International welfare comparisons and nonparametric testing of multivariate stochastic dominance

Brian McCaig and Adonis Yatchew[1]

October 1, 2005

Preliminary draft. Please do not quote.

Abstract

Cross-country welfare comparisons often involve data on more than one measure. For example, one might consider income, leisure, as well as indicators measuring education, health-related and environmental variables. Moreover, it is often the case that one country stochastically dominates in some dimensions while the other stochastically dominates in others. The purpose of this paper is to outline a class of statistical procedures that permit testing of a broad range of multi-dimensional stochastic dominance hypotheses and more generally, welfare hypotheses which rely upon multiple stochastic dominance conditions. We then apply the procedures to data on income and leisure from Germany, the U.K., and the U.S. For most population sub-groups examined, we find that no country stochastically dominates the others in both dimensions. Furthermore, while the U.S. dominates Germany and the U.K. with respect to income, in most periods Germany is dominant with respect to leisure.

[1] We are indebted to Gordon Anderson for numerous illuminating discussions on the subject matter of this paper.

I. Introduction

Cross-country welfare comparisons commonly use statistics. For example, one is often interested in comparing median or mean per-capita incomes, average hours worked, the proportion of the population living below the poverty line and so on. The statistical theory for testing hypotheses about one or more statistics is generally available. Individual statistics, however, capture only one feature or characteristic of a distribution. Often, there is interest in making point-wise comparisons of entire distributions to each other, for example using measures of stochastic dominance in the context of social welfare, inequality, and poverty. In the simplest case the distribution of income in country "*a*" stochastically dominates country "*b*" if for any income level "*x*" the proportion of the population with income at or below "*x*" is lower in country "*a*" than in country "*b*".

More generally, one may be interested in simultaneous comparisons along more than one dimension. For example, one might want to test whether one country stochastically dominates another with respect to several variables. Or, one might want to test whether one country dominates in some dimensions while another dominates in others. For example, while U.S. per capita GDP is substantially higher than in France, the French work fewer hours per week.[2]

Recently, several tests of stochastic dominance have been proposed in the literature. These tests can be grouped according to whether they test the distributions at a finite number of points or sub-regions of the support, or whether they test over the entire support of the distributions. In the former group are the Pearson goodness-of-fit type tests proposed by Anderson (1996) in the univariate case and Crawford (1999) in the bivariate case, as well as those proposed by Xu et al. (1995), Davidson and Duclos (2000) and Sahn et al. (2001). In the

[2] Such comparisons have filtered into the media and inform the debate on societal as well as individual choices. See e.g., Paul Krugman, "French Family Values", The New York Times, July 29, 2005.

latter group are procedures proposed by McFadden (1989), Klecan et al. (1991), Kaur et al. (1994), Barrett and Donald (2003), Linton et al. (2005), and Hall and Yatchew (2005). Maasoumi (2001) provides a survey of some of these stochastic dominance tests, while Tse and Zhang (2004) offer Monte Carlo results of the Kaur et al. (1994), Anderson (1996), and Davidson and Duclos (2000) tests.

The purpose of this paper is to outline a class of statistical procedures that permit testing of a broad range of multi-dimensional stochastic dominance hypotheses and more generally, hypotheses that rely upon multiple stochastic dominance conditions. Next, we conduct a small Monte Carlo study to examine the size and power properties of the test procedure. We then apply the procedures to data on income and leisure from Germany, the U.K. and the U.S. We find that for most population sub-groups no country stochastically dominates the others in both dimensions. Furthermore, while the U.S. stochastically dominates Germany and the U.K. with respect to income, in most periods Germany is stochastically dominant with respect to leisure.

The paper is organized as follows. Section II establishes notation and sets out the statistical procedures. These are an extension of tests found in Hall and Yatchew (2005). Section III describes the results of simulations and Section IV discusses empirical results.

II. Notation and Statistical Procedure

Let G_a and G_b denote two right-continuous *k*-dimensional cumulative distribution functions (CDFs). For convenience, assume that the support of the CDFs is Λ, the unit cube in $\mathbb{R}^k$.[3] We are interested in testing hypotheses of the form

$$H_o : G_a \succ_s G_b$$

[3] This rescaling of the data becomes important later when we introduce hypotheses involving more than one stochastic dominance condition.

where $\succ_s$ denotes stochastic dominance of order *s*. Let $D_a^1(\mathbf{z}) = G_a(\mathbf{z})$ and define

$$D_a^s(\mathbf{z}) = \int_0^z D_a^{s-1}(\mathbf{u})\, d\mathbf{u}$$

for integers $s \geq 2$. (An analogous definition applies to $D_b^s(\mathbf{z})$.) Adapting Davidson and Duclos (2000) to the *k*-dimensional case, we have $G_a \succ_s G_b$ iff $D_a^s(\mathbf{z}) \leq D_b^s(\mathbf{z})$ for all $\mathbf{z} \in \Lambda$. For $\boldsymbol{\lambda} \in \Lambda$, let $\psi(\boldsymbol{\lambda}) = \max\left\{D_a^s(\boldsymbol{\lambda}) - D_b^s(\boldsymbol{\lambda}), 0\right\}$. Then the null hypothesis is true iff $\psi(\boldsymbol{\lambda}) = 0$ for all $\boldsymbol{\lambda} \in \Lambda$. Define

$$T = \left\{\int_\Lambda \left[\psi^s(\boldsymbol{\lambda})\right]^2 d\,\boldsymbol{\lambda}\right\}^{1/2}. \qquad (2.1)$$

The objective is to estimate T and to test whether it is statistically different from zero. Let $\left(\mathbf{w}_{a1},...,\mathbf{w}_{an_a}\right)$ and $\left(\mathbf{w}_{b1},...,\mathbf{w}_{bn_b}\right)$ be independently and identically distributed observations from the two respective populations with corresponding empirical distribution functions $\hat{G}_a$ and $\hat{G}_b$. Our test statistic, say $\hat{T}$, is obtained by substituting numerical analogues of D_a^s and D_b^s, say $\widehat{D}_a^s$ and $\widehat{D}_b^s$ into T.

When one is examining k indicators of well-being one may want to form hypotheses based on subsets of the indicators. For example, one might want to test that the income distribution of country "*a*" dominates that of country "*b*", but the leisure-time distribution of country "*b*" dominates that of country "*a*". In this spirit, partition the *k*-dimensional vectors $\mathbf{w}_a$, $\mathbf{w}_b$ into sub-vectors of dimension k' and $(k-k')$, $0 < k < k'$ respectively; i.e., $\mathbf{w}_a = (\mathbf{x}_a, \mathbf{y}_a)$ and $\mathbf{w}_b = (\mathbf{x}_b, \mathbf{y}_b)$. A more general hypothesis is given by

$$H_o : G_{a_x} \succ_{s_x} G_{b_x} \quad and \quad G_{b_y} \succ_{s_y} G_{a_y}$$

where, we write $G_{a_x}(\mathbf{x}) = G_a(\mathbf{x},\mathbf{1})$ and the other distribution functions are defined similarly. We allow for the order of dominance to vary between the two subsets of indicators as denoted by s_x and s_y. In the bivariate case, for example, the above hypothesis asserts dominance with respect to marginal distributions without requiring dominance within the interior of the two-dimensional support. Let Λ_x be the unit cube in $\mathbb{R}^{k'}$ and let Λ_y be the unit cube in $\mathbb{R}^{(k-k')}$. For $\boldsymbol{\lambda}_x \in \Lambda_x$, let $\psi_x(\boldsymbol{\lambda}_x) = \max\left\{D_{a_x}^{s_x}(\boldsymbol{\lambda}_x) - D_{b_x}^{s_x}(\boldsymbol{\lambda}_x), \mathbf{0}\right\}$ and for $\boldsymbol{\lambda}_y \in \Lambda_y$, let $\psi_y(\boldsymbol{\lambda}_y) = \max\left\{D_{b_y}^{s_y}(\boldsymbol{\lambda}_y) - D_{a_y}^{s_y}(\boldsymbol{\lambda}_y), \mathbf{0}\right\}$. Define

$$T = \left\{ \int_{\Lambda_x} \left[\psi_x(\boldsymbol{\lambda}_x)\right]^2 d\boldsymbol{\lambda}_x \right\}^{1/2} + \left\{ \int_{\Lambda_y} \left[\psi_y(\boldsymbol{\lambda}_y)\right]^2 d\boldsymbol{\lambda}_y \right\}^{1/2}. \qquad (2.2)$$

Defining the support of $\mathbf{x}$ and $\mathbf{y}$ as unit cubes in $\mathbb{R}^{k'}$ and $\mathbb{R}^{(k-k')}$ ensures that the additive terms in (2.2) are not of radically different orders of magnitude. This becomes important for $s_x, s_y \geq 2$, as integrating over the CDFs would otherwise introduce the units of the variable into the integrated values.

