

# Correlates of County Per Capita Income Growth in Pennsylvania

by

David A. Latzko  
Business and Economics Division  
Pennsylvania State University, York Campus  
1031 Edgecomb Avenue  
York, PA 17403

phone: 717-771-4115  
fax: 717-771-4062  
e-mail: DXL31@psu.edu

July 2000

# Correlates of County Per Capita Income Growth in Pennsylvania

## Abstract

A substantial number of variables are partially correlated with the growth of county per capita income in Pennsylvania in at least one regression. The problem is that the coefficient of variable  $x_1$  is significant when the regression includes variables  $x_2$  and  $x_3$  but becomes insignificant when  $x_4$  rather than  $x_3$  is included. I adopt the methodology of Sala-i-Martin (1997a; 1997b) to identify which of 22 variables are strongly correlated with county income growth by examining the entire distribution of the estimates of the coefficient of variable  $x_i$ . Several variables are strongly correlated with county per capita income growth.

Many variables are potentially correlated with local area economic growth. A typical study such as Austin and Schmidt (1998), Duffy-Deno (1997; 1998), Glaeser, Kallal, Scheinkman, and Shleifer (1992), Carlino and Mills (1987), or Mills and Price (1984) regresses a set of variables such as taxes, industry mix, crime, and natural amenities on some measure of local area growth. The problem is that often the coefficient of variable  $x_1$  is significant when the regression includes variables  $x_2$  and  $x_3$  but becomes insignificant when  $x_4$  rather than  $x_3$  is included. Since economic theory is not specific enough about which variables belong in the regression, what variables are really correlated with regional growth? Sala-i-Martin (1997a; 1997b) examines the entire distribution of the estimates of the coefficient of variable  $x_i$  in order to determine which variables are strongly correlated with national economic growth. I adopt his methodology here to explore the correlates of county-level per capita income growth in Pennsylvania.

## I. Methodology

Suppose one is interested in knowing whether variable  $z$ , one of a pool of  $N$  variables identified as possibly being related to growth, is strongly correlated with county income growth. If one starts running regressions combining  $z$  with various combinations of the other  $N-1$  variables, the coefficient of variable  $z$  will eventually be found to be insignificant in some regression. Can variable  $z$  then be considered robustly correlated with growth?

Sala-i-Martin (1997a; 1997b) suggests computing the entire distribution of the estimators of the coefficient of variable  $z$ ,  $\beta_z$ , to decide whether variable  $z$  is strongly correlated with growth. If enough of the estimates of  $\beta_z$  are significantly different from zero then variable  $z$  is strongly correlated with county income growth.

“In particular, one might be interested in the fraction of the (cumulative) density function lying on each side of zero: if 95 percent of the density function for  $\beta_1$  lies to the right of zero and only 52 percent of the density function for  $\beta_2$  lies to the right of zero, one will probably think of variable 1 as being more likely to be correlated with growth than variable 2” (Sala-i-Martin 1997a, p. 179).

In order to obtain the cumulative density function of the estimators of the coefficient of variable  $z$ , I estimate regressions of the form

$$y = \alpha_i + \beta_{zi}z + \beta_{xi}X_i + \varepsilon, \quad (1)$$

where  $y$  is a measure of county income growth,  $z$  is the variable of interest, and  $X_i$  is a vector of three variables taken from the pool of  $N$  variables thought to possibly be related to income growth. This regression is estimated for all the  $M$  possible combinations of  $X_i$ . For each regression, an estimate of  $\beta_{zi}$  and a corresponding standard deviation,  $\sigma_{zi}$ , are obtained. I also calculate the likelihood function,  $L_i$ , for each regression.

Having obtained these estimates for all  $M$  regressions, the method of determining the cumulative density function of the estimates of  $\beta_z$  depends on whether or not these estimates are normally distributed. If the estimates of  $\beta_z$  are normally distributed the mean and standard deviation determine its cumulative density function. The mean estimates of  $\beta_z$  and  $\sigma_z$  are weighted averages of the individual point estimates:

$$\beta_z = \sum \omega_i \beta_{zi} \text{ and } \sigma_z = \sum \omega_i \sigma_{zi}. \quad (2)$$

The weights are designed to give greater weight to those regressions that are more likely to be “right” and are calculated by

$$\omega_i = L_i / \sum L_i, \quad (3)$$

which is regression  $i$ 's proportion of the sum of the likelihood functions of the  $M$  regressions for variable  $z$ . The cumulative density function for  $\beta_z$  is determined from the standard normal tables using the mean estimates from equation (2).

