

CUBAN SOCIOECONOMIC INDICATORS BEFORE THE REVOLUTION: AN INTERNATIONAL COMPARISON

Luis Locay¹

Interest in Cuban socioeconomic indicators is usually focused on pre- and post-Revolutionary comparisons. That is not the purpose of this paper. My focus here will be only on the pre-Revolutionary period. The purpose of this work is quite specific. It is to see whether the available data on socioeconomic indicators are consistent with the recent direct estimate in Devereux and Ward (2009) (henceforth DW) of purchasing power parity (PPP) adjusted per capita income for Cuba in the 1950s.

Briefly, in this paper I take a set of countries, not including Cuba, augmented with individual states of the United States, and construct statistical models of PPP adjusted per capita income for 1955, using various socioeconomic variables as predictors. The models are then used to predict income for Cuba, and the predicted values are compared with that provided by DW.

The paper is organized as follows. The first section briefly discusses the data, and in the second, I describe Cuba's relative position for several of the indicators. The next section treats income as the dependent variable, and uses regression analysis to derive predicted income for Cuba. There I estimate regressions for single and multiple indicators. In the following section I treat income as an explanatory variable, and I estimate models for each socioeconomic indicator separately. I then estimate Cuban income as the number that, given the estimated models, "best fits" the Cuban values for

all the indicators. All procedures used yield similar results, and all are quite consistent with the value derived by DW that Cuban per capita GDP in 1955 was 27% of that of the United States. I end with a brief conclusion in the last section.

THE DATA

The basic set of countries included in this study consists of those countries for which DW provides PPP measures of relative per capita income in 1955. These are the countries of Latin America, eight countries of Europe (Belgium, Denmark, France, Germany, France, Italy, the Netherlands, Norway, and the United Kingdom), and the United States.² I have augmented this set with individual American states by using relative personal per capita income in 1955 as the relevant relative income measure.

For convenience of exposition, the socioeconomic indicators for the late 1950s on which data has been obtained can be grouped into the categories of consumption, education, health, and nutrition. This categorization is only for convenience of exposition, as some variables can conceivably be included in more than one category.

Under consumption are included telephones, passenger cars, televisions, radios and energy consumed. In the category of education are newspaper circulation and school enrollment. For health, I have used the infant mortality rate (IMR), life expectancy, and number

1. I wish to thank Pan Yijun and Carl Mbao for able research assistance.

2. The latest version of their paper also includes the Soviet Union, but I worked from an earlier draft.

of physicians. Under nutrition I simply have used calories consumed. I will also use a group of demographic variables that consists of urbanization, percent of the population that is not white, and population density. For phones, cars, newspapers, IMR, physicians and the demographic variables I have data for the countries and the individual American states. For the other variables I only have data at the country level. Table 1 presents some summary statistics for the data.

Table 1. Summary Statistics

	Availability	Sample Means			
		Europe	Latin America	US	Cuba
Income/Capita Relative to US = 100, 1955	All	49	14	100	27
Phone Lines/1000 Persons, 1957	All	130	14	326	23
Passenger Cars/1000 Persons, 1957	All	54	9	332	23
TVs/1000 Persons, 1960	Countries	78	8	308	73
Radios/1000 persons, 1959	Countries	263	83	948	152
Energy Consumption/Capita, Coal Equivalent Tons 1959	Countries	3.15	0.52	8.60	0.86
Newspaper Circulation/1000 Persons, 1960	All	328	73	330	96
Sum of Primary and Secondary Enrollment Rates, 1960	Countries	166	99	204	123
Infant Mortality Rate, Deaths per 1000 Live Births, 1955	All	34	89	26 ^a	36
Life Expectancy at Birth, 1955	Countries	71	51	69	62
Physians/10000 Persons, 1955	All	10.7	4.3	13.3	10.0
Daily Kcal/Capita, 1954–56	Countries	3018	2340	3150	2875
Fraction Urban, 1965	All	0.76	0.47	0.72	0.57
Fraction Non-White, 1962	All	0.02	0.71	0.11	0.29 ^b
Population Density, 1960	All	0.54	0.07	0.05	0.15

a. 1959 for US and individual states.

b. Average of 1953 and 1962.