Social welfare theory can also imply multiple stochastic dominance restrictions. Atkinson and Bourguignon (1982) consider social welfare comparisons among bivariate distributions. Let $SW_a = \int U(\mathbf{z})\mathbf{d}G_a(\mathbf{z})$ represent social welfare in population "*a*" where the function $U(\mathbf{z})$ represents the social planner's valuation of welfare as a function of a vector of indicators $\mathbf{z}$. Define SW_b analogously. The motivation is to derive a set of conditions for which social welfare is unambiguously higher in population "*a*" than population "*b*," based upon properties of $U(\mathbf{z})$. For example, suppose the following null hypothesis is satisfied by a pair of bivariate distributions:

$$H_o : G_a \succ_2 G_b \quad and \quad G_{a_x} \succ_2 G_{b_x} \ and \ G_{a_y} \succ_2 G_{b_y} .$$

Atkinson and Bourguignon (1982, equations 15a-c) show that the above conditions imply $SW_a \geq SW_b$ so long as $U \in \{U : U_1, U_2 \geq 0, U_{11}, U_{22}, U_{12} \leq 0, U_{112}, U_{122} \geq 0, U_{1122} \leq 0\}$.

To test the above null hypothesis we define Λ_x and Λ_y as unit intervals in $\mathbb{R}$ and Λ as the unit square in $\mathbb{R}^2$. For $\lambda_x \in \Lambda_x$, let $\psi_x(\lambda_x) = \max\{D^2_{a_x}(\lambda_x) - D^2_{b_x}(\lambda_x), 0\}$; for $\lambda_y \in \Lambda_y$, let $\psi_y(\lambda_y) = \max\{D^2_{a_y}(\lambda_y) - D^2_{b_y}(\lambda_y), 0\}$; and for $\boldsymbol{\lambda} \in \Lambda$, let $\psi(\boldsymbol{\lambda}) = \max\{D^2_a(\boldsymbol{\lambda}) - D^2_b(\boldsymbol{\lambda}), \mathbf{0}\}$.

Define

$$T = \left\{ \int_{\Lambda_x} \left[\psi_x(\lambda_x)\right]^2 d\lambda_x \right\}^{1/2} + \left\{ \int_{\Lambda_y} \left[\psi_y(\lambda_y)\right]^2 d\lambda_y \right\}^{1/2} + \left\{ \int_{\Lambda} \left[\psi(\boldsymbol{\lambda})\right]^2 d\boldsymbol{\lambda} \right\}^{1/2} . \qquad (2.3)$$

The test statistics $\hat{T}$ in (2.1), (2.2) and (2.3) do not have known asymptotic distributions. However, following Hall and Yatchew (2005), we obtain consistent bootstrap critical values using the following algorithm. Combine the two datasets into one bootstrap dataset. Draw two samples of size n_a and n_b for bootstrap samples "*a*" and "*b*" respectively. The data generating mechanisms for the two bootstrap samples will weakly satisfy the null hypothesis since they are drawn from the same distribution. From the bootstrap samples calculate $\widehat{D}_a^{s*}(\mathbf{z})$, $\widehat{D}_b^{s*}(\mathbf{z})$ and insert these into (2.1), (2.2) or (2.3) to obtain T^*. Repeating this procedure many times, (we use 200 bootstrap iterations throughout the paper), allows one to bootstrap the distribution of $\hat{T}$ under the respective null hypothesis. From the bootstrap distribution of $\hat{T}$ we calculate the critical values.

III. Simulation results

To examine the properties of our testing procedure we conduct simulations using various data generating mechanisms (DGMs), which we set out below. In each case, the simulated data are generated using a bivariate lognormal pair (X, Y) where the underlying joint normal random variables (x, y) have means μ_x, μ_y, variances σ_x^2, σ_y^2 and covariance σ_{xy}.

For each DGM we run 1000 simulations, with sample sizes $n_a = n_b = 50$ and 500. We conduct tests of the following five hypotheses:

$$
\begin{aligned}
&H_0^A : G_a \succ_1 G_b \\
&H_0^B : G_{a_x} \succ_1 G_{b_x} \\
&H_0^C : G_{a_y} \succ_1 G_{b_y} \\
&H_0^D : G_{a_x} \succ_1 G_{b_x} \quad and \quad G_{b_y} \succ_1 G_{a_y} \\
&H_0^E : G_{b_x} \succ_1 G_{a_x} \quad and \quad G_{a_y} \succ_1 G_{b_y} \quad .
\end{aligned}
$$

Under the first DGM the two distributions are identical. The parameter values of the underlying normal distribution are set to $\left(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \sigma_{xy}\right) = \left(0.85, 0.85, 0.36, 0.36, 0.2\right)$ for both populations. We chose the means and variances to allow for easy comparability with the simulations of Barrett and Donald (2003). We expect the rate of rejection for all five null hypotheses to be approximately at the level of the test. The second DGM maintains the same parameter values for population "*b*" but those for population "*a*" change to $\left(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \sigma_{xy}\right)_a = \left(0.6, 0.6, 0.64, 0.64, 0.2\right)$. These parameter values imply that all five null hypotheses are false since the marginal distributions cross for both variables X and Y. The third DGM uses the original parameter values for population "*a*" and $\left(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \sigma_{xy}\right)_b = \left(0.85, 0.85, 0.36, 0.36, -0.2\right)$ for population "*b*." That is, both populations have the same parameters except for the covariance parameter. Thus, the marginal distributions are identical, in which case, hypotheses B, C, D and E are weakly true[4]; however, hypothesis A which tests bivariate stochastic dominance of "*a*" over "*b*" is false.[5] The fourth DGM uses the original parameter values for population "*b*" and $\left(\mu_x,\mu_y,\sigma_x^2,\sigma_y^2,\sigma_{xy}\right)_a=\left(0.65,2.1,0.36,0.36,0.2\right)$ for population "*a*." These parameters imply that hypotheses A, B and D are false while C and E are true. Table 1 summarizes the parameter values of the underlying normal distribution and its associated lognormal distribution for each of the DGMs.

Table 2 shows the results of the simulations described above. We find that our test procedure performs well for all four DGMs. The power of the test procedure improves substantially as the sample size increases from 50 to 500 observations. Our results for hypotheses B and C, which involve only the marginal distributions, are very similar to those of Hall and Yatchew (2005). Where hypothesis B or C is weakly true, as under DGMs 1 and 3, the null hypothesis is rejected at approximately the test level. Furthermore, when hypothesis B or C is false, as under DGM 2, the test procedure has substantial power even with a sample size of 50. Finally, the test procedure correctly fails to reject hypothesis C when it is strongly true, as is the case with DGM 4.

Our results concerning the combined marginal hypotheses, D and E, show similar patterns. When hypothesis D or E is weakly true, such as with DGMs 1 and 3, the rejection rate is approximately equal to the test level. Furthermore, when hypothesis D or E is false, there is substantial power even with a sample size of 50. Finally, when hypothesis E is strongly true, as under DGM 4, the rejection rate is close to zero.

---

[4] We use the term "weakly true" when the difference between the two functions is everywhere zero. We use "true" when the difference between the two functions is everywhere less than or equal to zero and less than zero over some portion of the support.

[5] Though, bivariate stochastic dominance of "*b*" over "*a*" is true.

We now turn to a discussion of tests of hypothesis A, which under the null asserts bivariate stochastic dominance. When this hypothesis is weakly true, as is the case for DGM 1, the rejection rates remain close to the nominal significance levels. Under DGM 2 where univariate stochastic dominance fails at both margins, it is perhaps not surprising that the bivariate test has substantial power. Thus, one could have rejected bivariate stochastic dominance by performing univariate tests on either or both margins.

