Poster Session #2
Time: Thursday, August 9, 2012 PM

Paper Prepared for the 32nd General Conference of
The International Association for Research in Income and Wealth

Boston, USA, August 5-11, 2012

Race & Gender Differences in the Experience of Earnings Inequality, 1995 to 2010

Markus P. A. Schneider

For additional information please contact:

Name: Markus P. A. Schneider
Affiliation: University of Denver

Email Address: Markus.Schneider@du.edu

NOT FOR PUBLIC RELEASE

Race & Gender Differences in the Experience of Earnings Inequality, 1995 to 2010

Markus P. A. Schneider
Economics Department - University of Denver

July 19, 2012

Abstract

This paper looks at the evolution of earnings inequality from 1995 to 2010 in order to investigate whether the increase in inequality documented by Thomas Piketty and Emmanuel Saez is the result of changes in income composition across the distribution or holds for income from labor. It turns out, earnings inequality has followed a qualitatively similar, though less extreme trend. In the process, the apparent disconnect between the Gini coefficient - which has not changes much - and inequality assessed via the share of income going to the top percent of income earners is clarified. The Gini is insufficiently sensitive to changes in the upper tail of the distribution to illustrate the relevant changes in the income distribution. Finally, the evolution of the earnings distribution of four major demographic groups are considered separately, which shows that there are important differences in the experience of inequality that imply that race and gender are not separable when it comes to understanding the distribution of earnings in the US. The main findings are that only white men have experienced changes in within-group inequality that parallel the changes in inequality seen in the overall distribution. By contrast, the black population (male and female) has seen no notable increase in within-group inequality. Some speculation how this might impact political attitudes towards the issue is provided to ground the analysis.

Subject Codes: D31; D63; C46
Keywords: Inequality; Earnings; Income Distribution; Dagum Distribution

1 Introduction

While the financial crisis that began to unfold in 2007 and the Great Recession that followed reinvigorated attention on economic inequality, unnerving trends in inequality driven by the concentration of income and wealth at the top of the distribution were well-documented before the crisis (Piketty and Saez, 2003, 2006; Gordon and Dew-Becker, 2007; Atkinson, Piketty, and Saez, 2011). What has not been addressed very much is that the trend in increased income concentration at the top of the distribution described by Piketty and Saez (2006) appears at odds with the trend in summary inequality measures like the Gini coefficient. By comparing changes in the Gini to a different measure of inequality that is more sensitive to changes in the upper tail of the distribution using data for the UK, Jenkins (2009) showed that this apparent disconnect is likely due to the choice of inequality measure. Burkhauser, Feng, Jenkins, and Larrymore (2009) and Burkhauser, Larrymore, and Simon (2011) also address the qualitatively different trends of the Gini and the share of income going to the top 1% to some extent, although they concentrate mostly on cautioning that Piketty and Saez (2006) may over-state the concentration at the top of the distribution by considering only pre-tax, pre-transfer income. While Burkhauser et al. (2011) argue that considering the economic resources available to households more broadly shows that households across the income distribution have made gains, their work illustrates that this has only happened through policy intervention and more relevantly to this paper, households in the top income brackets gained much more than those in the middle or bottom even when the broader measure of income is used. While redistribution through taxes and transfers ensures that economic gains are at least partially shared by households in the bottom and middle of the distribution, it is small compared to the upward transfer of income that occurs prior. These trends in inequality have had some effect on political attitudes across the political spectrum, as captured by Norton and Ariely (2011) who found an across the board consensus that the level of wealth inequality in the US is too high. This paper tackles the technical measurement of inequality and the differences in recent trends that changing the measure of inequality makes, the differing experiences of inequality among various demographic groups, and some speculation about the effects on political attitudes that these trends might have all with respect to earnings in the US.

There remains plenty of room to debate what rising inequality means and whether the type of inequality experienced changes the answer. Along the economic dimension, two extreme positions are that earnings inequality is the result of functioning markets rewarding income earners according to their marginal productivity and their choice of how much to work versus that the extreme growth in inequality drive by top income earner's remuneration reflects rent-seeking (i.e. the exploitation of market failures). The latter is often also tied to a positive feedback loop in which regulatory capture creates greater opportunities for rent-seeking and thus provides further resources to facilitate regulatory capture, thus suggesting that inequality is the symptom of a socially destabilizing threat to democracy. These two narratives underlie much of the political discussion around inequality, but are generally not addressed head on in more descriptive academic works on inequality. This paper is no exception and discerning between causes of inequality is not the central focus here. Gordon and Dew-Becker (2007) addresses the apparent imbalance in how the gains from increasing productivity are distributed, namely that they have been largely captured by top income earners while the majority of income earners do not share in the fruits of productivity gains (see also Gordon and Dew-Becker, 2005) (or at least not until redistributive policy intervenes, as Burkhauser et al. (2011) would have it). Gordon and Dew-Becker (2007) suggest that one of three mechanisms must explain the gains made by top income earners: they are "superstars" (athletes or entertainers who garner high incomes proportional to the audience they reach), they represent the incredible productivity gains made by highly skilled and highly paid professionals, and / or the disproportionate compensation of top-level managers (CEOs, CFOs, etc.) in the private sectors. The first and last category imply at least some contribution of rent-seeking to observed incomes, so perhaps it is worthwhile to think of some top-level managers also as "superstars" whose compensation is based on the size of the company they head and goes beyond remuneration of their individual talents. Surely top incomes reflect all elements of each of these explanations but Gordon and Dew-Becker (2007) suggest that the lavish compensation of top executives as a particularly significant factor has considerable evidence behind it. In so far as top-earnings are part of the compensation packages of executives, this study tangentially relates to these issues, but whether the reported trends reflect increased rent-seeking or "fair" compensation is not directly addressed. Since this paper looks at earnings the results surely also relate to how the gains from productivity gains are shared across the population (the subject of Gordon and Dew-Becker, 2005), but this is another story that is not directly explored here.

Earnings refer to income from labor in the form of wage and salary income, self-employment income, and farm self-employment income. The focus in this paper is that looking at earnings separates changes in income inequality that may be due to the composition of income more broadly, which includes both gains from labor productivity and the productivity of capital. Since people in the middle of the income distribution derive the majority of their income from earnings, while income earners at the top of the income distribution derive much less of their income from earnings and more from the ownership and management of assets, looking at only earnings, the first-order effect of differences in the composition of income is hopefully eliminated. In fact, the question whether the qualitative pattern of increasing inequality driven by top incomes even holds for earnings remains to be satisfactorily clarified in the literature. If earning inequality follows a qualitatively similar pattern as income inequality broadly, that finding would sharpen the question whether differential gains in labor productivity can really explain growing inequality since it would have to imply that those gains are also concentrated at the top. Honing in on this element of the question how changing inequality relates to productivity gains is one of the major contributions of this paper, and the evolution of the earnings distribution from 1995 to 2010 is indeed similar to the trends in pre-tax, pre-transfer income described by both Piketty and Saez (2006) (and summarized by Atkinson et al., 2011) and Burkhauser et al. (2009). Thus it appears that labor contracts have become institutional reinforcements of the top income earners garnishing the majority of the benefits from growth prior to the impact of policy; whether this is justified by productivity gains is left to subsequent work.

As mentioned above, this study extends the analysis to look at how some demographic groups experience changes in inequality differently, which picks up on the work of Gordon and Dew-Becker (2007) who looked at the income distribution of men and women separately. Smeeding and Thompson (2011) also shows that the racial composition of the top 10% and top 1% of the income distribution is quite different from that of the population overall. It stands to reason that the changes in the distribution of earnings therefore affect different groups quite differently. Monnat, Raffalovich, and Tsao (2012) describe in detail how key income levels - including incomes for the top 1% - evolved for different ethnic groups, and their results indeed suggest that different groups have experiences qualitatively different changes in their income distributions. Unfortunately neither paper considers the interplay of race (or ethnicity) and gender, while there is ample evidence that the average experience of working women of color is different than that of their white counterparts. By looking at the earnings distributions of white men, white women, black men, and black women separately, this study makes a novel contribution to how changes in earnings inequality differ across groups. The differences in changing within-group inequality experienced by these groups is illustrated by estimating the inequality measures for the earnings distribution of each group. The results will help shed light on why perceptions of inequality differ considerably among the groups considered in this study, which likely impacts political attitudes towards inequality. One finding is that white income earners in the US have experienced a considerable increase in within-group inequality driven by the concentration of earnings at the top, while black income earners have made modest gains in catching up the median white income earner. However, top white income earners have pulled away from both median white income earners and black income earners. These complicated changes in the distribution of earnings are likely what supports the multiple competing narratives about inequality and its meaning. Some novel conjectures along these lines are informally compared to changes in responses to relevant questions in the General Social Science (GSS) survey; these and all other results are presented in detail in section 5.

Before getting to the results, however, further discussion of which inequality measure to use and what the relevant qualitative trend in distributional change is are taken up in section 2. Aside from the choice of inequality measure and definition of income, there are several issues specific to the use of CPS data for assessing inequality. To deal with them, the method spelled out by Feng, Burkhauser, and Butler (2006) of fitting a synthetic parametric distribution to the data and estimating inequality statistics from it is adopted in this study. The discussion of which distribution to use as the synthetic distribution for earnings is also taken up in section 2, while data issues and estimation details are spelled out in sections 3 and 4 respectively.

2 Background

The changes in the income distribution that are the subject of Piketty and Saez's work are well illustrated by what happened to the incomes of the 400 largest (in terms of reported Adjusted Gross Income) tax returns filed with the IRS every year, summary statistics for which are made publicly available. In 1995, the top 400 income earners captured 0.49% of total income[1] and they accounted for 0.00034% of the total number of returns filed. By 2007, their share of total income reported had swelled to 1.59%, while they accounted for only 0.00028% of all returns.[2] For the top 400 income earners, salaries and wages account for only a small (and decreasing) portion of total income. While wage and salary income accounted for 14% of their total AGI in 1995 and 16.7% by 2000, it declined steadily thereafter. (In contrast, income from partnerships and S corporation, capital gains, and dividends account for 60% to 70% of the top 400's total AGI.) In 2007, only 306 of them even reported wage and salary income, accounting for 6.5% of their total reported income and 0.15% of total wage and salary income for all tax receipts. Given that the 306 top income earners who reported wage and salary earnings accounted for only 0.00021% of all returns filed, they still captured a disproportionate share of income from this source.

The changes in how much the top 400 increased their share of the pie are not necessarily apparent without normalizing the size of their group relative to the total number of tax returns filed. It is useful to translate the actual changes into equivalent hypothetical changes in the share going to the top 0.01% (the top 1% of the top 1%). The change in the share of total AGI captured by the top 400 is *as if* the top 0.01% captured 14.4% of total income in 1995 and 56.8% of total income in 2007. (Their share of net capital gains and dividend income increased four- and five-fold respectively.) The increase in their share of total salary and wage income is less dramatic, but substantial nonetheless: it is *as if* the top 0.01% captured 3.1% of total salary and wage income in 1995 and 7.1% in 2007. Looking at the top 400 income earners suggests that the qualitative results summarized in Atkinson et al. (2011) hold for earnings, even if the magnitudes involved are smaller in part because earnings constitute a relatively minor income source for the top income earners. If the earnings distribution followed the same pattern as the

[1] Based on AGI (Adjusted Gross Income) reported in the IRS publication "The 400 Individual Income Tax Returns Reporting the largest Adjusted Gross Income Each Year, 1992 - 2009" available at `www.irs.gov/taxstats/article/0,,id=203102,00.html`.

[2] The share of total income going to the top 400 income earners rose relatively steadily between 1995 and 2007, with a brief decline from in 2001 and 2002 when it dropped to 0.85% and then 0.69% respectively. However, the total number of returns increased steadily, so that even when the share of the top 400 declined modestly, the portion of the population they presented was smaller than in 1995.

overall income distribution, then this suggests a very broad redistribution of income from all sources (except transfers) towards the top of the income distribution. The question is that if this trend started in the late 1970s, why did it take so long for economists - plenty of whom are concerned with inequality and the distribution of income - to recognize what was happening? Part of the answer surely has to do with the availability of data (which Piketty and Saez solved tediously compiling information from IRS records), but another part of the answer has to do with how inequality is measured (which Piketty and Saez addressed by looking at the share of income going to different parts of the distribution).

