

Geography and the Determinants of Land Distribution: Evidence from the United States

Dietrich Vollrath[†]
University of Houston

September 12th, 2006

ABSTRACT: While the adverse impacts of land inequality have often been documented, the origins of this inequality are less commonly discussed. One hypothesis holds that the type of crops available for production dictated the initial land distribution within economies. Theoretically, though, it is entirely possible that farm size influences the choice of crops planted and that therefore the distribution of land determines the crop mix within an economy. This paper uses county level data from the 1860 Agricultural Census of the United States to identify the causal role of crop type on land distribution. Identification is obtained by using several county level geographic variables as instruments for the type of crops planted within a county. The results show that the percentage of output made up of cash crops (cotton, sugar, and tobacco) was highly significant in determining land inequality across U.S. counties in 1860. Geography, both indirectly through its influence on crop type and directly due to heterogeneity of land within counties, had a significant impact on land inequality. This supports the general hypothesis that geography affects economic development indirectly through institutions. However, it provides a much more specific example of the role of geography in development than previously identified.

JEL Codes:

Keywords: Land distribution

[†] Corresponding author. 201C McElhinney Hall, Houston, TX 77204. devollrath@uh.edu

The author would like to thank the University of Houston New Faculty Grant Program for financial support.

1 Introduction

Inequality, and in particular inequality of land holdings, has been cited many times as a drag on economic development. Research by Engerman and Sokoloff (1997) suggests that the inequality of land holdings in Central and South America led to a concentration of political power and the active repression of democracy and education. Galor, Moav, and Vollrath (2006) cite evidence from across U.S. states that higher land inequality limited growth in education spending in the early 20th century. Tomich, Kilby, and Johnston (1995) discuss a wide range of limitations on the agricultural sector that result from concentrated land distributions. Vollrath (2006) shows that across countries, land inequality significantly lowers agricultural output.

While there are many examples of the detrimental effects of land inequality, there is far less research regarding the origins of land inequality in the first place. Engerman and Sokoloff (1997) suggest that the type of agriculture available to the colonizers in the Americas affected the initial agrarian structure. Their hypothesis is that in those areas where cash crops, such as sugar, could be grown were more likely to be organized into plantations, given the high fixed costs of producing these cash crops. However, agricultural research has shown that farm size is a very important element of the choice of crop (see Tomich, Kilby, and Johnston 1995). Larger farms are capable of absorbing more uncertainty, and so may choose riskier crops. In addition, their larger scale means that crops with high fixed costs are available as well. So it is not necessarily true that the geographic conditions affecting crop choice were significant in determining land distribution.

This paper addresses the question of the origin of land inequality by using county level data from the 1860 Agricultural Census of the United States. From this data it is possible to calculate several measures of land distribution within counties, and in addition the Census included detailed information on the type of crops produced within each county. Thus it is possible to see clearly the close relationship of crop type and land inequality. To identify any causal role for crop type on land inequality, additional data on the geographic conditions of counties are used as instruments for crop type. The instruments are geographic characteristics that plausibly affect only the choice of crop (annual precipitation, variance in precipitation, and length of growing season) while additional

geographic characteristics are included as exogenous variables that could affect land distribution directly (heterogeneity in soil types, total land area).

The results show a very significant role for crop choice (measured as the percent of total output made up of cash crops) on land inequality in 1860. This effect is robust to the inclusion of other county level control variables, to the use of different measures of land inequality, and to the exclusion of northern counties from the sample. The results show quite clearly that the type of crops that can be produced have a major impact on the agrarian structure of a county. Given the importance of land inequality, this shows a very specific channel by which geography affects economic development.

This finding supports the general hypothesis that geography works indirectly on income levels through its effects on economic institutions. Acemoglu, Johnson, and Robinson (2001) find that the mortality experience of early colonizers (something closely associated with geographic factors) had a significant impact on the nature of political institutions installed by the colonizers. Easterly and Levine (2003) show that geographic factors are significant predictors of institutional quality, but do not appear to have a direct effect on income levels.

The paper proceeds as follows. Section 2 discusses in more detail how land distribution is measured using the data from the Agricultural Census of 1860. Section 3 then uses this land inequality data to show the relationship between crop choice and land distribution. Section 4 concludes.

2 Measuring Land Distribution

The method of measuring land inequality follows the methodology used by Deininger and Squire (1998) in their study of cross-country inequality. For each county reporting data in the 1860 Agricultural Census, the distribution of farms by size is available. The census lists the number of farms in the following categories: 3-9 acres, 10-19 acres, 20-49 acres, 50-99 acres, 100-499 acres, 500-999 acres, and greater than 1000 acres. No data are reported on farms less than three acres. This distribution, combined with assumptions about the average area of farms within each category,

allows for the estimation of a Gini coefficient. Figure 1 shows how this is accomplished. The data on farm and acreage distribution is combined to create a Lorenz curve, and the Gini is then found as the ratio of area A to the total area of the triangle defined by the 45 degree line.