DGM 3 is more interesting. In this case, univariate stochastic dominance holds (weakly) at both margins but bivariate dominance does not. The bivariate test has substantial power even for a sample size of 50.

For DGM 4, bivariate dominance does not hold and indeed univariate dominance fails on one of the margins. In this case, bivariate dominance is rejected with greater power by performing univariate tests.   Under DGM 4, bivariate dominance actually holds for a substantial portion of the support, which would appear to underlie the weaker power of the bivariate test.

IV. Empirical results

Our empirical analysis is motivated by the literature on the differences in time spent working between continental European countries, (in particular Germany and France), and the U.S. This literature largely focuses on trying to understand the determinants of the differences in average hours worked (see Alesinsa, Glaeser and Sacerdote (2005), Prescott (2004), and Schettkat (2003)). It thus far has not been concerned with trying to robustly measure the welfare consequences of the associated differences in income and leisure time across the countries.

We compare the distributions of income and leisure in Germany[6], the United Kingdom, and the United States using data drawn from the Cross-National Equivalent File (CNEF)[7]. The German data within the CNEF dataset originate from the German Socio-Economic Panel Study (GSOEP), the U.K. data come from the British Household Panel Survey (BHPS), and the U.S. data come from the Panel Study of Income Dynamics (PSID). The CNEF data reports the variables with standardized definitions wherever possible. For some series, this has required imputation of variables not directly reported in the original survey. In particular, the GSOEP does not directly report annual hours worked by the individual in the previous calendar year. As such, the annual hours worked series for Germany is constructed using information on employment status during the survey year, average number of hours worked per week during the survey year, and the number of months worked in the previous year. Similarly, the PSID ceased providing an estimate of household post-government income as of survey year 1992. Thus, the CNEF dataset provides a constructed series.

We compare income and leisure for the years 1983, 1990, and 2000 for Germany and the U.S., and 1990 and 2000 for the U.K. (Consistent earlier data for the U.K. were not available.) Throughout the paper, we define leisure as the number of hours spent outside the formal labour market. We restrict the CNEF dataset in the following ways. The CNEF-PSID data do not provide income or hours worked data on individuals who are not present in the household at the time of the survey. We remove these households from the dataset. Furthermore, we remove individuals for whom the income variable or the annual hours worked variable contains an invalid response. We also modify the CNEF-GSOEP annual hours worked series, as the German data – unlike the U.K. and U.S. data – include holiday, vacation, and sick leave in the estimate of

[6] For data prior to 1991 Germany refers to the former territory of West Germany. From 1991 onward Germany refers to the unified territories of East and West Germany.
[7] Please refer to http://www.human.cornell.edu/pam/gsoep/equivfil.cfm for further information.

annual hours worked. Since the latter method is a better reflection of actual time worked, we adjust the German data using the yearly averages of these values provided by the Institute for Employment Research of the Federal Employment Services, Germany (IAB).

We focus our attention on the welfare of three different population subsets, using post-government income and annual hours worked by the individual[8]. We convert from local currency units to U.S. dollars using purchasing power parities for actual individual consumption taken from the OECD's National Accounts database. Table 3 provides a summary of the weighted means of post-government income per adult equivalent in U.S. dollars[9] for all household heads and partners. As expected, per adult equivalent income is highest in the U.S., followed by the U.K. and Germany for 2000 and 1990. Similarly, for 1983 per adult equivalent income is higher in the U.S. than in Germany. All three countries show growth in per adult equivalent income over their respective time periods, but it is greatest in the U.S. Table 3 also displays the weighted average annual hours of leisure time per individual for all household heads and partners. In contrast to the trends in income, Germany displays the highest average annual hours of leisure for household heads and partners, followed by the U.K. and then the U.S. Moreover, workers in Germany and the U.K. displayed increases in leisure hours over the years observed, while their U.S. counterparts were working more hours in 2000 than in 1983 or 1990.

Figure 1 and Figure 2 present the empirical marginal distributions of income and leisure, respectively, for Germany and the U.S. in 2000. Not only are the means substantially different, but the U.S. income distribution appears to first-order stochastically dominate that of Germany,

---

[8] Specifically, we employ series I11102XX for total household post-government income for Germany and the U.K. and I11113XX for the U.S., and E11101XX for annual work hours of the individual, where XX refers to the last two digits of the survey year.

[9] To convert household income to per adult equivalent income we divide by $e = 1 + 0.7(A-1) + 0.5K$, where A and K represent the number of adults and children, respectively, in the household, and a child is defined as being between 0 and 14 years of age inclusive. This equivalence scale formula is provided with the CNEF data.

while the German leisure distribution appears to first-order stochastically dominate that of the U.S. Indeed, we find this pattern of stochastic dominance for 1983, 1990, and 2000 when comparing Germany and the U.S. Table 4 displays our results for tests of first- and second-order stochastic dominance of the income, leisure, and joint distributions respectively for each possible cross-country comparison in the years 1983, 1990, and 2000. The table reports the estimated p-value of the indicated hypothesis in each case. We find that the income distribution of the U.S. first-order stochastically dominates the income distribution of Germany in each year at a five percent test level. In contrast, we find strong evidence that the leisure distribution of Germany first-order stochastically dominates that of the U.S. in each year. Not surprisingly, we reject both null hypotheses concerning bivariate first-order stochastic dominance. Since stochastic dominance of order s implies stochastic dominance of order $s+1$, the sufficient stochastic dominance conditions to identify a non-negative change in social welfare (see Atkinson and Bourguignon (1982)) cannot be established when comparing Germany and the U.S. When the first-order stochastic dominance rankings imply that a robust ranking of bivariate social welfare is impossible, we do not run tests for higher-order stochastic dominance.

Table 4 also displays the results of stochastic dominance tests between Germany and the U.K., and between the U.K. and the U.S. In general, we find the same pattern of stochastic dominance with these cross-country comparisons. The U.S. income distribution tends to stochastically dominate that of the U.K., while the U.K.income distribution stochastically dominates that of Germany. When comparing leisure distributions the opposite pattern emerges. Germany’s leisure distribution stochastically dominates that of the U.K., while the U.K.leisure distribution stochastically dominates that of the U.S. A few interesting observations though can be made from the estimated p-values reported in Table 4. First, in 1990, the U.S. and U.K.

income distributions cross, as shown by the strong rejection of stochastic dominance in both directions. It is the case though, that the U.S. income distribution in 1990 second-order stochastically dominates that of the U.K. A second interesting observation from the 1990 comparison between the U.K. and the U.S. is the estimated p-value of 0.290 for the null hypothesis that the U.K. bivariate distribution stochastically dominates that of the U.S. This results implies that the null hypothesis cannot be rejected, which is in contrast to the rejection of the null hypothesis concerning the income distribution. This is indicative of the weaker power of the test in bivariate situations when the null hypothesis is true over a substantial portion of the bivariate support.

Next, we check the robustness of our results concerning household heads and partners by focusing on single-person households. This avoids any complications arising from unequal sharing of income or leisure time within a household or calculations involving equivalence scales. Table 5 presents summary statistics for all single-person households. The mean incomes for single-person households are quite similar to those shown for household heads and partners in per adult equivalent terms. For leisure, the magnitudes and trends in average hours are also similar for single-person households and for all household heads and partners. The patterns for the standard deviation of income and leisure are also roughly consistent with those for all household heads and partners, as the standard deviation of income grows within each country over time while the standard deviation of leisure remains roughly constant.

Results of tests for stochastic dominance for single-person households, shown in Table 6, are very similar to those for household heads and partners. The U.S. income distribution either first- or second-order dominates those of Germany and the U.K., while the U.K. income distribution first-order stochastically dominates Germany's income distribution in 1990 and

2000. Furthermore, the American leisure distribution is strongly dominated at first-order by the leisure distributions of both Germany and the U.K., while Germany's leisure distribution first-order stochastically dominates that of the U.K. at the five percent test level in both years. These patterns of stochastic dominance amongst the marginal distributions again make it impossible to establish the stochastic dominance conditions sufficient for identifying a robust difference in social welfare between the countries for single-person households.

As a further check on our previous results, we test for stochastic dominance amongst distributions associated with working single-person households. In the previous two population subsets examined, we were implicitly assuming that anyone who did not work during the year was enjoying the maximum amount of leisure possible in a given year. This is a contestable assumption. Table 7 presents summary statistics for working single-person households. We report the same broad trends as for the previous two population subsets. Per capita income grows within each country over time, with American average income being the highest and average leisure time does not change much with Germany or the U.K., while it decreases within the U.S. from 1983 to 2000.