There are numerous ways of measuring inequality and economists from different parts of the discipline tend to favor one or another. What is perhaps often overlooked is that different indexes of inequality privileges changes in one part of the distribution over another. In particular some very popular measures like the Gini coefficient may be particularly ill-suited to capturing the changes in distribution documented by Piketty and Saez (2006), leading to the puzzling observation that the official Gini (calculated from the uncensored CPS income data) has not increased very much over the period from 1995 to 2010 compared to the share of income going to the top. The reason for this is that the Gini is sensitive to changes in the distribution around the mode (as pointed out by Atkinson, 1970; Jenkins, 2009). Specifically, if income is systematically re-distributed from the bottom 99% of the distribution to the top 1% without changing the relative position of individuals among the bottom majority, then the Gini might change relatively little compared to other measures. Using UK income data, Jenkins (2009) illustrates that the Gini understates inequality (compared to an alternate measure that is more sensitive to changes in the upper tail of the distribution) when the change in the distribution is driven by an increasing share of income being captured by the top 1% or top 0.1%. There is ever more evidence that the US income distribution (as well as the UK's) has experienced exactly this kind of change since the late 1970s (not the least of it presented by Piketty and Saez, 2006; Atkinson et al., 2011).

Anthony B. Atkinson was among the first to link this to different weights being assigned in a social welfare function and thus suggesting that the choice of inequality measure is a normative decision made by the researcher (see Atkinson, 1970). Atkinson proposed his own inequality index that allows the desired weighting of the implied social welfare function to be specified by a single parameter, and Atkinson's index is subsumed by the generalized inequality index used by Jenkins (2009) and in the present work. More importantly, the evidence presented by Piketty and Saez (2006) suggests that how the distribution of income evolved experienced a qualitative change compared to the two decades after World War II. Not taking this into consideration by adjusting the weighting of one's social welfare function (i.e. changing which inequality measure one pays attention to or ideally considering multiple measures) and thus missing a break in the evolution of the income distribution - which appears to indeed have been missed by many economists - goes beyond something that can be dismissed as making different normative choices.[3]

The alternative measure of inequality used by Jenkins (2009) and in this paper is the generalized entropy index, $I[\alpha]$, given by (1) where $F[y]$ is the *cdf* of the distribution of y (see Jenkins, 2009, for details and further citations). The $I[\alpha]$ index is based fundamentally on the

[3]If there are qualitative changes in how the distribution of earnings evolves over time, then sticking to one inequality index actually implies that the researcher is changing his or her normative prerogatives by not adjusting to changes in how earnings are distributed. Using the same index under the false assumption that it always reflects the same normative position is an error that goes beyond normative decision making at the researcher's discretion. Consider whether x% of total income is take from the bottom half of income earners and given to the top half, or the same percentage of total income is stripped from the incomes of all but the top income earner and his income is increased by the sum taken from all the others. The Gini will increase the same amount in both cases. However, the two redistributions of income have a very different character, and it is hard to imagine under which circumstances viewing them as the same reflects consistent normative priorities.

distance of different income observations, y, relative to the mean income, $\mu \equiv \mathrm{E}[y]$, weighted by their respective probability of being observed, $dF[y]$. Thanks to the parameter α, $I[\alpha]$ can be calibrated to weight income below the mean ($\alpha < 1$), around the mean ($\alpha = 1$)[4], or above the mean ($\alpha > 1$) more heavily. In other words, the weights of the implied social welfare function can be adjusted to reflect a greater distaste for inequality at the bottom, in the middle, or at the top of the distribution.

$$I[\alpha] = \frac{1}{\alpha\,(\alpha - 1)} \left(\int \left(\frac{y}{\mu}\right)^{\alpha} dF[y] - 1 \right),\ \alpha \neq 0, 1 \tag{1}$$

In particular, $I[2]$ (equivalently $\frac{1}{2}CV^2$) amplifies the impact of observations larger than μ even if these observations are not very likely, placing additional emphasis on inequality driven by very large incomes going to a few top earners. This study will compare the evolution of the Gini coefficient to changes in $I[0]$, $I[1]$, and $I[2]$ to show how the perception of how inequality has changed is affected by the normative weighting implied by the chosen measure of inequality, where the divergence (or not) between the Gini and $I[2]$ is of special interest.

The choice of index is not the only complication to how inequality is measured, even once a definition of income has been settled on. There are several specific issues with the CPS income data that have been discussed variously by Feng et al. (2006); Burkhauser, Feng, Jenkins, and Larrymore (2011). The method of fitting a synthetic distribution to the data and calculating inequality measures from it is directly borrowed from Feng et al. (2006), who showed that this can provide consistent estimates of inequality measures using the public-use CPS data. Feng et al. (2006) also showed that procedural changes in the early 1990s make it practically impossible to estimate consistent measures of inequality through this period. The period from 1995 to 2010 was chosen to avoid issues associated with these changes.

The major issue addressed by using a synthetic distribution instead of calculating inequality measures directly from the data is top-coding: the replacement of large observations with some pre-defined limit or average value, which will be discussed in a section below. Rather than using the multiple imputation method suggested by Jenkins, Burkhauser, Feng, and Larrymore (2011) to deal with top-coded values that were replaced by cell-means, a conservative likelihood function is built for the fitting of the synthetic distribution to the data in which all top-coded values are treated as censored. All observations not subject to top-coding are thus taken as if they come from a truncated sample. One reason for taking this particular approach is that the central concern of this study is the relative trend in different inequality indexes, not the exact estimation of their levels. There is no reason to believe that the strategy chosen would produce spurious results in the relative trends, even if the level estimates turn out to be biased.

Feng et al. (2006) use the generalized beta distribution of the second kind (GB2) as the synthetic distribution that is fit to the CPS income data, while the analysis presented here uses the Dagum distribution.[5] In general, the GB2 has recently gained popularity for fitting the observed income distribution (Burkhauser, Butler, Feng, and Houtenville, 2004; Feng et al., 2006; Jenkins, 2009) because of its great flexibility in modeling heavy tail behavior and a single positive mode (features which are shared to a limited extent by the Dagum distribution) and because Parker (1999) has shown that the GB2 distribution can be derived from neoclassical micro-foundations. Many other popular distributions used in the distant and recent past can be derived from the GB2 distribution (i.e. presented as nested models of the GB2), including the Dagum distribution (McDonald, 1984; Borzadaran and Behdani, 2009). The good fit of the GB2 to the income distribution be seen as a testament to the flexibility of the GB2 (it should certainly not be interpreted as a validation of Parker's model). Fitting a continuous distribution

[4]Strictly speaking the parameter α is restricted to not being equal to 0 or 1, but it is possible to derive $I[0] = \lim_{\alpha \to 0} I[\alpha]$ and $I[1] = \lim_{\alpha \to 1} I[\alpha]$ (Jenkins, 2009).

[5]This distribution was named after Camilo Dagum, who proposed it as a size distribution of incomes (Dagum, 1977).

- like the GB2 or any of the distribution derived from it - to the earnings data is an inventive use of a functional representation to describe the shape of the data; the chosen distribution is a convenient tool to summarize the shape of the observed distribution using a limited set of parameters and nothing more. To emphasize this point, the parametric distribution fit to the data and used to estimate various inequality indexes has been referred to as the synthetic distribution.

The reason for rejecting a theoretical argument for the GB2 is that it is very difficult in general to make a believable case that labor market outcome are generated by some process that has a well-defined continuous stationary distribution. The non-homogeneity of labor, search frictions, etc. all imply that aggregate labor outcomes are at best represented by a mixture of different distributions.[6] The model proposed by Parker (1999) is illustrative on this point. At its heart is a representative firm hiring identical workers who choose to obtain different levels of human capital according to what the firm will pay each skill-level in order to obtain the optimal distribution of skills across its workforce. There is no unemployment in this model, nor search frictions or any other real behavior that afflict employer and employees. Under specific assumptions - namely constant elasticity of income returns to changes in human capital, constant elasticity of costs to changes in income, and constant labor elasticity of output - the GB2 distribution arises as the optimal distribution of income paid by the representative firm. By extension, the model also suggests that this optimal distribution could be of the type identified by Dagum if the income elasticity of human capital was smaller than the income elasticity of employment costs and that their ratio satisfied a specific relationship[7] to the labor elasticity of output. Alas, the model proposed by Parker (1999) hardly offers a convincing theoretical motivation for using the GB2 distribution.

Given the lack of guidance provided by economic theory, the choice of which distribution to use as a functional description of the data must balance fit with parsimony, and will invariably reflect the priorities of the researcher. That other authors (Burkhauser et al., 2004; Feng et al., 2006; Jenkins, 2009) prefer the GB2 reflects their preferences for fit over parsimony. But aside from fit, it is important to consider whether the synthetic distribution used in the analysis has the flexibility to capture the relevant features of the earnings (or income) distribution. Additional flexibility provided by more parameters - even if they result in better fit - does not necessarily contribute to the analysis. The analysis presented in this paper uses the best-fitting, 3-parameter distribution - the Dagum distribution - which is capable of capturing the relevant features of the earnings distribution. The impact of this choice is explored in a subsequent section and while it appears that using the Dagum leads to a estimating a thinner tail for the earnings distribution than choosing the GB2, the qualitative trends do not appear to change.

The features that distinguish 3-parameter distributions from popular 2-parameter distribution - like the log-normal or Weibull distributions - that appear most relevant to the distribution of earnings are the ability to fit data with a mode at zero or a positive non-zero mode (and bi-modality in a limited sense), and that they can model fat-tails. The log-normal precludes concentrations of observation at or near zero and cannot produce infinite second (or higher) moments. The Weibull distribution can produce some of these features, but is not flexible enough to allow all the desirable combinations thereof. The Singh-Maddala distribution is aanother popular 3-parameter distribution used for incomes that can be derived from the GB2 distribution, and there is a direct relationship between the Signh-Maddala distribution and the Dagum distribution.[8] Yet, McDonald (1984) and Kleiber (2008) have argued that the Dagum distri-

[6]To be specific, search frictions and heterogeneity surely are relevant features (Rogerson, Shimer, and Wright, 2005), not to mention the theoretical issues associated with representative agent models (Kirman, 1992).

[7]Letting $\alpha \in (0,1)$ be the labor elasticity of output, $\gamma \in (0,1)$ the income elasticity of human capital, and $b \in (0,1)$ the income elasticity of employment costs, then $\alpha = \frac{1+\gamma}{1+b}$ must be satisfied for the optimal distribution of incomes paid to be the Dagum distribution in Parker's model.

[8]If X is Singh-Maddala distributed, then $1/X$ is Dagum distributed.

bution fits observed income data better than the Singh-Maddala distribution in part because the Dagum distribution allows for tail behavior to be fitted using two parameters, and Feng et al. (2006) also explicitly argue against the use of the Singh-Maddala distribution based on goodness-of-fit. In combination with the author's own assessment of fit, this study is based on the earlier statement that the Dagum distribution appears to be the best-fitting 3-parameter distribution that accurately captures the relevant features of the observed earnings distribution (see also Kleiber and Kotz, 2003), and it is therefore deemed an appropriate tool for the analysis presented here.

Thanks to the wealth of demographic information available in the CPS data, the distribution for four demographic groups are analyzed separately to investigate further whether the experience of how earnings inequality is changing is shared among them. Gordon and Dew-Becker (2007) suggested that men and women's income distributions might be changing in quite different ways, and that the experience of evolving inequality thus differs. Smeeding and Thompson (2011) and Monnat et al. (2012) investigate distributional issues across racial / ethnic groups. All rely heavily on percentile ratios (90/10, 90/50, and 50/10), which may not be appropriate if most of the action is happening within the top 1%, as will become apparent in light of the results presented in this paper. While all three studies look at the share of income going to the top 1% or the 99$^{\text{th}}$ percentile income to round out their analysis, there are few observations in the CPS of some ethnic groups in the top 1% (as Monnat et al., 2012, points out explicitly) and the top-coding of income reports must be addressed explicitly.

Unfortunately, Monnat et al. (2012) specifically do not appear to deal with top-coding (despite citing Burkhauser et al., 2009), which calls into question their results for the median income of the top 1% across racial/ethnic groups. They nonetheless show that income gains varied more across family income quartiles than race for the bottom 99%, although there is a notable increase in the black-white income gap. More precisely, they show median incomes each of the bottom two quartiles remaining largely stagnant in real terms while incomes in the top two quartiles excluding the top 1% rose appreciably after 1995. (In fact, the creates rise in real income in the top quartiles appears to be from 1995 to 2001.) The same patterns are seen among black and white families and persist in family income from employment. However, median income for black families in each quartile are consistently considerably below median income of white families and the gains made appear to be smaller. Monnat et al. (2012) thus report an increasing black-white family income gap after the late 1990s especially in income from employment.