A more formal definition is as follows. There are eight size categories, including a category that measures farms of size zero, numbered from 1 to 8 in order of increasing size of farms. Let f_i be the share of all farms that are in category i . Let a_i be the share of all acreage that is in category i . Now let $F_i = \sum_{s=1}^i f_s$, which denotes the share of farms that are of size i or smaller. Similarly, $A_i = \sum_{s=1}^i a_s$. By definition, $F_8 = A_8 = 1$. It can be shown that the Gini coefficient, G , can be calculated as follows

$$G = 1 - \sum_{i=1}^7 (F_{i+1} - F_i) (A_{i+1} + A_i). \quad (1)$$

This method requires data on the share of acreage in each farm size category, which is not reported in the 1860 Census. To address this, it is assumed that within each category, each farm is of the average size for that category. For example, each farm in the 50-99 acres category is assumed to be 74.5 acres. For the over 1000 acre category, each farm is assumed to be 2000 acres.¹

The data allow for Gini calculations on 1988 counties within the United States. The average Gini coefficient is 0.475, and the median is 0.478. The standard deviation is 0.082. The minimum value calculate is zero (for several counties in which all farms are listed in the same category) and the maximum value is 0.85 (Starr county, Texas). Table 1 lists the mean values for the Gini by census division. As can be seen, the average Gini is higher among those divisions in the South, and lowest in the Midwest.

There are several issues with the Gini coefficient. First, it is calculated using the area of farms, not their value. This may create the illusion of inequality of land where none exists. If small farmers have higher quality land within a county and large farmers have poor land, then the inequality in farm area is greater than the inequality in farm value. Without data on farm values, there is no way to correct for this. However, it is hoped that land values within counties (as opposed to within states or countries) are relatively homogenous and that this issue is not debilitating. A

¹The choice of the value of 2000 acres matches the average acreage of farms over 1000 acres found in the 1900 Census of Agriculture. Using different values for this category does not materially impact the results of the paper.

second issue with the Gini is that it only captures inequality in the distribution of existing farms, without regard for the actual number of farms relative to the rural population.² The Gini thus cannot account for the presence of landlessness among the rural population. There is data available to address, in part, this shortcoming.

The first is simply the rural inhabitants relative to the number of farms. The rural population is defined as all people not living in towns of 2,500 or more (the Census' definition of an urban area). This will give a crude measure of the availability of farms, regardless of size. We have 1941 observations of this ratio, with a mean of 17.41, a median value of 12.63 and a standard deviation of 27.73. The minimum value is 3.8 and the maximum is 698.5. The value by Census division is located in Table 1, showing that the number of rural persons per farm was generally lower in the Northeast and Midwest than in the South and West.

One issue with using total rural population is that differences in family size across counties may be obscuring the true availability of land to the active agricultural population. For a more refined measure, we construct a measure of the number of rural adult males relative to the number of farms. This is intended to give a more accurate measurement of how the number of farms relates to the number of potential primary farmers. The number of rural adult males is calculated as follows. The total number of adult males between the ages of 20 and 69 is obtained from the Census of 1860. This number is then scaled down by the ratio of rural population (defined above) to the total population (from the Census), a method which presumes that the overall distribution of the population is mirrored exactly by the adult male population. There may be some issue with this if adult males are felt to be more likely to live in cities, although there is no clear indication the bias should run one direction over the other. The number of rural adult males per farm has a mean value of 4.60 and a median of 2.91. Values by Census Division are found in Table 1.

It is possible that inequality of farm sizes (the Gini coefficient) is related to inequality in the number of farms (adult rural males per farm). The two measures do have a significantly negative correlation, with a coefficient of -0.066 . Figure 2 plots the log of rural adult males per farm

²An example highlights the issue. Consider county A, which has 1000 acres and 10 people. The 1000 acres are operated by two people, each with 500 acres. The Gini for this county is zero, because both farms of equal size. In contrast, county B has 1000 acres and 10 people, but each person owns 100 acres. In county B the Gini is zero as well. So the Gini coefficient will not indicate any difference in land distribution between these counties.

against the Gini coefficient, and this shows no obvious correlation between the two measures. It appears that these two different series are capturing different aspects of land distribution.

Finally, it is possible to create a more sophisticated version of the Gini coefficient that incorporates the information available from population/farm ratios. Vollrath and Erickson (2006) describe the method more precisely, but the formula for the overall Gini (G_{OV}) is as follows

$$G_{OV} = G \times \left(\frac{\text{Farms}}{\text{Adult Rur Males}} \right) + \left(1 - \frac{\text{Farms}}{\text{Adult Rur Males}} \right). \quad (2)$$

The G_{OV} measure is essentially a weighted average of the original Gini G and one. The weighting is based on the number of farms per rural adult male. As the number of farms per rural adult male approaches one, then the overall Gini is simply equal to the original Gini. As the number of farms declines the overall Gini approaches one, regardless of the distribution of the existing farms. The G_{OV} captures two dimensions of land distribution in a single measure.

3 Crop Choice and Land Distribution

The simple correlation between the types of crops grown and the distribution of farmland can be seen in figure 3. This plots overall Gini coefficient, G_{OV} , against the percentage of total crop output made up of cotton, tobacco, or sugarcane. The positive relationship is apparent, showing that U.S. counties that had cash crops as a larger percent of their total output also had more unequal distributions of farmland. While the correlation is apparent, the causal nature of this relationship is not clear. To address this problem, we will use several geographic variables as instruments for the type of crops produced, identifying the causal role that crop types had on the degree of inequality across counties.

