The Power Law Distribution of Agricultural Land Size

Sherzod B. Akhundjanov
Assistant Professor
Department of Applied Economics
Utah State University
sherzod.akhundjanov@usu.edu

Lauren Chamberlain
Graduate Student
Department of Applied Economics
Utah State University
briana.thomas@aggiemail.usu.edu

Selected Poster prepared for presentation at the 2019 Agricultural & Applied Economics Association Annual Meeting, Atlanta, Georgia, July 21-July 23

The Power Law Distribution of Agricultural Land Size

Sherzod B. Akhundjanov, Lauren Chamberlain

UtahState University

Background, Motivation, Objectives

- Agricultural land size plays an important role in understanding U.S. farm productivity and wealth.

- Large farms are now leading crop production in the United States: in 1987, only 15% of cropland was on farms with 2,000 acres or greater, by 2012, this number had increased to 36% **[1]**.
  - Although land area is only one input of agricultural production, the improved productivity and efficiency of larger farms is a result that holds even after accounting for land scarcity, soil, geography, agrarian structure, and varying forms of agriculture **[2]**.

- The U.S. is also seeing a dramatic increase in the number of very small farms and the disappearance of midsize farms **[3]**.
  - These small farms, often accounting for very little production, have been gradually skewing the distribution of U.S. farm size and increasing a disparity between small and large farms.

- The presence of very large farms, the very wide dispersion in farm size, and the role of agricultural sector in the U.S. economy make it crucial for policymakers to better understand farmland distribution for effective planning and policy design as well as efficient use of government subsidies and oversight.

- **Previous literature**:
  - distribution of production and sales among farms **[1]**
  - determinants of farm size distribution in country-level wealth **[2]**
  - growth process of farm size (i.e., Gibrat's law) **[4, 5]**

- Many existing studies focusing on empirical analysis of agricultural land size assume that agricultural land size is *normally* distributed [see, for instance, **6**].
  - The normality assumption is often made on the basis of convenience in estimation and inferences, with little a priori justification.

- **Objective**: investigate whether power law distribution can be used to describe the size distribution of U.S. county-level agricultural land.

- **Data**:
  - County-level agricultural land (in acres) for 1997, 2002, 2007, and 2012 from the USDA Census of Agriculture.
  - Multiple year data to demonstrate the robustness of power law analysis to time period, and specifically to various transformations (e.g., policy and technology changes) that can take place over a long period of time that could potentially affect agricultural land size and hence its distribution.

Empirical Methodology

- Let X represent a random variable of interest (e.g., agricultural land acres) whose distribution is a continuous power-law distribution

$$f(x) = \frac{\alpha-1}{x_{min}}\left(\frac{x}{x_{min}}\right)^{-\alpha}$$

  - x is an outcome of random variable X for $x \in \mathbb{R}_+$, where $\mathbb{R}_+ = \{x \in \mathbb{R} | x \geq 0\}$
  - x_{min} is the minimum value of the outcome of random variable X beyond which (i.e., for $x \geq x_{min}$) power law behavior takes hold
  - α is the power-law exponent (*the parameter of interest*)

- Given the observed sample $(x_1, \dots, x_n)$, the joint log likelihood function is

$$\ln \mathcal{L}(\alpha; x) = \sum_{i=1}^{n}\left[\ln(\alpha-1) - \ln x_{min} - \alpha \ln \frac{x_i}{x_{min}}\right]$$

- Maximum likelihood estimate (MLE) of α is

$$\alpha^{MLE} = 1 + n\left(\sum_{i=1}^{n} \ln \frac{x_i}{x_{min}}\right)^{-1} \qquad SE(\alpha^{MLE}) = \frac{\alpha^{MLE}-1}{\sqrt{n}}$$

- The Hill estimator of the counter-cumulative parameter $\gamma = \alpha^{MLE} - 1$ is

$$\gamma^{Hill} = \frac{n-2}{\sum_{i=1}^{n-1}(\ln x_i - \ln x_{min})} \qquad SE(\gamma^{Hill}) = \frac{\gamma^{Hill}}{\sqrt{n-3}}$$