The present study considers person-level earnings and is the first study to consider that race and gender interact while accounting for top-coding of the CPS. Specifically, it is plausible and probable that there may be important qualitative differences between the distribution of earnings and experience of inequality among white men, black men, white women, and black women. It is mostly left to the imagination of the reader - and further research - how the gender and racial composition of certain occupations might connect these results to the mechanisms of distribution identified by Gordon and Dew-Becker (2007). A suggestive reminder will conclude this section: white men still dominate leadership positions in many corporations (especially large ones), and while white women have made some in roads (predominantly in smaller companies), men and women of color remain largely underrepresented. They are, however, heavily over-represented at the bottom of the earnings distribution.

3 Data

The present study looks at the evolution of inequality in person-level earnings, where earnings are understood as pre-tax income from time spent working (as opposed to the ownership and management of assets). The data used for this study is the publicly available ASEC supplement[9] to the CPS collected annually by the census bureau. It contains over 60,000 person responses[10] plus analytical weights indicating how representative observations are of the population based on the previous decennial census. The analysis was conducted for 19 - 69 year old respondents identified as in the civilian workforce who reported non-zero positive earnings. The earnings variable used for this study includes wage and salary income from all jobs, as well as business and net farm self-employment income.[11]

The period of investigation was primarily chosen to capture the most recent year available at the time of writing and to go back as far as reasonable. As Feng et al. (2006) showed convincingly, procedural changes in the early 1990s make it difficult to create series of consistent estimators of inequality measures from the CPS data. There appears to be a break between 1994 and 1995 that is likely caused by procedural changes. In order to avoid drawing incorrect conclusions about the trend in inequality, the earliest year chosen for this study was therefore 1995 (incomes reported in 1996 as earned during the previous year).

A particular problem with the public-use ASEC data is that responses to questions about earnings are top-coded both at the point of collection and when the full restricted-access data is modified for public use. When the survey data is collected, responses to questions regarding earnings that exceed the top-code limit are recorded as that limit. For example, the maximum amount that was recorded for income from the longest job held in the previous year[12] was \$999,999 in 1995, and responses greater than that were recorded as \$999,999 (see Burkhauser et al., 2004, for a fuller discussion). A second round of top-coding occurs when the data is prepared for release to the public. Starting in 1995, observations that exceed the top-code limit in one or more of the categories that contribute to total earnings are replaced by the mean of all income reports by respondents with similar demographic characteristics whose income also exceeded the top-code limit. The top-code limits for the relevant income categories appear in table 1.

Income Category	**Top-Code Limits**		
	1995 - 2001	**2002 - 2009**	**2010**
ERN_VAL	\$150,000	\$200,000	\$250,000
WS_VAL	\$25,000	\$35,000	\$47,000
SE_VAL	\$40,000	\$50,000	\$60,000
FRM_VAL	\$25,000	\$25,000	\$30,000

Table 1: *Top-Code limits for different income categories that make up earnings.*

While few respondents earn incomes that fall above the recording limit, their effect on the distribution has direct implications for this study. The practice of limiting the maximum response distorts the shape of the upper-tail of the earnings distribution by truncating it but also adding an exaggerated point mass at the top-code limit. Secondly, imposing a lower limit on income categories when the restricted-access data is prepared for public use propagates the

[9]Often referred to as the March Supplement to the CPS.

[10]Notable is that between 2000 and 2001, the census switched to an electronic data collection system and consequently the sample size increased substantially. Before 2001, samples ranged from 61,000 to 64,000 observations, while from 2001 onward, between 92,000 and 102,000 observations are available. Actual sample sizes can be found in the Appendix.

[11]The specific variable used is PERN_VAL, which is the sum of ERN_VAL, WS_VAL, SE_VAL, and FRM_VAL. The first two variables capture wages & salary earnings, and the latter capture self-employment and farm self-employment earnings respectively.

[12]The ASEC CPS asks respondents to report their income from the longest held job in the previous year, and to classify that income as wage / salary, self-employment, or farm self-employment income.

distortion caused by the recording limit downward. When responses above this lower limit are replaced by the mean of responses that fit the same demographic characteristics, the recorded top-code limits are included in the the cell-means. Worse, the public use data does not include a top-code flag when the recording limit was included. It is therefore impossible to directly estimate the effect of this distortion. The good news is that the top-code limits used to create the public-use data are high enough to affect less than 5% of the observations (which raises the question why Jenkins et al. (2011) use only the densest 70% of the observed distribution to fit the synthetic distribution used for their multiple imputation procedure).

Despite these issue with the data, the sheer number of observations makes it possible to robustly fit distributional models to it. Burkhauser et al. (2004) and Feng et al. (2006) have shown that the procedure of fitting a synthetic distribution can produce a time-consistent series for the Gini coefficient from 1995 onward that smoothes out changes in top-coding procedures and limits, which is why their approach is adopted for this investigation.[13] The range of earnings observations not affected by top-coding almost reaches 99% most years. It is, therefore, reasonable to report estimated threshold incomes up to the 99$^{\text{th}}$ percentile income (which implies a small extrapolation beyond the data).

This study also considers differences among income earners who identify themselves as either white or black. After 2002, respondents to the CPS could identify as multiple races. To reconcile the racial identification variable (A_RACE) before and after 2002, the current study take respondents who identified only as white as white, and respondents who identified as black alone or black plus one other race as black. This is likely a controversial departure from the convention of comparing respondents who identify as white only and those who identify as black only. However, it is not clear that racial identity is symmetric and it is assumed here that a respondent who chooses black together with another race does so because they feel treated as black by society at least some of the time. In other words, it is presumed that while the choice of "white" reflects a level of racial unawareness often referred to as "white privilege", the identification as black (alone or together with another race) reflects socially reinforced racial awareness, which is more likely to have lead to identification as black when the choices were restricted to white, black, or other prior to 2002. In any case, there does not appear to be a break in the series of inequality measures for the distribution of earnings among black respondents in 2002, suggesting that this particular choice did not create an inconsistency in which group was identified.

4 Estimation

The analysis presented in this paper rests on a synthetic distribution being fit to the data and various inequality measures being calculated from the fitted distribution rather than the data itself. The parameters of the synthetic distribution are estimated using maximum likelihood estimation (MLE), which is the standard practice. Implicitly the possibility of observational errors are ignored[14] - i.e. the earnings observations are treated as if they are i.i.d. draws from the distribution being fitted - and that only the parameter values are unknown. The likelihood is built in a way that accounts for the the top-coding issues discussed above by treating the top-coded observations as censored and all other observations as coming from a truncated sample. By treating the top-coded observations as censored, information contained in the cell-means with which top-coded observations are replaced in the CPS post-1995 is disregarded. However,

[13] While Monnat et al. (2012) cite Burkhauser et al. (2009), it is not clear they head the warnings about procedural changes during the early '90s, thus finding an incredible jump in median income for the top 1% among white respondents from 1995 to 1996. It seems probable that this jump is at least in part an artifact of changes in top-coding procedures which are not explicitly dealt with in Monnat et al. (2012).

[14] It would be a worthwhile addition to this literature to treat reporting errors and possible respondent bias explicitly in the likelihood, but this daunting task is not taken up in this study.

the censoring limit is informative in two ways that are taken advantage of in the likelihood. First, observations that are not top-coded are known to be below some maximum above which all observations are top-coded for one reason or another. For example, the top-code limits in 2000 suggest that the maximum a non-top-coded observation of PERNVAL could be is \$230,000 (\$150,000 + \$25,000 + \$40,000 + \$25,000). Every observation that is not top-coded, therefore must have come from the portion of the distribution truncated at this maximum value. Second, the top-code limits provide a lower bound for censored observations. If an observation was top-coded because primary earnings (ERN_VAL) exceeded their respective top-code limit, then this determines the censoring limit that the observation must have exceeded.[15] This is an alternative approach to choosing either a constant truncation limit below the most constraining top-code limit, or dropping a fixed percentage of the observations (described and criticized in Feng et al., 2006).

The synthetic distribution fit to the data is assumed to be continuous with $f[x|\theta]$ being the *pdf* and $F[x|\theta]$ the *cdf*, where θ is the parameter vector that specifies the distribution. A general version of the likelihood used to fit the synthetic distribution to the data is given by (2). Of the N total observations, n are uncensored but come from a truncated sample, m are censored because they exceeded the ERN_VAL top-code limit, l exceeded the SE_VAL limit, k exceeded the WS_VAL limit, and h exceeded the FRM_VAL limit. If - as is the case with the CPS data - the data is weighted, then w_i denotes the appropriate weight for each uncensored observation, and m, l, k, and h are the respective sums of weights for the censored values.

$$L = \prod_{j}^{\{m,l,k,h\}} \left(1 - F\left[X_{TC_j}|\theta\right]\right)^j \cdot \prod_{i=1}^{n} \left(\frac{f\left[x_i|\theta\right]}{F\left[X_{TC\_MAX}|\theta\right]}\right)^{w_i} \tag{2}$$

The reason for choosing this formulation for the likelihood is that it concisely deals with the double top-coding that the data is subject to. The public use data does not include a top-code flag for values that were censored at the point of recording (unlike the restricted-access data that Feng et al. (2006) have access to). Rather than include the cell-means, which are potentially downward biased by the inclusion of top-censored earnings records, the decision here is to use only the unbiased information in the data. That means disregarding much of the information provide by the cell-means and likely makes the estimates presented here quite conservative; it seems particularly likely that this procedure underestimates the true weight of the upper tail. Not using the cell-means should also mean that the additional variability that must be accounted for if they are included - which Jenkins et al. (2011) deal with using a multiple imputation approach - is not an issue in this analysis.

Parameters were estimated by maximizing the log-likelihood using *Mathematica*'s built-in numerical maximization function, `NMaximize`. The corresponding standard errors were estimated by taking the inverse of the Hessian of the log-likelihood function and evaluating it at the ML estimators, as is usual practice. Tables listing the estimates and their standard errors can be found in the Appendix.

The Dagum distribution has the convenient feature that both the *pdf*, (3), and *cdf* have simple closed-form expressions and consequently that many of the inequality measures relevant to this paper also have functional forms that are easy to calculate using the estimated parameters. For the Dagum distribution, the Gini has a relatively straightforward formula that can be found in the Appendix.

$$p[y; a, b, p] = \frac{ap}{b} \left(\frac{y}{b}\right)^{ap} \left(1 + \left(\frac{y}{b}\right)^{a}\right)^{-(p+1)} \tag{3}$$

[15]Observations that where top-coded based on multiple income categories use the largest top-code limit as the censoring limit.

It is illustrative to look at the fitted Dagum distributions for a couple of sample years (1996 and 2006 were chosen relatively arbitrarily). Since the focus of this paper are trends in inequality measures which are scale-free, the *pdf*s shown in figure 1 were fitted to nominal earnings. It is therefore surprising that there is no obvious shift to the right in the mode of the fitted distribution from 1996 to 2006. Rather, almost all of the changes appear to be a fattening of the tail and the associated squashing of the lower portion of the distribution. This turns out to be the general case for the evolution of the earnings distribution.