3.1 Geographic Instruments and Controls

The GEOECOLOGY database of Olson, Emerson, and Nungesser (2003) is a county-level environmental dataset for the United States, based on observations from several sources over the years

1964-1979. While the time period in which they observe the geographic data does not overlap the period of analysis (1860), the data they collect is assumed to be very stable over time. One reason for this assumption is that the data they collect is on characteristics that are not immediately affected by human activity. For example, while Olson et al (2003) have data on the type of soil within a county (a static item), they do not have data on the depth of soil (a variable item affected by human activity).

The geographic data chose for use as instruments for the type of crops produced consist of several series. They are

1. *Length of growing season*: This is measured as the number of days between the average last day of frost in the spring and the average last day of frost in the fall.
2. *Total annual rainfall*: In centimeters, the total amount of precipitation received over the course of an average year.
3. *Standard deviation of average monthly rainfall*: This captures the variability of rainfall over the course of a year.

These three elements are chosen because they are crucial to the type of agriculture pursued. In addition, it is assumed that they do not have any direct bearing on how the farmland of a county has been allocated, except through the choice of, and productivity of, crops.

While these characteristics are assumed not to affect the distribution of land, several other geographic characteristics may have a direct affect on distribution. The Gini coefficient presumes that all acres of land are homogenous, something that is clearly not true. If land within a county is heterogeneous, then farms may differ by size but not by quality or value. Geographic diversity within a county may lead mechanically to higher Gini coefficients while not revealing anything important economically. To control for this heterogeneity, two further characteristics are included. The first is the *log of total farm land (ha)*, on the presumption that the larger a county is, the more likely there are to be variations in land quality within the county

The second characteristic is more complex. The *relative entropy of soil order* measures the diversity of soil types within a county. Olson et al (2003) provide, by soil order, the number of

hectares within in county. There are nine soil orders, representing the broadest classification of soil types.³ The soil entropy of a county can be calculated as follows

$$E = \sum_{i=1}^9 p_i \ln(1/p_i). \quad (3)$$

In (3), the term p_i is the proportion of all land in the county of soil type i . For any $p_i = 0$, it is assumed that $p_i \ln(1/p_i) = 0$. The maximum entropy that can be achieved in any county, given that there are nine orders of soil, is $E_{\max} = \ln(9) = 2.197$. The relative entropy of soil order is then defined as

$$RE = \frac{E}{E_{\max}} \in [0, 1]. \quad (4)$$

3.2 Instrumental Variables Regressions

With the geographic data in hand we can now address the causality of crop type for land inequality. The regressions will use one of the measures of land distribution (the Gini, overall Gini, or adult rural males per farm) as the dependent variable. The percent of total crop output coming from cotton, tobacco, or sugarcane is the explanatory variable of interest. The log of total farm land and the relative soil entropy will be included as additional geographic control variables for land inequality. The combination of the datasets provides a sample of 1828 counties. The summary statistics for all variables that will be used are in table 2.

The initial results are in tables 3, 4, and 5. The first table regresses the overall Gini coefficient, G_{OV} , on the percent of total crop output from cotton, tobacco, and sugar. The first two columns are simple OLS regressions, showing a very significant positive relationship between the percentage of cash crops and the degree of land inequality within counties, even when controlling for within county heterogeneity in soil and total county area. Both of those factors have significant effects on land inequality, but do not disturb the result on cash crops.

As noted in the introduction, this relationship is not surprising given our knowledge of U.S.

³The nine orders are Alfisol, Aridisol, Entisol, Histosol, Inceptisol, Mollisol, Spodosol, Ultisol, and Vertisol. The orders differ by their clay content (which affects water retention), pH balance, cation exchange capability (a measure of the ability of plants to absorb nutrients), and organic matter.

history. What this OLS regression cannot address, though, is the issue of causation. The final columns of table 3 control for the endogeneity of the cash crop measure by instrumenting it using the three geographic instruments described above: annual precipitation, the standard deviation of precipitation, and length of the growing season. In each regression, the coefficient on cash crops is positive and highly significant. The size of the coefficient has actually increased from the OLS regressions, indicating that the causal effect of cash crops on land inequality is stronger than the overall correlation. This implies that the effect of land inequality on the production of cash crops was actually negative.

The results for cash crops are robust to the inclusion of further controls for county characteristics. In particular, the percent urbanization (calculated as the share of population within a county residing in cities over 2,500 people, as reported in the 1860 Census), and access to rail or water transport (binary variables from the 1860 Census), do not alter the results.

A final control is the log of total agricultural output per hectare. If the land within a county is generally poor, then it may be that the only economical way of farming is to organize production of cash crops with high fixed costs. If this were the case, then the presence of cash crops would simply be a proxy for land quality. Total agricultural output per hectare is measured using data reported in the Agricultural Census of 1860 on the quantity and price of goods produced within each county. The total number of goods is thirty-three. The nominal value of agricultural output is used in the regressions here, but measures of real output do not significantly alter the results. There is very little variation reported across counties in the prices of different goods, so that the nominal and real output series are almost identical.

