Finance and Economics Discussion Series
Divisions of Research & Statistics and Monetary Affairs
Federal Reserve Board, Washington, D.C.

Comparing Cross-Country Estimates of Lorenz Curves Using a Dirichlet Distribution Across Estimators and Datasets

Andrew C. Chang, Phillip Li, and Shawn M. Martin

2017-062

Please cite this paper as:
Chang, Andrew C., Phillip Li, and Shawn M. Martin (2017). "Comparing Cross-Country Estimates of Lorenz Curves Using a Dirichlet Distribution Across Estimators and Datasets," Finance and Economics Discussion Series 2017-062. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2017.062.

Comparing Cross-Country Estimates of Lorenz Curves Using a Dirichlet Distribution Across Estimators and Datasets

Andrew C. Chang* Phillip Li† Shawn M. Martin‡

April 18, 2017

Abstract

Chotikapanich and Griffiths (2002) introduced the Dirichlet distribution to the estimation of Lorenz curves. This distribution naturally accommodates the proportional nature of income share data and the dependence structure between the shares. Chotikapanich and Griffiths (2002) fit a family of five Lorenz curves to one year of Swedish and Brazilian income share data using unconstrained maximum likelihood and unconstrained non-linear least squares. We attempt to replicate the authors' results and extend their analyses using both constrained estimation techniques and five additional years of data. We successfully replicate a majority of the authors' results and find that some of their main qualitative conclusions also hold using our constrained estimators and additional data.

JEL Codes: C24; C51; C87; D31

Keywords: Constrained Estimation; Dirichlet; Gini Coefficient; Income Distribution; Lorenz Curve; Maximum Likelihood; Non-linear Least Squares; Replication; Share Data

*Chang: Board of Governors of the Federal Reserve System. 20th St. NW and Constitution Ave., Washington DC 20551 USA. +1 (657) 464-3286. a.christopher.chang@gmail.com. https://sites.google.com/site/andrewchristopherchang/.

†Li: Office of Financial Research. phil.li@gmail.com.

‡Martin: University of Michigan, Ann Arbor. smm332@georgetown.edu.

§The views and opinions expressed here are those of the authors and are not necessarily those of the Board of Governors of the Federal Reserve System, Department of the Treasury, or the Office of Financial Research. We thank conference participants at the 2015 International Atlantic Economic Conference, especially Tiago Pires and Keshab Bhattarai, for helpful comments. We are responsible for any errors.

Introduction

The Lorenz curve is a commonly used tool to illustrate income distributions and income inequality. It is constructed by relating ordered cumulative proportions of income to ordered cumulative population shares. The curve is then used to estimate income inequality measures, such as the Gini coefficient or Atkinson's inequality measure.

Unfortunately, estimates of inequality from Lorenz curves can depend crucially on distributional assumptions, functional form assumptions, and estimation methodologies (Cheong, 2002; Chotikapanich and Griffiths, 2002, 2005; and Abdalla and Hassan, 2004). Therefore, the literature proposes different functional forms and re-parameterizations for both the Lorenz curve and income distributions.[1] Estimation is commonly based on least squares techniques, with more recent studies using Bayesian and maximum likelihood estimation.[2]

We have three main objectives in this paper. For our first objective, we attempt a narrow replication of Chotikapanich and Griffiths (2002), hereafter CG, who propose using a Dirichlet distribution to model cumulative income share data. The Dirichlet distribution naturally accommodates the proportional nature and dependence structure of income share data, which are characteristics of income share data that often lack recognition (Chotikapanich and Griffiths, 2002). CG estimate five Lorenz curves using both maximum likelihood (ML) and non-linear least squares (NL) on one year of Brazilian and Swedish data, obtaining implied Gini coefficients. CG have three main findings: (1) the point estimates of the parameters and of the Gini coefficients are generally insensitive to the choice of Lorenz curve specification and estimator, (2) the standard errors are sensitive to the specification and estimator, and (3) ML under the Dirichlet distributional assumption performs better than NL for all Lorenz curve specifications.

We replicate a majority of CG's three main findings. For less parameterized Lorenz curves, our point estimates and standard errors match CG. We experience considerable instability in estimating the more parameterized Lorenz curves, consistent with CG. Our successful narrow replication contributes to the current push for replication and robustness

[1] For example, Kakwani (1980), Rasche et al. (1980), Ortega et al. (1991), Chotikapanich (1993), Sarabia et al. (1999, 2001, 2005), Rohde (2009), Helene (2010), and Wang and Smyth (2015).

[2] See Chotikapanich and Griffiths (2002, 2008), Hasegawa and Kozumi (2003).

in economics research (Chang and Li, 2015; Welch, 2015; Zimmermann, 2015).

Our second objective is to extend CG by using constrained estimators. We apply constrained maximum likelihood (CML) and constrained non-linear least squares (CNL) to the same functional forms and data as CG. We use constrained estimators because the parameters from the Lorenz curve specifications in CG should be constrained to ensure that the curves are invariant to increasing convex exponential and power transformations (Sarabia et al. (1999)). Although these restrictions are mentioned in CG, some of CG's estimates violate the constraints. We find that some parameter estimates differ between constrained and unconstrained estimators, but the implied Gini coefficients are similar between constrained and unconstrained estimators.

Our third objective is to fit the various Lorenz curve specifications with both constrained and unconstrained estimators on five additional years of Swedish and Brazilian income distribution data from the World Bank: data not used by CG. We find that a few of the main conclusions from CG also hold using the constrained estimators and these additional data. Similar to Abdalla and Hassan (2004), who apply the methodologies from CG to data from the Abu Dhabi Emirate and their own Lorenz curve form, we find that Gini coefficient point estimates are robust to different functional forms and estimation methods when applied to additional data.

