABLE 7

Panel A: Disability Insurance Payments

Period	Counties with little or no coal compared to counties with large reserves	Counties with moderate coal reserves compared to counties with large reserves	Counties with little or no coal compared to counties with moderate reserves	Three-region comparison
1970-1977	-0.828	-0.852	-1.296	-0.887
	(4.61)*	(2.50)*	(3.11)*	(5.17)*
	N=2,072	N=944	N=2,264	N=2,640
1978-1982	-2.128	-5.574	-2.120	-2.075
	(1.14)	(0.22)	(0.89)	(1.31)
	N=777	N=354	N=849	N=990
1983-1993	0.112	0.053	0.198	0.124
	(1.27)	(0.29)	(1.62)	(1.63)
	N=2,849	N=1,298	N=3,113	N=3,630
1970-1993	-0.333	-0.353	-0.234	-0.324
	(4.41)*	(2.17)*	(2.31)	(4.88)*
	N=5698	N=2596	N=6226	N=7260

Panel B: Supplemental Security Income Payments

Period	Counties with little or no coal compared to counties with large reserves	Counties with moderate coal reserves compared to counties with large reserves	Counties with little or no coal compared to counties with moderate reserves	Three-region comparison
1970-1977	-0.479	-0.704	-0.241	-0.460
	(2.96)	(1.82)*	(0.66)	(3.07)
	N=2,056	N=9,44	N=2,248	N=2,624
1978-1982	-0.267	3.014	1.921	1.733
	(0.11)	(0.48)	(1.03)	(0.94)
	N=1,295	N=590	N=1,415	N=1,650
1983-1993	-0.382	-0.453	-0.257	-0.371
	(3.97)	(2.07)*	(1.87)	(4.39)*
	N=2,849	N=1,298	N=3,113	N=3,630
1970-1993	-0.424	-0.532	-0.209	-0.398
	(5.47)	(2.70)*	(1.54)	(5.64)*
	N=6,200	N=2,832	N=6,776	N=7,904

Notes: Absolute values of t-statistics are in parentheses. * if data differs from the original table in the last digit.

Replication of TABLE 8

	(1) Coefficient on difference of log earnings	(2) Moderate coal region fixed effect	(3) Large coal region fixed effect
(1) Full sample (N=7,260)	0.386	-0.010	-0.017
	(5.46)*	(5.24)*	(6.40)
Two-way comparisons by seam:			
(2) Counties with little or no coal and counties with moderate reserves only (N=6,226)	-0.408	-0.011	-
	(3.52)*	(5.29)*	
(3) Counties with little or no coal and counties with large reserves only (N=5,698)	-0.374	-	-0.017
	(4.76)*		(6.11)*
(4) Counties with little or no coal and counties with large reserves only (N=2,596)	-0.413	-	-0.008
	(2.41)*		(2.98)*
Two-way comparisons by time period:			
(5) Peak and bust only (N=4,620)	-0.408	-0.010	-0.018
	(3.54)	(3.97)*	(4.22)*
(6) Peak and bust only (N=4,620)	-0.385	-0.010	-0.017
	(5.29)*	(4.96)*	(6.10)*
(7) Peak and bust only (N=3,630)	-0.295	-0.011	-0.020
	(1.30)	(3.20)	(3.00)*
Alternative 2SLS estimates:			
(8) 2SLS with change in value of coal reserves and two lagged values as instruments and fixed effect for large and moderate coal reserves (N = 7,260)	-0.275	-0.010	-0.017
	(3.68)*	(5.31)*	(6.82)*
(9) 2SLS with change in value of coal reserves and two lagged values as instruments and county fixed effects (N = 7,260)	-0.271	-	-
	***		
(10) 2SLS with seam and time interactions as instruments and county fixed effects (N = 7,260)	-0.386	-	-
	***		

Notes: Absolute values of t-statistics are in parentheses. * if data differs from the original table in the last digit (difference within 0.002 is not labelled), ** if data differs in two last digits, *** if data differs completely.