The inclusion of agricultural output as a control does not alter the result that cash crops significantly influenced inequality within counties. So even among counties with equal productivity overall, those counties which had the conditions for cash crop production found themselves with a less equitably distributed endowment of land.

In each of the two-stage least squares regressions in table 3, the equation is actually overidentified (i.e. there are three instruments for the cash crop percentage). The final rows of the table report the test statistics of the Hansen J test for overidentification. In no case can we reject the null

hypothesis that all the instruments are valid. Statistics regarding the first stage are also reported, showing that a very large portion of the variation in the cash crop percentage is explained by the full set of exogenous variables, including instruments.

The estimated coefficient indicates that a one standard deviation increase in the share of cash crops in all output, an increase of about 27 percentage points, would lead to an increase in G_{OV} of approximately 0.047. This is approximately two-thirds of a standard deviation in the sample of G_{OV} . The effect of cash crops appears to be economically meaningful as well as statistically significant. Considering that the difference in the mean cash crop percentage between north and south in the United States is 32 percentage points, the estimates imply a difference of 0.056 in the overall Gini coefficient between these two regions. The actual difference in mean G_{OV} is in fact 0.059.

As noted previously, the overall Gini coefficient (G_{OV}) combines two different dimensions of data on land distribution. The regressions reported in tables 4 and 5 replicate the specifications for G_{OV} , except using the basic Gini coefficient (G), or the log of rural adult males per farm as the dependent variable. In both cases, the results follow the same pattern as for G_{OV} . Instrumenting with the geographic data, we see that the percent of cash crops produced in a county has a highly significant impact on the level of inequality within a county. More cash crops mean that not only is the existing land more concentrated on large farms (i.e. a high G), but there are fewer farms relative to the population (i.e. a higher number of rural adult males per farm). The tests of overidentification again show that the instruments are valid within these specifications and the first stage results are again very strong.

One concern may be that the results are driven primarily by the different nature of agriculture between the north and south, and that the regressions are only capturing this broad difference. In particular, the variation in the cash crop percentage in the North is very small, so it is possible to argue that they only represent a single observation. To address this concern, table 6 presents results from regressions using the three different measures of land inequality as dependent variables, but including either a dummy variable for Southern states or excluding the Northern states entirely.⁴

⁴The states defined as Southern are: Alabama, Arkansas, Delaware, Florida, Georgia, Kentucky, Louisiana,

While the size of the coefficient estimate on the cash crop percentage decreases when the Southern dummy variable is included (in columns 1-3), the estimates are nearly identical when the North is excluded entirely. This implies that some structural difference existed between the north and south that led to higher land inequality in the South, outside of the differences in geography. However, within the South, geographic differences still show themselves as having a significant impact on land inequality through their influence on the choice of crops.

Additional confirmation of the viability of the geographic instruments is available due to the over-identification. We test the excludability of each individual instrument by running regressions with each as an included exogenous variable while the remaining two instruments are excluded exogenous variables. A statistically significant coefficient for any of the instruments would suggest that they are not correctly excluded from the main regressions. Table 7 shows the coefficient estimate and significance for each instrument when it is included in a fully specified regression. As can be seen, there is no case where any of the three instruments have a significant impact directly on land inequality. The effect of these geographic variables on inequality is solely through their effect on the percentage of cash crops in output.

The overall implication of the regressions presented here is that geographic characteristics can have very powerful effects on meaningful economic institutions, such as land distribution. Within the United States, those counties that had the right conditions for cotton, tobacco, and sugar were much more likely to have unequal distributions of land in all dimensions. Given the long-lasting effects of land inequality on many economic outcomes, these results show the power of geography to indirectly influence economic development.

4 Conclusion

The importance of land inequality in determining economic outcomes has been discussed at length, but rarely has that discussion touched on the origins of land inequality in question. This paper has attempted to bring to light one particularly important determinant of initial land inequality, Mississippi, Maryland, North Carolina, South Carolina, Tennessee, Texas, and Virginia.

the type of crops available for production. Using a broad county level dataset from the United States, it was shown that the greater the importance of cash crops, the higher was land inequality. The causality was established by using geographic data on annual rainfall and the length of the growing season as instruments for the choice of crop types. Further geographic features, such as the heterogeneity of soil types, were found to have significant impacts directly on the distribution of land within counties as well.

The importance of this result is in showing the deep effects that geography can have on subsequent economic development. While sunlight, temperature, and rain may be inconsequential for many modern economic activities, they were crucial in determining the initial distribution of land within U.S. counties. The general relationship of geography and institutions has been established in the literature, but this paper has provided a much more specific mechanism by which geography influences the long run performance of economies.