If the estimates of  $\beta_z$  are not normally distributed, the aggregate cumulative density function,  $\Phi_z$ , is calculated as the weighted average of the individual cumulative density functions,  $\Phi_{zi}$ , for each of the  $M$  regressions, where the weights are given by equation (3):

$$\Phi_z = \sum \omega_i \Phi_{zi}. \quad (4)$$

I also compute the aggregate cumulative density function using an unweighted average of all the regressions as a check against just one or two regressions capturing all the weight in equations (2) and (4).

## II. Data

I want to explain the growth of income per capita among Pennsylvania counties between 1970 and 1996. My interest is in determining the correlates of the growth of income from economic activity within these counties, but the appropriate income measure is not available: a county-level version of gross domestic product. Personal income data is available on a county basis. But use of personal income is problematic if people work in one county and live in another or if people tend to own capital in other counties because the personal income accounts reported by the Commerce Department's Bureau of Economic Analysis assign income to the county in which the owner of the inputs resides not to the county in which the income was earned. For instance, Lackawanna County received 7033 commuters, amounting to 6 percent of the county work force, from Luzerne County in 1990 while sending there 5175 commuters, and nearly 60,000 Delaware County residents worked in Philadelphia County with an even larger

number of Philadelphia commuters coming from out of state (Pennsylvania State Data Center web site). Also, the personal income measure includes transfer payments. So, personal income is not a good measure of county economic activity because it includes both unearned income and income earned outside the county.

The Bureau of Economic Analysis also tracks “total earnings by place of work”. This I use as the measure of county income because it attributes income to the county in which it was earned. Total earnings includes wages and salaries, other labor income, contributions for social insurance, and proprietors’ income. It excludes dividends, interest, rent, and transfer payments. I divide total earnings by total population for each county. Data for total earnings and population are taken from the Regional Economic Information System web page.<sup>1</sup> The dependent variable, measuring the growth of per capita county income, in all regressions is the log of real county total earnings per capita for 1996 minus the log of real county total earnings per capita for 1970. Its mean value is 0.16 with a maximum of 0.71 and a minimum of -0.32. I chose 1970 as the start year because the dependent variable is available only as far back as 1969 and the values of many explanatory variables are taken from the 1970 Census.

I calculate the aggregate cumulative density function for 22 variables. This admittedly is a “fishing expedition”. I selected variables that can be viewed as representing a county’s initial conditions; the variables were chosen on the basis of plausibility and data availability.

(a) Political Variables: fraction of the total number of county voters registered Democratic, municipality density, and total taxes collected by the county government as a percentage of county total earnings.

---

<sup>1</sup> <<http://fisher.lib.virginia.edu/reis>>. The Bureau of Economic Analysis produces the REIS database.

(b) Economic Variables: log of 1970 real county total earnings per capita, rate of investment in new capital, agricultural income as a percentage of county total earnings, fraction of total county employment in manufacturing, manufacturing exports divided by county total earnings, highway density, and bank deposits as a percentage of county personal income.

(c) Location Variables: distance from the county seat to Philadelphia and distance from the county seat to Pittsburgh.

(d) Demographic Variables: average annual rate of population growth between 1970 and 1996, percent of total year-round housing in the county renter occupied, population density, percent of county population living in urban areas, liquor sales (excluding beer) per capita, and public assistance expenditures as a percentage of county total earnings.

(e) Human Capital Variables: proportion of the county population of persons 25 years and over with a bachelor's degree or higher and per pupil public school expenditures in the county.

(f) Scale Variables: land area in thousands of acres and total employment in 1970.

If income growth influences the independent variables, then estimates of equation (1) using ordinary least squares are potentially inconsistent. To reduce the simultaneity problem, I take the values of the explanatory variables from as close as possible to the beginning of the study period, 1970. The data appendix provides more detailed descriptions and the sources of the variables.