CUBA'S RELATIVE INTERNATIONAL POSITION

This section illustrates the relationship in the raw data between some of the socioeconomic indicators and per capita income across countries and individual states, and Cuba's place among them. Needless to say, one does not wish to rely too heavily on a single indicator, as unaccounted-for country specific idiosyncratic factors can affect a given value of an indicator for a specific country at a point in time. Instead, one should look at the overall picture that emerges as one compares income to a wide variety of indicators.

Figure 1 shows the log of phone lines per 1000 persons in 1957 against the log of relative income in 1955. As

can be seen, phones and income were highly correlated (0.98). Based on the intensity of phone lines, the DW measure of Cuban income appears to be somewhat high, but probably not significantly so.

Figure 1. Phone Lines

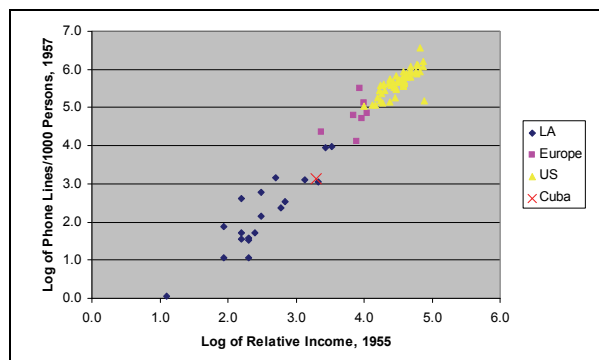
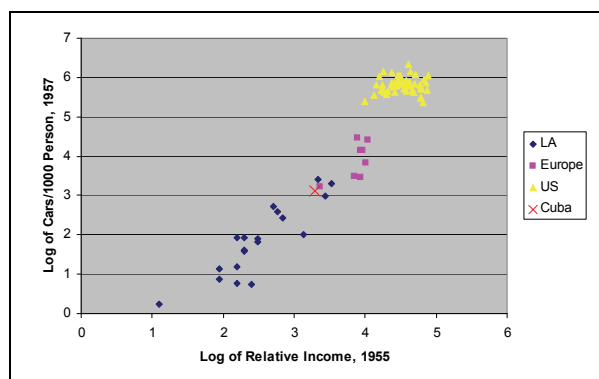


Figure 2 shows the relationship between passenger cars and income. Once again it is evident that the correlation between the two is very high (0.97). This time the DW measure of income is pretty much what one would expect given the intensity of passenger cars in 1957 Cuba.

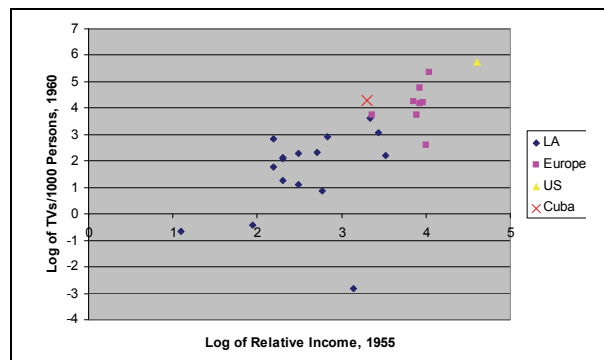
Figure 2. Passenger Cars



The third of the consumption indicators shown is the availability of television sets. I did not obtain data by state, so only country data are shown in Figure 3. Furthermore, Bolivia, Ecuador, and Paraguay are excluded because they reported no televisions in 1960. I presume that for reasons other than low income those countries were extremely slow in starting up television broadcasts.³ In terms of televisions, Cuba appears to have had higher levels of ownership than the DW measure of income would imply. Alternatively, the

availability of televisions in Cuba implies that DW's measure of income is too low.

Figure 3. Televisions



Turning to health indicators, Figure 4 shows the log of the IMR against the log of relative income. As with televisions, Cuba had an IMR that was lower than would be expected given the DW measure of Cuban income, or that the implied level of income for Cuba given the country's IMR is higher.