Replication of TABLE 9

	(1)	(2)
Controls:	OLS	2SLS
State-year dummies	Yes	Yes
Instruments:		
Region and time-period interactions	No	Yes
Change in county's earnings	-0.586	0.031
	(9.06)	(0.18)
N	7,867	7,867

Stata Commands

```
clear all
set mem 500m
set matsize 10000
set more off
capture log close

use "bdsdata.dta",clear

// drop non-county-level data.
drop if (county=="USA")
drop if (county =="KENTUCKY")
* 39000-Ohio, 42000-Penn, 54000-WestVergi
drop if(fips ==39000)|(fips == 42000)|(fips ==54000)

tsset fips year, yearly

***************************************************
*** Logarithmic difference in SSI & DI payments
***************************************************
* generate logithmetic difference of SSI payment in real term (p7)
gen ldSSI = ln(dSSI/p7)
bys fips (year): gen dldSSI = d.ldSSI

* generate logithmetic difference of DI payment in real term (p7)
gen lpay = ln(pay/p7)
bys fips (year): gen dlpay = d.lpay

****************************************************
*** Logarithmic difference in county earnings
****************************************************
gen lctyInc = ln(EarnPOW/p7)
bys fips (year): gen dlctyInc = d.lctyInc

****************************************************
*** Logarithmic difference in population
*** Population
****************************************************
* logarithm of county's population
gen lpop = ln(Pop)
* log difference in county's population
bys fips (year): gen dlpop = d.lpop

****************************************************
```

```
*** Logarithmic difference in real price of Coal
*** Mean coal reserves
****************************************************
// measure of coal price: (July index values of each year are used)
// PPI for coal/Consumer Price Index of Urban wage earners
gen pcoal = pcoal7/p7
gen lpcoal = ln(pcoal)
bys fips (year): gen dlpcoal = D.lpcoal

****************************************************
*** Logarithmic difference in Coal Value Instrument
****************************************************
gen d_resval = dlpcoal*ln(coalres)
replace d_resval=0 if d_resval==.

****************************************************
*** Fraction of economy in manufacturing
****************************************************
gen propmanu = Manufact/EarnPOW
replace propmanu=0 if missing(propmanu)

****************************************************
** scale of coal-reserve
****************************************************
// >1000m, 100m<x<1000m, <100m
gen lgcoal= (coalres>=1000) & !missing(coalres)
gen mdcoal= (coalres>=100 & coalres<1000 )
gen oscoal= (coalres<100)

// easier for tabstat
gen coalsize=0
replace coalsize=2 if coalres>=1000 & !missing(coalres)
replace coalsize=1 if coalres>=100 & coalres<1000

// boom, peak and bust periods
gen boom=0
replace boom=1 if year>=1970 & year<=1977
gen peak=0
replace peak=1 if year>=1978 & year<=1982
gen bust=0
replace bust=1 if year>=1983 & year<=1993

*****************************************************
*** Table 2 Output
*****************************************************
tabstat dldSSI dlpay dlctyInc dlpop dlpcoal d_resval coalres propmanu msa ///
```

```
if boom==1, s(mean sd) by(coalsize)
tabstat propmanu if year==1969, s(mean sd) by(coalsize)
tabstat dldSSI dlpay dlctyInc dlpop dlpcoal d_resval Pop if  bust==1, ///
s(mean sd) by(coalsize)

****************************************************
*** Table 3 Output
****************************************************
/* Regression (1) - OLS using year FE
Controlled for:
1. State-year dummies
2. MSA
3. Population
4. Changes in popn
5. fraction of earnings from manufacturing, 1969
*/
// generate constant term of propmanu for each county
gen manu69=propmanu if year==1969
by fips: replace manu69 = manu69[_n-1] if missing(manu69)
by fips: replace manu69 = manu69[_n+1] if missing(manu69)
replace manu69=0 if missing(manu69)

// Panel A - DI
reg dlpay dlctyInc state#year msa lpop dlpop manu69 if year>=1970, ///
cluster(fips)

// Panel B - SSI
reg dldSSI dlctyInc state#year msa lpop dlpop manu69 if year>=1970, ///
cluster(fips)

*******************************************************
/* Regression (3) - OLS using year FE
Controlled for:
1. State-year dummies
*/
// Panel A - DI
reg dlpay dlctyInc state#year if year>=1970, cluster(fips)

// Panel B - SSI
reg dldSSI dlctyInc state#year if year>=1970, cluster(fips)

*******************************************************
/* Regression (2) - 2SLS using year FE
Controlled for:
1. State-year dummies
2. MSA
```

```
3. Population
4. Changes in popn
5. fraction of earnings from manufacturing, 1969
*/
// Panel A - DI
