International welfare comparisons and nonparametric testing of multivariate stochastic dominance

Brian McCaig[1]
Department of Economics
University of Toronto
150 St. George Street
Toronto, ON, M5S 3G7
http://www.chass.utoronto.ca/~bmccaig

and

Adonis Yatchew[2]
Department of Economics
University of Toronto
150 St. George Street
Toronto, ON, M5S 3G7

November 24, 2005[3]

---

[1] Corresponding author. Tel.: +1 416-978-4603; Fax: +1 416-978-6713; E-mail: brian.mccaig@utoronto.ca.
Financial support from the Social Sciences and Humanities Research Council of Canada is greatly appreciated.
[2] Financial support from the Social Sciences and Humanities Research Council of Canada is greatly appreciated.
[3] We are indebted to Gordon Anderson for numerous illuminating discussions on the subject matter of this paper.

Summary

This paper outlines a class of statistical procedures that permit testing of a broad range of multi-dimensional stochastic dominance hypotheses and more generally, welfare hypotheses that rely upon multiple stochastic dominance conditions. We apply the procedures to data on income and leisure hours for individuals in Germany, the U.K., and the U.S. We find that no country first-order stochastically dominates the others in both dimensions for all years of comparison. Furthermore, while in general the U.S. stochastically dominates Germany and the U.K. with respect to income, in most periods Germany stochastically dominates with respect to leisure hours. Finally, we find evidence that bivariate poverty is lower in Germany than in either the U.K. or the U.S. for each year of comparison, while poverty rankings between the U.K. and the U.S. are sensitive to the sub-population of individuals considered.

I. Introduction

Cross-country welfare comparisons commonly use statistics such as 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 such statistics is generally available. Individual statistics, however, capture only one 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*" first-order 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.[4]

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

[4] Such comparisons have filtered into the media and inform the debate on societal as well as individual choices. See e.g., Krugman (2005).

by Anderson (1996) in the univariate case and Crawford (2005) in the multivariate case, as well as those proposed by Xu *et al*. (1995), Davidson and Duclos (2000) and Duclos *et al*. (2004). In the 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 procedures suggested by Kaur *et al*. (1994), Anderson (1996), and Davidson and Duclos (2000).

This paper outlines 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. 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 hours from Germany, the U.K. and the U.S. For individuals 25 years of age or older we find that no country first-order 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 hours. In addition, we find evidence of lower bivariate poverty in Germany as compared to both the U.K. and the U.S. We check the robustness of our main results using alternative sub-population groups and find similar results.

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 $\mathbb{R}^k$.[5] We are interested in testing hypotheses of the form

$$H_0 : G_a \succeq_s G_b$$

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

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

for integers $s \geq 2$. (An analogous definition applies to $D_b^s(\mathbf{z})$.) Weak stochastic dominance of order s holds iff $D_a^s(\mathbf{z}) \leq D_b^s(\mathbf{z})$ for all $\mathbf{z} \in \Lambda$ while strong stochastic dominance requires a strict inequality over some region of Λ, denoted by $\succ_s$. 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 $(\mathbf{w}_{a1}, ..., \mathbf{w}_{an_a})$ and $(\mathbf{w}_{b1}, ..., \mathbf{w}_{bn_b})$ be independently and identically distributed observations from the two respective populations with corresponding empirical distribution functions $\hat{G}_a$

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

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.

The above test procedure can be used to distinguish instances of weak and strong stochastic dominance by reversing the role of D_a^s and D_b^s in the null hypothesis. Thus, testing the two respective null hypotheses, $G_a \succeq_s G_b$ and $G_b \succeq_s G_a$, together produces four possible outcomes. First, when both null hypotheses are rejected we conclude that neither distribution stochastically dominates the other at order s. Second, when both null hypotheses fail to be rejected we conclude that the functions are statistically coincident. Third, when $H_0: G_a \succeq_s G_b$ is rejected but $H_0: G_b \succeq_s G_a$ is not rejected we conclude that distribution "*b*" strongly stochastically dominates distribution "*a*" at order s. Finally, when the pattern of rejection is reversed we conclude that distribution "*a*" strongly stochastically dominates distribution "*b*" at order s.