Each regression contains four independent variables: the variable of interest and all the possible groups of three of the other 21 independent variables. 1330 regressions were run for each variable for a total of 29,260 regressions. There is a tradeoff between computation time and

the omitted variable bias. Combining the other independent variables into groups of 4 would have required a total of 131,670 regressions

### III. Results

Table 1 reports the main results from the regressions. The variables are ordered by weighted non-normal cumulative density function. Only one variable, the fraction of adults with college degree, is significantly different from zero at the 0.05 level in all 1330 regressions. Several variables were not significant in any regression. The correlation between the column reporting the cumulative density function, CDF, under the assumption that the distribution of the estimates of  $\beta_z$  is normal and the column reporting the weighted CDF assuming non-normality is 0.986, indicating that the density function of the estimates of  $\beta_z$  is nearly but not normal.

As an example of how to interpret Table 1, the “distance from Philadelphia” variable is significant at the 0.05 level in 18 percent of the regressions in which it is the variable of interest. The point estimate of the coefficient, calculated as in equation (2), is  $-0.0003$  with a standard deviation of 0.0002. Column (4) provides the cumulative density function assuming a normal distribution of the estimates of the coefficient. From the standard normal tables 0.83, or 83 percent, of the cumulative density function for the variable lies to the left of zero. Column (5) reports the density function assuming a non-normal distribution of  $\beta_z$ . 0.81 of the CDF lies to the left of zero. Column (6) presents a simple arithmetic average of the density functions from the 1330 regressions using “distance from Philadelphia” as the variable of interest.

Focusing on the weighted non-normal cumulative density functions in column (5) and taking 0.95 as the usual level of confidence, three variables are strongly correlated with county-level per capita income growth in Pennsylvania: the fraction of adults with a college degree

(positively related to growth), the growth rate of county population (negatively related to growth), and the fraction of total earnings in agriculture (positively related to growth). Two additional variables have a CDF greater than 0.90 and can be considered weakly correlated with county-level growth: the fraction of voters registered Democratic (negatively related to growth) and the percentage of the county population living in urban areas (positively related to growth). Note, however, that just the college education variable has an unweighted CDF greater than 0.95.

The college education, a proxy for human capital, and population growth variables have standard extended neoclassical growth model interpretations (Mankiw, Romer, and Weil 1992). The positive correlation between the importance of agriculture to the county economy in 1970 and subsequent per capita income growth may be due to the farmland being available for suburban sprawl; a relatively urban population was good for county growth. I cannot offer an adequate explanation of why having had a largely Democratic electorate was bad for income growth. Perhaps, Democratic voters were correlated with unionization and unions were concentrated in declining manufacturing industries.

#### IV. Conclusions

A county in Pennsylvania has probably experienced rapid per capita income growth since 1970 if

- it had a college educated labor force,
- agricultural income was a large share of county total earnings, and
- its population was largely urban.

Since 1970, income per capita probably grew slowly, or not at all, if

- the county's population grew relatively rapidly and

- voters were mostly registered Democrats.

**1. fraction of voters registered Democratic** - Fraction of total number of registered voters registered Democratic on November 3, 1970. See Table 146 in the 1973 *Pennsylvania Statistical Abstract*.

**2. municipality density** - Total number of municipalities levying taxes under Act 511 in 1969 per thousand of acres of county land area. See Table 178 in the 1971 *Pennsylvania Statistical Abstract*.

**3. local government tax revenue** - Total taxes collected (includes real estate, occupations, per capita, and personal property taxes) in 1969 by the county government as a percentage of county total earnings. See Table 149 in the 1971 *Pennsylvania Statistical Abstract*.

**4. log of 1970 real total earnings per capita** - Total earnings by place of work and population data from the Regional Economic Information System web site.

**5. new capital investment rate** - Fraction of county total earnings invested in manufacturing capital goods. Estimated for each county by dividing manufacturing capital expenditures in 1970 by county total earnings for 1970. Capital expenditures for each county are found in Table 56 of the 1973 *Pennsylvania Statistical Abstract*.

**6. fraction of total income in agriculture** - Estimated cash receipts from the sale of agriculture and livestock products from farms and government crop payments in 1968 as a percentage of county total earnings. Data from Table 222 in the 1973 *Pennsylvania Statistical Abstract*.

**7. manufacturing employment share** - Fraction of total county employment in manufacturing in 1969. Manufacturing employment from the 1973 *Pennsylvania Statistical Abstract*, Table 56.

**8. share of income in manufacturing exports** - Total value of exports in 1970 by Pennsylvania manufacturing establishments as a fraction of county total earnings. See Table 53 in the 1971 *Pennsylvania Statistical Abstract*.