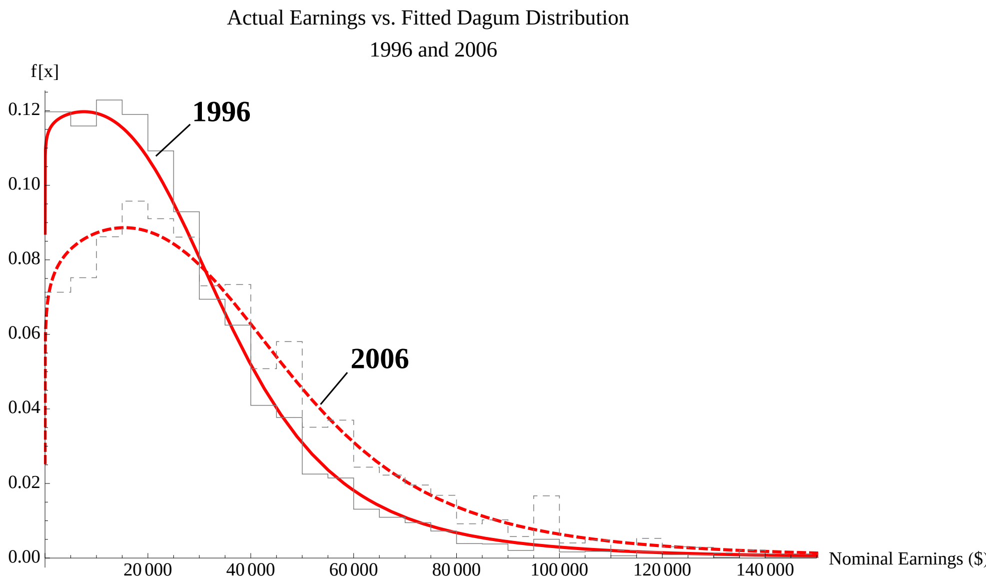

Figure 1: *Fitted Dagum distribution fit to nominal earnings in 1996 and 2006. The histogram of the actual observations is outlined in gray. Visually, the mode and median of the distribution changes little while more weight shifts to the upper tail.*

The GB2 distribution favored by Feng et al. (2006) does provide a statistically significantly better fit of the observed income distribution. However, this author contends that the difference in fit is not economically significant, to borrow from Ziliak and McCloskey (2004). To illustrate the point, the GB2 was fit to two years (1996 and 2006) as a "spot check". According to the log-likelihood ratio test or fit criteria like AIC or BIC, the fit of the Dagum distribution can be rejected in favor of the GB2, but this is based on estimating the sample variability under the assumption that the observations are error-free draws from one or the other distribution. (Can it really be assumed that the answer to the question "How much did your earn last year?" is answered error-free during a phone conversation?) Without more work to take this into account in the likelihood, blindly trusting statistical criteria of fit likely invites over-fitting of the data. As argued previously, there is also no economic reason to believe either distribution is the actually generating distribution of the data. To reiterate the justification for using the Dagum in this study: it appears to be the simplest distribution that fits the data sufficiently well to capture all relevant features.

The difference in estimated statistics is small and reasonably predictable. The mean (in nominal \$) of the Dagum fitted to the 1996 data is \$27,787 versus \$29,070 for the GB2 fitted to the data. According to the fitted Dagum, 1.97% of the distribution falls above \$100,000 whereas according to the GB2, 2.64% of the distribution is above this arbitrary limit. Comparative summary statistics are shown in Table 2 and interpreted as supporting the claim that little is gained by fitting the GB2 to the data.

It is worth noting, however, that the Dagum consistently puts less weight on in the upper tail

Distribution	Median	Mean	90^{th} Percentile	99^{th} Percentile
		1996 Nominal Earnings		
Dagum	\$21,625	\$27,787	\$55,965	\$125,293
GB2	\$22,004	\$29,070	\$56,071	\$151,931
		2006 Nominal Earnings		
Dagum	\$31,054	\$41,298	\$82,020	\$200,863
GB2	\$31,791	\$45,258	\$84,531	\$269,690

Table 2: *Difference in income statistics due to choice of synthetic distribution.*

of the distribution. A rough calculation based on the WTI data suggests that 8.4% of earnings[16] were captured by the top 1% in 1996. Based on the Dagum fitted to the 1996 data, the top 1% captured 6.7% of earnings and based on the GB2 they captured 9.6%.[17] Similar discrepancies are seen when these two distributions are fit to the 2006 data. The main point is that using the Dagum as the synthetic distribution of choice likely implies that the weight of the upper tail is under-estimated, suggesting that reported discrepancies between the Gini coefficient and $I[2]$ should be interpreted as conservative lower bounds; the actual divergent trends are likely to be more dramatic.

Three GE inequality measures are contrasted with the Gini coefficient in this paper: the mean logarithmic deviation or MLD ($I[0]$), the Theil index ($I[1]$), and the half coefficient of variation squared ($\frac{1}{2}CV^2$ or $I[2]$). Each of these has an explicit formulation for the Dagum distribution in terms of the estimated shape parameters p and a (see Appendix). As suggested above, $I[0]$ and $I[1]$ are sensitive to changes below or around the mean of the distribution, while $I[2]$ is more sensitive to changes in the upper tail. Comparing the trends in Gini versus $I[0]$ and $I[1]$ provides a baseline of how the distribution of earnings has evolved around the middle of the distribution. The trend in $I[2]$ will hopefully reveal the nature of changes in inequality in US earnings when top earnings are weighted more heavily and answer the question whether relevant changes in the earnings distribution are neglected if the Gini is the exclusive indicator of inequality. The findings by Piketty and Saez (2006) suggest there should be a notable divergence between the Gini coefficient and the $I[2]$ measure of inequality.

The results for total income that are the subject of Thomas Piketty and Emmanuel Saez's numerous articles are often framed in terms of shares of income going to a particular percentage of income earners in the upper tail. While it is straightforward to produce similar statistics based on the synthetic distribution fit to the data, doing so could dangerously extrapolate beyond where uncensored observations are available. For this reason, this approach is restricted to providing estimates of the 90^{th} and 99^{th} percentile earnings which can be compared to similar threshold incomes provided in the World Top Incomes (WTI) dataset.[18] To overcome the difference in income definition (AGI versus earnings) between the present work and the data provided by the WTI, the relative trends in the mean earnings / income of the bottom 90%, 90^{th} percentile earnings / income, and the 99^{th} percentile earnings / income will be used to see whether the general results presented by Piketty and Saez (2006) hold true for earnings.

[16] Imperfectly estimated by noting that the top 1% captured 14.1% of total AGI, while earnings (wage, salary, and pensions) accounted for 60% of their income. This probably understates the portion of earnings captured by the top 1%, but provides a sufficient baseline for the purposes of this paper.

[17] The cursory analysis performed here does not provide conclusive evidence that the GB2 systematically places too much weight in the upper tail, but anecdotally it appears the truth lies in the middle.

[18] `http://g-mond.parisschoolofeconomics.eu/topincomes/`

5 Results

The trends in inequality indexes estimated from the synthetic earnings distribution are show in figure 2. Clearly, the measure of inequality that are less sensitive to changes in the upper tail of the distribution - Gini, $I[0]$, and $I[1]$ - all suggest that earnings inequality has not changed very substantially over the period from 1995 to 2010. By contrast, the measure most sensitive to changes to in the upper tail of the distribution ($I[2]$) suggests a dramatic increase inequality that was a consistent phenomena throughout the period.

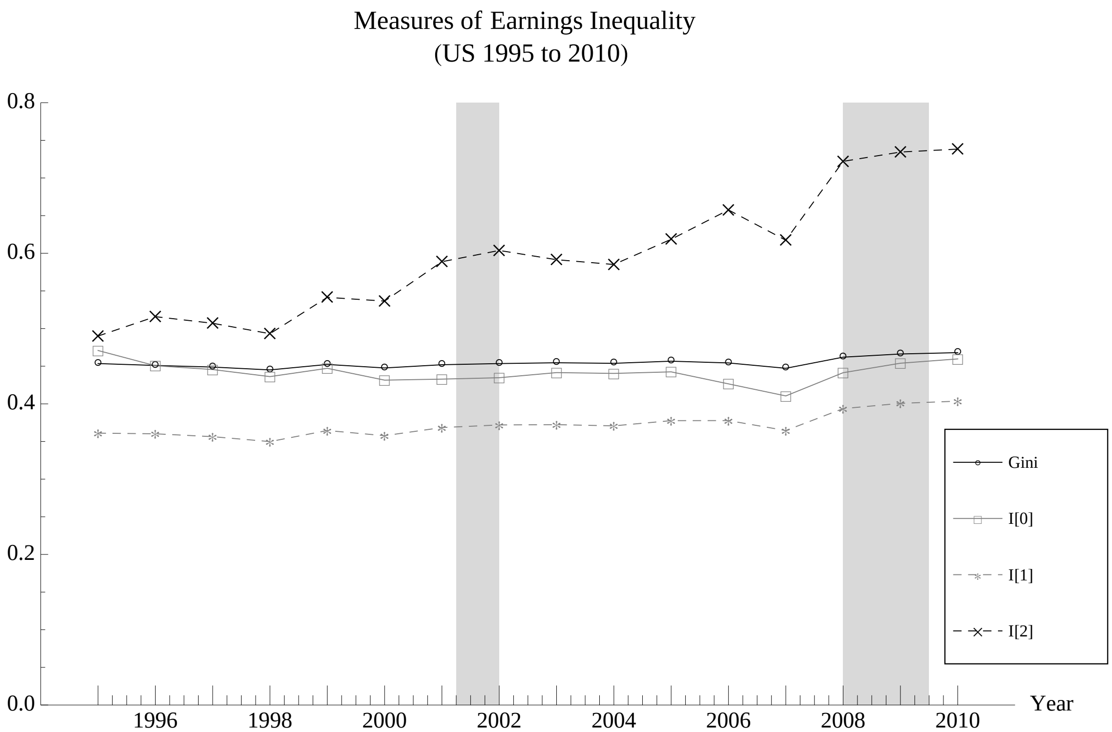

Figure 2: *Gini and $I[2]$ inequality measures calculated from a Dagum distribution fit to US earnings for 1995 to 2010. The gray bars indicate official recessions.*

A closer look suggests that there was some relative earnings compression near the bottom of the distribution leading to a 2.4% decline in $I[0]$. The Gini showed a very mild increase in inequality of about 3.2%. Based on the standard errors estimated from the specification of the likelihood shown in (2), this is a statistically significant change in the Gini (see Appendix), but by contrast $I[1]$ increased by 11.8% and $I[2]$ by 50.7%. Since $I[1]$ is sensitive to changes relative to the mean, which itself is more sensitive to changes in the upper tail of the distribution than the mode, corroborates the trend in inequality amplified by $I[2]$. What the picture provided by figure 2 suggests is that changes in inequality over the period from 1995 to 2010 are predominantly characterized by the share of total earnings going to top earners increasing, rather than a more widely shared dispersion of earnings among workers. It is consistent with those at the very top earning larger incomes, while even the upper middle gained little. The majority of the population near the bulk of the distribution (mode or median) saw, if anything, a slight compression in the distribution of their incomes; modest real gains are dwarfed by a growing distance from the top of the income ladder. In so far as workers sensed this change in the earnings distribution, it supports at least some aspects of the competing appeals to populism that have emerged across the political spectrum.

These results are broadly consistent with the conclusions of Thomas Piketty and Emmanuel Saez's work using IRS data (summarized in Atkinson et al., 2011). Burkhauser et al. (2011) suggest that Piketty and Saez's work may overstate these trends, but it is unclear how conclusive their results are. For one, their multiple imputation approach to filling in the upper income distribution is very novel for providing a reasonable estimate of the sampling variability that the replacement of observations with cell means introduces. However, the synthetic distribution they use is based on the richest 70% of the distribution (Jenkins et al., 2011), which almost certainly means that they are understating the weight of the upper tail. Given the results summarized in Atkinson et al. (2011) and the results of this study, it appears that while Burkhauser et al. (2011) reasonably cautions that the degree to which top incomes have driven increased inequality may have been exaggerated, but they have not made the case that the qualitative story is incorrect.

5.1 Differences in the Experience of Inequality

The next step to consider is whether different demographic groups experience these changes in similar ways. Since the $I[0]$ and $I[1]$ indexes have trends relatively similar to the Gini coefficient, the focus in this section will be the divergence between the Gini and $I[2]$. Figure 3 presents the general results. Smeeding and Thompson (2011) points out that of the population as a whole, almost 70% identify as non-hispanic white and just under 14% identify as black. Meanwhile, the top 1% is 92% white and less than 2% identify as black. Given this demographic composition of the top 1%, it is not surprising that the dramatic increase in $I[2]$ is predominantly experienced among white male income earners. White female income earners saw some of the same trends as white male income earners, though to a much lesser degree, which is likely the result of some increased representation in finance and as the heads of corporations, though they remain very much a minority in those positions. The trends in $i[2]$ for white men and women - and what they imply about how the distribution of earnings is changing for these groups - seems to support Monnat et al. (2012) findings of a growing black-white family income gap in income from employment driven by earnings towards the top of the income distribution.

Second, only white male earners experience inequality mirroring the level of inequality similar to the population at large (in part because they are the single largest group in the civilian workforce). Women as a group experience considerably less within-group inequality than white men, and income earners who identify as black also experience less within-group inequality than white men or white income earners broadly. Given the sizable and persistent black-white wage-gap, these observations all reinforce a sense that for the black population, inequality is primarily issue of inequality between races. For white men, the inequality within their demographic group resembles inequality broadly, and this likely reinforces beliefs that inequality is not a racial phenomena. It is perhaps noteworthy that the two political movements most associated with populist narratives - the Tea Party movement on the right and the Occupy movement on the left - draw predominantly white male participants.[19]

In general, white women, black men, and black women experience inequality more as between-group inequality. Their experience of rising inequality is therefore predominantly a phenomena of rising between-group inequality, specifically of white male top income earners pulling away. It is notable that especially among income earners who identified as black, there is relatively little divergence between the Gini and $I[2]$.