The stochastic dominance test results for working single-person households are shown in Table 8. The patterns of stochastic dominance are similar to those of the previous two population subsets. However, there are some notable exceptions. In particular, when comparing Germany and the U.K. in 1990, we find that sufficient conditions for higher social welfare in Germany hold. Specifically, at first-order, Germany's income distribution weakly stochastically dominates the U.K. income distribution, and Germany's leisure and bivariate distributions stochastically dominate those of the U.K. These are the requirements presented by Atkinson and Bourguinon (1982) for higher social welfare in Germany for all social welfare functions, as defined previously, for which $U_1, U_2 \geq 0$ and $U_{12} \leq 0$ for all (z_1, z_2). The results can also be seen by referring to Figure 4 through Figure 6. The empirical income distributions appear to cross, but this crossing is not statistically significant. The same crossing appears in the contour plots of the empirical bivariate distribution functions, but again this crossing is statistically insignificant. Finally, the German leisure distribution lies below or to the right of the U.K. leisure distribution over the majority of the support.

V. Concluding Remarks

In this paper, we introduce testing procedures for multi-dimensional stochastic dominance. In particular, we present a framework for testing stochastic dominance relationships both for the entire multi-dimensional distribution and over subsets of dimensions. We use this testing procedure to compare welfare between three different population subsets in Germany, the U.K, and the U.S., using income and leisure as measures of well-being. We find that sufficient conditions for evaluating differences in social welfare using stochastic dominance relationships do not hold in general. Furthermore, we find that the U.S. income distribution stochastically dominates both those of Germany and the U.K., usually at first-order, while the U.K. income distribution similarly stochastically dominates that of Germany. However, when comparing the leisure distributions the directions of dominance are reversed. The German leisure distribution dominates that of the U.K. and the U.S., usually at first-order, while the U.K. leisure distribution dominates that of the U.S.

Tests of stochastic dominance embody more stringent conditions than tests based on summary statistics such as means or medians. As a result, it is often not possible to robustly rank social welfare across comparison groups. However, in some cases, interest may lie in subsets of the supports of the distributions, in which case the tools outlined here may readily be employed. For example, one may be interested in evaluating differences in bivariate poverty (Sahn et al., 2001). If one considers an intersection definition of bivariate poverty, that is, an individual is "poor" if both income and leisure for the individual fall below certain thresholds (i.e., the working poor), then this is equivalent to looking at the bivariate distribution over the lower-left quadrant of the support. Thus, while the two bivariate distributions may cross, there might be a region of the support, originating at the origin, over which stochastic dominance holds. Indeed, contour plots of the bivariate distributions for Germany and the U.S. suggest that the German distribution stochastically dominates the American distribution over almost all regions of the support, with the major exception being near the upper limit of leisure hours. This can be seen in Figure 3, which is a contour plot of the bivariate CDFs for Germany and the U.S. in 2000. It shows that the contours for Germany lie to the northeast of the corresponding contours for the U.S. over most of the support. This suggests that the bivariate poverty headcount ratio is lower in Germany as compared to the U.S., whenever the bivariate poverty frontier lies within this interior area of the support.

Data Appendix

The GSOEP data does not include estimates of time off work due to holidays, vacations, sick leave, or other reasons for each year of the survey. As such, the CNEF-GSOEP uses an estimate of actual weekly time spent working extrapolated over the year. This procedure overestimates the amount of time actually spent working. We adjust for this by subtracting the average number of days Germans spent away from work, as reported by the IAB. We checked this procedure using the 1985 and 2000 GSOEP surveys, which contain estimates of individual time away from work during the previous year. For both years, we found that the two distributions of annual hours work, one using the IAB estimates of time away from work and the other using GSOEP estimates, were very similar. Given the large differences in leisure hours distributions across countries, the small difference between these two methods for estimating the German leisure hours distribution is unlikely to affect our conclusions.

References

Alesina, Roberto, Glaeser, Edward and Sacerdote, Bruce. 2005. Work and leisure in the U.S. and Europe. Why so different? Harvard Institute of Economic Research, Discussion paper no. 2068.

Anderson, Gordon. 1996. Nonparametric tests of stochastic dominance in income distributions. *Econometrica* **64**: 1183-1193.

Atkinson, A. B. and Bourguignon, F. 1982. The comparison of multi-dimensioned distributions of economic status. *Review of Economic Studies* **XLIX**: 183-201.

Bach, Hans-Uwe and Koch, Susanne. 2003. Working time and the volume of work in Germany: The IAB concept. IAB Labour Market Research Topics, No. 53.

Barrett, Garry F. and Donald, Stephen G. 2003. Consistent tests for stochastic dominance. *Econometrica* **71**: 71-104.

Crawford, Ian. 1999. Nonparametric tests of stochastic dominance in bivariate distributions, with an application to UK data. The Institute for Fiscal Studies, Working Paper 28/99.

Davidson, Russell and Duclos, Jean-Yves. 2000. Statistical inference for stochastic dominance and for the measurement of poverty and inequality. *Econometrica* **68**: 1435-1464.

Frey, Bruno S. and Stutzer, Alois. 2002. What can economists learn from happiness research? *Journal of Economic Literature* **40**: 402-435.

Hall, Peter and Yatchew, Adonis. 2005. Unified approach to testing functional hypotheses in semiparametric contexts. *Journal of Econometrics* **127**: 225-252.

Justel A, Peña D, Zamar R. 1997. A multi-variate Kolmogorov-Smirnow test of goodness of fit. *Statistics and Probability Letters* **35**: 251-259.

Kaur A, Rao BLSP, Singh H. 1994. Testing for second-order stochastic dominance. *Econometric Theory* **10**: 849-866.

Klecan L, McFadden R, McFadden D. 1991. A robust test for stochastic dominance. Working paper, Dept. of Economics, MIT.

Linton, Oliver, Maasoumi, Esfandiar, and Whang, Yoon-Jae. 2005. Consistent testing for stochastic dominance under general sampling schemes. *Review of Economic Studies*. **72**: 735-765.

Maasoumi E. 2001. Parametric and nonparametric tests of limited domain and ordered hypotheses in economics. In *A Companion to Econometric Theory*, Baltagi B (ed). Blackwell: Basil.

Maasoumi E, Heshmati A. 2000. Stochastic dominance among Swedish income distributions. *Econometric Reviews* **19**: 287-320.

McFadden D. 1989. Testing for stochastic dominance. In *Studies in the Economics of Uncertainty. Part II*, Fomby T, Seo TK (eds). Springer-Verlag.

Prescott, Edward C. 2004. Why do Americans work so much more than Europeans? *Federal Reserve Bank of Minneapolis Quarterly Review*. **28**: 2-13.

Sahn, David, Duclos, Jean-Yves, and Younger, Stephen S. 2001. Robust multidimensional poverty comparisons. *Economic Journal* forthcoming.

Schettkat, Ronald. 2003. Differences in U.S.-German time-allocation: Why do Americans work longer hours than Germans? IZA discussion paper no. 697.

Tse YK, Zhang X. 2004. A monte carlo investigation of some tests of stochastic dominance. *Journal of Statistical Computation and Simulation* **74**: 361-378.

Xu K, Fisher G, Wilson D. 1995. New distribution-free tests for stochastic dominance. Working paper No. 95-02, Dept. of Economics, Dalhousie University.