**9. highway density** - Highway mileage (excluding Turnpike) in 1970 divided by total county land area. Highway mileage from Table 267 of the 1973 *Pennsylvania Statistical Abstract*.

**10. deposits as a percentage of personal income** - Total bank deposits as of June 1970 divided by personal income. Bank deposits from the 1972 *County and City Data Book*, pages 397 and 409. Personal income data from REIS web site.

**11. distance from Philadelphia** - Highway mileage from the county seat to Philadelphia.

**12. distance from Pittsburgh** - Highway mileage from the county seat to Pittsburgh.

**13. population growth rate** - Average annual rate of county population growth between 1970 and 1996. Population data taken from the REIS web site.

**14. percent of housing renter occupied** - Percentage of total year-round housing occupied by renters. See Table 2 of the 1971 *Pennsylvania Statistical Abstract*. Data for 1970.

**15. population density** - Table 5 in the 1971 *Pennsylvania Statistical Abstract*.

**16. percent urban population** - Calculated from data in Table 5 in the 1971 *Pennsylvania Statistical Abstract*.

**17. liquor sales per capita** - Sales of Pennsylvania state liquor stores for July 2, 1969 to June 30, 1970 inclusive divided by county population in 1970. Liquor sales by county taken from Table 37 of the 1971 *Pennsylvania Statistical Abstract*.

**18. public assistance as a fraction of earnings** - Estimated distribution of state and federal public assistance funds for the fiscal year ending June 30, 1970 divided by county total earnings in 1970. See Table 121 of the 1973 *Pennsylvania Statistical Abstract*.

**19. fraction of adults with a college degree** - Proportion of the county population of persons 25 years and over with a bachelor's degree or higher calculated using data from pages 391 and 403 of the 1972 *County and City Data Book*.

**20. public school expenditures per pupil** - Per pupil expenditures during the 1969-70 school year for county public schools based on average daily attendance of public school districts in the county. Data from Table 97 of the 1971 *Pennsylvania Statistical Abstract*.

**21. land area** - Land area in thousands of acres in 1965. See Table 222 of the 1971 *Pennsylvania Statistical Abstract*.

**22. total employment in 1970** - Total full and part-time employment. From the REIS web site.

---

Table 1 – Main Results from Regressions

| <u>variable</u>                             | (1)<br><u>fraction</u><br><u>significant</u> | (2)<br><u>coefficient</u> | (3)<br><u>standard</u><br><u>deviation</u> | (4)<br><u>CDF</u><br><u>normal</u> | (5)<br><u>CDF non-</u><br><u>normal</u><br><u>(weighted)</u> | (6)<br><u>CDF non-</u><br><u>normal</u><br><u>(not weighted)</u> |
|---|--|---------------------------|--|------------------------------------|--|--|
| fraction of adults with college degree      | 100.00%                                      | 3.2442                    | 0.6911                                     | 0.9999                             | 0.9999   | 0.9970   |
| population growth rate                      | 22.93%                                       | -7.5749                   | 2.2884                                     | 0.9984                             | 0.9966   | 0.8163   |
| fraction of total earnings in agriculture   | 28.80%                                       | 912.78                    | 277.73                                     | 0.9983                             | 0.9964   | 0.8500   |
| fraction of voters registered Democratic    | 31.28%                                       | -0.0579                   | 0.0307                                     | 0.9360                             | 0.9302   | 0.9264   |
| percent urban population                    | 33.38%                                       | 0.2646                    | 0.1190                                     | 0.9702                             | 0.9224   | 0.7876   |
| distance from Philadelphia                  | 17.67%                                       | -0.0003                   | 0.0002                                     | 0.8309                             | 0.8122   | 0.7269   |
| share of income in manufacturing exports    | 0.00%  | -0.3220                   | 0.2572                                     | 0.7847                             | 0.7819   | 0.6071   |
| log 1970 county total earnings per capita   | 4.96%  | -0.1706                   | 0.1079                                     | 0.8811                             | 0.7676   | 0.5385   |
| manufacturing employment share              | 6.84%  | -0.2429                   | 0.2256                                     | 0.7142                             | 0.6742   | 0.6898   |
| deposits as a percentage of personal income | 13.68%                                       | -0.2017                   | 0.1906                                     | 0.7059                             | 0.6357   | 0.7623   |
| municipality density                        | 31.65%                                       | 0.5798                    | 0.6391                                     | 0.6322                             | 0.5973   | 0.8555   |
| new capital investment rate                 | 0.00%  | -0.4324                   | 0.5374                                     | 0.5759                             | 0.5712   | 0.4389   |
| public assistance as a fraction of earnings | 0.00%  | 570585                    | 677044                                     | 0.5974                             | 0.5486   | 0.4179   |
| distance from Pittsburgh                    | 6.09%  | -0.0002                   | 0.0003                                     | 0.5623                             | 0.4980   | 0.3821   |
| liquor sales per capita                     | 1.35%  | -0.0019                   | 0.0024                                     | 0.5684                             | 0.4711   | 0.4638   |
| total employment in 1970                    | 9.47%  | 0.0000                    | 0.0000                                     | 0.3747                             | 0.3558   | 0.6317   |
| population density                          | 2.48%  | -0.0000                   | 0.0000                                     | 0.2340                             | 0.2828   | 0.4883   |
| percent of housing renter occupied          | 7.67%  | 0.1096                    | 0.5183                                     | 0.1668                             | 0.2776   | 0.6987   |
| highway density                             | 7.22%  | 0.0029                    | 0.0103                                     | 0.2208                             | 0.2641   | 0.5448   |
| local government tax revenue                | 0.00%  | 1.1754                    | 9.2056                                     | 0.1012                             | 0.1190   | 0.3280   |
| public school expenditures per pupil        | 9.85%  | 0.0000                    | 0.0003                                     | 0.0266                             | 0.1773   | 0.7545   |
| land area                                   | 0.00%  | 0.0000                    | 0.0000                                     | 0.0809                             | 0.1550   | 0.4803   |