Narrow Replication

The data are the cumulative proportions of income ($\eta_1, \eta_2, ..., \eta_M$ with $\eta_M = 1$) and corresponding cumulative population shares ($\pi_1, \pi_2, ..., \pi_M$ with $\pi_M = 1$).[3] Let $q_i = \eta_i - \eta_{i-1}$ be the income shares. CG assume that $(q_1, \ldots, q_M)$ has a Dirichlet distribution with parameters $(\alpha_1, \ldots, \alpha_M)$, where $\alpha_i = \lambda[L(\pi_i; \beta) - L(\pi_{i-1}; \beta)]$. $L(\cdot)$ is the Lorenz curve specification with an associated vector of unknown parameters β, and $\lambda > 0$ is an unknown scalar parameter from the Dirichlet distribution.

[3]For this paper, we conduct the replications without assistance from the authors and without their code, using data from the original source (Jain, 1975). We use Matlab R2013a and Stata 13MP on the Windows 7 Enterprise (64-bit) and OS X Version 10.9.5 operating systems respectively.

CG apply five Lorenz curve specifications to one year of Brazilian and Swedish data:

$$L_1(\pi_i; k) = \frac{e^{k\pi} - 1}{e^k - 1}, \qquad k > 0 \tag{1}$$

$$L_2(\pi_i; \alpha, \delta) = \pi^\alpha[1 - (1 - \pi)^\delta], \qquad \alpha \geq 0, 0 < \delta \leq 1 \tag{2}$$

$$L_3(\pi_i; \delta, \gamma) = [1 - (1 - \pi)^\delta]^\gamma, \qquad \gamma \geq 1, 0 < \delta \leq 1 \tag{3}$$

$$L_4(\pi_i; \alpha, \delta, \gamma) = \pi^\alpha[1 - (1 - \pi)^\delta]^\gamma, \qquad \alpha \geq 0, \gamma \geq 1, 0 < \delta \leq 1 \tag{4}$$

$$L_5(\pi_i; a, d, b) = \pi - a\pi^d(1 - \pi)^b. \qquad a > 0, 0 < d \leq 1, 0 < b \leq 1 \tag{5}$$

Each specification is then estimated with ML based on the Dirichlet distributional assumption or with NL without the distributional assumption. Functions L_2 and L_3 are nested in function L_4 when $\gamma = 1$ and $\alpha = 0$. L_5 is the "beta" function, see Kakwani (1980), and can yield L_2 when a and d are 1 in L_5 and $\alpha = 1$ in L_2.

The log-likelihood of the j-th Lorenz curve specification and the Dirichlet distribution is

$$\log[f(q|\theta)] = \log\Gamma(\lambda) + \sum_{i=1}^{M}(\lambda[L_j(\pi_i;\beta) - L_j(\pi_{i-1};\beta)] - 1) \times \log q_i \tag{6}$$
$$- \sum_{i=1}^{M} \log\Gamma(\lambda[L_j(\pi_i;\beta) - L_j(\pi_{i-1};\beta)]).$$

ML standard errors are derived from the negative inverse of the numeric Hessian matrix evaluated at the maximum. We use the Matlab function *fminunc* to perform the optimizations.

The NL objective function is

$$R = \sum_{i=1}^{M}(\eta_i - L_j(\pi_i;\beta))^2. \tag{7}$$

We use the Matlab function *lsqcurvefit* and the Stata command *nl* for the optimizations. For NL, CG suggest using Newey and West (1987) standard errors.[4]

Tables 1 and 2 show our narrow replication results. For Lorenz curves L_1 to L_3 and

[4]We implement *nl* in Stata with different lag values for the Newey-West standard errors and find that a lag of 2 matches the standard errors reported by CG. These are the standard errors we report. We use the Stata option *vce(hac nwest 2)* in the *nl* command.

for both countries, our ML point estimates and standard errors more or less match those from CG. Our ML estimation for L_4 is unstable, with more stable estimation using Brazilian data than Swedish data, consistent with CG. However, the Swedish ML point estimates for α fluctuate around values that are often greater than CG's estimates. When we perform ML with random starting values on Swedish data, the point estimates are similar to CG's but the standard errors are unstable.[5] This instability may indicate that the area around the maximum is flat, yielding point estimates and variances that are not unique (Gill and King, 2003). In addition, the numeric variance-covariance matrix evaluated at the converged values is not positive definite for over 50% of the random starting values. As a result, we do not report ML standard errors for L_4 with Swedish data. For L_4 with Brazilian data, our point estimates and standard errors more or less match those from CG.

We are unable to replicate CG's ML results for L_5 for both countries, despite attempting estimation using a grid of starting values. As noted in Ortega et al. (1991) and Sarabia et al. (1999), L_5 can result in a negative income share η_i for a population share π_i, leading to the difference $L_5(\pi_i;\beta)-L_5(\pi_{i-1};\beta)$ being negative and the term $\log\Gamma(\lambda[L_5(\pi_i;\beta)-L_5(\pi_{i-1};\beta)])$ from (6) being computationally infeasible.

We use NL for each Lorenz curve, initialized over a grid of starting values that spans the support of the parameters. We find that all Lorenz curve specifications except L_1 display some instability.[6] Instability is most frequent for L_4 and L_5. However, the parameter estimates that minimize the NL objective function and the corresponding standard errors are equivalent to CG's estimates.