Table 1 - Parameter values of the simulated bivariate normal and lognormal distributions

	DGM 1		DGM 2		DGM 3		DGM 4	
	a	*b*	*a*	*b*	*a*	*b*	*a*	*b*
Normal distribution								
μ_x	0.850	0.850	0.600	0.850	0.850	0.850	0.650	0.850
μ_y	0.850	0.850	0.600	0.850	0.850	0.850	2.100	0.850
σ_x^2	0.360	0.360	0.640	0.360	0.360	0.360	0.360	0.360
σ_y^2	0.360	0.360	0.640	0.360	0.360	0.360	0.360	0.360
σ_{xy}	0.200	0.200	0.200	0.200	0.200	-0.200	0.200	0.200
Lognormal distribution								
μ_X	2.801	2.801	2.509	2.801	2.801	2.801	2.293	2.801
μ_Y	2.801	2.801	2.509	2.801	2.801	2.801	9.777	2.801
σ_X^2	3.400	3.400	5.645	3.400	3.400	3.400	2.279	3.400
σ_Y^2	3.400	3.400	5.645	3.400	3.400	3.400	41.419	3.400
σ_{XY}	1.737	1.737	1.394	1.737	1.737	-1.422	4.964	1.737

Table 2 - Level and power of testing procedure

	$n_a = n_b = 50$			$n_a = n_b = 500$		
	10%	**5%**	**1%**	**10%**	**5%**	**1%**
Hypothesis			**DGM 1**			
A: weakly true	0.105	0.055	0.014	0.121	0.068	0.021
B: weakly true	0.093	0.044	0.008	0.105	0.055	0.016
C: weakly true	0.074	0.044	0.010	0.101	0.055	0.018
D: weakly true	0.098	0.044	0.010	0.093	0.048	0.013
E: weakly true	0.088	0.040	0.009	0.104	0.058	0.016
			DGM 2			
A: false	0.605	0.458	0.214	1.000	1.000	1.000
B: false	0.605	0.452	0.233	1.000	1.000	0.998
C: false	0.496	0.350	0.144	1.000	1.000	0.997
D: false	0.480	0.324	0.159	1.000	1.000	0.997
E: false	0.381	0.252	0.102	1.000	1.000	0.996
			DGM 3			
A: false	0.489	0.344	0.149	1.000	0.997	0.956
B: weakly true	0.113	0.050	0.015	0.112	0.066	0.014
C: weakly true	0.090	0.041	0.013	0.095	0.051	0.015
D: weakly true	0.141	0.071	0.022	0.114	0.056	0.013
E: weakly true	0.081	0.041	0.010	0.104	0.049	0.017
			DGM 4			
A: false	0.120	0.058	0.012	0.973	0.914	0.739
B: false	0.598	0.443	0.236	1.000	1.000	0.998
C: true	0.000	0.000	0.000	0.000	0.000	0.000
D: false	1.000	1.000	1.000	1.000	1.000	1.000
E: true	0.001	0.001	0.000	0.000	0.000	0.000

Table 3 – Summary statistics for all household heads and partners

	1983	1990	2000
Germany			
Mean of income	8443	12766	17268
Mean of leisure	7867	7910	7978
St. dev. of income	4911	7419	9849
St. dev. of leisure	947.9	867.6	901
Cor. of income and leisure	-0.288	-0.289	-0.285
No. of observations	9550	7403	14310
United Kingdom			
Mean of income	n.a.	13658	20858
Mean of leisure	n.a.	7612	7705
St. dev. of income	n.a.	7493	14196
St. dev. of leisure	n.a.	1062	1040
Cor. of income and leisure	n.a.	-0.255	-0.165
No. of observations	n.a.	6143	6801
United States			
Mean of income	11234	16765	26551
Mean of leisure	7484	7397	7263
St. dev. of income	7909	13104	27338
St. dev. of leisure	1044	1074	1061
Cor. of income and leisure	-0.193	-0.220	-0.116
No. of observations	10893	11477	11621

Note: All income values are reported in current year U.S. dollars.
"n.a." denotes not available due to lack of data.

Table 4 – Estimated p-values for tests of stochastic dominance -- all household heads and partners

H_o	First-order			Second-order		
	1983	**1990**	**2000**	**1983**	**1990**	**2000**
	Germany and the U.K.					
$Ger_{income} \succ_1 UK_{income}$	n.a.	0.000	0.000	n.a.		
$UK_{income} \succ_1 Ger_{income}$	n.a.	0.300	0.760	n.a.		
$Ger_{leisure} \succ_1 UK_{leisure}$	n.a.	1.000	1.000	n.a.		
$UK_{leisure} \succ_1 Ger_{leisure}$	n.a.	0.000	0.000	n.a.		
$Ger \succ_1 UK$	n.a.	0.105	0.000	n.a.		
$UK \succ_1 Ger$	n.a.	0.000	0.000	n.a.		
	Germany and the U.S.					
$Ger_{income} \succ_1 US_{income}$	0.000	0.000	0.000			
$US_{income} \succ_1 Ger_{income}$	0.345	0.210	0.635			
$Ger_{leisure} \succ_1 US_{leisure}$	1.000	1.000	1.000			
$US_{leisure} \succ_1 Ger_{leisure}$	0.000	0.000	0.000			
$Ger \succ_1 US$	0.000	0.000	0.000			
$US \succ_1 Ger$	0.000	0.000	0.000			
	The U.K. and the U.S.					
$UK_{income} \succ_1 US_{income}$	n.a.	0.000	0.000	n.a.	0.000	
$US_{income} \succ_1 UK_{income}$	n.a.	0.000	0.065	n.a.	0.545	
$UK_{leisure} \succ_1 US_{leisure}$	n.a.	1.000	1.000	n.a.	0.930	
$US_{leisure} \succ_1 UK_{leisure}$	n.a.	0.000	0.000	n.a.	0.000	
$UK \succ_1 US$	n.a.	0.290	0.045	n.a.	0.005	
$US \succ_1 UK$	n.a.	0.000	0.000	n.a.	0.000	

Table 5 – Summary statistics for all single-person households

	1983	1990	2000
Germany			
Mean of income	8686	13630	17325
Mean of leisure	8096	8063	8096
St. dev. of income	5827	8678	11512
St. dev. of leisure	887.6	864.1	883.9
Cor. of income and leisure	-0.432	-0.452	-0.404
No. of observations	1284	996	2318
United Kingdom			
Mean of income	n.a.	14331	22357
Mean of leisure	n.a.	8030	8057
St. dev. of income	n.a.	7390	18797
St. dev. of leisure	n.a.	1032	1018
Cor. of income and leisure	n.a.	-0.411	-0.214
No. of observations	n.a.	977	1173
United States			
Mean of income	11564	16092	26092
Mean of leisure	7583	7512	7363
St. dev. of income	7662	13416	25176
St. dev. of leisure	1043	1113	1085
Cor. of income and leisure	-0.409	-0.396	-0.264
No. of observations	1557	1831	1628

Note: All income values are reported in current year U.S. dollars.
"n.a." denotes not available due to lack of data.

Table 6 – Estimated p-values for tests of stochastic dominance -- all single-person households

H_o	First-order			Second-order		
	1983	**1990**	**2000**	**1983**	**1990**	**2000**
	Germany and the U.K.					
$Ger_{income} \succ_1 UK_{income}$	n.a.	0.000	0.000	n.a.		
$UK_{income} \succ_1 Ger_{income}$	n.a.	0.865	0.485	n.a.		
$Ger_{leisure} \succ_1 UK_{leisure}$	n.a.	0.055	0.080	n.a.		
$UK_{leisure} \succ_1 Ger_{leisure}$	n.a.	0.005	0.000	n.a.		
$Ger \succ_1 UK$	n.a.	0.030	0.000	n.a.		
$UK \succ_1 Ger$	n.a.	0.005	0.005	n.a.		
	Germany and the U.S.					
$Ger_{income} \succ_1 US_{income}$	0.000	0.000	0.000			
$US_{income} \succ_1 Ger_{income}$	0.945	0.220	0.825			
$Ger_{leisure} \succ_1 US_{leisure}$	1.000	1.000	1.000			
$US_{leisure} \succ_1 Ger_{leisure}$	0.000	0.000	0.000			
$Ger \succ_1 US$	0.025	0.275	0.030			
$US \succ_1 Ger$	0.000	0.000	0.000			
	The U.K. and the U.S.					
$UK_{income} \succ_1 US_{income}$	n.a.	0.000	0.000	n.a.	0.020	
$US_{income} \succ_1 UK_{income}$	n.a.	0.000	0.105	n.a.	0.225	
$UK_{leisure} \succ_1 US_{leisure}$	n.a.	1.000	1.000	n.a.	1.000	
$US_{leisure} \succ_1 UK_{leisure}$	n.a.	0.000	0.000	n.a.	0.000	
$UK \succ_1 US$	n.a.	0.290	0.340	n.a.	1.000	
$US \succ_1 UK$	n.a.	0.000	0.000	n.a.	0.000	

Table 7 – Summary statistics for working single-person households

	1983	1990	2000
Germany			
Mean of income	11372	17682	21991
Mean of leisure	7064	7130	7082
St. dev. of income	6787	10360	14148
St. dev. of leisure	511.3	474.9	521.5
Cor. of income and leisure	-0.329	-0.270	-0.346
No. of observations	599	477	971
United Kingdom			
Mean of income	n.a.	17699	26698
Mean of leisure	n.a.	6807	6833
St. dev. of income	n.a.	9414	27575
St. dev. of leisure	n.a.	679	696.5
Cor. of income and leisure	n.a.	-0.281	-0.148
No. of observations	n.a.	410	474
United States			
Mean of income	12885	18793	29136
Mean of leisure	6986	6807	6797
St. dev. of income	7627	14634	24317
St. dev. of leisure	761.1	749.8	736.6
Cor. of income and leisure	-0.445	-0.362	-0.232
No. of observations	1052	1181	1162

Note: All income values are reported in current year U.S. dollars.
“n.a.” denotes not available due to lack of data.