When 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

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

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 multivariate case, the above hypothesis asserts stochastic dominance with respect to marginal distributions (which may be multi-dimensional) without requiring stochastic dominance everywhere. 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

$\psi_x\left(\boldsymbol{\lambda}_x\right)=\max\left\{D_{a_x}^{s_x}\left(\boldsymbol{\lambda}_x\right)-D_{b_x}^{s_x}\left(\boldsymbol{\lambda}_x\right),\mathbf{0}\right\}$ and for $\boldsymbol{\lambda}_y \in \Lambda_y$, let

$\psi_y\left(\boldsymbol{\lambda}_y\right)=\max\left\{D_{b_y}^{s_y}\left(\boldsymbol{\lambda}_y\right)-D_{a_y}^{s_y}\left(\boldsymbol{\lambda}_y\right),\mathbf{0}\right\}$. Define

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

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 $SW_a=\int U\left(\mathbf{z}\right)\mathbf{d}G_a\left(\mathbf{z}\right)$ represent social welfare in population "*a*" where the function $U\left(\mathbf{z}\right)$ 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 $U\left(\mathbf{z}\right)$. For example, suppose a pair of bivariate distributions satisfy:

$$H_o : G_a \succeq_2 G_b \quad and \quad G_{a_x} \succeq_2 G_{b_x} \quad and \quad G_{a_y} \succeq_2 G_{b_y} .$$

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

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 $\psi_x(\lambda_x) = \max\left\{ D_{a_x}^2(\lambda_x) - D_{b_x}^2(\lambda_x), 0 \right\}$; for $\lambda_y \in \Lambda_y$, let $\psi_y(\lambda_y) = \max\left\{ D_{a_y}^2(\lambda_y) - D_{b_y}^2(\lambda_y), 0 \right\}$; and for $\boldsymbol{\lambda} \in \Lambda$, let $\psi(\boldsymbol{\lambda}) = \max\left\{ D_a^2(\boldsymbol{\lambda}) - D_b^2(\boldsymbol{\lambda}), \mathbf{0} \right\}$. Define

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

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). 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:

$$
\begin{aligned}
&H_0^A : G_a \succeq_1 G_b \\
&H_0^B : G_{a_x} \succeq_1 G_{b_x} \\
&H_0^C : G_{a_y} \succeq_1 G_{b_y} \\
&H_0^D : G_{a_x} \succeq_1 G_{b_x} \quad \mathit{and} \quad G_{b_y} \succeq_1 G_{a_y} \\
&H_0^E : G_{b_x} \succeq_1 G_{a_x} \quad \mathit{and} \quad G_{a_y} \succeq_1 G_{b_y} \quad .
\end{aligned}
$$

Under the first DGM the two distributions are identical. The parameter values of the underlying normal distribution are set to $(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \sigma_{xy}) = (0.85, 0.85, 0.36, 0.36, 0.2)$ 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 $(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \sigma_{xy})_a = (0.6, 0.6, 0.64, 0.64, 0.2)$. 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 $(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \sigma_{xy})_b = (0.85, 0.85, 0.36, 0.36, -0.2)$ 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 true; however, hypothesis A which tests bivariate stochastic dominance of "*a*" over "*b*" is false.[6] 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 DGM.

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,

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

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.

We now turn to a discussion of tests of hypothesis A, which 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. In this instance, 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 social welfare and poverty consequences of the associated differences in income and non-labour market time across the countries.

We compare the joint distributions of income and non-labour market time in Germany[7], the United Kingdom, and the United States using data from the Cross-National Equivalent File (CNEF)[8]. 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 years of comparison are 1983, 1990 and 2000 for Germany and the U.S. but only 1990 and 2000 for comparisons involving the U.K. (Consistent earlier data for the U.K. are not available in the CNEF.) We define leisure as the residual from the difference between total hours in a year and the reported number of annual hours spent in the formal labour market.

We focus on the welfare of individuals aged 25 years or older, using post-government income and leisure time as our indicators of well-being. Table 3 provides summary statistics for individuals 25 years of age or older. Not surprisingly, average income is highest in the U.S. for all three years of comparison and the variation in incomes is also greatest in the U.S. Germany displays the highest average annual hours of leisure time. The average number of leisure hours increased in Germany and the U.K. over time, but it decreased by over 200 hours in the U.S. between 1983 and 2000.

Figure 1 and Figure 2 present the empirical marginal distributions of income and leisure time, respectively, for individuals 25 years of age and older in Germany and the

[7] 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.
[8] Please refer to http://www.human.cornell.edu/pam/gsoep/equivfil.cfm for further information.