[19]Based on a 2010 New York Times/CBS News poll of backers of the emerging Tea Party movement and a survey using visitors to the Occupy Wall Street movements website by Hctor Codero-Guzmn, a sociology professor at the City University of New York (CUNY).

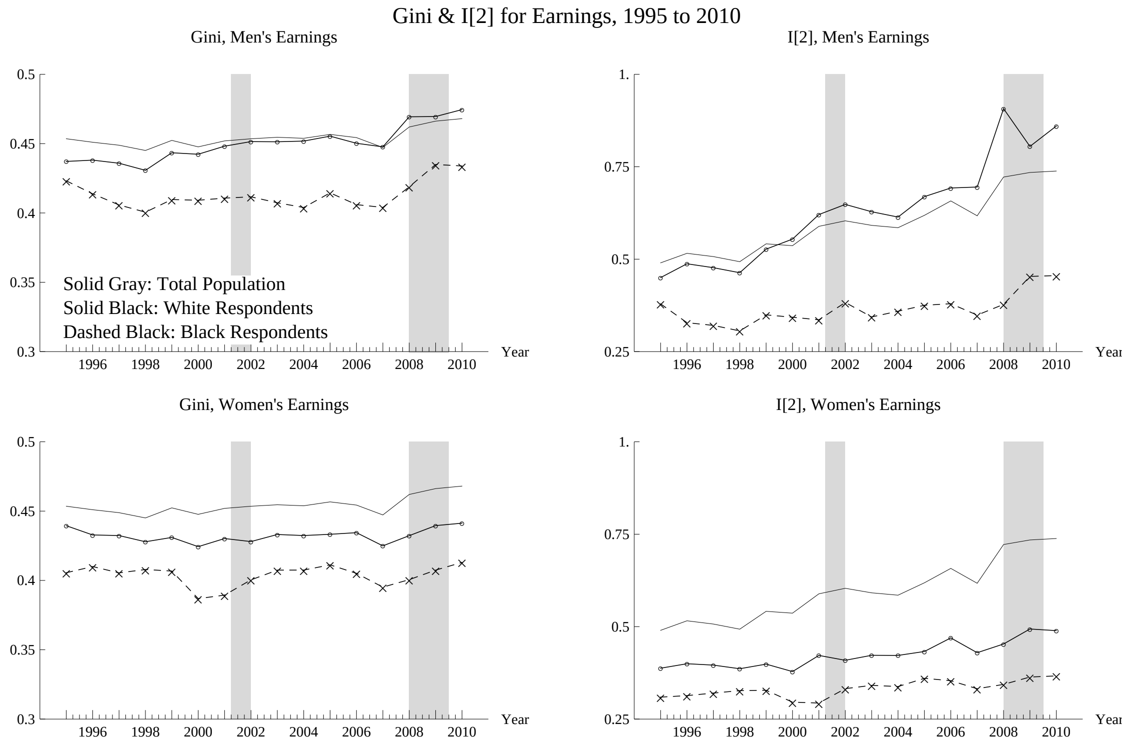

Figure 3: *Changes in Gini and $I[2]$ for male and female respondents who identified as white or black.*

5.2 Changes in Incomes

To help understand what the divergent trend in Gini and $I[2]$ is capturing, it is useful to look at changes in key earnings statistics. Specifically, looking at median earnings compared to the 90$^{\text{th}}$ percentile and 99$^{\text{th}}$ percentile earnings reveals a lot about the changing nature in inequality that is captured by $I[2]$ but not the Gini. Each of these can be easily calculated from the fitted Dagum distribution, which conveniently has a well-defined inverse cumulative distribution function, (4), where F is the % of the distribution that falls below $y[F]$, and a, b, and p are the parameters of the Dagum distribution. While not exactly the same, this is comparable exercise to looking at the income share going to some top percent of income earners that Piketty and Saez (2006) rely upon.

$$y[F] = b\left(F^{-\frac{1}{p}} - 1\right)^{\frac{1}{a}} \tag{4}$$

Since the Dagum distribution was fit to nominal earnings, the calculated values for earnings have to be adjusted for inflation to make them comparable. This was done using the CPI to convert them to 2010 \$US, which is consistent with the data provided by in the WTI database. The evolution of these threshold earnings levels compared to the average earnings for the bottom 90% of income earners in shown in figure 4. To illustrate the difference in the evolution of these key earnings statistics compared to the comparable statistics for income more broadly provided in the WTI data, the latter is also shown. (To make the income reported in the WTI which is based on AGI comparable to earnings thresholds reported here, both are indexed to be 100 in 1995.)

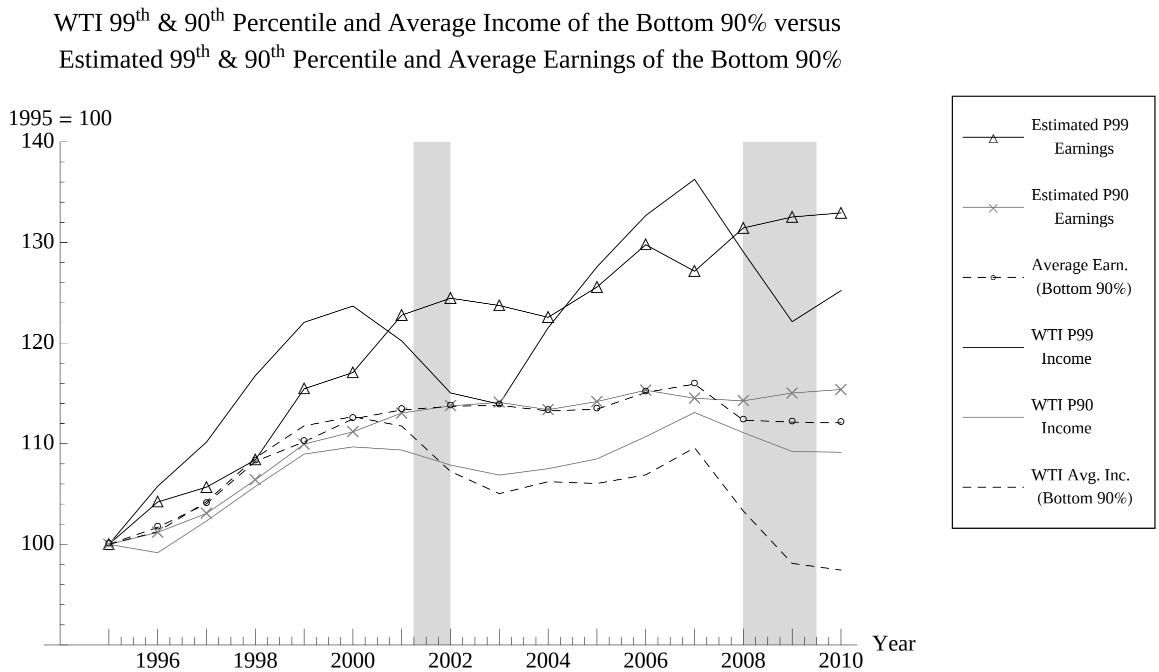

Figure 4: *The evolution of total income and earnings illustrated by looking at the average for the bottom 90% versus the 90th and 99th percentile.*

Real incomes grew broadly throughout the late 1990s. Since the turn of the millennium, however, only incomes and earnings at the very top continued to grow consistently. Average earnings for the bottom 90% and 90$^{\text{th}}$ percentile earnings grew very little compared to 99$^{\text{th}}$ percentile earnings. Top incomes - including all sources of income reported to the IRS as part of AGI - where more volatile than earnings, showing clear fluctuations with the business cycle, but they followed the same overall trend as 99$^{\text{th}}$ percentile earnings. The cyclical movements in top earnings - aside from being smaller - also appear to lag the cyclical movements in top incomes. The 90$^{\text{th}}$ percentile earnings and income followed roughly the same pattern, which most notably meant that both more or less stagnated since the turn of the millennium. Mean total income for the bottom 90% actually declined since 2000, falling below their level in 1995. Mean earnings of the bottom 90% show expected cyclical movements but appear to have held more or less steady; no appreciable gains but also none of the loss seen in income more broadly. The losses for the bottom 90% then have to have come largely in the decline of non-earnings income coordinated with the business cycle (since cyclical movements appear exaggerated in the income series) from which they never recovered. Presumably declines in social transfers and pension payments are part of this pattern.

GDP per capita grew somewhat more sluggishly in the 2000s than in the 1990s, though it did grow, but where did the gains from that growth go? The consistent and substantial growth in 99$^{\text{th}}$ percentile earnings while even 90$^{\text{th}}$ percentile earnings stagnated suggests that most of the gains from that growth went to the top 1% (the gains were indeed privatized while losses were socialized!). The fact that the average income going to the bottom 90% lost all the gains made in the late 1990s suggests that the 2000s have been a period of de facto redistribution from the bottom majority to the top that funneled non-labor income at the bottom into the earnings at the top. Unless one believes that all productivity gains were concentrated among the top 1% of income earners, it is difficult to reconcile these observations with the idea these changes in inequality are not substantially the result of rent-seeking. It also illustrates why some popular measures of inequality like the 90/10 or 90/50 ratio (used in Gordon and Dew-Becker, 2007, for example) may not capture the relevant changes in income or earnings because they do not capture the disproportionate gains made by the top 1%.

To put the experiences of different demographic groups into context, the key threshold incomes estimated from the synthetic distribution for each group are shown in table 3. The estimates indicate that white men are across the board the group with the highest earnings, followed by white women. Median white female earnings are estimated to be \$0.74 per \$1 earned by a white man, where as the 99$^{\text{th}}$ percentile earnings for white women amounts to \$0.54 per \$1 earned by their male peers. Overall, black earnings lag even further behind white men's earnings, though the gender earnings gap is somewhat smaller among black respondents.[20]

	Key Threshold Earnings, 2010		
	Median	90$^{\text{th}}$ Percentile	99$^{\text{th}}$ Percentile
Total	\$32,472	\$88,679	\$222,104
White Men	\$38,913	\$108,312	\$281,583
White Women	\$28,674	\$71,027	\$154,478
Black Men	\$28,054	\$70,901	\$157,429
Black Women	\$25,705	\$60,001	\$121,288

Table 3: *Important income thresholds across demographic groups estimated from the synthetic income distribution arrived at by fitting the Dagum distribution to person-level earnings reports from the CPS ASEC dataset.*

Tables 4 and 5 show the annualized percent changes in real earnings[21] across the income distribution and the groups covered in this study. During the expansion of the late 1990s, earnings grew substantially across groups and income levels. Notable is that overall, 90$^{\text{th}}$ percentile earnings did not grow as fast as either median and 99$^{\text{th}}$ percentile earnings, and phenomena most clearly seen among the black population. For white men, growth of top earnings outpaced all other groups' earnings growth, and there was a clear pattern of faster earnings growth as one moved higher up the income ladder. White women experienced the opposite pattern, with median and 90$^{\text{th}}$ percentile earnings outpacing top earnings. It appears that in the lead up to the .com bubble bursting, white men's top earnings were leaving the rest of the population behind, while white women in the middle and upper-middle of the income distribution were catching up to their top peers. Within white households, this may well have resulted in the same hollowing out of the (upper) middle seen in the growth patterns of black earnings. Despite these variations among groups, it is important to emphasize the relatively robust earnings growth across all groups and income levels.

During and in the aftermath of the burst of the .com bubble, all real earnings growth effectively halted. Top white male income earners saw some of the biggest declines in earnings over the period from 2001 to 2004, though the losses were small compared to the gains made previously. The only group that saw a bigger decline in earnings were black women. By definition, the rapid growth in 99$^{\text{th}}$ percentile earnings effects a very small portion of the population, whereas the decline in white mens' and black women's median incomes is necessarily felt very broadly. A final point to make is that top female income earners - white and black - saw little change

[20]The literatures on how much of these differentials are explained by differences in educational attainment and to what extent that excuses their existence are well-developed, and will not be discussed here.

[21]The CPI was used to convert nominal earnings for each year into base-year dollars before calculating % changes.

in the trajectory of their earnings from 1995 through 2004, suggesting that they continued to make incremental gains relative to the majority of their white male peers.