Table 8 – Estimated p-values for tests of stochastic dominance -- working single-person households

H_o	First-order 1983	First-order 1990	First-order 2000	Second-order 1983	Second-order 1990	Second-order 2000
Germany and the U.K.						
$Ger_{income} \succ_1 UK_{income}$	n.a.	0.125	0.000	n.a.		
$UK_{income} \succ_1 Ger_{income}$	n.a.	0.405	0.485	n.a.		
$Ger_{leisure} \succ_1 UK_{leisure}$	n.a.	0.865	0.805	n.a.		
$UK_{leisure} \succ_1 Ger_{leisure}$	n.a.	0.000	0.000	n.a.		
$Ger \succ_1 UK$	n.a.	0.260	0.000	n.a.		
$UK \succ_1 Ger$	n.a.	0.000	0.000	n.a.		
Germany and the U.S.						
$Ger_{income} \succ_1 US_{income}$	0.000	0.005	0.000	0.005	0.195	
$US_{income} \succ_1 Ger_{income}$	0.175	0.050	0.600	0.340	0.210	
$Ger_{leisure} \succ_1 US_{leisure}$	0.000	0.545	0.780	1.000	1.000	
$US_{leisure} \succ_1 Ger_{leisure}$	0.000	0.000	0.000	0.000	0.000	
$Ger \succ_1 US$	0.000	0.000	0.000	0.000	1.000	
$US \succ_1 Ger$	0.000	0.000	0.000	0.115	0.000	
The U.K. and the U.S.						
$UK_{income} \succ_1 US_{income}$	n.a.	0.000	0.000	n.a.	0.020	
$US_{income} \succ_1 UK_{income}$	n.a.	0.000	0.105	n.a.	0.225	
$UK_{leisure} \succ_1 US_{leisure}$	n.a.	1.000	1.000	n.a.	1.000	
$US_{leisure} \succ_1 UK_{leisure}$	n.a.	0.000	0.000	n.a.	0.000	
$UK \succ_1 US$	n.a.	0.290	0.340	n.a.	1.000	
$US \succ_1 UK$	n.a.	0.000	0.000	n.a.	0.000	

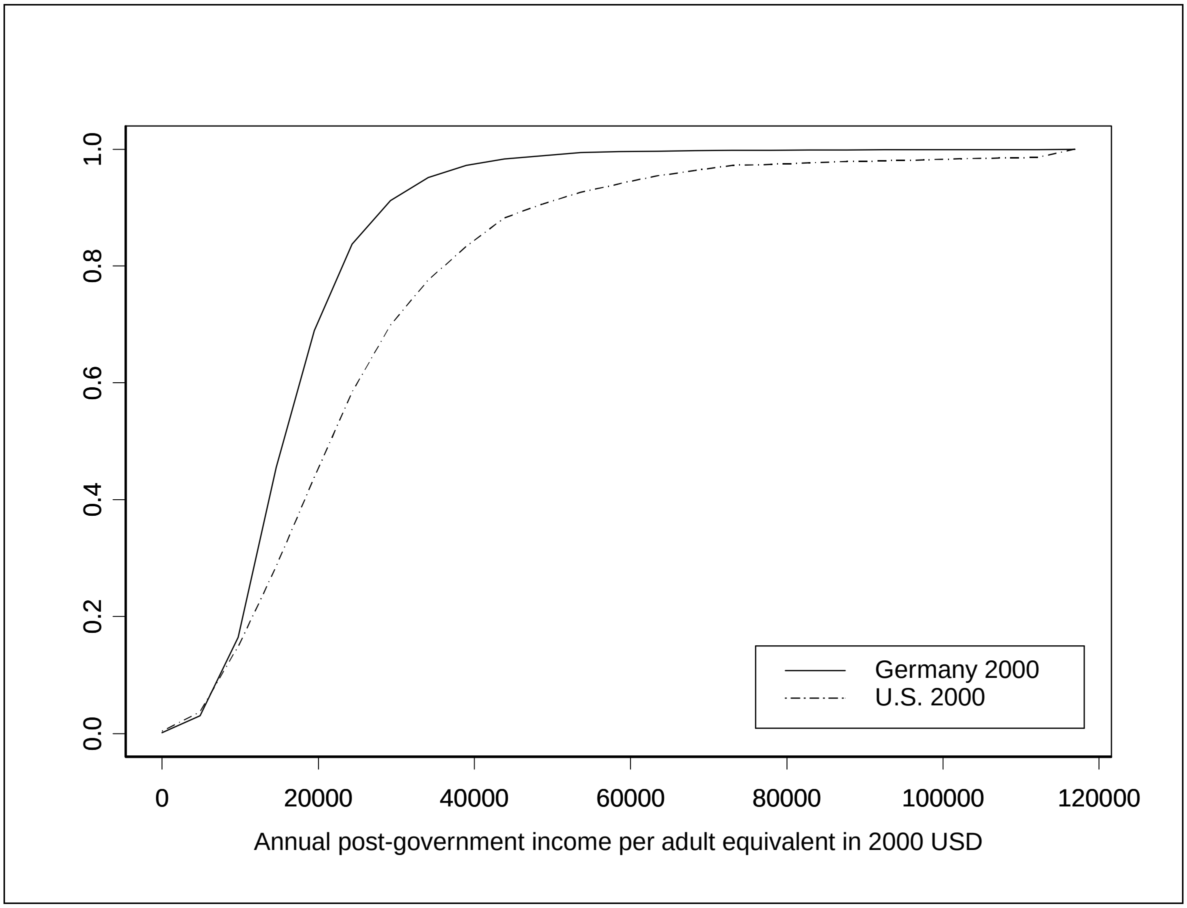

Figure 1 - Empirical income distributions for household heads and partners for Germany and the U.S., 2000