U.S. in 2000. The U.S. income distribution appears to be first-order stochastically dominant over most of the income support, although the empirical income distributions appear to cross at the lower end of the income support. The integral test allows us to check whether this crossing is statistically significant. As for the leisure distribution, the German distribution appears to first-order stochastically dominate that of the U.S. The distributions are discontinuous at the upper limit of leisure time, reflected by the sharp jumps in the plots, due to non-working individuals. Table 4 displays our results for tests of first- and second-order stochastic dominance of the income, leisure, and joint distributions for each possible cross-country comparison in the years 1983, 1990, and 2000 for individuals 25 years of age and older. The table reports the estimated p-value of each indicated hypothesis. Some consistent patterns emerge from the results. First, the German leisure distribution is strongly first-order stochastically dominant with respect to the U.K. and U.S. leisure distributions for each comparison while the U.K. leisure distribution is strongly first-order stochastically dominant with respect to the U.S. Second, the U.S. income distribution is strongly stochastically dominant at first-order except for in comparison to the U.K. in 2000, while the U.K. income distribution strongly first-order stochastically dominates the German distribution. Finally, the integral test strongly rejects first-order stochastic dominance of the bivariate surfaces for each comparison except for Germany and the U.K. in 1990. In this instance, see Figure 3, the German empirical bivariate distribution lies everywhere below that of the U.K. except along the marginal income distribution. The p-value of 0.110 for the null hypothesis of German stochastic dominance is indicative of the low power of the integral test under the alternative hypothesis when the violation of the null hypothesis is not very persistent, as discussed previously regarding our fourth simulation. This claim is supported by the strong rejection of German stochastic dominance at first-order when the integral test is applied strictly to the income distribution. Overall, the results of tests for first-order stochastic dominance suggest that bivariate social welfare cannot be robustly ranked for individuals 25 years of age and older.

Bivariate poverty dominance (see Duclos *et al.*, 2004) provides an alternative way to interpret the stochastic dominance results presented in Table 4. In particular, consider all bivariate poverty indices of the form $P(\lambda) = \iint_{\Lambda} \pi(x,y)\, dF(x,y)$ where Λ is the poverty set defined by a poverty frontier and $\pi(x,y)$ is an individual's contribution to poverty, given well-being indicators x and y (income and leisure in our case). The "poverty focus axiom" induces $\pi(x,y) \geq 0$ if $(x,y) \in \Lambda$ and $\pi(x,y) = 0$ otherwise.

Duclos *et al*. (2004) show the equivalence between first-order stochastic dominance and poverty rankings for all poverty indices with $\pi_x, \pi_y \leq 0$, $\pi_{xy} \geq 0$, and that are continuous along the poverty frontier, regardless of how the poverty frontier is defined along the bivariate support. Moreover, if we strengthen the conditions on $\pi(x,y)$ to include $\pi_{xx}, \pi_{yy} \geq 0$ (i.e., the poverty index obeys the transfer principle in both dimensions), $\pi_{xxy}, \pi_{yyx} \leq 0$ (i.e., the transfer principle is stronger in one dimension the lower then level of the other indicator), and $\pi_{xxyy} \geq 0$ then second-order stochastic dominance of the bivariate distributions implies robust poverty rankings for all poverty indices that feature these properties. As an example, this class of poverty indices includes the two-dimensional poverty gap measure, an extension of the FGT index (Foster *et al*., 1984) to two dimensions. Thus, based on the test results for bivariate second-order stochastic dominance, we conclude that bivariate poverty is robustly lower in Germany than in either the U.K. or the U.S. in the sense described above and is also lower in the U.K. as compared to the U.S. for individuals 25 years of age and older.

The above poverty orderings are quite general, as they are for *all* poverty frontiers that one could care to define within the bivariate support. Stronger poverty orderings, which place fewer restrictions on $\pi(x, y)$, are possible if the poverty frontier is restricted to lie within a sub-region of the support. The obvious drawback is that poverty frontiers that extend outside of this sub-region are excluded from the analysis. Table 5 displays our results of restricted first-order stochastic dominance tests for individuals 25 years of age and older. We restrict the domain to a lower-left quadrant based on an income upper limit of 25000, 32000, and 40000 USD for the years 1983, 1990, and 2000, respectively, and a time-invariant leisure upper limit of 7000 hours annually. Our results indicate that within these restricted supports the German bivariate distribution first-order stochastically dominates the U.K. and the U.S. bivariate distributions, while the U.K. joint distribution first-order stochastically dominates the U.S. in 2000.