- Regression-based estimate of counter-cumulative parameter can be obtained from:

$$\ln(i) = \beta_O - \gamma^{OLS} \ln x_i + \varepsilon_i$$

  - (i) is the observation's rank in the distribution
  - γ^{OLS} is the parameter of interest
  - The associated standard error of γ^{OLS} is the asymptotic standard error of the form $\gamma^{OLS}(n/2)^{-1/2}$

- x_{min} is determined based on is the data-driven procedure **[7]**:
  - Step 1: Set $x_{min} = x_1$;
  - Step 2: Estimate power-law exponent (γ^{Hill} and γ^{OLS}) using $x \geq x_{min}$;
  - Step 3: Calculate the KS statistic;
  - Step 4: Repeat steps 1-4 for all x_i for $i = 1, \dots, n$;
  - Step 5: Choose x_{min} with the lowest Kolmogorov-Smirnov (KS) statistic

$$KS = \max_{x \geq x_{min}} \left|E(x) - \hat{F}(x)\right|$$

- *Diagnostics 1:* Gabaix and Ibragimov [**8**] suggest "rank – 1/2" test:

$$\ln\left(i - \frac{1}{2}\right) = \alpha + \zeta \ln x_i + q(\ln x_i - x^*)^2 + \epsilon_i$$

  - $x^* = \frac{\mathrm{Cov}[(\ln x_i)^2, \ln x_i]}{2\mathrm{Var}(\ln x_i)}$
  - The test statistic of interest is given by $\frac{q}{\zeta^2}$
  - The null hypothesis that agricultural land size is distributed according to a power law is rejected if $\frac{q}{\zeta^2} > 1.95(2n)^{-1/2}$

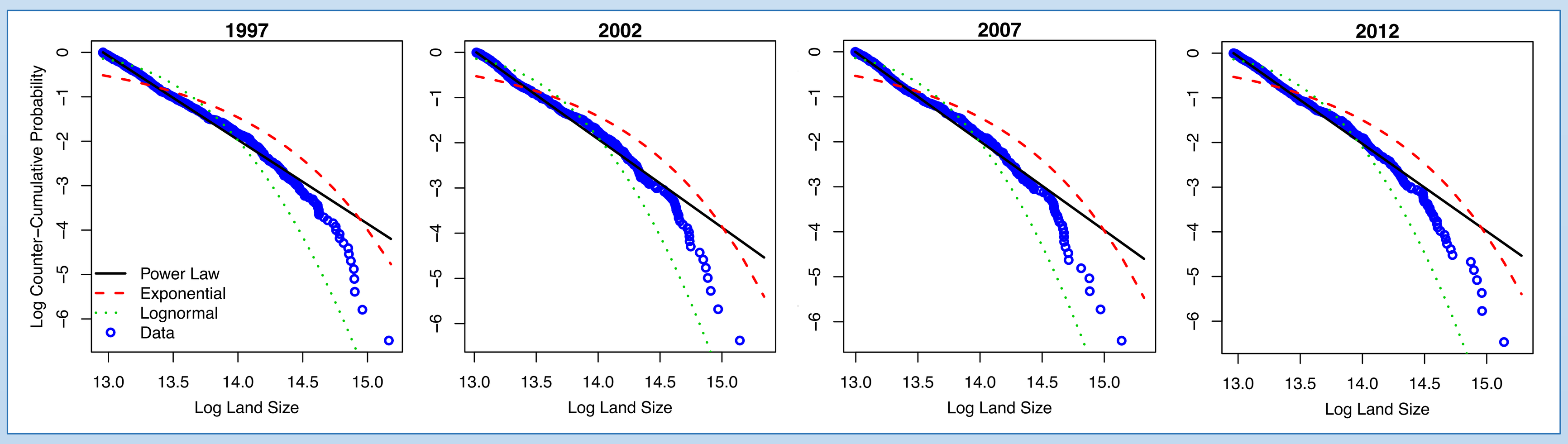
- *Diagnostics 2:* Alternative distributions: the lognormal and exponential
  - The relative fit of alternative distributions can be compared more rigorously using the likelihood ratio test **[7]**:

$$\mathcal{R} = \sum_{i=1}^{n}\left[\ln \hat{f}_1(x_i) - \ln \hat{f}_2(x_i)\right]$$