For both ML and NL, we also attempt to replicate the Gini coefficient $G = 1-2\int_0^1 L_j(\pi;\beta)\, d\pi$, which is an income inequality measure. Following CG, we obtain point estimates of G by replacing β with the ML or NL $\hat{\beta}$s for each Lorenz curve specification. With the exception of L_1, we successfully replicate the Gini point estimates and standard errors for all estimation techniques and Lorenz curve specifications. Our initial inability to replicate the ML standard errors for the L_1 Gini coefficients led us to analytically verify the formula for the variance of the Gini coefficient, $var(\hat{G})$. We find a typo in CG's L_1 formula for $var(\hat{G})$ but are able to

[5] We use 2000 sets of random starting values from a standard normal distribution.

[6] A majority of the parameter estimates are similar. However, some initial values lead to NL point estimates with larger residual sum of squares, and in some cases infinite Gini coefficients.

replicate the ML standard errors for the Gini coefficient with our corrected formula.[7] Also, we discover a minor computational issue in the calculation of the NL standard errors for L_1 by CG.[8] We report the corrected quantities in our tables.

Similar to CG, we find that the Gini point estimates are insensitive to the choice of Lorenz curve specification and estimator, although L_1 fitted with Brazilian data is an exception. Given our inability to estimate L_5 using ML and the non-positive definite numeric Hessian for L_4 using Swedish data, we do not report an ML Gini coefficient for L_5 or standard errors of the Gini coefficient for both L_4 and L_5.

We also successfully replicate the information inaccuracy measures suggested by Theil (1967) and the likelihood ratio test (LRT) results except for L_5 vs. L_2 (with α=1) for Brazil.[9] We obtained 51.355 as the test statistic compared to 31.355 from CG. Both likelihood ratio statistics, however, lead to the same conclusion that the functional form L_2, with $\alpha = 1$, is rejected relative to L_5. The L_5 LRT and information inaccuracy measure for L_5 are calculated using CG's reported point estimates.

Scientific Replication: Constrained Optimization

Although the parameters for each Lorenz curve specification in (1) to (5) should be constrained to ensure that the Lorenz curves are invariant to increasing convex exponential and power transformations, we believe that CG did not enforce the constraints as some of their estimates violate the ranges. Therefore, we reestimate the models with the constraints imposed.[10] Our results are detailed in Tables 1 and 2.

---

[7] CG report $\text{var}(\hat{G}) = \left[\frac{2(e^{\hat{k}}(e^2-\hat{k}^2-2)+1)}{(\hat{k}(e^{\hat{k}}-1))^2}\right]^2 \text{var}(\hat{k})$ but we analytically find $\text{var}(\hat{G}) = \left[\frac{2(e^{\hat{k}}(e^{\hat{k}}-\hat{k}^2-2)+1)}{(\hat{k}(e^{\hat{k}}-1))^2}\right]^2 \text{var}(\hat{k})$.

[8] We find that the CG standard errors for the L_1 NL Gini coefficient are calculated as $\text{var}(\hat{G}) = \frac{\partial G}{\partial \beta} \text{var}(\hat{k})$ when the correct formula is $\text{var}(\hat{G}) = \frac{\partial G}{\partial \beta'} \text{var}(\hat{k}) \frac{\partial G}{\partial \beta}$. We verify this using CG's reported Brazilian values for $\text{SE}(\hat{G})$ and $\text{SE}(\hat{k})$, .1647 and .6726, in the formula of $\text{var}(\hat{G})$ corrected for the typo detailed in footnote 7: $.1647^2 = \frac{\partial G}{\partial \beta} \times .6726^2 \times \frac{\partial G}{\partial \beta}$, which implies $[\frac{\partial G}{\partial \beta}]^2 = .0600$ and $\frac{\partial G}{\partial \beta} = .2449$, however $\frac{\partial G}{\partial \beta}$ evaluated at $\hat{k} = .0600$. Therefore the variance of $\hat{G}$ should be $\text{var}(\hat{G}) = .0600 \times .6742^2 \times .0600 = .0016$ and $\text{SE}(\hat{G}) = .0403$. A similar computational error occurs for Swedish data.

[9] The Theil (1967) information inaccuracy measure, $I = \sum_{i=1}^{M} q_i log(\frac{q_i}{\hat{q}_i})$, compares actual income shares, q_i, to predicted income shares, $\hat{q}_i$. Smaller values of I indicate a better fit.

[10] We use the Matlab functions *fmincon* and *lsqcurvefit*. In unreported results we also attempt to use the Matlab function *patternsearch* to apply ML and CML. *Patternsearch* yields parameter estimates that are

Point estimates for L_1 to L_3 are identical for constrained and unconstrained estimation. For L_4 with Swedish data, the CNL estimates deviate the most from the NL estimates. For example, the CNL estimate for α is close to 0 while the NL estimate is -0.7549. For L_4 with Brazilian data, the CNL estimates are close to the NL estimates. In terms of CML results for L_4, the constrained estimates and standard errors match the unconstrained quantities. Though we were unable to replicate unconstrained ML point estimates for L_5, with parameter constraints imposed the CML point estimates are close to the unconstrained estimates from CG; we were unable to generate standard error estimates as the numeric hessians were quite unstable across different sets of starting values. The CNL point estimates for L_5 are either identical to or very close to the NL quantities.

Overall, we find that the CML and CNL estimates of the model parameters can differ from their ML and NL counterparts. However, the implied Gini coefficients are similar even when the unconstrained and constrained parameter estimates differ.