We check the robustness of our main results by examining whether the differences in employment rates are heavily influencing our conclusions by restricting the data to only those individuals that reported positive hours of work. Table 6 displays summary statistics for all working individuals. Among all working individuals, average income and leisure tend to show the same pattern as for all individuals 25 years of age or older. Our results for tests of first- and second-order stochastic dominance among all working individuals are presented in Table 7. We find that first-order stochastic dominance of the bivariate distributions holds only once, in favour of the U.S. versus the U.K. in 1990. This result is puzzling, however, when one considers the results amongst the two marginal distributions. The integral test concludes that the U.S. income distribution strongly first-order stochastically dominates while the U.K. leisure distribution strongly first-order stochastically dominates. Hence, the null of bivariate stochastic dominance should be rejected in both directions. This contradiction is similar to that noted above when comparing the bivariate distributions of Germany and the U.K. in 1990 for all individuals 25 years of age or older. Other results suggest that the German leisure time distribution second-order stochastically dominates both the U.K. and U.S. distributions, the U.S. income distribution first-order stochastically dominates those of Germany and the U.K., and the U.K. is an intermediate case.

The second-order stochastic dominance results concerning the bivariate distributions reverse some of the poverty orderings arrived at previously for all individuals 25 years of age and older. German poverty remains lowest by all poverty indices with the properties associated with tests of second-order stochastic dominance outlined above, while U.S. bivariate poverty is robustly lower than in the U.K., contrasting the results for the previous group of individuals. These results hold for any poverty frontier. Stronger poverty orderings are presented in Table 8. The poverty rankings implied by these results mirror those suggested by tests of second-order stochastic dominance over the entire support: bivariate poverty is lowest in Germany, followed by the U.S.

An alternative check on our primary results is offered by looking only at single-person households, where the individuals are 25 years of age or older. For this population subset we need not be concerned with choice of income equivalence scale, nor with our implicit assumption thus far that there are no economies of scale associated with leisure time for a multi-person household. We present results for tests of first- and second-order stochastic dominance over the entire support in Table 10. We find that U.S. income is strongly first-order stochastically dominant with respect to Germany and second-order with respect to the U.K. Similarly, Germany is first-order stochastically dominant with respect to the U.S. for leisure and second-order in comparison to the U.K. Results of tests for bivariate second-order stochastic dominance also follow a similar pattern to those of our primary population group. Germany is stochastically dominant with respect to both the U.K. and the U.S., although only weakly in 2000 with the U.K., and the U.K. is stochastically dominant in comparison to the U.S. Our results for tests of restricted first-order stochastic dominance, see Table 11, likewise displays the same patterns as for all individuals 25 years of age or older. For singles 25 years or age or older, Germany is first-order stochastically dominant over the restricted bivariate surface, while the U.K. is first-order stochastically dominant in comparison to the U.S.

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 social welfare and poverty for individuals 25 years of age or older 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. 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.

Though robust rankings of social welfare are not possible for most comparisons, using stochastic dominance conditions we find bivariate poverty to be lowest in Germany, followed by the U.K. and then the U.S. However, the poverty orderings for the U.K. and the U.S. are sensitive to whether or not we are looking solely at workers or at the larger population. In particular, we find that the German bivariate distribution first-order stochastically dominates its counterparts over a restricted portion of the support. If one is uncomfortable with this region defining the largest possible poverty frontier, then one can employ weaker stochastic dominance conditions and perform poverty comparisons for a smaller class of poverty indices. In particular, we find that the German joint distribution second-order stochastically dominates its peers over the entire support, implying lower poverty for any possible poverty frontier, but for a smaller class of poverty indices than for the previous results.

Data Appendix

We remove all individuals for whom the “income” variable or the “annual hours worked” variable contains an invalid response.

INCOME:

We employ the CNEF series I11102XX for Germany and the U.K. and I11113XX for the U.S. as our measure of post-government household income, where XX refers to the year of the survey. We convert household income to per adult equivalent units by dividing total household income 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. Next, 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. The conversion factors are 2.1434, 1.8731, and 1.8162 DM per USD for Germany in 1983, 1990, and 2000 respectively, and 0.5688 and 0.6066 pounds per USD for the U.K. in 1990 and 2000.