References

- Acemoglu, D., Johnson, S. & Robinson, J. A. (2001), ‘The colonial origins of comparative development: An empirical investigation’, American Economic Review **91**(5), 1369–1401.
- Deininger, K. & Squire, L. (1998), ‘New ways of looking at old issues: Inequality and growth’, Journal of Development Economics **57**(2), 259–87.
- Easterly, W. & Levine, R. (2003), ‘Tropics, germs, and crops: the role of endowments in economic development’, Journal of Monetary Economics **50**(1).
- Galor, O., Moav, O. & Vollrath, D. (2003), ‘Land inequality and the origin of divergence and overtaking in the growth process: Theory and evidence’, Brown Economics Working Paper .
- Haines, M. (2004), Historical, demographic, economic, and social data: The united states 1790-2000. ICPSR Study No. 2896.
- Olson, R. J., Emerson, C. J. & Nungesser, M. K. (2003), Geoecology: County-level environmental data for the united states, 1946-1979. Oak Ridge National Laboratory Distributed Active Archive Center.
- Tomich, T. P., Kilby, P. & Johnston, B. F. (1995), Transforming Agrarian Economies, Cornell University Press, Ithaca, NY.
- Vollrath, D. (2006), Land distribution and international agricultural productivity. Working Paper.
- Vollrath, D. & Erickson, L. (2006), Land distribution and financial system development. Working Paper.