Scientific Replication: Extension to World Bank Data

We further extend CG using data from the World Bank Poverty and Equity Database (World Bank, 2015b).[11] We construct a dataset of seven quantiles of cumulative income shares for Brazil in 1987, 1992, 1995, 2001 and 2005 and for the equivalent years for Sweden, with 2001 replaced by 2000.[12]

Tables 3 and 4 show our results using these World Bank data. Unconstrained and constrained estimation applied to World Bank data yield qualitative conclusions similar to those reported by CG, who use data from Jain (1975). With the exception of L_4, the point estimates of the parameters for all Lorenz curve specifications are similar across estimation techniques, but there are differences in the standard errors. ML and NL point estimates for L_4 differ for all years of Brazilian and Swedish World Bank data. Similar to our narrow

either identical to *fmincon* or imply a smaller log-likelihood; it also tends to be less stable than *fmincon*.

[11] We have agreed to the terms of use as described at http://go.worldbank.org/OJC02YMLA0.

[12] World Bank Poverty and Equity Database variables used include income share held by lowest 10%, lowest 20%, second 20%, third 20%, fourth 20%, highest 20%, and highest 10%. Swedish data for 2001 and Brazilian data for 2000 are unavailable. At the time of submission, Swedish data for these years and variables were no longer available in the World Bank DataBank, but they are available from the authors upon request.

replication, we experience the same computational instability with unconstrained ML for L_4 and computational infeasibility for L_5 with World Bank data. We employ methods from our narrow replication and constrained estimation to obtain point estimates for L_4 and L_5.

We also find that, for a given year, Gini coefficients are similar across Lorenz curve specifications and estimators, with the exception of L_1 with Brazilian data. Although some unconstrained parameter estimates violate the restricted ranges and are different from the constrained estimates, the estimates still yield similar point estimates of the Gini coefficients. For Brazil, ML estimation of L_1 results in Gini coefficients that are lower than other functional forms, and NL estimation results in higher Gini coefficients. In addition, the point estimates of the Gini coefficients obtained in our analysis are similar to those officially reported by the World Bank (see Table 5). World Bank Gini coefficients are based on the generalized quadratic and beta parameterizations of the Lorenz curves, suggested by Villasenor and Arnold (1989) and Kakwani (1980).[13]

Table 6 compares the fit using the Theil (1967) information inaccuracy measure. Similar to CG, we find ML estimation with Swedish data provides a better fit than NL for all Lorenz curve specifications, with the largest differences observed for L_4 and L_5. CG's conclusion is also consistent for the Swedish World Bank data with the exception of 2001 and 2005 for L_1 and 2005 for L_4. For Brazil, CG find that ML provides a better fit than NL for L_2, L_3, and L_4, a worse fit for L_1 and an equivalent fit for L_5. We find that NL is a better fit in 4 of the 5 years of World Bank data for L_1 and in all years for L_4, but ML is a better fit for all years with functions L_2, L_3 and L_5. For both L_4 fit to Swedish data and L_5 fit to Brazilian data, CNL has a smaller information measure than NL, suggesting that CNL provides a better fit relative to NL.[14] A closer examination shows that for both Brazilian and Swedish data NL overpredicts q_1 and underpredicts q_2 and q_3, relative to q_i, by a larger margin than CNL.

[13]Data are from nationally representative household surveys conducted by national statistical offices or by private agencies under the supervision of government or international agencies. Parametric Lorenz curves are used with groups distributional data when they are expected to provide close estimates to the micro data. If estimation using parametric Lorenz curves is unlikely to work well, estimation is done directly from micro data obtained from nationally representative household surveys (World Bank, 2015a).

[14]The NL objective function, however, is lower for unconstrained NL.

Conclusion

Our narrow replication of CG verifies a majority of their results. However, we discover a few minor computational and presentational issues in CG. These issues do not affect CG's qualitative conclusions. Our scientific replication extends the analysis from CG to constrained estimators and additional data. We conclude that some of the qualitative results from CG also hold with constrained estimators and additional data. Some of our constrained parameter estimates are different than CG's corresponding unconstrained estimates. However, the Gini coefficient estimates from both sets of estimates are similar.

Although we have explored different functional forms and estimators for modeling Lorenz curves, it is difficult for us to make a sweeping recommendation as to which estimator and functional form that researchers should use. However, assuming you only care about the Gini coefficient, and not the fit of actual income shares, then we feel the parsimonious L_1 is the best option. L_1's implied Gini coefficient is relatively, though not completely, invariant to estimator choice and also is stable across initialized starting values.

References

Abdalla, I. M., Hassan, M. Y., 2004. Maximum likelihood estimation of lorenz curves using alternative parametric model. Metodoloski Zvezki 1 (1), 109–118.

Chang, A. C., Li, P., 2015. Is economics research replicable? sixty published papers from thirteen journals say "usually not". Finance and Economics Discussion Series 2015-083. Washington: Board of Governors of the Federal Reserve System.

Cheong, K. S., 2002. An empirical comparison of alternative functional forms for the lorenz curve. Applied Economics Letters 9 (3), 171–176.

Chotikapanich, D., 1993. A comparison of alternative functional forms for the lorenz curve. Economics Letters 41 (2), 129–138.

Chotikapanich, D., Griffiths, W. E., 2002. Estimating lorenz curves using a dirichlet distribution. Journal of Business & Economic Statistics 20 (2), 290–295.

Chotikapanich, D., Griffiths, W. E., 2005. Averaging lorenz curves. The Journal of Economic Inequality 3 (1), 1–19.

Chotikapanich, D., Griffiths, W. E., 2008. Estimating Income Distributions Using a Mixture of Gamma Densities. Springer.

Gill, J., King, G., 2003. Numerical issues involved in inverting hessian matrices. Numerical Issues in Statistical Computing for the Social Scientist.

Hasegawa, H., Kozumi, H., 2003. Estimation of lorenz curves: a bayesian nonparametric approach. Journal of Econometrics 115 (2), 277–291.

Helene, O., 2010. Fitting lorenz curves. Economics Letters 108 (2), 153–155.