Results

	1997	*2002*	*2007*	*2012*
γ^{Hill}	1.883 (0.074)	1.946 (0.081)	1.973 (0.079)	1.951 (0.077)
γ^{OLS}	2.049 (0.014)	2.097 (0.015)	2.136 (0.015)	2.147 (0.014)
x_{min}	424,121	449,671	440,462	426,329
Observations	656	587	614	643
The Gabaix and Ibragimov goodness-of-fit test				
Goodness of fit test statistic	-0.139	-0.139	-0.136	-0.130
Goodness of fit threshold	0.054	0.057	0.056	0.054
The likelihood ratio test: Power law vs exponential				
Likelihood ratio statistic	503.284	466.060	493.321	509.232
P-value	0.000	0.000	0.000	0.000
The likelihood ratio test: Power law vs lognormal				
Likelihood ratio statistic	-5.631	-16.137	-30.479	-42.397
P-value	0.799	0.444	0.157	0.051

Note: Estimation is based on upper-tail observations ($x \geq x_{min}$), where x_{min} is determined based on the minimization of the KS statistic. Standard errors are in parentheses. For the Gabaix and Ibragimov **[8]** test, the null hypothesis that agricultural land size is distributed according to a power law is rejected if test statistic is greater than a threshold. Clauset et al. **[7]** recommend to have at least 50 observations for accurate power law analysis, a condition satisfied here. A positive value of the likelihood ratio statistic indicates that the power law is the better fitting distribution. A negative value indicates the alternative distribution fits the data more closely. P-values are calculated using the methods detailed in **[7]**.



Conclusion

- Our analysis provides evidence in favor of Pareto distribution, with estimates remaining robust across different periods, estimation methods, and diagnostic tests, and the distribution fitting the data as good or better than a series of alternative distributions.

- The robust Pareto fit to farmland indicates that U.S. agricultural land size is "heavy-tailed," with a handful of counties accounting for the majority of farmland.

- This finding is significant for two reasons.
  - It becomes inconsequential to talk about average agricultural land size as this statistic is no longer representative of the majority of counties; the total farmland is essentially determined by the largest farms. Focusing on quantile analysis and order statistics instead would be more appropriate in this case.
  - On a more technical concern, "fat tails" of agricultural land size—as suggested by power-law distribution—have significance for empirical research. Statistical analysis based on thin-tailed distributions (such as the normal) might dismiss extremely large farm sizes as an outlier or improbable observation. It is impossible to make sound empirical inferences from the distribution of agricultural land unless it is correctly specified.

References

[1] J.M. MacDonald, R. Hoppe, and D. Newton, *Three Decades of Consolidation in U.S. Agriculture*, U.S. Department of Agriculture, Economic Research Service, EIB-189, 2018.
[2] T. Adamopoulos and D. Restuccia, *The Size Distribution of Farms and International Productivity Differences*, American Economic Review 104 (2014), pp. 1667-1697.
[3] J.M. MacDonald, P. Korbe, and R. Hoppe, *Farm Size and the Organization of U.S. Crop Farming*, U.S. Department of Agriculture, Economic Research Service, ERR-152, 2013.
[4] J.S. Clark, M. Fulton, and D.J. Brown, *Gibrat's Law and Farm Growth in Canada*, Canadian Journal of Agricultural Economics 40 (1992), pp. 55-70.
[5] D. Shapiro, R.D. Bollman, and P. Ehrensaft, *Farm Size and Growth in Canada*, American Journal of Agricultural Economics 69 (1987), pp. 477-483.
[6] H.K. Gibbs, A.S. Ruesch, F. Achard, M.K. Clayton, P. Holmgren, N. Ramankutty, and J.A. Foley, *Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s*, Proceedings of the National Academy of Sciences, 107 (2010), pp. 16732-16737.
[7] A. Clauset, R. Shalizi, and M. Newman, *Power Law Distributions in Empirical Data*, SIAM Review 51 (2009), pp. 661-703.
[8] X. Gabaix and R. Ibragimov, *Rank-1/2: A Simple Way to Improve the OLS Estimation of Tail Exponents*, Journal of Business and Economic Statistics 29 (2011), pp. 24-39.