ivregress 2sls dlpay state#year msa lpop dlpop manu69 (dlctyInc=d_resval ///
L.d_resval L2.d_resval) if year>=1970, cluster(fips) first
estat firststage
// Panel B - SSI
ivregress 2sls dldSSI state#year msa lpop dlpop manu69 (dlctyInc=d_resval ///
L.d_resval L2.d_resval) if year>=1970, cluster(fips) first
estat firststage

*********************************************************
/* Regression (4) - 2SLS using year FE
Controlled for:
1. State-year dummies
*/
// Panel A - DI
ivregress 2sls dlpay state#year (dlctyInc=d_resval L.d_resval L2.d_resval) ///
if year>=1970, cluster(fips) first
estat firststage
// Panel B - SSI
ivregress 2sls dldSSI state#year (dlctyInc=d_resval L.d_resval L2.d_resval) ///
if year>=1970, cluster(fips) first
estat firststage


******************************************************
*** Table 6 Output
******************************************************

************************************************
// Diff in log of real earnings
reg dlctyInc state#year mdcoal lgcoal if year>=1970 & year<=1977, cluster(fips)
reg dlctyInc state#year mdcoal lgcoal if year>=1978 & year<=1982, cluster(fips)
reg dlctyInc state#year mdcoal lgcoal if year>=1983, cluster(fips)

************************************************
// Diff in log of real Disability Insurance payments
reg dlpay state#year mdcoal lgcoal if year>=1970 & year<=1977, cluster(fips)
reg dlpay state#year mdcoal lgcoal if year>=1978 & year<=1982, cluster(fips)
reg dlpay state#year mdcoal lgcoal if year>=1983, cluster(fips)

************************************************
// Diff in log of real SSI payment
```

```
reg dldSSI state#year mdcoal lgcoal if year>=1970 & year<=1977, cluster(fips)
reg dldSSI state#year mdcoal lgcoal if year>=1978 & year<=1982, cluster(fips)
reg dldSSI state#year mdcoal lgcoal if year>=1983, cluster(fips)

****************************************************
*** Table 7 Output
****************************************************
****************************************************
// Panel A: DI payment
* Column 1: counties with little -->large reserves
ivregress 2sls dlpay (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year if mdcoal==0 & boom, cluster(fips)
ivregress 2sls dlpay (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year if mdcoal==0 & peak, cluster(fips)
ivregress 2sls dlpay (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year if mdcoal==0 & bust, cluster(fips)
ivregress 2sls dlpay (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year if mdcoal==0, cluster(fips)

* Column 2: counties with moderate -->large reserves
ivregress 2sls dlpay (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year if oscoal==0 & boom, cluster(fips)
ivregress 2sls dlpay (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year if oscoal==0 & peak, cluster(fips)
ivregress 2sls dlpay (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year if oscoal==0 & bust, cluster(fips)
ivregress 2sls dlpay (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year if oscoal==0, cluster(fips)

* Column 3: counties with little --> moderate
ivregress 2sls dlpay (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year if lgcoal==0 & boom, cluster(fips)
ivregress 2sls dlpay (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year if lgcoal==0 & peak, cluster(fips)
ivregress 2sls dlpay (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year if lgcoal==0 & bust, cluster(fips)
ivregress 2sls dlpay (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year if lgcoal==0, cluster(fips)

* Column 4: three region comparison
ivregress 2sls dlpay (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year if boom, cluster(fips)
ivregress 2sls dlpay (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year if peak, cluster(fips)
ivregress 2sls dlpay (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year if bust, cluster(fips)
```

```
ivregress 2sls dlpay (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year, cluster(fips)

*****************************************************
// Panel B: SSI payment
* Column 1: counties with little -->large reserves
ivregress 2sls dldSSI (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year if mdcoal==0 & boom, cluster(fips)
ivregress 2sls dldSSI (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year if mdcoal==0 & peak, cluster(fips)
ivregress 2sls dldSSI (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year if mdcoal==0 & bust, cluster(fips)
ivregress 2sls dldSSI (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year if mdcoal==0, cluster(fips)

* Column 2: counties with moderate -->large reserves