Jain, S., 1975. Size distribution of income: A compilation of data.

Kakwani, N., 1980. On a class of poverty measures. Econometrica 48 (2), 437–446.

Newey, W. K., West, K. D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix.

Ortega, P., Martin, G., Fernandez, A., Ladoux, M., Garcia, A., 1991. A new functional form for estimating lorenz curves. Review of Income and Wealth 37 (4), 447–452.

Rasche, R., et al., 1980. Functional forms for estimating the lorenz curve: Comment. Econometrica 48 (4), 1061–62.

Rohde, N., 2009. An alternative functional form for estimating the lorenz curve. Economics Letters 105 (1), 61–63.

Sarabia, J. M., Castillo, E., Pascual, M., Sarabia, M., 2005. Mixture lorenz curves. Economics Letters 89 (1), 89–94.

Sarabia, J.-M., Castillo, E., Slottje, D. J., 1999. An ordered family of lorenz curves. Journal of Econometrics 91 (1), 43–60.

Sarabia, J.-M., Castillo, E., Slottje, D. J., 2001. An exponential family of lorenz curves. Southern Economic Journal 67 (3), 748–756.

Theil, H., 1967. Economics and Information Theory. North-Holland Amsterdam.

Villasenor, J., Arnold, B. C., 1989. Elliptical lorenz curves. Journal of econometrics 40 (2), 327–338.

Wang, Z., Smyth, R., 2015. A hybrid method for creating lorenz curves. Economics Letters 133, 59–63.

Welch, I., 2015. Plausibility. Available at SSRN 2570577.

World Bank, 2015a. Povcalnet: an online analysis tool for global poverty monitoring. http:\iresearch.worldbank.org/PovcalNet, accessed 2015-08-05.

World Bank, 2015b. Poverty & equity data. http://data.worldbank.org/data-catalog/poverty-and-equity-database, accessed 2015-06-04.

Zimmermann, C., 2015. On the need for a replication journal. FRB St Louis Paper No. FEDLWP2015-016.

Tables

Table 1: Sweden estimates using data from Jain (1975)[a]

		Chotikapanich & Griffiths				Our Unconstrained & Constrained Results			
		α	δ	γ	Gini	α	δ	γ	Gini
L_2	NL	.5954	.6352		.3880	.5954	.6352		.3880
		(.0136)	(.0052)		(.0013)	(.0136)	(.0052)		(.0013)
	ML	.6068	.6412		.3872	.6068	.6412		.3872
		(.0206)	(.0085)		(.0041)	(.0206)	(.0084)		(.0040)
L_3	NL		.7269	1.5602	.3871		.7269	1.5602	.3871
			(.0032)	(.0076)	(.0007)		(.0032)	(.0076)	(.0007)
	ML		.7335	1.5767	.3877		.7335	1.5767	.3877
			(.0072)	(.0176)	(.0036)		(.0081)	(.0190)	(.0038)
L_4	CNL	–	–	–	–	.0000	.7269	1.5602	.3871
	NL	-.7552	.7931	2.2893	.3864	-.7549	.7931	2.2890	.3865
		(.5638)	(.0366)	(.5458)	(.0000)	(.5643)	(.0366)	(.5462)	(.0006)
	CML	–	–	–	–	.0050	.7330	1.5720	.3877
	ML	.0048	.7330	1.5721	.3876	.0045	.7330	1.5724	.3877
		(.6612)	(.0756)	(.6369)	(.0036)	–	–	–	–
L_1		k				k			
	NL	2.5029			.3792	2.5029			.3792
		(.0826)			(.0292)	(.0825)			(.0103)
	ML	2.5313			.3828	2.5313			.3828
		(.1831)			(.0228)	(.1830)			(.0228)
L_5		a	d	b		a	d	b	
	NL	.7664	.9397	.5929	.3876	.7664	.9397	.5929	.3876
		(.0148)	(.0138)	(.0108)	(.0010)	(.0148)	(.0138)	(.0108)	(.0011)
	CML	–	–	–	–	.7492	.9200	.5862	.3870
		–	–	–	–	–	–	–	–
	ML	.7492	.9199	.5862	.3870	–	–	–	–
		(.0143)	(.0093)	(.0109)	(.0031)	–	–	–	–

[a] ‘ML’: maximum likelihood. ‘NL’: non-linear least squares. ‘CML’ : constrained maximum likelihood. ‘CNL’: constrained non-linear least squares. We report constrained estimates only when they differ from the unconstrained estimates. ‘–’ represents estimates we are unable to obtain.

Table 2: Brazil estimates using data from Jain (1975)[b]