HOURS WORKED:

For annual hours worked, we employ E11101XX. 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 Bach and Koch (2003). 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 worked, one using the Bach and Koch (2003) estimates of time away from work and the other using GSOEP estimates, were very similar. Given the large differences in leisure distributions across countries, the small difference between these two methods for estimating the German leisure distribution is unlikely to affect our conclusions.

WEIGHTING:

The GSOEP, BHPS, and PSID are stratified samples. To draw inferences on the respective populations we use weighting factors available in the CNEF. We employ W11101XX for Germany, the multiple of W11107XX and W11110XX for the U.K., and the multiple of W11101XX and W11104XX for the U.S.

References

Alesina R, Glaeser E, Sacerdote B. 2005. Work and leisure in the U.S. and Europe. Why so different? Harvard Institute of Economic Research, Discussion Paper 2068.

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

Atkinson AB, Bourguignon F. 1982. The comparison of multi-dimensioned distributions of economic status. *Review of Economic Studies* **49**: 183-201.

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

Barrett GF, Donald SG. 2003. Consistent tests for stochastic dominance. *Econometrica* **71**: 71-104.

Crawford I. 2005. A nonparametric test of stochastic dominance in multivariate distributions. University of Surrey, Discussion Papers in Economics, DP 12/05.

Davidson R, Duclos J-Y. 2000. Statistical inference for stochastic dominance and for the measurement of poverty and inequality. *Econometrica* **68**: 1435-1464.

Duclos J-Y, Sahn D, Younger SD. 2004. Robust multidimensional poverty comparisons. CIRPEE Working Paper No. 03-04.

Foster JE, Greer J, Thorbecke E. 1984. A class of decomposable poverty measures. *Econometrica*. **52**: 761-776.

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

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

Kaur A, Prakasa Rao BLS, Singh H. 1994. Testing for second-order stochastic dominance of 2 distributions. *Econometric Theory* **10**: 849-866.

Klecan L, McFadden R., McFadden D. 1991. A robust test for stochastic dominance. Mimeo, MIT.

Krugman, P. 2005. French family values. *The New York Times*, July 29.

Linton O, Maasoumi E, Whang Y-J. 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.

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

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

Schettkat R. 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 individuals 25 years of age and older

	1983	1990	2000
Germany			
Mean of income	8533	12916	17133
Mean of leisure time	7875	7904	7966
St. dev. of income	4927	7402	9500
St. dev. of leisure time	953	870	906
Correlation	-0.278	-0.285	-0.274
Percentage working	52.2	54.9	48.4
No. of observations	9549	7522	18692
United Kingdom			
Mean of income	n.a.	13945	21165
Mean of non-labour market time	n.a.	7575	7647
St. dev. of income	n.a.	7524	13811
St. dev. of non-labour market time	n.a.	1073	1063
Correlation	n.a.	-0.231	-0.159
Percentage working	n.a.	64.7	60.1
No. of observations	n.a.	6469	7224
United States			
Mean of income	11446	17046	26852
Mean of non-labour market time	7507	7415	7279
St. dev. of income	7985	13197	27521
St. dev. of non-labour market time	1053	1079	1073
Correlation	-0.193	-0.230	-0.120
Percentage working	70.1	73.4	75.1
No. of observations	10394	11563	11444

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

Table 4 – Estimated p-values for tests of stochastic dominance for individuals 25 years of age and older