ivregress 2sls dldSSI (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year if oscoal==0 & boom, cluster(fips)
ivregress 2sls dldSSI (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year if oscoal==0 & peak, cluster(fips)
ivregress 2sls dldSSI (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year if oscoal==0 & bust, cluster(fips)
ivregress 2sls dldSSI (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year if oscoal==0, cluster(fips)

* Column 3: counties with little --> moderate
ivregress 2sls dldSSI (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year if lgcoal==0 & boom, cluster(fips)
ivregress 2sls dldSSI (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year if lgcoal==0 & peak, cluster(fips)
ivregress 2sls dldSSI (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year if lgcoal==0 & bust, cluster(fips)
ivregress 2sls dldSSI (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year if lgcoal==0, cluster(fips)

* Column 4: three region comparison
ivregress 2sls dldSSI (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year if boom, cluster(fips)
ivregress 2sls dldSSI (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year if peak, cluster(fips)
ivregress 2sls dldSSI (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year if bust, cluster(fips)
ivregress 2sls dldSSI (dlctyInc= lgcoal#boom lgcoal#peak lgcoal#bust ///
 mdcoal#boom mdcoal#peak mdcoal#bust) state#year, cluster(fips)

*****************************************************
```

```
*** Table-8 output
****************************************************
// seam and time interactions
gen mdboom=1 if mdcoal==1 & boom==1
replace mdboom=0 if mdboom==.
gen mdpeak=1 if mdcoal==1 & peak==1
replace mdpeak=0 if mdpeak==.
gen mdbust=1 if mdcoal==1 & bust==1
replace mdbust=0 if mdbust==.

gen lgboom=1 if lgcoal==1 & boom==1
replace lgboom=0 if lgboom==.
gen lgpeak=1 if lgcoal==1 & peak==1
replace lgpeak=0 if lgpeak==.
gen lgbust=1 if lgcoal==1 & bust==1
replace lgbust=0 if lgbust==.

gen osboom=1 if oscoal==1 & boom==1
replace osboom=0 if osboom==.
gen ospeak=1 if oscoal==1 & peak==1
replace ospeak=0 if ospeak==.
gen osbust=1 if oscoal==1 & bust==1
replace osbust=0 if osbust==.

***************************************************
// Regression - row-1
ivregress 2sls dlpay state#year lgcoal mdcoal (dlctyInc=mdboom mdpeak ///
 mdbust lgboom lgpeak lgbust) if year>=1970, cluster(fips)

***************************************************
// Regression - Two-way comparison by seam:
ivregress 2sls dlpay state#year lgcoal mdcoal (dlctyInc=mdboom mdpeak ///
 mdbust lgboom lgpeak lgbust) if year>=1970 & lgcoal!=1, cluster(fips)
ivregress 2sls dlpay state#year lgcoal mdcoal (dlctyInc=mdboom mdpeak ///
 mdbust lgboom lgpeak lgbust) if year>=1970 & mdcoal!=1, cluster(fips)
ivregress 2sls dlpay state#year lgcoal mdcoal (dlctyInc=mdboom mdpeak ///
 mdbust lgboom lgpeak lgbust) if year>=1970 & oscoal!=1, cluster(fips)

***************************************************
// Regression - Two-way comparison by time period:
ivregress 2sls dlpay state#year lgcoal mdcoal (dlctyInc=mdboom mdpeak ///
 mdbust lgboom lgpeak lgbust) if year>=1970 & boom!=1, cluster(fips)
ivregress 2sls dlpay state#year lgcoal mdcoal (dlctyInc=mdboom mdpeak ///
 mdbust lgboom lgpeak lgbust) if year>=1970 & peak!=1, cluster(fips)
ivregress 2sls dlpay state#year lgcoal mdcoal (dlctyInc=mdboom mdpeak ///
 mdbust lgboom lgpeak lgbust) if year>=1970 & bust!=1, cluster(fips)
```

```
**************************************************
// Regression - Alternative 2SLS estimates:
ivregress 2sls dlpay mdcoal lgcoal state#year (dlctyInc=d_resval L.d_resval ///
 L2.d_resval) if year>=1970, vce(cluster fips)
ivregress 2sls dlpay i.fips state#year (dlctyInc=d_resval L.d_resval ///
 L2.d_resval) if year>=1970, cluster(fips)
ivregress 2sls dlpay i.fips state#year (dlctyInc=lgboom lgpeak lgbust ///
 mdboom mdpeak mdbust) if year>=1970, cluster(fips)


**************************************************
*** Table 9 output
**************************************************
// generate difference of log Unemployment Insurance exp
gen ldUnemploy = ln(dUnemply)
bys fips (year): gen dldUnemploy = d.ldUnemploy
// OLS
reg dldUnemploy dlctyInc state#year if year>=1970 & year<=1993, cluster(fips)
// 2SLS
ivregress 2sls dldUnemploy state#year ///
(dlctyInc=mdboom mdpeak mdbust lgboom lgpeak lgbust) ///
if year>=1970 & year<=1993, cluster(fips)
```

References

[1] Dan Balck, Kermit Daniel, and Seth Sanders LaTeX: *The Impact of Economic Conditions on Participation in Disability Programs: Evidence from the Coal Boom and Bust* The American Economic Review, Vol.92, No.1, 2002.