		Chotikapanich & Griffiths				Our Unconstrained & Constrained Results			
		α	δ	γ	Gini	α	δ	γ	Gini
L_2	NL	.5727	.2876		.6361	.5727	.2876		.6361
		(.0223)	(.0019)		(.0012)	(.0223)	(.0019)		(.0012)
	ML	.5270	.2857		.6326	.5270	.2857		.6326
		(.0383)	(.0053)		(.0052)	(.0382)	(.0053)		(.0052)
L_3	NL		.3782	1.4357	.6328		.3782	1.4357	.6328
			(.0038)	(.0127)	(.0010)		(.0038)	(.0127)	(.0010)
	ML		.3721	1.4160	.6325		.3721	1.4160	.6325
			(.0068)	(.0225)	(.0040)		(.0069)	(.0228)	(.0039)
L_4	CNL	–	–	–	–	.2170	.3467	1.2674	.6340
	NL	.2169	.3467	1.2674	.6339	.2169	.3467	1.2674	.6340
		(.1950)	(.0289)	(.1473)	(.0013)	(.1954)	(.0289)	(.1474)	(.0013)
	ML	.0262	.3683	1.3950	.6325	.0262	.3683	1.3950	.6326
		(.2148)	(.0318)	(.1734)	(.0039)	(.2229)	(.0330)	(.1800)	(.0039)
		k				k			
L_1	NL	5.3685			.6368	5.3685			.6368
		(.6726)			(.1647)	(.6726)			(.0403)
	ML	3.8438			.5234	3.8438			.5234
		(.8237)			(.0747)	(.8237)			(.0747)
		a	d	b		a	d	b	
L_5	CNL	–	–	–	–	.9150	1.0000	.2698	.6349
	NL	.9151	1.0001	.2698	.6349	.9151	1.0001	.2698	.6349
		(.0030)	(.0024)	(.0016)	(.0003)	(.0030)	(.0024)	(.0016)	(.0003)
	CML	–	–	–	–	.9131	.9991	.2685	.6350
		–	–	–	–	–	–	–	–
	ML	.9131	.9990	.2685	.6349	–	–	–	–
		(.0044)	(.0024)	(.0021)	(.0013)	–	–	–	–

[b] 'ML': maximum likelihood. 'NL': non-linear least squares. 'CML' : constrained maximum likelihood. 'CNL': constrained non-linear least squares. We report constrained estimates only when they differ from the unconstrained estimates. '–' represents estimates we are unable to obtain.

Table 3: Sweden Estimates using World Bank Data[c]

		1987				1992				1995			
		α	δ	γ	Gini	α	δ	γ	Gini	α	δ	γ	Gini
L_2	NL	.3128	.7802		.2366	.3279	.7563		.2539	.3429	.7628		.2550
		(.0097)	(.0051)		(.0011)	(.0085)	(.0042)		(.0010)	(.0051)	(.0025)		(.0006)
	ML	.3202	.7849		.2365	.3346	.7601		.2539	.3461	.7648		.2549
		(.0189)	(.0116)		(.0051)	(.0169)	(.0098)		(.0044)	(.0107)	(.0062)		(.0028)
L_3	NL		.8209	1.3060	.2365		.8017	1.3192	.2538		.8091	1.3342	.2548
			(.0047)	(.0086)	(.0010)		(.0037)	(.0072)	(.0008)		(.0018)	(.0037)	(.0004)
	ML		.8257	1.3145	.2368		.8059	1.3271	.2542		.8112	1.3386	.2551
			(.0111)	(.0186)	(.0050)		(.0096)	(.0166)	(.0044)		(.0060)	(.0105)	(.0027)
L_4	CNL	.0000	.8209	1.3060	.2365	.0000	.8017	1.3192	.2538	.0001	.8091	1.3342	.2548
	NL	-3.3747	.9403	4.6606	.2364	-1.0974	.8780	2.4016	.2536	-.4101	.8452	1.7376	.2547
		(9.1283)	(.1068)	(9.1052)	(.0014)	(1.9457)	(.0819)	(1.9253)	(.0008)	(.6439)	(.0450)	(.6328)	(.0004)
	CML	.0000	.8257	1.3145	.2367	.0000	.8059	1.3271	.2542	.0547	.8051	1.2849	.2551
	ML	.0000	.8257	1.3145	.2367	.0000	.8059	1.3271	.2542	.0551	.8051	1.2846	.2551
		–	(.0111)	(.0186)	–	–	(.0096)	(.0166)	–	–	–	–	–
		k				k				k			
L_1	NL	1.4565			.2346	1.5746			.2522	1.5790			.2528
		(.0428)			(.0064)	(.0530)			(.0078)	(.0487)			(.0072)
	ML	1.4691			.2365	1.5803			.2530	1.5873			.2541
		(.1295)			(.0195)	(.1404)			(.0207)	(.1366)			(.0202)
		a	d	b		a	d	b		a	d	b	
L_5	NL	.4744	.8576	.6715	.2368	.5010	.8650	.6481	.2541	.5073	.8628	.6588	.2552
		(.0188)	(.0254)	(.0210)	(.0010)	(.0160)	(.0206)	(.0168)	(.0008)	(.0076)	(.0099)	(.0077)	(.0004)
	CML	.4605	.8337	.6616	.2363	.4891	.8455	.6401	.2538	.5001	.8518	.6535	.2550

Continued on next page

[c] We use income data from the World Bank Poverty and Equity Database (http://data.worldbank.org/data-catalog/poverty-and-equity-database). The variables include income share held by lowest 10%, lowest 20%, second 20%, third 20%, fourth 20%, highest 20%, and highest 10%. 'ML': maximum likelihood. 'NL': non-linear least squares. 'CML' : constrained maximum likelihood. 'CNL': constrained non-linear least squares. We report constrained estimates only when they differ from the unconstrained estimates. '–' represents estimates we are unable to obtain.