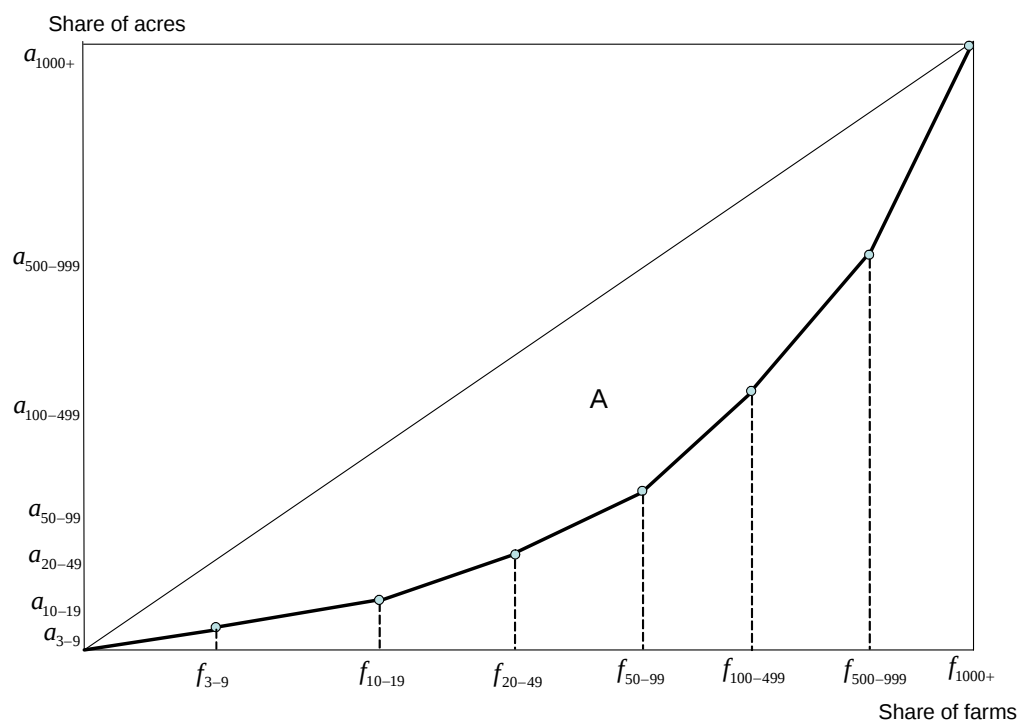


Figure 1: Calculation of the Gini Coefficient

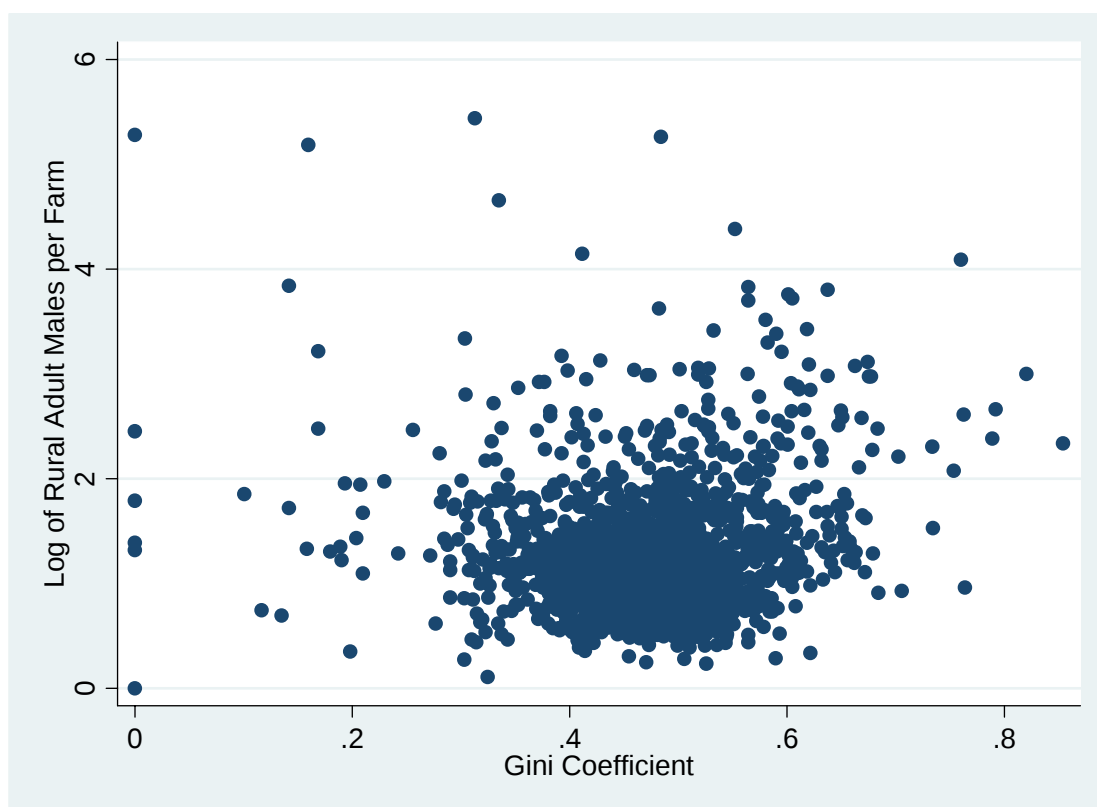


Figure 2: Comparison of the Log of Adult Rural Males per farm and the Gini coefficient, by county

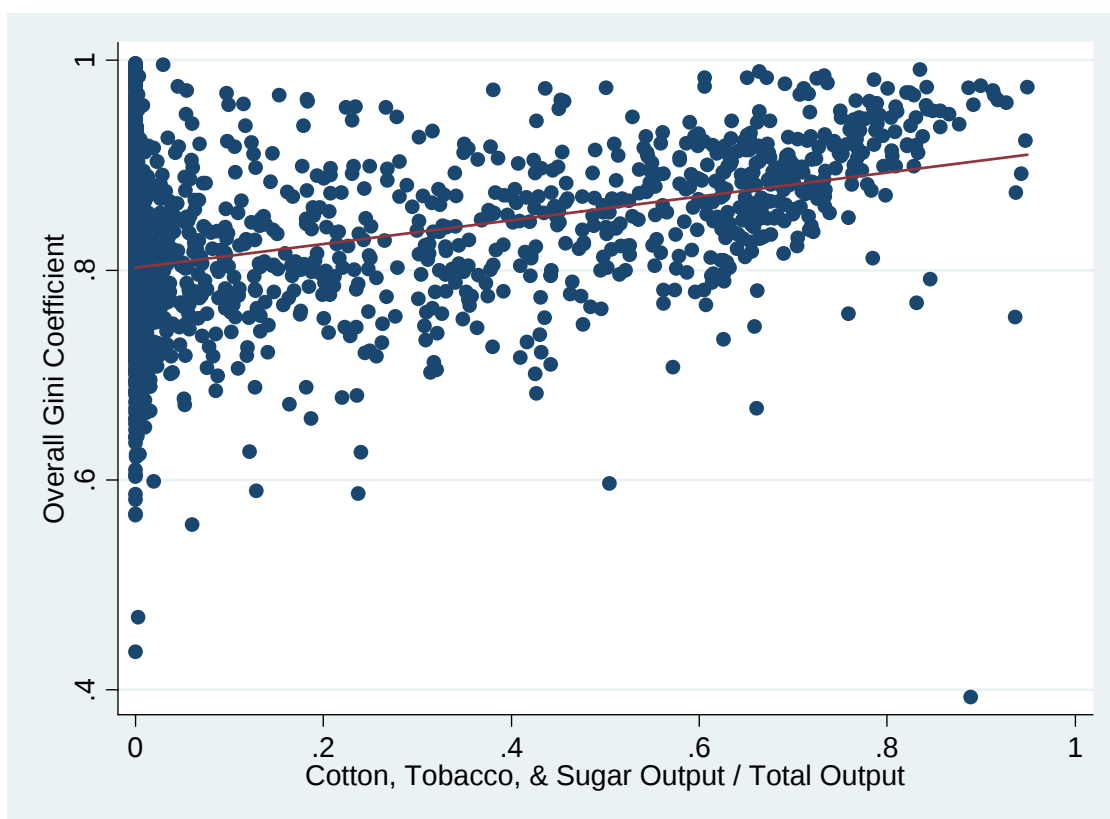


Figure 3: Relationship of the Overall Gini Coefficient to Type of Crops Produced

Division	Gini (G)	Rural Pop/Farm	Rur Adult Males/Farm	Overall Gini (G_{OV})
New England	0.456	11.29	2.92	0.797
Middle Atlantic	0.451	15.20	3.71	0.825
East North Central	0.449	15.08	4.29	0.787
West North Central	0.438	11.44	3.07	0.780
South Atlantic	0.486	19.97	4.44	0.846
East South Central	0.497	16.94	3.78	0.825
West South Central	0.536	25.39	6.58	0.869
Pacific	0.500	27.10	16.03	0.901
All	0.475	17.41	4.60	0.821