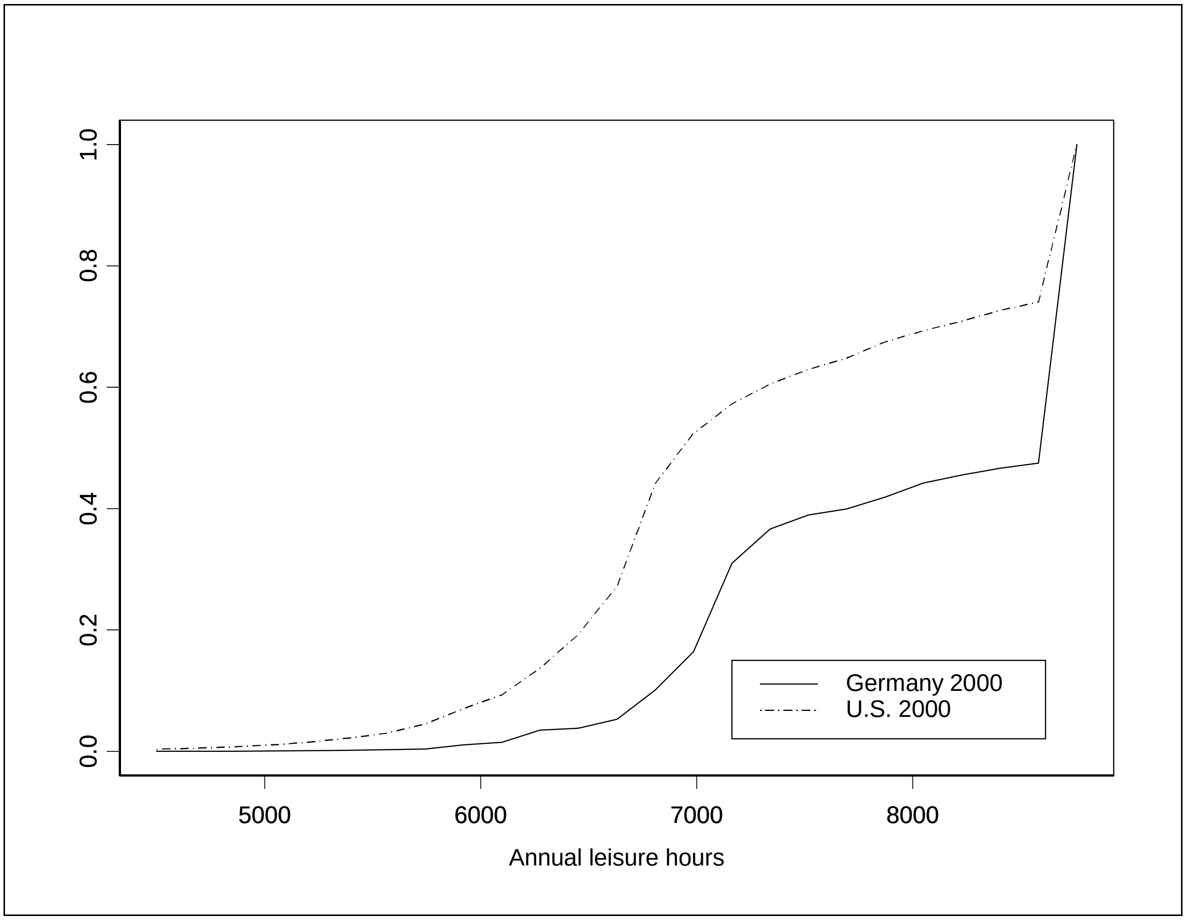

Figure 2 - Empirical leisure distributions for household heads and partners for Germany and the U.S., 2000

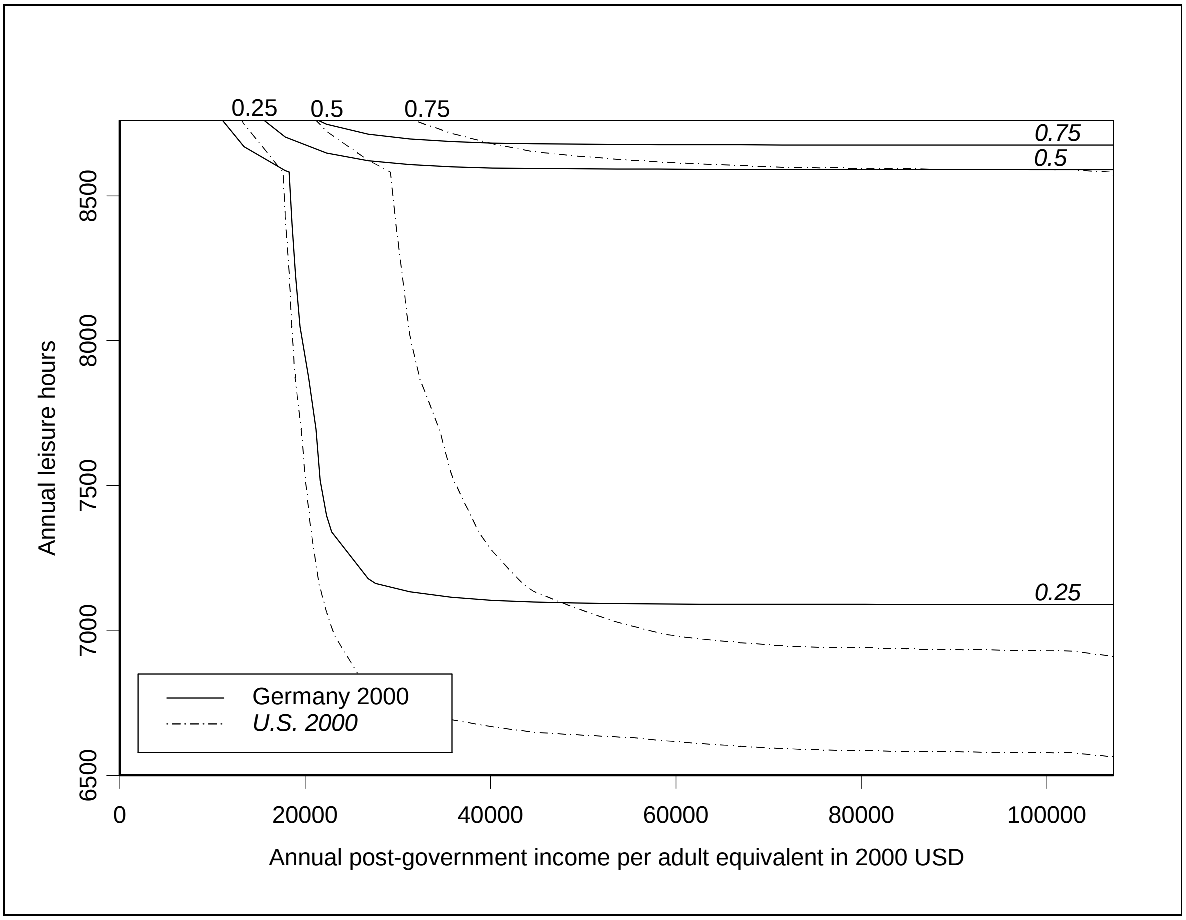

Figure 3 – Contour plots of the empirical CDFs of income and leisure for household heads and partners in Germany and the U.S., 2000

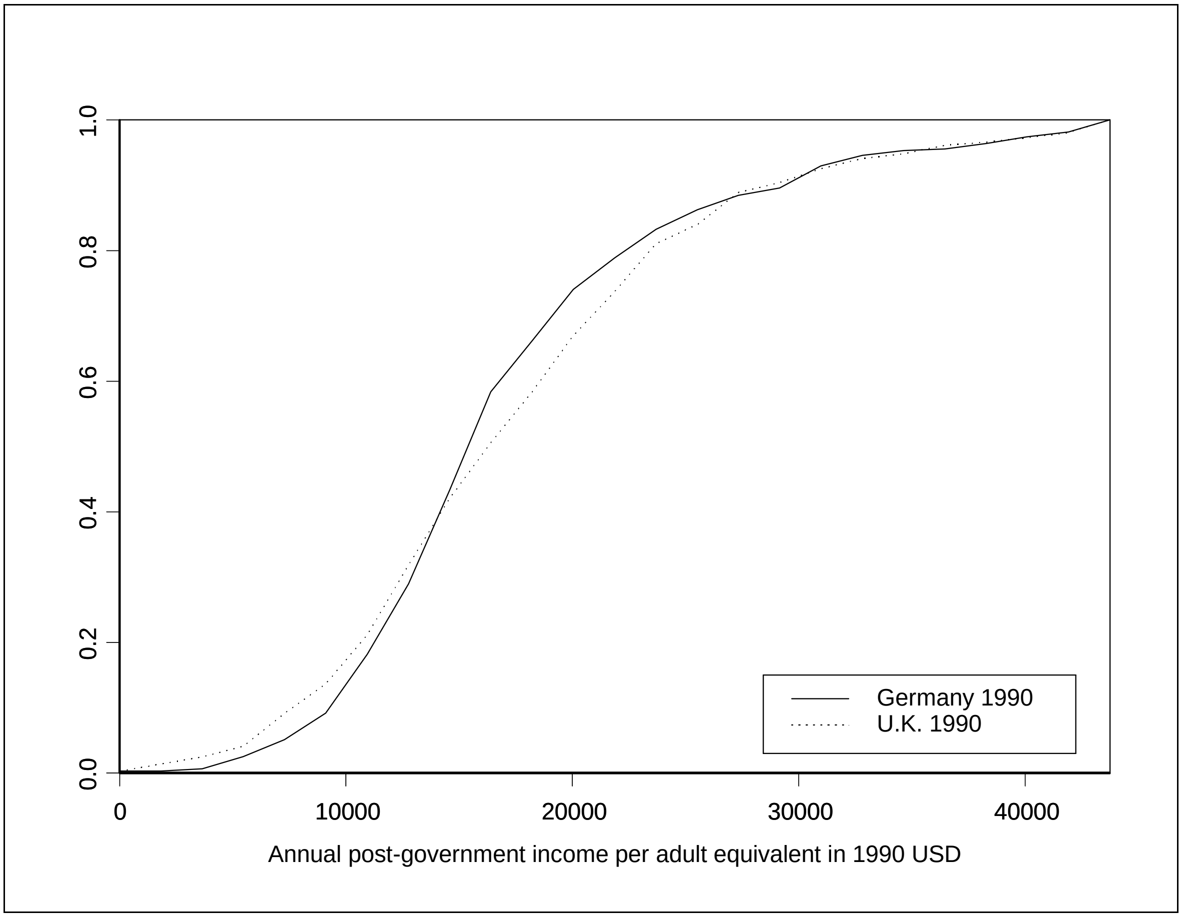

Figure 4 - Empirical income distributions for working single-person households in Germany and the U.K., 1990