	Percent Changes in Earnings		
	Median	90^{th} Percentile	99^{th} Percentile
1995 - 2000	2.30	2.15	3.22
White Men	2.06	2.56	4.31
White Women	3.15	2.43	2.11
Black Men	2.69	2.00	2.59
Black Women	2.67	1.96	2.65
2001 - 2004	-0.03	0.10	-0.05
White Men	-0.39	-0.15	-0.40
White Women	0.05	-0.04	2.22
Black Men	0.40	0.55	0.40
Black Women	-0.75	0.87	2.47

Table 4: *Annualized growth rates in median, 90^{th}, and 99^{th} percentile earnings leading up to, during, and after the burst of the .com bubble.*

Table 5 shows what happened leading up to the bursting of the housing bubble that triggered the Great Recession. During the expansion from 2004 to 2007, the qualitative patterns of the expansion at the end of the 1990s are broadly replicated although the actual growth rates are much smaller. A closer look, reveals that the lowest growth rates are accorded to the majority of white income earners. Only top white male earners' earnings showed appreciable growth, while white men in the middle and upper middle of the income distribution saw almost no earnings growth. White women across the income distribution saw their earnings grow modestly at best compared to other groups, with earners in the middle still doing the best of the group.

It is unclear what was going on economically that produced such different patterns for the majority of white and black income earners, but during the 2004 to 2007 expansion, median black income and top black incomes made decent gains. Even 90^{th} percentile earnings saw decent growth in a period of generally anemic increases in incomes, though they grew less than incomes in the middle or the top. Much like the pattern seen in white women's incomes, when the Great Recession hit, median black incomes suffered the most (the differences between black women and men is likely partially explained by differences in unemployment). Top incomes among black income earners continued to grow from 2008 to 2010, which is consistent with the observation of increasing inequality by both measures (Gini and $I[2]$) and within each group seen in figure 3

During and in the aftermath of the Great Recession, the losses in terms of declines in earnings where shared broadly across the population. Median earnings for all but black men declined, and this does not account for the large increase in unemployment. In fact, it is probable that the only reason black mens' earnings appeared to grow is that black unemployment soared to a staggering 16.2% in 2010. This highlights that the changes in earnings reported here only apply to respondents who continued to be identified as being in the civilian workforce and reported non-zero earnings. Considering the declines in median real earnings in combinations with the increases in unemployment, which is known to disproportionally fall on individuals in the bottom to middle of the income distribution, the losses associated with the Great Recession were clearly felt very broadly. Top white male income earners did share in those losses to an extent (and compared to their black and / or female peers), although both the results presented here and

	Percent Changes in Earnings		
	Median	90^{th} Percentile	99^{th} Percentile
2004 - 2007	0.71	0.34	1.26
White Men	0.18	0.09	1.55
White Women	0.86	0.72	0.19
Black Men	1.22	0.79	1.68
Black Women	1.66	0.96	1.79
2008 - 2010	-0.21	0.49	0.57
White Men	-0.91	-0.40	-1.14
White Women	-2.41	-0.11	3.58
Black Men	0.95	2.16	3.25
Black Women	-0.10	1.30	1.69

Table 5: *Annualized growth rates in median, 90^{th}, and 99^{th} percentile earnings leading up to, during, and after the Great Recession.*

elsewhere suggest that they did not wipe out the gains made previously. (Figure 4 illustrates this most concisely.)

On the one hand, declining earnings and increases unemployment were shared experiences across race and gender groups in the middle of the income distribution. On the other hand, top white male income earners were the only top income earners to experience much of a relative decline in their earnings. It is possible that this sustains the narrative of the pain they share with the masses (versus *other* high income earners, who appear to be weathering the storm better). This narrative, of course, rests on the convenient amnesia about the disproportionate gains made by this group previously. A quick review of the threshold earnings for 2010 reported in table 3 shows how much ground other groups have to make up to catch up with white men in term of earnings.

The group that experienced the biggest decline in earnings as a result of the Great Recession are white women in the middle of the income distribution. This is likely at least partially explained by reductions in public sector employment due to state and local level budget cuts over this period. This group also experienced peak unemployment somewhat later than their male counterparts for the same reason.

5.3 Hints of Political Consequences of Inequality

A superficial look at changes in attitudes using data from the General Social Science Survey (GSS) suggests that changes in the distribution of earnings (and income more broadly) have some effect on attitudes. Unfortunately the sample sizes are too small to lead to statistically sound conclusions, but they provide at least anecdotal evidence. In total, there were 18,111 observations spread across surveys in 1998, 2000, 2002, 2004, 2006, 2008, and 2010 (the breakdown of number of observations by race and gender is given in table 6. The weights provided as part of the GSS to correct for survey design and oversampling of black respondents were used in the calculation of all statistics reported in this section.

On the question whether respondents were satisfied with their financial situation, the proportion of lower or working class[22] white men who answered “not at all” increased relatively

[22] Class identification was provided by the respondents and is correlated with income (statistically significant), but contains fewer missing values. More than 70% of respondents who choose lower or working class to describe they

	White	Black
Male	6,921	1,065
Female	8,348	1,777

Table 6: *Demographic breakdown of respondents to the GSS questions regarding their financial situation and the importance of improving conditions for the black population in the US.*

steadily from 2000 to 2010 from 27.6% to 45%.[23] Similarly, there was a jump in dissatisfied lower and working class white women from the early 2000s to 2010 (33% to 43.6% respectively). Middle and upper class white respondents (male and female) saw no real change in the proportion of respondents who were dissatisfied with their financial situation, which hovered around 11.8% and 12.8% respectively for the period.

The marked difference in dissatisfaction between classes among white respondents was much smaller among black respondents. Of those who identified as lower or working class, around 45% reported not all being satisfied with their financial situation. Of middle and upper class black respondents, around 30% reported not being satisfied. These proportions showed no notable - or statistically significant - change across the period considered. (Differences between black men and women were not considered to preserve a reasonable number of observations in each cell.)

Unfortunately, the closest proxy to testing whether the different experiences of inequality directly translate into changing attitudes about race is the question whether spending on improving the conditions of the black population was too little, just right, or too much showed little change in attitudes among most of the groups considered from 2000 to 2010. The proportion of responses is fairly predictable along racial lines. On average, roughly a fifth of the white population believes too little is being done and a fifth believes too much is being done. Among the black population, roughly three quarters believe too little is being spent and less than 5% believe too much is being spent. Only among middle and upper class white male respondents was there a consistent and notable decline in the proportion of "too little" responses from 28.8% in 2000 to 21.6% in 2010, but even this change does not appear to be statistically significant. This is however the core population attracted to both the Tea Party and Occupy movements, so perhaps this presents some tacit evidence that the changes in inequality have political consequences.

It is also quite possible that there are long lags before the changes in the earnings distribution manifest changes in attitudes and finally political action. Such changes are difficult to asses from the perspective of the individual income earner, especially if s/he does not have frequent contact with individuals outside of his or her socioeconomic class. Norton and Ariely (2011)'s survey results show that while Americans have a sense of high and - by consensus across political lines among their respondents - unacceptable levels of wealth inequality in the US, they substantially underestimate just how unequal wealth is distributed. The same likely holds for both income and earnings, though both are less unequal in their distribution than wealth and it seems plausible that income and especially earnings inequality are easier to assess than wealth inequality. Popular reporting of growing inequality by the press saw a sharp increase only in the second half of 2011[24] while the trends in inequality reported here had clearly been an underlying phenomena for a full decade and been reported by researchers at least half a decade ago. It is perhaps not surprising then that the GSS survey data does not provide solid support for the

social position reported a family income of less than $50,000.

[23]The proportion of lower and working class white men not satisfied with their financial situation in 2008 was 34.7%, which represents a statistically significant increase from 2000.

[24]As indicated by Google Trends (http://www.google.com/trends/?q=inequality) accessed July 12, 2012.

conjectures about political attitudes presented in this paper, though it notably also does not appear to refute them.

6 Conclusion

The results presented in this study highlight a couple of points about the nature of the change in inequality experienced over the past decade and a half. For one, the pattern of increases in income inequality driven by an increasing share of income going to top income earners while the remainder is distributed among the rest much as before appears to hold true for earnings. In fact, the changes in earnings inequality consistent with this pattern appear much more stable than the changes in income inequality more broadly, because more volatile income streams - like dividends and capital gains - are excluded. These patterns lend further support to the relevance of Thomas Piketty and Emmanuel Saez's work; in fact, looking at earnings emphasizes how some of these patterns have become institutionalized in the form of contractual compensation. The fact that these patterns hold for earnings, which make up only a small portion of income at the top of the distribution, suggests that the observations made by Piketty and Saez (2006) are not just the results of differences in income composition. It appears that those at the top of the income distribution - the top 1% and above - are gaining relatively to the bottom from almost all[25] income sources.

The impact and meaning of rising inequality remains debatable. If, however, inequality in earnings is the result of functioning market processes and therefore relatable to changes in the productivity of different workers, it remains to be shown that only the top 1% of income earners made gains in productivity over the past decade. If instead labor contracts at the top of the income scale are more likely to reflect market imperfections, then the gains in the share of earnings accruing to them are better categorized as successful rent-seeking. The results provided in this paper do not help differentiate which story offers the better explanation, although the finding are certainly consistent with the hypothesis of lavish CEO compensation being a driving force summarized by Gordon and Dew-Becker (2007). The fact that the evolution of the earnings distribution exhibits the qualitatively similar pattern as the distribution of income suggests that labor contracts are also an institutional mechanism for redistributing from the bottom to the top.

The particular changes in the earnings distribution - which match changes in the distribution of income more broadly - are not done justice by the Gini, which is more sensitive to changes around the mode of a distribution and less sensitive to changes in the upper tail. By contrasting the evolution of the Gini from 1995 to 2010 with the evolution of a measure that is more explicitly weighted to emphasize changes in the upper tail of a distribution, this paper has illustrated that the increased share of income going to the top 1% (and top 0.1%) is not inconsistent with a ostensibly stagnant Gini coefficient. It also highlights that sticking with a chosen measure of inequality for "normative" reasons may preclude researchers from realizing important qualitative changes in how incomes are distributed.

Furthermore, the growth in earnings inequality has primarily benefitted top white male income earners. As a result, white income earners - and especially white men - have experienced a significant increase in within-group inequality, while for black income earners increasing inequality has continued to take the form of increases in between-group inequality. Since the late 1990s, white women have experienced the most complicated evolution in their earnings, because in some sense median earners made progress in catching up to top earners within that group, while at the same time the distance between them and top white male income earners grew. Since 2004, median and 90^{th} percentile white men's earnings have stagnated or fallen, while the increasing within-group inequality experienced by them appears to mirror the level and trend in

[25] Probably social transfers as a source of income is the only notable exception.

inequality broadly. This experience likely fuels much of the political dissatisfaction voiced both on the right and the left by this group. (While white men may have some material reason to be more agitated by rising within-group inequality, it is important to remember that this group remains far ahead of the other groups studied in this paper, as the threshold incomes listed in table 3 illustrate.) Black income earners continued to experience inequality predominantly as a between-group phenomena. Without down-playing their political agitation on the issue, it likely has not changed as acutely in part because their experience fits a long-established narrative.

All of these results suggest that much more is to be discovered about the changes in inequality experienced among and between different groups, and what the effects of these changes are.

References

Atkinson, A. B. (1970). On the measurement of Inequality. *Journal of Economic Theory 2*, 244–263.

Atkinson, A. B., T. Piketty, and E. Saez (2011). Top Incomes in the Long Run of History. *Journal of Economic Literature 49*, 3–71.

Borzadaran, G. R. M. and Z. Behdani (2009). Maximum Entropy and the Entropy of Mixing for Income Distributions. *Journal of Income Distribution 18*, 179–186.

Burkhauser, R. V., J. S. Butler, S. Feng, and A. J. Houtenville (2004). Long term trends in earnings inequality: what the CPS can tell us. *Economics Letters 82*, 295–299.

Burkhauser, R. V., S. Feng, S. Jenkins, and J. Larrymore (2009). Recent trends in top income shares in the usa: Reconciling estimates from the march cps and irs tax return data. CES Working Paper.

Burkhauser, R. V., S. Feng, S. Jenkins, and J. Larrymore (2011). Estimating trends in u.s. income inequality using the current population survey: the importance of controlling for censoring. *Journal of Economic Inequality 9*, 393–415.

Burkhauser, R. V., J. Larrymore, and K. I. Simon (2011). A 'second opinion' on the economic health of the american middle class. NBER Working Papers: 17164.

Dagum, C. (1977). A New Model of Income Distribution: Specification and Estimation. *Economie Appliquée 30*, 413–367.