Null hypothesis	First-order			Second-order		
	1983	1990	2000	1983	1990	2000
Germany and the U.K.						
$Ger_{income} \succeq_s UK_{income}$	n.a.	0.000	0.000	n.a.	0.000	0.000
$UK_{income} \succeq_s Ger_{income}$	n.a.	0.325	0.935	n.a.	0.545	0.690
$Ger_{leisure} \succeq_s UK_{leisure}$	n.a.	1.000	1.000	n.a.	1.000	1.000
$UK_{leisure} \succeq_s Ger_{leisure}$	n.a.	0.000	0.000	n.a.	0.000	0.000
$Ger \succeq_s UK$	n.a.	0.110	0.000	n.a.	0.955	0.995
$UK \succeq_s Ger$	n.a.	0.000	0.000	n.a.	0.000	0.000
Germany and the U.S.						
$Ger_{income} \succeq_s US_{income}$	0.000	0.000	0.000	0.000	0.000	0.000
$US_{income} \succeq_s Ger_{income}$	0.325	0.155	0.470	0.380	0.375	0.560
$Ger_{leisure} \succeq_s US_{leisure}$	1.000	1.000	1.000	1.000	1.000	1.000
$US_{leisure} \succeq_s Ger_{leisure}$	0.000	0.000	0.000	0.000	0.000	0.000
$Ger \succeq_s US$	0.000	0.000	0.000	1.000	1.000	1.000
$US \succeq_s Ger$	0.000	0.000	0.000	0.000	0.000	0.000
The U.K. and the U.S.						
$UK_{income} \succeq_s US_{income}$	n.a.	0.000	0.000	n.a.	0.000	0.000
$US_{income} \succeq_s UK_{income}$	n.a.	0.350	0.010	n.a.	0.455	0.440
$UK_{leisure} \succeq_s US_{leisure}$	n.a.	1.000	1.000	n.a.	0.680	1.000
$US_{leisure} \succeq_s UK_{leisure}$	n.a.	0.000	0.000	n.a.	0.000	0.000
$UK \succeq_s US$	n.a.	0.000	0.000	n.a.	0.355	1.000
$US \succeq_s UK$	n.a.	0.000	0.000	n.a.	0.055	0.000

Table 5 – Estimated p-values for tests of restricted first-order stochastic dominance for individuals 25 years of age and older

	1983	**1990**	**2000**
Income cut-off	25000	32000	40000
Leisure cut-off	7000	7000	7000
Null hypothesis		**P-value**	
$Ger \succeq_1 UK$	n.a.	0.985	0.990
$UK \succeq_1 Ger$	n.a.	0.000	0.000
$Ger \succeq_1 US$	1.000	1.000	1.000
$US \succeq_1 Ger$	0.000	0.000	0.000
$UK \succeq_1 US$	n.a.	0.040	1.000
$US \succeq_1 UK$	n.a.	0.000	0.000

Table 6 – Summary statistics for all working individuals

	1983	1990	2000
Germany			
Mean of income	9471	14488	19135
Mean of leisure time	7095	7219	7132
St. dev. of income	4908	8109	10197
St. dev. of leisure time	605	534	559
Correlation	-0.162	-0.161	-0.224
No. of observations	6557	5090	10150
United Kingdom			
Mean of income	n.a.	14615	21840
Mean of non-labour market time	n.a.	6970	6946
St. dev. of income	n.a.	7794	14584
St. dev. of non-labour market time	n.a.	793	742
Correlation	n.a.	-0.213	-0.175
No. of observations	n.a.	5122	5536
United States			
Mean of income	11709	17921	27609
Mean of non-labour market time	7083	6968	6898
St. dev. of income	7752	13281	24451
St. dev. of non-labour market time	826	820	808
Correlation	-0.202	-0.190	-0.110
No. of observations	9382	10077	10292

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

Table 7 – Estimated p-values for tests of stochastic dominance for all working individuals

Null hypothesis	First-order			Second-order		
	1983	1990	2000	1983	1990	2000
Germany and the U.K.						
$Ger_{income} \succeq_s UK_{income}$	n.a.	0.000	0.000	n.a.	0.095	0.000
$UK_{income} \succeq_s Ger_{income}$	n.a.	0.020	0.900	n.a.	0.270	0.740
$Ger_{leisure} \succeq_s UK_{leisure}$	n.a.	0.000	0.005	n.a.	1.000	1.000
$UK_{leisure} \succeq_s Ger_{leisure}$	n.a.	0.000	0.000	n.a.	0.000	0.000
$Ger \succeq_s UK$	n.a.	0.000	0.000	n.a.	0.985	0.170
$UK \succeq_s Ger$	n.a.	0.000	0.000	n.a.	0.000	0.000
Germany and the U.S.						
$Ger_{income} \succeq_s US_{income}$	0.000	0.000	0.000	0.000	0.000	0.000
$US_{income} \succeq_s Ger_{income}$	0.105	0.100	0.430	0.400	0.430	0.515
$Ger_{leisure} \succeq_s US_{leisure}$	0.000	0.000	0.000	1.000	1.000	1.000
$US_{leisure} \succeq_s Ger_{leisure}$	0.000	0.000	0.000	0.000	0.000	0.000
$Ger \succeq_s US$	0.000	0.000	0.000	0.000	0.370	0.075
$US \succeq_s Ger$	0.000	0.000	0.000	0.015	0.000	0.000
The U.K. and the U.S.						
$UK_{income} \succeq_s US_{income}$	n.a.	0.000	0.000	n.a.	0.000	0.000
$US_{income} \succeq_s UK_{income}$	n.a.	0.780	0.385	n.a.	0.570	0.475
$UK_{leisure} \succeq_s US_{leisure}$	n.a.	0.140	0.225	n.a.	0.585	1.000
$US_{leisure} \succeq_s UK_{leisure}$	n.a.	0.000	0.000	n.a.	0.205	0.000
$UK \succeq_s US$	n.a.	0.000	0.000	n.a.	0.000	0.005
$US \succeq_s UK$	n.a.	0.110	0.005	n.a.	0.825	0.235