Figure 4. Infant Mortality Rate

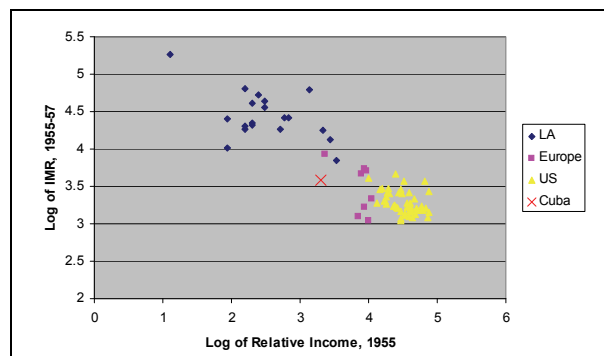
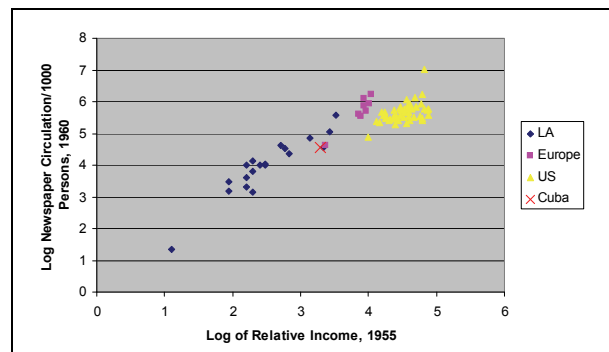


Figure 5 shows the relationship between newspaper circulation and income. Here Cuba appears to be low given the DW measure of income. Interestingly, with

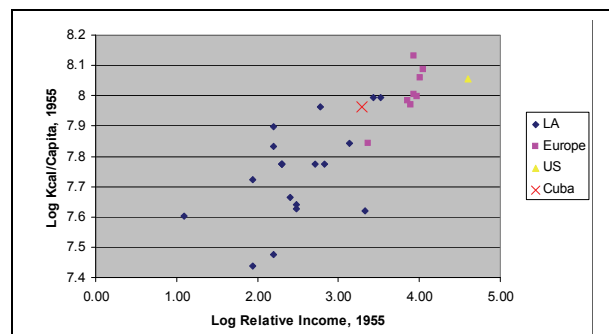
the exception of the District of Columbia, all of the U.S. also appears low.

Figure 5. Newspaper Circulation



Finally, Figure 6 shows the relationship between calories consumed and income. If anything, Cuba appears to have had higher consumption of calories than would be expected given the DW measure of income.

Figure 6. Calories



These six socioeconomic indicators are representative of the relationship in the raw data between all the indicators and income. The correlations (in the logs) are always high (from a low of 0.78 for calories to a high of 0.98 for phones and energy consumption). For some, the DW measure of income appears to be somewhat high given the international variation, but for others it seems somewhat low. For yet still others, it looks just about right. For none of the indicators does Cuba appear to be an outlier. This casual inspection of the data suggests that the DW measure of per capita income for Cuba in 1955 is broadly consistent with socioeconom-

3. It should be noted that the outlier with the lowest number of televisions per capita in Figure 3 is Chile. Neighbor Peru was also among the countries with few televisions.

ic variables from the period 1955–60. In the next section I provide more systematic evidence in support of this proposition.

INCOME AS DEPENDENT VARIABLE

Consider the following statistical model for the log of per capita income in country or state i , y_i :

$$y_i = \alpha_j + \beta_j x_{ij} + \lambda_j z_i + \varepsilon_{ij} \quad (1)$$

where x_{ij} is country i 's value of indicator j , z_i is a set demographic variables, and ε_{ij} is an error term with $E(\varepsilon_{ij}) = 0$. This error term captures unobserved variables whose effects are orthogonal to the other variables on the right hand side of (1). Equation (1) is to be interpreted merely as a statistical relationship, not a causal one. Its use is solely predictive.

The first procedure in this section is to estimate equations such as (1) for each socioeconomic indicator.

The estimated coefficients, α_j , β_j and λ_j , can then be used along with Cuban values for x_{Cj} and z_{Ci} , to obtain a point estimate of Cuban income, y_C . This estimate of the log of relative per capita income can then be compared with the DW value of $y_C = \ln(27) = 3.30$.

Table 2 shows the estimates of the coefficients of equations of type (1) for 11 different socioeconomic indicators. For each indicator two regressions were estimated. The first type of regression uses the indicator as the sole regressor. The second type includes the demographic variables as additional regressors. In all regressions, the indicator variable is highly significant. In all but one of the second type of regressions at least one of the demographic variables is significant at conventional levels. The fit is high in all cases, with R^2 s exceeding 0.80 in all of the regressions that include the demographic variables.