Feng, S., R. V. Burkhauser, and J. S. Butler (2006). Levels and Long-Term Trends in Earnings Inequality: Overcoming Current Population Survey Censoring Problems Using the GB2 Distribution. *Journal of Business & Economic Statistics 24*, 57–62.

Gordon, R. J. and I. Dew-Becker (2005). Where did the productivity growth go? inflation dynamics and the distribution of income. NBER Working Paper 11842.

Gordon, R. J. and I. Dew-Becker (2007). Selected issues in the rise of income inequality. *Brookings Paper on Economic Activity 2007*, 169–190.

Jenkins, S., R. V. Burkhauser, S. Feng, and J. Larrymore (2011). Measuring inequality using censored data: A multiple-imputation approach to estimation and inference images. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 174*, 63–81.

Jenkins, S. P. (2009). Distributionally-Sensitive Inequality Indices and the GB2 Income Distribution. *The Review of Income and Wealth 22*, 392–398.

Kirman, A. P. (1992). Whom or what does the representative individual represent? *Journal of Economic Perspectives 6*, 117–136.

Kleiber, C. (1996). Dagum vs. Singh-Maddala Income Distributions. *Economics Letters 53*, 265–268.

Kleiber, C. (2008). *A Guide to the Dagum Distribution*, Chapter 6, pp. 392–398. Springer.

Kleiber, C. and S. Kotz (2003). *Statistical Size Distributions in Economics and the Actuarial Sciences*. John Wiley.

McDonald, J. B. (1984). Some Generalized Size Functions for the Size Distribution of Income. *Econometrica 52*, 647–663.

Monnat, S. M., L. E. Raffalovich, and H.-s. Tsao (2012). Trends in the Family Income Distribution by Race/Ethnicity and Income Source, 1988-2009. *Population Review 51*, 85–115.

Norton, M. I. and D. Ariely (2011). Building a Better America - One Wealth Quintile at a Time. *Perspectives on Psychological Science 6*, 9–12.

Parker, S. C. (1999). The Generalized Beta as a Model for the Distribution of Earnings. *Economics Letters 62*, 197–200.

Piketty, T. and E. Saez (2003). Income Inequality in the United States, 1913 - 1998. *Quarterly Journal of Economics 118*, 1–39.

Piketty, T. and E. Saez (2006). The Evolution of Top Incomes: A Historical and International Perspective. *American Economic Review 96*, 200–205.

Rogerson, R., R. Shimer, and R. Wright (2005). Search-Theoretic Models of the Labor Market: A Survey. *Journal of Economic Literature 43*, 959–988.

Smeeding, T. M. and J. P. Thompson (2011). Recent Trends in Income Inequality: Labor, Wealth and More Complete Measures of Income. *Research in Labor Economics 32*, 1–50.

Ziliak, S. T. and D. McCloskey (2004). Size matters: the standard error of regressions in the American Economic Review. *Journal of Socio-Economics 33*, 527–546.

Appendix

Comparing fit: GB2 vs. Dagum

The GB2 - the most common alternative to the Dagum as synthetic distribution for fitting the income or earnings data - was fit to earnings in 1996 and 2006 as a "spot check". Figure 5 shows that the *pdf*s of the fitted distribution are roughly similar. However, the GB2 does seem to put even more weight in the tail of the distribution, suggesting that the results presented above under-estimate the divergence between median and high earnings (this is illustrated by the inset in the figure, which shows the *complementary cumulative distribution function* or *ccdf* on a linear-log plot).

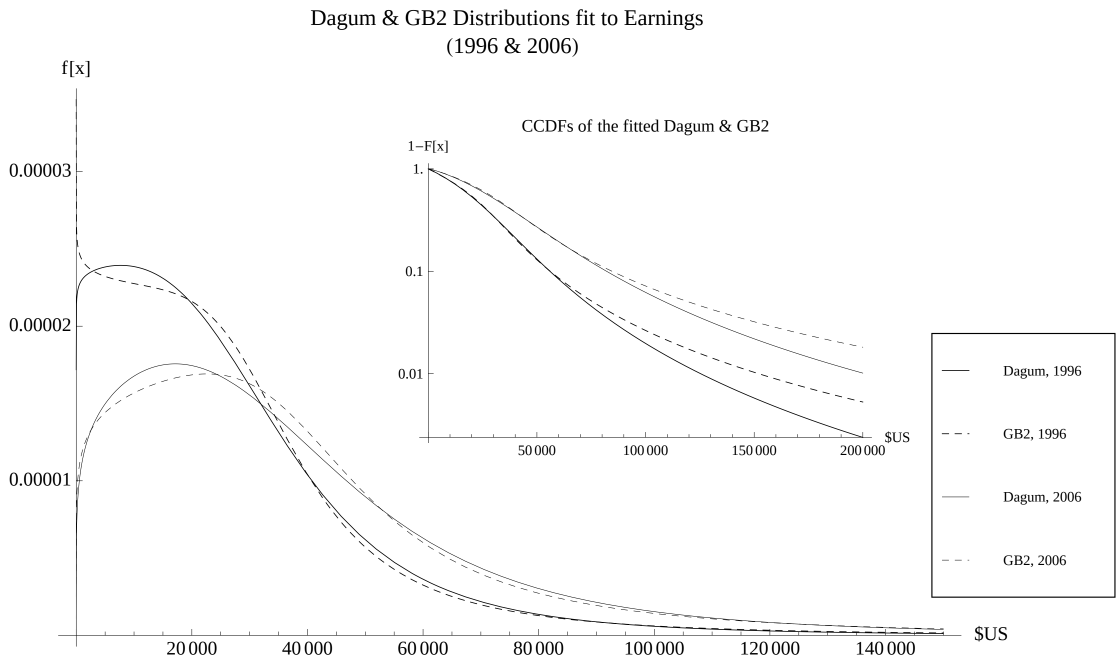

Figure 5: *The pdfs of the GB2 and Dagum distributions fit to earnings in 1996 and 2009.*

Raw Estimates

The following tables contain the raw MLE parameter estimates and associated standard errors as well as the inequality indexes estimated from the fitted synthetic distribution and their associated standard errors. The first table lists the sample sizes.

Year	All	White & Male	Black & Male	White & Female	Black & Female
1996	60,836	28,328	3,147	23,398	3,440
1997	62,230	28,839	3,185	23,816	3,633
1998	62,433	28,680	3,251	23,806	3,784
1999	62,896	28,702	3,314	24,099	3,841
2000	64,313	29,184	3,465	24,623	4,023
2001	61,967	27,934	3,396	23,720	3,876
2002	102,058	46,145	5,337	38,782	6,125
2003	100,627	45,183	5,162	38,060	6,088
2004	98,833	44,412	5,078	37,099	6,088
2005	97,381	43,642	5,087	36,417	5,969
2006	97,040	43,702	5,161	35,948	5,930
2007	97,007	43,503	5,152	35,953	5,963
2008	97,075	43,243	5,185	36,095	5,991
2009	96,860	42,977	5,135	36,210	6,037
2010	95,030	42,013	4,991	35,772	5,848
2011	91,864	40,625	4,862	34,464	5,684

Table 7: *Sample sizes across demographic groups. The year listed above is the year associated with the actual survey. In all other tables and graphs, the year listed is the year for which income was reported.*

Year	$\hat{p}$	SE[$\hat{p}$]	$\hat{a}$	SE[$\hat{a}$]	$\hat{b}$	SE$\left[\hat{b}\right]$
1995	0.296	0.0035	3.24	0.028	41,320	247
1996	0.330	0.0040	3.10	0.026	40,830	254
1997	0.330	0.0039	3.12	0.026	42,590	261
1998	0.333	0.0039	3.14	0.026	44,640	268
1999	0.347	0.0041	3.02	0.025	45,790	282
2000	0.363	0.0044	3.00	0.025	46,960	297
2001	0.388	0.0037	2.88	0.019	47,090	242
2002	0.393	0.0036	2.86	0.018	47,710	247
2003	0.379	0.0035	2.89	0.018	49,930	258
2004	0.377	0.0035	2.90	0.018	51,110	265
2005	0.390	0.0037	2.84	0.018	52,040	276
2006	0.429	0.0042	2.75	0.018	51,470	289
2007	0.434	0.0043	2.79	0.018	52,690	296
2008	0.435	0.0044	2.69	0.018	53,530	314
2009	0.423	0.0044	2.69	0.018	54,350	326
2010	0.417	0.0043	2.69	0.018	55,780	337

Table 8: *MLE parameter estimates for the whole sample.*

Year	$\hat{p}$	SE[$\hat{p}$]	$\hat{a}$	SE[$\hat{a}$]	$\hat{b}$	SE$\left[\hat{b}\right]$
1995	0.320	0.0055	3.27	0.040	49,240	412
1996	0.357	0.0062	3.10	0.037	48,510	426
1997	0.356	0.0062	3.13	0.038	50,830	442
1998	0.364	0.0062	3.14	0.038	53,110	454
1999	0.373	0.0065	3.00	0.036	55,050	487
2000	0.403	0.0074	2.91	0.037	55,520	530
2001	0.432	0.0062	2.79	0.027	55,040	427
2002	0.437	0.0062	2.76	0.025	55,750	443
2003	0.422	0.0060	2.79	0.026	57,610	457
2004	0.409	0.0058	2.82	0.027	60,060	468
2005	0.433	0.0062	2.74	0.026	60,030	483
2006	0.476	0.0071	2.68	0.026	58,950	502
2007	0.492	0.0076	2.67	0.026	59,590	528
2008	0.497	0.0079	2.52	0.026	60,820	573
2009	0.446	0.0071	2.61	0.028	63,590	581
2010	0.449	0.0071	2.57	0.026	64,600	608

Table 9: *MLE parameter estimates for white men.*

Year	$\hat{p}$	SE[$\hat{p}$]	$\hat{a}$	SE[$\hat{a}$]	$\hat{b}$	SE$\left[\hat{b}\right]$
1995	0.286	0.0143	3.61	0.129	36,150	835
1996	0.240	0.0125	4.10	0.159	40,830	866
1997	0.261	0.0131	4.01	0.145	41,510	880
1998	0.254	0.0123	4.14	0.146	42,800	854
1999	0.297	0.0144	3.69	0.127	44,340	973
2000	0.288	0.0140	3.76	0.130	46,200	1,000
2001	0.268	0.0103	3.89	0.109	48,150	791
2002	0.345	0.0135	3.40	0.090	44,340	843
2003	0.298	0.0118	3.71	0.104	48,860	875
2004	0.344	0.0130	3.49	0.089	46,130	835
2005	0.319	0.0127	3.51	0.097	49,030	923
2006	0.369	0.0145	3.36	0.087	48,290	928
2007	0.324	0.0133	3.59	0.101	53,160	1,008
2008	0.302	0.0121	3.55	0.101	54,980	1,026
2009	0.333	0.0139	3.23	0.094	52,920	1,107
2010	0.340	0.0142	3.21	0.090	51,800	1,112

Table 10: *MLE parameter estimates for black men.*

Year	$\hat{p}$	SE[$\hat{p}$]	$\hat{a}$	SE[$\hat{a}$]	$\hat{b}$	SE$\left[\hat{b}\right]$
1995	0.233	0.0045	3.85	0.056	35,220	301
1996	0.275	0.0054	3.58	0.050	34,330	312
1997	0.271	0.0053	3.61	0.051	36,060	323
1998	0.276	0.0053	3.63	0.050	37,650	329
1999	0.281	0.0053	3.56	0.048	38,890	344
2000	0.279	0.0053	3.64	0.050	40,980	357
2001	0.315	0.0048	3.36	0.036	41,090	303
2002	0.307	0.0046	3.43	0.036	42,500	306
2003	0.303	0.0045	3.40	0.036	44,400	325
2004	0.306	0.0048	3.39	0.037	45,120	343
2005	0.315	0.0049	3.33	0.036	46,200	353
2006	0.354	0.0057	3.15	0.034	45,990	380
2007	0.349	0.0055	3.26	0.035	47,830	379
2008	0.344	0.0055	3.21	0.035	48,690	398
2009	0.357	0.0059	3.09	0.034	48,890	421
2010	0.345	0.0057	3.12	0.035	51,020	441

Table 11: *MLE parameter estimates for white women.*

Year	$\hat{p}$	SE$[\hat{p}]$	$\hat{a}$	SE$[\hat{a}]$	$\hat{b}$	SE$\left[\hat{b}\right]$
1995	0.239	0.0126	4.23	0.164	31,100	669
1996	0.227	0.0111	4.30	0.156	32,630	637
1997	0.262	0.0127	4.01	0.140	32,280	664
1998	0.264	0.0133	3.96	0.143	34,180	742
1999	0.271	0.0124	3.91	0.129	35,570	705
2000	0.300	0.0138	3.95	0.126	35,700	705
2001	0.281	0.0104	4.06	0.107	38,120	584
2002	0.310	0.0113	3.71	0.094	38,350	633
2003	0.293	0.0105	3.74	0.095	39,440	637
2004	0.287	0.0102	3.79	0.095	40,840	653
2005	0.309	0.0117	3.60	0.094	42,430	752
2006	0.328	0.0121	3.56	0.090	42,220	732
2007	0.338	0.0126	3.62	0.091	43,270	745
2008	0.332	0.0125	3.59	0.092	43,850	768
2009	0.334	0.0130	3.50	0.092	43,910	813
2010	0.312	0.0118	3.56	0.093	46,460	830

Table 12: *MLE parameter estimates for black women.*

Inequality Indexes

Below are the specific formulas for the various inequality indexes used in this study in terms of the parameters of the Dagum distribution - p, a, b - as defined in (3).

$$G = \frac{\Gamma[p]\,\Gamma[2p+\frac{1}{a}]}{\Gamma[2p]\,\Gamma[p+\frac{1}{a}]} - 1$$

where $\Gamma[\cdot]$ is the Gamma function.