No data is available from the Mountain Division

Data are from authors calculations from data described in text

Table 1: Mean Values of Land Distribution Measures, by Census Division

Variable	Number	Mean	SD	Min	Max
G	1828	0.47	0.07	0.10	0.82
G_{OV}	1828	0.82	0.07	0.39	0.99
Rural adult males / Farm	1828	3.79	3.64	1.11	63.5
Percent cotton, tobacco, sugarcane	1828	0.18	0.26	0.00	0.95
Growing season (days)	1828	191.4	40.3	92	345
Total annual precipitation (cm)	1828	107.5	20.4	19.1	176.7
Standard deviation of monthly precipitation	1828	7.57	2.90	2.06	24.92
Log of total farm acreage (ha)	1828	11.18	1.43	3.71	13.70
Relative soil entropy	1828	0.11	0.14	0.00	0.53
Log of total output per acre (\$)	1828	0.04	0.59	-6.10	2.98
Percent urban	1828	0.05	0.13	0.00	0.99
Percent slaves	1828	0.17	0.22	0.00	0.93
Rail Access (0=No, 1=Yes)	1828	0.40	0.49	0	1
Water Access (0=No, 1=Yes)	1828	0.41	0.49	0	1

Table 2: Summary Statistics

Exp Variables	Dep Variable: Overall Gini Coefficient (G_{OV})				
	(1)	(2)	(3)	(4)	(5)
<i>Endogenous</i>					
% Cotton, Tobacco, and Sugar	0.123** (11.30)	0.122** (11.50)	0.155** (8.04)	0.164** (10.71)	0.176** (10.79)
<i>Exogenous</i>					
Log Output per Ha					-0.013** (2.40)
Rel Soil Entropy		0.061** (3.63)	0.055** (3.37)	0.035** (2.08)	0.039** (2.33)
Log Total Area (ha)		-0.005 (1.41)	-0.006 (1.62)	-0.012** (3.51)	-0.014** (4.15)
Percent Urban				0.050** (3.97)	0.055** (4.57)
Rail Access				0.022** (3.90)	0.021** (3.61)
Water Access				0.038** (7.55)	0.037** (7.44)
Method ^a	OLS	OLS	2SLS	2SLS	2SLS
First stage F-test p-value ^b			<0.01	<0.01	<0.01
First stage R-squared ^c			0.51	0.53	0.54
Hansen J Statistic ^d			2.38	0.47	0.29
Hansen p-value ^e			0.30	0.79	0.86

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

Standard errors are clustered by state

* denotes significance at 10%, ** denotes significance at 5%.

a) instruments are length of growing season, annual precipitation, and the SD(monthly precipitation)

b) p-value of the test of joint significance of the three instruments in the 1st stage regression for the % cotton, tobacco, and sugar in total output

c) R-squared of the first stage regression for % cotton, tobacco, and sugar

d) Test statistic, distributed $\chi^2(2)$ in columns 3-5, $\chi^2(2)$ in column 6, for Hansen's J test of overidentifying restrictions.

e) p-value of Hansen's J test

Table 3: **Determinants of the Overall Gini Coefficient**

Exp Variables	Dep Variable: Gini Coefficient (G)				
	(1)	(2)	(3)	(4)	(5)
<i>Endogenous</i>					
% Cotton, Tobacco, and Sugar	0.101** (5.98)	0.098** (6.28)	0.192** (8.17)	0.189** (8.44)	0.171** (8.76)
<i>Exogenous</i>					
Log Output per Ha					0.019** (4.11)
Rel Soil Entropy		0.060** (3.76)	0.044** (2.51)	0.038** (2.33)	0.034** (1.98)
Log Total Area (ha)		0.001 (0.20)	-0.001 (0.06)	0.001 (0.15)	0.003 (0.70)
Percent Urban				0.080** (5.80)	0.073** (5.30)
Rail Access				-0.015** (2.30)	-0.013** (2.10)
Water Access				0.005 (0.93)	0.006 (1.15)
Method ^a	OLS	OLS	2SLS	2SLS	2SLS
First stage F-test p-value ^b			<0.01	<0.01	<0.01
First stage R-squared ^c			0.51	0.53	0.54
Hansen J Statistic ^d			0.94	1.35	1.86
Hansen p-value ^e			0.62	0.51	0.39

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

Standard errors are clustered by state

* denotes significance at 10%, ** denotes significance at 5%.

a) instruments are length of growing season, annual precipitation, and the SD(monthly precipitation)

b) p-value of the test of joint significance of the three instruments in the 1st stage regression for the % cotton, tobacco, and sugar in total output

c) R-squared of the first stage regression for % cotton, tobacco, and sugar

d) Test statistic, distributed $\chi^2(2)$ in columns 3-5, $\chi^2(2)$ in column 6, for Hansen's J test of overidentifying restrictions.