Table 2. Income Regressions Using Individual Indicators

Dependent Variable: Log of Relative per Capita Income		Independent Variables				R2	Obs.	Predicted Log of Income
Socioeconomic Indicator	Socio- Economic Indicator	Fraction Urban	Fraction Non-White	Population Density				
	0.57***				0.96	76	3.05	
Log of Phone Lines/1000 Persons	0.57***	0.25	0.022	-0.19	0.96	75	3.06	
	0.50***				0.92	78	3.20	
Log of Passenger Cars/1000 Persons	0.41***	0.79***	-0.37***	0.05	0.96	77	3.28	
	0.47***				0.75	24	3.78	
Log of TVs/1000 Persons ^a	0.25***	0.89	-0.88***	-0.37	0.85	24	3.59	
	0.69***				0.82	28	3.26	
Log of Radios/1000 Persons	0.42***	0.72	-0.58**	0.20	0.88	28	3.24	
	0.58***				0.91	28	3.12	
Log of Energy Consumption/ Capita, Coal Equivalent Tons	0.45***	0.14	-0.53**	-0.15	0.93	28	3.20	
	0.93***				0.83	78	3.32	
Log of Newspaper Circulation/1000 Persons	0.69***	0.06	-0.91***	-0.38*	0.86	77	3.45	
	2.09***				0.81	27	3.12	
Log of Sum of Primary and Secondary Enrollment Rates	1.18***	1.40**	-0.34	-0.03	0.84	27	3.13	
	-1.55***				0.83	76	3.95	
Log of Infant Mortality Rate, Deaths per 1000 Live Births	-1.14***	1.42***	-0.52**	-0.69***	0.91	75	3.87	
	3.54***				0.87	26	3.33	
Log of Life Expectancy at Birth ^b	2.49***	1.16***	-0.02	0.12	0.90	26	3.22	
	1.28***				0.80	76	4.18	
Log of Physicians/10000 Persons	0.92***	-0.32	-1.11***	-0.17	0.86	75	4.08	
	3.62***				0.62	28	3.45	
Log of Daily Kcal/Capita	1.42***	1.84***	-0.69**	-0.22	0.84	28	3.32	
Average of Predicted Values of Log of Relative Income with Indicator as Sole Regressor							3.43	
Average of Predicted Values of Log of Relative Income with Demographic Variables							3.40	
Log of Devereux-Ward Measure of Relative Income							3.30	

***, **, * Significant at the 1%, 5% and 10% levels, respectively.

a. Excludes Bolivia, Ecuador, Chile and Paraguay

b. Excludes Haiti and US

The last column of Table 2 shows the predicted values for the log of Cuban per capita GDP in 1955. While there are slightly more instances for which the predicted values are below the DW measure than for which they exceed it, the average predicted value exceeds the DW measure for both types of regressions. Focusing on the predictions from the regressions that include the demographic variables, the average predicted level of relative income is about 10% above that of DW. Under the assumption that the error terms are uncorrelated across equations (very unlikely), the average predicted value of 3.40 is within two standard errors of the DW measure of 3.30. In any case, we have no evidence that the DW measure is too low.

An alternative approach is to estimate an equation such as (1), but using multiple indicators. The corresponding equation would be:

$$y_i = \alpha + \beta x_i + \lambda z_i + \varepsilon_i \quad (2)$$

where x_i is now a vector of socioeconomic indicators. The primary limitation with this approach is the relatively small number of observations that I currently have for several of the indicators. I have data for countries and individual states for phone lines, passenger cars, newspaper circulation, IMR, and physicians. For the other indicators I have data only at the country level. Given the small number of observations in the latter group, and the high collinearity among the indicators, it is not possible to estimate (2) using all the indicators. Instead, I have carried out two separate regressions. The first involves countries and states, and includes as independent variables those indicators that I have available for the states (phone lines, passenger cars, newspaper circulation, IMR, and physicians), along with the demographic variables. The results of this estimation are shown in the first column of Table 3.