Given below are the explicit formulas for calculating the generalized entropy indexes used in this study. As stated in a footnote above, $I[0]$ and $I[1]$ are derived by taking the appropriate limit of (1) and applying l'Hopital's rule (Jenkins, 2009). Euler's constant appears below as γ and $\psi[\cdot]$ is the digamma function.

$$I[0] = \frac{1}{a}\left(\gamma - \psi[p]\right) - \ln\Gamma[p] + \ln\Gamma[p+\tfrac{1}{a}] + \ln\Gamma[1-\tfrac{1}{a}]$$

$$I[1] = \frac{1}{a}\left(\psi[p+\tfrac{1}{a}] - \psi[1-\tfrac{1}{a}]\right) + \ln\Gamma[p] - \ln\Gamma[p+\tfrac{1}{a}] - \ln\Gamma[1-\tfrac{1}{a}]$$

$$I[2] = -\frac{1}{2} + \frac{\Gamma[p]\Gamma[p+\frac{2}{a}]\Gamma[1-\frac{2}{a}]}{2\,\Gamma^2[p+\frac{1}{a}]\Gamma^2[1-\frac{1}{a}]}$$

Statistical Significance of Parameter Changes

The standard errors of the inequality indexes were estimate using the "delta method" approximation, which uses the gradient of the inequality index with respect to the parameters (evaluated at the ML point estimates) to capture the sensitivity of the index to variability in the estimated parameters. Multiplying the gradient by the estimated variance-covariance matrix produces a standard error estimate. For example, letting $G[\theta]$ be an expression for the Gini coefficient given parameter θ, the standard error is estimated as:

$$SE[G] = \sqrt{\triangledown G[\hat{\theta}]^T \cdot \hat{\Omega} \cdot \triangledown G[\hat{\theta}]}$$

where $\hat{\Omega}$ is the estimated variance-covariance matrix evaluated at $\hat{\theta}$.

Year	Gini	SE[Gini]	$I[2]$	SE[$I[2]$]
1995	0.4535	0.0014	0.4900	0.0080
1996	0.4510	0.0015	0.5158	0.0094
1997	0.4489	0.0015	0.5070	0.0091
1998	0.4451	0.0015	0.4932	0.0086
1999	0.4523	0.0015	0.5415	0.0105
2000	0.4477	0.0016	0.5365	0.0108
2001	0.4520	0.0013	0.5888	0.0105
2002	0.4535	0.0012	0.6037	0.0104
2003	0.4546	0.0012	0.5915	0.0100
2004	0.4538	0.0013	0.5851	0.0099
2005	0.4567	0.0013	0.6186	0.0112
2006	0.4543	0.0014	0.6576	0.0134
2007	0.4472	0.0013	0.6173	0.0122
2008	0.4620	0.0014	0.7221	0.0165
2009	0.4662	0.0015	0.7345	0.0172
2010	0.4680	0.0014	0.7384	0.0168

Table 13: *Estimated Gini & $I[2]$, and their standard errors, for all respondents.*

Year	Gini	SE[Gini]	$I[2]$	SE[$I[2]$]
1995	0.4372	0.0021	0.4502	0.0106
1996	0.4381	0.0022	0.4876	0.0131
1997	0.4359	0.0022	0.4770	0.0127
1998	0.4307	0.0022	0.4634	0.0122
1999	0.4434	0.0023	0.5272	0.0158
2000	0.4424	0.0025	0.5543	0.0188
2001	0.4482	0.0021	0.6207	0.0187
2002	0.4515	0.0020	0.6481	0.0191
2003	0.4513	0.0020	0.6283	0.0181
2004	0.4518	0.0020	0.6139	0.0171
2005	0.4554	0.0021	0.6692	0.0206
2006	0.4502	0.0022	0.6925	0.0235
2007	0.4477	0.0022	0.6951	0.0246
2008	0.4694	0.0025	0.9071	0.0416
2009	0.4696	0.0024	0.8050	0.0326
2010	0.4745	0.0024	0.8601	0.0357

Table 14: *Estimated Gini & $I[2]$, and their standard errors, for white men.*

Year	Gini	SE[Gini]	$I[2]$	SE[$I[2]$]
1995	0.4232	0.0056	0.3792	0.0209
1996	0.4138	0.0052	0.3282	0.0149
1997	0.4058	0.0051	0.3209	0.0146
1998	0.4006	0.0049	0.3060	0.0130
1999	0.4097	0.0052	0.3494	0.0180
2000	0.4092	0.0052	0.3429	0.0171
2001	0.4107	0.0041	0.3360	0.0126
2002	0.4116	0.0045	0.3829	0.0185
2003	0.4073	0.0043	0.3438	0.0143
2004	0.4039	0.0044	0.3592	0.0161
2005	0.4145	0.0044	0.3753	0.0174
2006	0.4060	0.0045	0.3796	0.0189
2007	0.4040	0.0044	0.3488	0.0158
2008	0.4188	0.0044	0.3777	0.0172
2009	0.4350	0.0050	0.4533	0.0262
2010	0.4339	0.0049	0.4557	0.0261

Table 15: *Estimated Gini & $I[2]$, and their standard errors, for black men.*

Year	Gini	SE[Gini]	$I[2]$	SE[$I[2]$]
1995	0.4395	0.0020	0.3876	0.0071
1996	0.4328	0.0021	0.3997	0.0085
1997	0.4324	0.0021	0.3960	0.0083
1998	0.4279	0.0020	0.3862	0.0079
1999	0.4310	0.0021	0.3986	0.0084
2000	0.4244	0.0021	0.3785	0.0078
2001	0.4302	0.0017	0.4222	0.0081
2002	0.4280	0.0017	0.4090	0.0073
2003	0.4331	0.0017	0.4225	0.0078
2004	0.4324	0.0017	0.4222	0.0080
2005	0.4333	0.0018	0.4328	0.0085
2006	0.4344	0.0019	0.4696	0.0108
2007	0.4249	0.0018	0.4295	0.0091
2008	0.4323	0.0019	0.4531	0.0101
2009	0.4395	0.0020	0.4936	0.0122
2010	0.4413	0.0019	0.4892	0.0117

Table 16: *Estimated Gini & $I[2]$, and their standard errors, for white women.*

Year	Gini	SE[Gini]	$I[2]$	SE[$I[2]$]
1995	0.4056	0.0048	0.3094	0.013
1996	0.4100	0.0047	0.3134	0.0122
1997	0.4054	0.0047	0.3203	0.0138
1998	0.4079	0.0048	0.3269	0.0147
1999	0.4068	0.0047	0.3279	0.0142
2000	0.3869	0.0046	0.2958	0.0128
2001	0.3894	0.0037	0.2934	0.0098
2002	0.4004	0.0039	0.3323	0.0126
2003	0.4074	0.0039	0.3412	0.0127
2004	0.4075	0.0038	0.3380	0.0122
2005	0.4112	0.0040	0.3602	0.0148
2006	0.4051	0.0040	0.3538	0.0147
2007	0.3952	0.0039	0.3321	0.0135
2008	0.4006	0.0040	0.3438	0.0142
2009	0.4074	0.0042	0.3637	0.0162
2010	0.4130	0.0041	0.3667	0.0155

Table 17: *Estimated Gini & $I[2]$, and their standard errors, for black women.*

The standard error of a change in a particular estimate - e.g. the change in the Gini coefficient - was estimate using (5), which assumes independence between the estimates being compared. While technically probably not a reasonable assumption, the standard error calculated this way should provide a conservative benchmark. In almost all relevant cases where this approach in used in the paper, the differences are many time (>> 3) greater than the standard error estimate, so that even making no assumptions about the shape of the sampling distributions (and using Chebychev's Theorem to find critical values for given significance levels[26]), the null hypothesis that the difference is zero can be rejected.

$$SE[\Delta G] \approx \sqrt{SE[G_{y_0}]^2 + SE[G_{y_1}]^2} \tag{5}$$

The estimated change in the Gini and $I[2]$ are shown in table 18. The standard errors are estimated assuming that the observation of the Gini (and $I[2]$) in 2010 can be considered independent of the observation in 1995, i.e. that their lagged dependence has died out over this 15 year period. What table 18 illustrates is that while the Gini saw statistically significant increase under any reasonable distributional assumption for the whole sample from 1995 to 2010 as well as for white men, the increase was small ($\Delta \sim 0.015$ for the whole sample). (It might even be deemed economically insignificant by some.) For black respondents and white women, there was no statistically significant increase in inequality according to the Gini coefficient. If instead one looks at $I[2]$, there is a statistically significant increase in inequality across all groups under common assumptions, and for white respondents the increase would be statistically significant under very general assumptions.[27]

	Gini Coefficient			**I[2]**		
Group	Δ	SE[Δ]	t	Δ	SE[Δ]	t
All	0.015	0.002	7.18	0.248	0.019	13.31
White Men	0.037	0.003	11.85	0.410	0.037	10.99
Black Men	0.011	0.007	1.44	0.077	0.033	2.29
White Women	0.002	0.003	0.66	0.102	0.014	7.39
Black Women	0.007	0.006	1.17	0.057	0.020	2.83

Table 18: *Changes in inequality measures between 1995 and 2010, and their statistical significance. The t-statistic listed corresponds $H_0 : \Delta = 0$.*

[26] According to Chebyshev's inequality, for any distribution with finite variance, 90% of the distribution lies within 3.2 standard deviations of the mean; and 95% lies within 4.5 standard deviations.

[27] The differences in assumptions boils down to whether the sample sizes are sufficient to guarantee asymptotic normality, or whether one assumes only finite variances of the sampling distribution and applies Chebyshev's inequality.

Lorenz Dominance

It is also possible to consider the stochastic ordering of the Lorenz curves of two distributions to help gain some insight about what particular parameter constellations imply about changes in inequality. If the Lorenz curve of the distribution of Y_A does not intersect that of Y_B lies everywhere below the second it, then $Y_A \geq_L Y_B$, indicating that the distribution of Y_A is unequivocally more unequal.

Given a well-defined synthetic distribution, the Lorenz ordering can be easily assessed from the estimated parameters. (If the Lorenz curves intersect, then two distributions cannot be ranked using this simple ordering. While other stochastic orderings are available for such cases - though most of them are not scale-free - they are thus not considered here.) According to Kleiber (1996), a Dagum distribution A has a Lorenz curve that is everywhere below the Lorenz curve of another Dagum distribution B - i.e. exhibits greater inequality including a larger Gini coefficient - if $a_A \leq a_B$ and $a_A p_A \leq a_B p_B$. These constitute the necessary and sufficient conditions for Lorenz dominance[28] of A over B and imply greater inequality in distribution A.

The estimates presented in this paper suggest that the earnings distributions in 2008, 2009, and 2010 Lorenz dominate the earnings distributions prior to 2006 (DOUBLE CHECK). In other words, a broad increase in earnings inequality among those who are reporting earnings is associated with the unfolding of the Great Recession, which is also seen in figures 2 and 3. For no other years is there a clear Lorenz ranking of the income distributions, consistent with the observation that while parts of the distribution contributed to greater inequality (the upper tail), other parts saw modest declines in inequality (the lower portion of the distribution as captured by $I[0]$).

[28]It should be noted that this is consistent with the convention in economics and opposite to the convention found in statistics according to Kleiber and Kotz (2003)