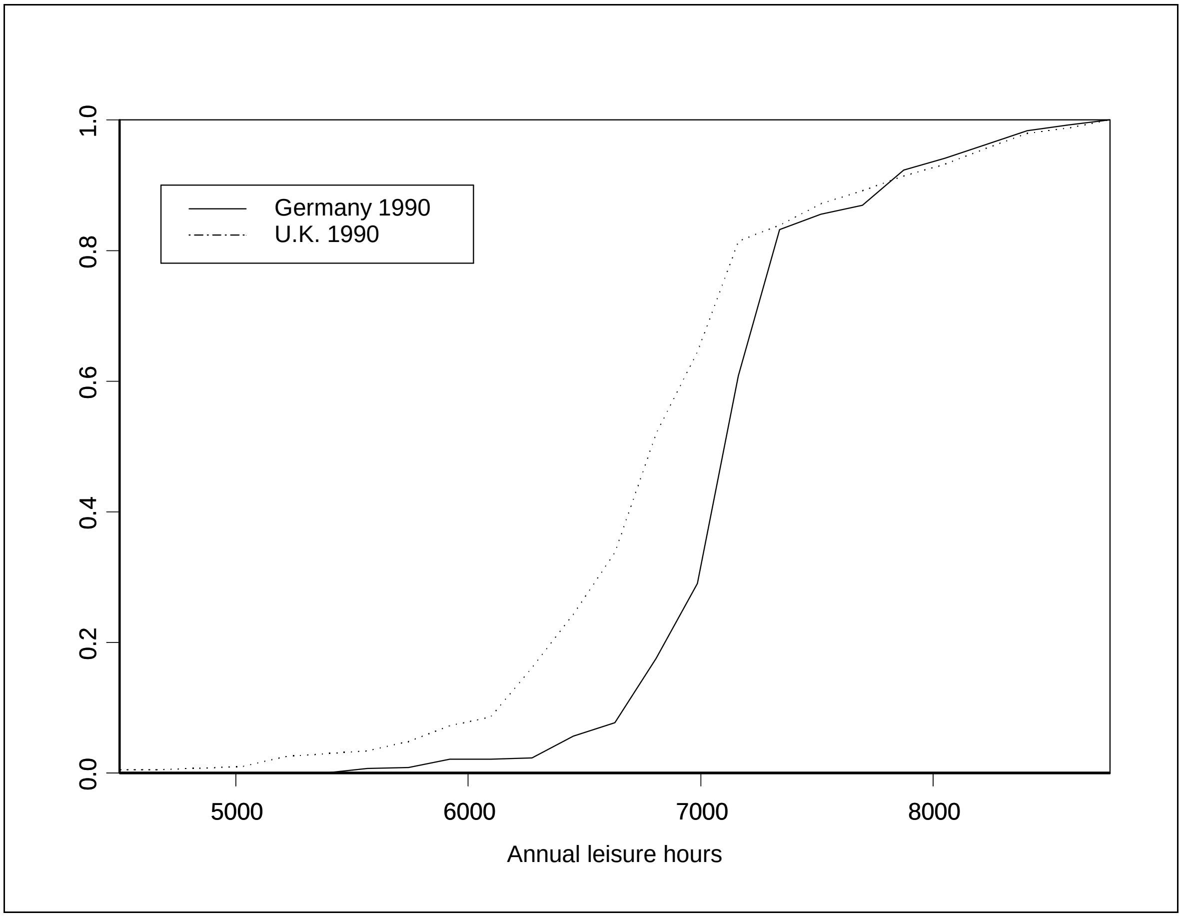

Figure 5 – Empirical leisure distributions for working single-person households in Germany and the U.K., 1990

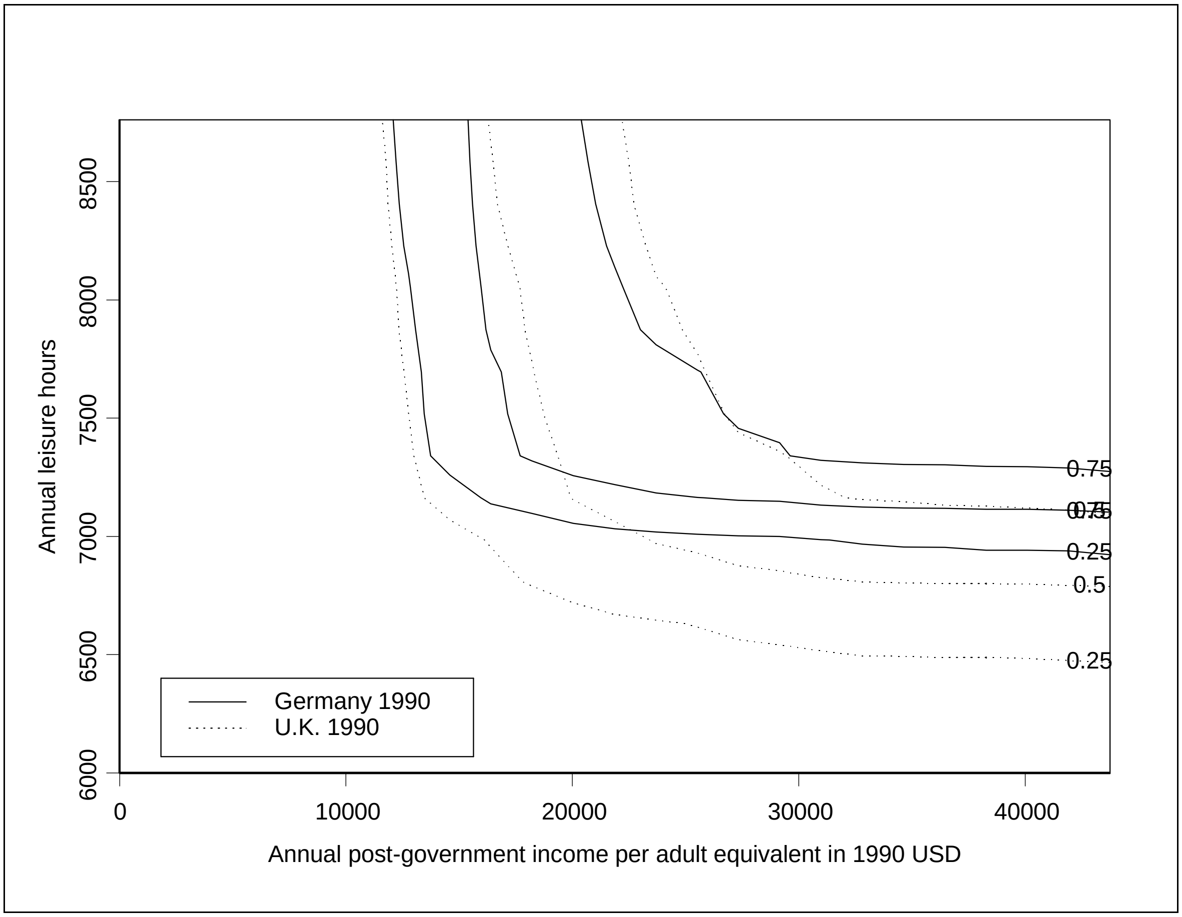

Figure 6 - Contour plots of the empirical bivariate CDFs of income and leisure for working single-person households in Germany and the U.K., 1990