The second estimation involves only country level data. For this regression I included as explanatory variables those independent variables from the first regression that were significant at conventional levels, plus the indicators that were not included in the first re-

Table 3. Income Regressions with Multiple Indicators

Dependent Variable:		
Log of Relative per Capita Income		
Variables	Countries and States	Countries Only
	0.13**	-0.05
Log of Phone Lines/1000 Persons	(0.07)	(0.12)
	0.25***	0.10804
Log of Passenger Cars/1000 Persons	(0.04)	(0.09)
		0.01
Log of Radios/1000 persons		(0.10)
		0.25**
Log of Energy Consumption/Capita, Coal Equivalent Tons		(0.10)
	0.09	
Log of Newspaper Circulation/1000 Persons	(0.06)	
		0.28
Log of Sum of Primary and Secondary Enrollment Rates		(0.29)
	-0.22**	-0.18
Log of Infant Mortality Rate, Deaths per 1000 Live Births	(0.09)	(0.12)
		0.47
Log of Life Expectancy at Birth		(0.60)
	0.08	
Log of Physicians/10000 Persons	(0.09)	
		0.52
Log of Daily Kcal/Capita		(0.34)
	0.56***	0.37
Fraction Urban	(0.19)	(0.39)
	0.02	
Fraction Non-White	(0.12)	
	-0.05	
Population Density	(0.01)	
	1.99***	-3.88
Intercept	(0.065)	(2.70)
R2	0.98	0.97
Obs.	74	27
Predicted Relative Income	3.30	3.33

***, **, * Significant at the 1%, 5% and 10% levels, respectively. Standard errors in parentheses.

gression because they were not available for the states.⁴ The results from the regression are shown in the second column of Table 3.

At the bottom of each column of Table 3 appears the predicted value of the log of Cuban per capita relative income. As can be seen, the predicted income values from both regressions (3.30 and 3.33) are very close to the DW calculation (3.30). In both cases the point estimates for Cuba are well within two standard errors of the DW measure.

4. The indicator of television availability was not included because three countries report zero televisions for 1955, and a fourth, Chile, nearly zero.

Table 4. Indicator Regressions with Income as Explanatory Variable

Dependent Variable	Independent Variables				R2	Obs.
	Log of Income	Fraction Urban	Fraction Non-White	Population Density		
	1.68***				0.96	76
Log of Phone Lines/1000 Persons	1.44***	-0.04	-0.77***	0.12	0.97	75
	1.84***				0.93	78
Log of Passenger Cars/1000 Persons	2.03***	-1.70***	-0.03	-0.03	0.95	77
	1.63***				0.76	24
Log of TVs/1000 Persons ¹	1.32***	2.25	0.79	0.96	0.81	24
	1.18***				0.84	28
Log of Radios/1000 persons	1.04***	0.97	-0.24	-0.70	0.87	28
	1.59***				0.92	28
Log of Energy Consumption/Capita, Coal Equivalent Tons	1.44***	1.30	0.35	0.21	0.93	28
	0.89***				0.81	78
Log of Newspaper Circulation/1000 Persons	0.61***	0.86***	-0.58**	-0.06	0.84	77
	0.39***				0.9	27
Log of Sum of Primary and Secondary Enrollment Rates	0.27***	0.01	-0.33**	-0.07	0.92	27
	-0.54***				0.83	76
Log of Infant Mortality Rate, Deaths per 1000 Live Births	-0.54***	0.85***	0.29**	-0.37***	0.88	75
	0.25***				0.88	26
Log of Life Expectancy at Birth ²	0.19***	-0.11	-0.17**	0.003	0.91	26
	0.63***				0.80	76
Log of Physicians/10000 Persons	0.36***	1.20***	-0.43**	-0.32**	0.88	75
	0.17***				0.62	28
Log of Daily Kcal/Capita	0.18***	-0.21	-0.07	0.07	0.65	28

***, **, * Significant at the 1%, 5% and 10% level, respectively. ¹ Excludes Bolivia, Ecuador, Chile and Paraguay. ² Excludes Haiti and US.