Notes: Dependent variable is the log of real county total earnings per capita for 1996 minus the log of real county total earnings per capita for 1970. Column (1) reports the percentage of regressions in which the variable's coefficient is significantly different from zero at the 0.05 level of significance. The coefficient and standard deviation in columns (2) and (3) are calculated using equation (2) in the text. The cumulative density function assuming that the estimates of the coefficient are normally distributed is reported in column (4). Column (5) is calculated using equation (4) in the text. Column (6) is the arithmetic average of the individual cumulative density functions. Sample size is 1330 for each coefficient.

## References

- Austin, John S. and James R. Schmidt, "Convergence Amid Divergence in a Region," *Growth and Change* 29 (Winter 1998), 67-89.
- Carlino, Gerald A., and Edwin S. Mills, "The Determinants of County Growth," *Journal of Regional Science* 27 (February 1987), 39-54.
- Duffy-Deno, Kevin T., "The Effect of State Parks on the County Economies of the West," *Journal of Leisure Research* 29 (2nd Quarter 1997), 201-224.
- Duffy-Deno, Kevin T., "The Effect of Federal Wilderness on County Growth in the Intermountain Western United States," *Journal of Regional Science* 38 (February 1998), 109-136.
- Department of Commerce, Commonwealth of Pennsylvania, *Pennsylvania Statistical Abstract*, (1971, 1973).
- Glaeser, Edward L., Hedi D. Kallal, Jose A. Scheinkman, and Andrei Shleifer, "Growth in Cities," *Journal of Political Economy* 100 (December 1992), 1126-1152.
- Mankiw, N. Gregory, David Romer, and David N. Weil, "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics* 107 (May 1992), 407-437.
- Mills, Edwin S. and Richard Price, "Metropolitan Suburbanization and Central City Problems," *Journal of Urban Economics* 15 (March 1985), 1-17.
- Pennsylvania State Data Center, "Most Common County of Residence for Commuters to PA Counties, 1990," <[http://psdc.hbg.psu.edu/psdc/data/pa\\_county\\_level\\_data/48b.html](http://psdc.hbg.psu.edu/psdc/data/pa_county_level_data/48b.html)>.
- Sala-i-Martin, Xavier X., "I Just Ran Two Million Regressions," *American Economic Review* 87 (May 1997a), 178-183.
- Sala-i-Martin, Xavier X., "I Just Ran Four Million Regressions," Working paper, National Bureau of Economic Research, Cambridge, MA, (November 1997b).
- U.S. Department of Commerce, Bureau of the Census. *1972 County and City Data Book*, Washington, DC: U.S. Government Printing Office, (1972).