Table 3: Sweden Estimates using World Bank Data[d]

		2001				2005			
Continued from previous page									
		α	δ	γ	Gini	α	δ	γ	Gini
L_2	NL	.2949	.7021		.2743	.2802	.7183		.2605
		(.0055)	(.0024)		(.0007)	(.0055)	(.0026)		(.0007)
	ML	.2962	.7035		.2739	.2807	.7196		.2599
		(.0084)	(.0044)		(.0022)	(.0074)	(.0040)		(.0020)
L_3	NL		.7495	1.2831	.2741		.7623	1.2702	.2603
			(.0017)	(.0033)	(.0004)		(.0019)	(.0034)	(.0004)
	ML		.7511	1.2859	.2741		.7635	1.2720	.2602
			(.0041)	(.0073)	(.0020)		(.0034)	(.0058)	(.0016)
L_4	CNL	.0000	.7495	1.2830	.2741	.0000	.7623	1.2702	.2603
	NL	-.2552	.7798	1.5305	.2739	-.4223	.8075	1.6819	.2602
		(.2821)	(.0287)	(.2730)	(.0004)	(.3089)	(.0263)	(.3015)	(.0003)
	CML	.0000	.7511	1.2859	.2742	.0000	.7635	1.2720	.2602
	ML	.0000	.7511	1.2859	.2742	.0000	.7635	1.2720	.2602
		–	(.0041)	(.0073)	–	–	(.0034)	(.0058)	–
		k				k			
L_1	NL	1.7263			.2744	1.6295			.2603
		(.0892)			(.0129)	(.0801)			(.0118)
	ML	1.6940			.2697	1.6043			.2566
		(.1681)			(.0244)	(.1539)			(.0227)
		a	d	b		a	d	b	
L_5	NL	.5175	.8865	.5844	.2745	.4965	.8833	.5972	.2607
		(.0094)	(.0118)	(.0095)	(.0005)	(.0089)	(.0113)	(.0096)	(.0004)
	CML	.5121	.8771	.5815	.2743	.4928	.8761	.5959	.2605

[d] We use income data from the World Bank Poverty and Equity Database (http://data.worldbank.org/data-catalog/poverty-and-equity-database). The variables include income share held by lowest 10%, lowest 20%, second 20%, third 20%, fourth 20%, highest 20%, and highest 10%. 'ML': maximum likelihood. 'NL': non-linear least squares. 'CML' : constrained maximum likelihood. 'CNL': constrained non-linear least squares. We report constrained estimates only when they differ from the unconstrained estimates. '–' represents estimates we are unable to obtain.

Table 4: Brazil Estimates using World Bank Data[e]

		1987				1992				1995			
		α	δ	γ	Gini	α	δ	γ	Gini	α	δ	γ	Gini
L_2	NL	.8719	.3684		.6028	.8555	.4627		.5357	.8382	.3642		.6020
		(.0528)	(.0064)		(.0027)	(.0379)	(.0068)		(.0023)	(.0535)	(.0065)		(.0028)
	ML	.8124	.3667		.5970	.8216	.4619		.5314	.7834	.3632		.5962
		(.0675)	(.0111)		(.0083)	(.0466)	(.0105)		(.0065)	(.0665)	(.0109)		(.0083)
L_3	NL		.5009	1.7035	.5997		.5952	1.7369	.5336		.4928	1.6747	.5990
			(.0055)	(.0205)	(.0012)		(.0048)	(.0160)	(.0010)		(.0060)	(.0216)	(.0013)
	ML		.4974	1.6844	.5977		.5938	1.7276	.5322		.4904	1.6597	.5971
			(.0074)	(.0256)	(.0036)		(.0060)	(.0188)	(.0028)		(.0076)	(.0258)	(.0037)
L_4	CNL	.0000	.5009	1.7035	.5997	.0000	.5952	1.7369	.5336	.0000	.4928	1.6747	.5990
	NL	-.9886	.5973	2.5619	.5976	-1.1085	.6935	2.7521	.5321	-1.0078	.5925	2.5478	.5968
		(.0494)	(.0038)	(.0441)	(.0002)	(.2198)	(.0146)	(.2049)	(.0002)	(.1827)	(.0144)	(.1621)	(.0004)
	ML	.0000	.4974	1.6844	.5978	.0000	.5938	1.7276	.5322	.0000	.4904	1.6597	.5971
		–	(.0074)	(.0256)	–	–	(.0060)	(.0188)	–	–	(.0076)	(.0258)	–
		k				k				k			
L_1		5.1102			.6208	4.0814			.5443	5.1140			.6210
	NL	(.4550)			(.0293)	(.2837)			(.0242)	(.4675)			(.0301)
		4.1721			.5519	3.7155			.5116	4.1368			.5490
	ML	(.6625)			(.0550)	(.4157)			(.0390)	(.6797)			(.0570)
		a	d	b		a	d	b		a	d	b	
L_5	CNL	.9688	1.0000	.3629	.6017	.9579	1.0000	.4580	.5346	.9618	1.0000	.3578	.6009
	NL	1.0025	1.0286	.3784	.6005	.9758	1.0150	.4666	.5341	.9914	1.0255	.3715	.5999
		(.0223)	(.0164)	(.0108)	(.0031)	(.0217)	(.0155)	(.0113)	(.0012)	(.0245)	(.0181)	(.0120)	(.0015)
	CML	.9622	1.0000	.3624	.5979	.9534	.9971	.4586	.5327	.9575	.9999	.3587	.5976

Continued on next page

[e] We use income data from the World Bank Poverty and Equity Database (http://data.worldbank.org/data-catalog/poverty-and-equity-database). The variables include income share held by lowest 10%, lowest 20%, second 20%, third 20%, fourth 20%, highest 20%, and highest 10%. 'ML': maximum likelihood. 'NL': non-linear least squares. 'CML' : constrained maximum likelihood. 'CNL': constrained non-linear least squares. We report constrained estimates only when they differ from the unconstrained estimates. '–' represents estimates we are unable to obtain.