INCOME AS EXPLANATORY VARIABLE

In this section I reverse the question of the previous section. I now ask, “What level of income would best explain the values of Cuba’s socioeconomic indicators?” To answer this question for each indicator, I first estimate equations of the following form:

$$x_{ij} = a_j + b_j y_i + c_j z_i + u_{ij} \quad (3)$$

where u_{ij} is an error term with $E(u_{ij}) = 0$. Let y_C be the log of per capita Cuban relative income. Using the estimated coefficients for (3), a_j , b_j and c_j , I construct for each indicator a residual for Cuba as follows:

$$u_{Cj} = x_{Cj} - (a_j + b_j y_C + c_j z_C) \quad (4)$$

I then search for the y_C that “best fits,” where that term is made clear below.

Ordinary least squares (OLS) estimates of equation (3) for each indicator are shown in Table 4. Two equations were estimated for each indicator. The first uses the log of relative income as the only regressor. The second equation includes the demographic variables. As with the reverse of these equations in Table 2, the goodness of fit for all the indicators is quite high, and

the coefficient of the log of relative income is always highly significant.

For a set of weights corresponding to the indicators, $w = (w_1, \dots, w_J)$, and a given level of the log of income, y_C , define the following sum of squared errors:

$$S(w, y_C) = \sum_j w_j (u_{Cj})^2 \quad (5)$$

where the residuals, u_{Cj} , $j = 1, \dots, J$, are determined according to equation (4). I now define the level of relative income that best explains the values of Cuba’s socioeconomic indicators (the one that “best fits”) as the value of y_C that minimizes equation (5).

To be able to minimize equation (5) we need to determine a set of weights. I consider two cases. In the first case I set all weights equal to one. In the second case I use as weights the inverse of the standard error of the coefficients of the log of relative income from the two types of regressions in Table 4. Thus there are four different estimates. These are shown in Table 5. When the coefficient estimates from the regressions with the log of relative income as the sole independent variable are used, the best fitting values for the log of Cuban relative income are 3.34 with equal weights, and 3.26

with the inverse of the standard errors as weights. These estimates imply levels of income for Cuba that are 4% above and 4% below the DW measure. When the coefficient estimates from the regressions that include the demographic variables are used, the corresponding values are 3.35 and 3.23, or 5% above and 7% below the DW measure.

Table 5. Best Fitting Cuban Relative Income

Regressors	Weights	
	Equal	Inverse of Standard Error
Log of Relative Income Only	3.34	3.26
Log of Relative Income and Demographic Variables	3.35	3.22

To see how likely such deviations are, I computed the same type of estimates for each of the countries in the data set using the coefficient estimates from the regressions that included the demographic variables and the weighted sum of squares. Considering only the 21 countries that were included in the eleven indicator regressions, I found that for eight of them, the best fitting income level was more than 8% below the actual level, and for five of them it was more than 8% above.

REFERENCES

- Abouharb, M. R. and A. L. Kimball (2007). "A New Dataset on Infant Mortality Rates, 1816–2002," *Journal of Peace Research*, vol. 44, no. 6, 2007, pp. 745–756.
- Bureau of the Census, U.S. Department of Commerce (1956–62). *Statistical Abstract of the United States*. Washington: U.S. Government Printing Office.
- Bureau of the Census, U.S. Department of Commerce (1961). *Census of Population: 1960*. Washington: U.S. Government Printing Office.
- Devereux, J. and M. Ward (2009). "The Road Not Taken: Pre-Revolutionary Cuban Living Standards in Comparative Perspective," unpublished, Queens College, CUNY.
- Oxford Latin American Economic History Database at: <http://oxlad.qeh.ox.ac.uk/>

Among all 28 countries there is even more dispersion. Eighteen of them have best fitting income levels that are more than 8% above or below their actual measures. Consequently, the deviations between the estimates of y_C and the DW measure are well within the variation we observe among the various countries.

CONCLUSION

In this paper I employ a variety of techniques for estimating Cuban per capita relative income in 1955 from socioeconomic indicators. The eight estimates nicely bracket the DW measure of 27% of U.S. per capita income. If we consider only those estimates that use the demographic variables, the estimates are within $\pm 8\%$ of the DW measure. The two most direct measures—multiple regression with income as the dependent variable—actually give estimates that are equal to, and 3% above, the DW measure. I believe this is significant evidence that the measure arrived at in Devereux and Ward (2009) through direct measurement is quite consistent with the values of Cuba's socioeconomic indicators.