e) p-value of Hansen's J test

Table 4: **Determinants of the Gini Coefficient**

Exp Variables	Dep Variable: Log Rural Adult Males per Farm				
	(1)	(2)	(3)	(4)	(5)
<i>Endogenous</i>					
% Cotton, Tobacco, and Sugar	0.685** (6.24)	0.700** (6.38)	0.742** (4.16)	0.805** (5.65)	0.968** (8.43)
<i>Exogenous</i>					
Log Output per Ha					-0.172** (4.59)
Rel Soil Entropy		0.303** (3.10)	0.297** (2.97)	0.154 (1.49)	0.198* (1.96)
Log Total Area (ha)		-0.072** (2.36)	-0.073** (2.50)	-0.123** (4.42)	-0.146** (5.31)
Percent Urban				0.198** (2.06)	0.269** (2.96)
Rail Access				0.182** (4.96)	0.166** (4.45)
Water Access				0.286** (6.10)	0.278** (6.72)
Method ^a	OLS	OLS	2SLS	2SLS	2SLS
First stage F-test p-value ^b			<0.01	<0.01	<0.01
First stage R-squared ^c			0.51	0.53	0.51
Hansen J Statistic ^d			4.23	1.37	0.67
Hansen p-value ^e			0.12	0.51	0.72

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

Standard errors are clustered by state

* denotes significance at 10%, ** denotes significance at 5%.

a) instruments are length of growing season, annual precipitation, and the SD(monthly precipitation)

b) p-value of the test of joint significance of the three instruments in the 1st stage regression for the % cotton, tobacco, and sugar in total output

c) R-squared of the first stage regression for % cotton, tobacco, and sugar

d) Test statistic, distributed $\chi^2(2)$ in columns 3-5, $\chi^2(2)$ in column 6, for Hansen's J test of overidentifying restrictions.

e) p-value of Hansen's J test

Table 5: **Determinants of Rural Adult Males per Farm**

Exp Variables	Dep Variable:					
	G_{OV} (1)	G (2)	$\ln \frac{\text{Males}}{\text{Farm}}$ (3)	G_{OV} (4)	G (5)	$\ln \frac{\text{Males}}{\text{Farm}}$ (6)
<i>Endogenous</i>						
% Cotton, Tobacco, and Sugar	0.126** (4.95)	0.217** (7.06)	0.701** (3.01)	0.160** (10.34)	0.144** (7.66)	0.933** (7.34)
<i>Exogenous</i>						
Log Output per Ha	-0.017** (3.37)	0.020** (4.42)	-0.194** (5.12)	-0.021** (2.97)	0.014* (1.77)	-0.253** (6.52)
Rel Soil Entropy	0.043** (2.67)	0.028 (1.54)	0.222** (2.23)	0.026* (1.90)	0.020 (0.90)	0.154* (1.71)
Log Total Area (ha)	-0.014** (4.33)	0.003 (0.52)	-0.148** (5.43)	-0.013** (2.37)	-0.019** (3.76)	-0.129** (2.73)
Percent Urban	0.056** (4.74)	0.091** (6.48)	0.270** (2.98)	0.079** (4.87)	0.073** (2.36)	0.643** (3.65)
Rail Access	0.024** (4.61)	-0.016** (2.51)	0.182** (5.41)	0.026** (4.71)	-0.016** (2.37)	0.210** (3.87)
Water Access	0.038** (7.85)	0.005 (0.87)	0.284** (6.85)	0.033** (5.28)	0.007 (0.86)	0.268** (6.11)
South Dummy	0.029** (2.39)	-0.023** (2.18)	0.151* (1.97)			
Include Northern States?	Yes	Yes	Yes	No	No	No
Method ^a	2SLS	2SLS	2SLS	2SLS	2SLS	2SLS
Observations	1828	1828	1828	951	951	951
First stage F-test p-value ^b	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
First stage R-squared ^c	0.55	0.55	0.55	0.48	0.48	0.51
Hansen J Statistic ^d	1.05	0.74	0.98	0.77		3.97
Hansen p-value ^e	0.59	0.69	0.61	0.38		0.14

Notes: Absolute values of t-statistics are given in parentheses. Standard errors are clustered by state * denotes significance at 10%, ** denotes significance at 5%

a) instruments are length of growing season, annual precipitation, and the SD(monthly precip) for columns 1-3 and column 6, growing season and ann. precip for column 4 and growing season only for column 5.

b) p-value of the test of joint significance of the three instruments in the 1st stage regression for the % cotton, tobacco, and sugar in total output

c) R-squared of the first stage regression for % cotton, tobacco, and sugar

d) Test statistic, distributed $\chi^2(2)$ in columns 3-5, $\chi^2(2)$ in column 6, for Hansen's J test of overidentifying restrictions.

e) p-value of Hansen's J test

Table 6: **Regression Specifications Controlling for the South**

Instrument	Dependent Variable:		
	G_{OV}	G	$\frac{Rur\ Adult\ Males}{Farm}$
Annual Precipitation (cm) (x1000)	-0.125 ^a (0.45) ^b	0.170 (0.39)	-0.709 (0.51)
SD(Monthly Precip) (x1000)	-0.587 (0.43)	1.870 (1.56)	6.153 (0.69)
Length of Growing Season (x1000)	0.088 (0.23)	0.013 (0.03)	0.136 (0.06)

a) Each cell represents the coefficient estimate for the instrument when it is included as an exogenous variable, and the other two instruments act as IV's for the cash crop percentage.

b) Absolute value of t-stats reported in parentheses. Standard errors are clustered by state.

Note: The other explanatory variables included are the cash crop percent, log of agric. output per hectare, relative soil entropy, the log of total county area, the percent urban, and rail and water dummies

Table 7: Tests of Exclusion for Instruments