Table 4: Brazil Estimates using World Bank Data[f]

		2000				2005			
Continued from previous page									
		α	δ	γ	Gini	α	δ	γ	Gini
L_2	NL	.8259	.3668		.5985	.7159	.3849		.5708
		(.0422)	(.0053)		(.0023)	(.0306)	(.0043)		(.0019)
	ML	.7868	.3664		.5941	.6855	.3845		.5670
		(.0521)	(.0088)		(.0066)	(.0400)	(.0075)		(.0057)
L_3	NL		.4942	1.6662	.5956		.5004	1.5861	.5683
			(.0035)	(.0124)	(.0007)		(.0019)	(.0060)	(.0005)
	ML		.4940	1.6622	.5947		.5000	1.5826	.5677
			(.0047)	(.0157)	(.0023)		(.0030)	(.0095)	(.0015)
L_4	CNL	.0000	.4942	1.6662	.5956	.0000	.5004	1.5861	.5683
	NL	-.5239	.5517	2.1143	.5943	-.2725	.5338	1.8194	.5676
		(.1680)	(.0161)	(.1453)	(.0004)	(.0772)	(.0088)	(.0670)	(.0002)
	ML	.0000	.4940	1.6622	.5947	.0000	.5000	1.5826	.5677
		–	(.0047)	(.0157)	–	–	(.0030)	(.0095)	–
		k				k			
L_1		5.0577			.6174	4.6512			.5893
	NL	(.4662)			(.0304)	(.4410)			(.0322)
		4.0997			.5459	3.8252			.5218
	ML	(.6744)			(.0571)	(.6300)			(.0574)
		a	d	b		a	d	b	
L_5	CNL	.9595	1.0000	.3604	.5976	.9294	1.0000	.3740	.5698
	NL	.9766	1.0149	.3685	.5969	.9314	1.0018	.3750	.5698
		(.0204)	(.0152)	(.0102)	(.0012)	(.0149)	(.0114)	(.0079)	(.0009)
	CML	.9554	.9974	.3610	.5956	.9178	.9897	.3703	.5689

[f] We use income data from the World Bank Poverty and Equity Database (http://data.worldbank.org/data-catalog/poverty-and-equity-database). The variables include income share held by lowest 10%, lowest 20%, second 20%, third 20%, fourth 20%, highest 20%, and highest 10%. ‘ML’: maximum likelihood. ‘NL’: non-linear least squares. ‘CML’ : constrained maximum likelihood. ‘CNL’: constrained non-linear least squares. We report constrained estimates only when they differ from the unconstrained estimates. ‘–’ represents estimates we are unable to obtain.

Table 5: World Bank estimates of Gini coefficients[g]

	1987	1992	1995	2000	2001	2005
WB Brazil[h]	0.5969	0.5317	0.5957		0.5933	0.5665
Our Brazil[i]	0.5979	0.5327	0.5976		0.5956	0.5683
WB Sweden	0.2371	0.2542	0.2554	0.2748		0.2608
Our Sweden	0.2365	0.2539	0.2549	0.2741		0.2603

[g] We use income data from the World Bank Poverty and Equity Database (http://data.worldbank.org/data-catalog/poverty-and-equity-database). The variables include income share held by lowest 10%, lowest 20%, second 20%, third 20%, fourth 20%, highest 20%, and highest 10%. World Bank Gini coefficients are from Povcalnet, an online tool for poverty measurement developed by the Development Research Group of the World Bank (http://iresearch.worldbank.org/PovcalNet, World Bank (2015a)).

[h] For the above year-country combinations Povcalnet utilizes income-based data in the format of household level ('unit record') data.

[i] Calculated as the median of implied Gini coefficients for unconstrained and constrained ML and NL point estimates of functions L_1 to L_5.

Table 6: Information Inaccuracy Measures, World Bank Data[j]

		Brazil					Sweden				
		1987	1992	1995	2001	2005	1987	1992	1995	2001	2005
L_1	ML	.043372	.021109	.045875	.046204	.044179	.003607	.004171	.003947	.005840	.004976
	NL	.043296	.021403	.045706	.045773	.043774	.003607	.004172	.003948	.005835	.004972
L_2	ML	.000986	.000568	.000989	.000622	.000431	.000260	.000200	.000079	.000050	.000040
	NL	.001126	.000625	.001109	.000687	.000479	.000268	.000207	.000080	.000051	.000041
L_3	ML	.000185	.000107	.000195	.000074	.000032	.000250	.000192	.000075	.000038	.000025
	NL	.000204	.000113	.000208	.000076	.000033	.000259	.000199	.000078	.000040	.000026
L_4	ML	.000185	.000107	.000195	.000074	.000032	.000250	.000192	.000075	.000038	.000025
	CML	.000185	.000107	.000195	.000074	.000032	.000250	.000192	.000075	.000038	.000025
	NL	.000003	.000020	.000023	.000037	.000012	.000268	.000209	.000082	.000042	.000025
	CNL	.000204	.000113	.000208	.000076	.000033	.000259	.000199	.000078	.000040	.000026
L_5	CML	.000517	.000332	.000559	.000400	.000214	.000175	.000131	.000035	.000046	.000037
	NL	.000771	.000538	.000851	.000640	.000313	.000225	.000169	.000047	.000056	.000044
	CNL	.000563	.000345	.000583	.000414	.000288	.000225	.000169	.000047	.000056	.000044

[j] We use income data from the World Bank Poverty and Equity Database (http://data.worldbank.org/data-catalog/poverty-and-equity-database). The variables include income share held by lowest 10%, lowest 20%, second 20%, third 20%, fourth 20%, highest 20%, and highest 10%. The information inaccuracy measure compares actual income shares, q_i, to the predicted income shares, $\hat{q}_i$. As the difference between $\hat{q}_i$ and q_i decreases this measure decreases, suggesting a better fit. 'ML': maximum likelihood. 'NL': non-linear least squares. 'CML' : constrained maximum likelihood. 'CNL': constrained non-linear least squares. We report constrained estimates only when they differ from the unconstrained estimates. '–' represents estimates we are unable to obtain.