Table 8– Estimated p-values for tests of restricted first-order stochastic dominance for all working individuals

	1983	**1990**	**2000**
Income cut-off	25000	32000	40000
Leisure cut-off	7000	7000	7000
Null hypothesis		**P-value**	
$Ger \succeq_1 UK$	n.a.	1.000	0.985
$UK \succeq_1 Ger$	n.a.	0.000	0.000
$Ger \succeq_1 US$	0.945	1.000	1.000
$US \succeq_1 Ger$	0.000	0.000	0.000
$UK \succeq_1 US$	n.a.	0.000	0.045
$US \succeq_1 UK$	n.a.	0.415	0.115

Table 9 – Summary statistics for all singles 25 years of age and older

	1983	1990	2000
Germany			
Mean of income	8998	13914	17460
Mean of non-labour market time	8121	8072	8084
St. dev. of income	5937	8777	11219
St. dev. of non-labour market time	886	867	898
Correlation	-0.444	-0.462	-0.392
Percentage working	37.1	41.8	39.6
No. of observations	1166	918	2712
United Kingdom			
Mean of income	n.a.	14500	23008
Mean of non-labour market time	n.a.	8074	8082
St. dev. of income	n.a.	7408	18713
St. dev. of non-labour market time	n.a.	1015	1015
Correlation	n.a.	-0.433	-0.211
Percentage working	n.a.	35.5	34.3
No. of observations	n.a.	930	1082
United States			
Mean of income	11940	16543	26973
Mean of non-labour market time	7674	7567	7389
St. dev. of income	7918	13881	25931
St. dev. of non-labour market time	1051	1124	1105
Correlation	-0.435	-0.415	-0.270
Percentage working	61.3	60.2	69.1
No. of observations	1313	1653	1432

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

Table 10 – Estimated p-values for tests of stochastic dominance for all singles 25 years of age or older

Null hypothesis	First-order 1983	First-order 1990	First-order 2000	Second-order 1983	Second-order 1990	Second-order 2000
Germany and the U.K.						
$Ger_{income} \succeq_s UK_{income}$	n.a.	0.000	0.000	n.a.	0.025	0.000
$UK_{income} \succeq_s Ger_{income}$	n.a.	0.735	0.825	n.a.	1.000	0.680
$Ger_{leisure} \succeq_s UK_{leisure}$	n.a.	0.040	0.015	n.a.	0.745	1.000
$UK_{leisure} \succeq_s Ger_{leisure}$	n.a.	0.020	0.005	n.a.	0.020	0.015
$Ger \succeq_s UK$	n.a.	0.010	0.000	n.a.	0.555	0.145
$UK \succeq_s Ger$	n.a.	0.030	0.005	n.a.	0.040	0.130
Germany and the U.S.						
$Ger_{income} \succeq_s US_{income}$	0.000	0.000	0.000	0.000	0.000	0.000
$US_{income} \succeq_s Ger_{income}$	0.830	0.170	0.800	0.640	0.375	0.605
$Ger_{leisure} \succeq_s US_{leisure}$	1.000	1.000	1.000	1.000	1.000	1.000
$US_{leisure} \succeq_s Ger_{leisure}$	0.000	0.000	0.000	0.000	0.000	0.000
$Ger \succeq_s US$	0.045	0.205	0.055	1.000	1.000	1.000
$US \succeq_s Ger$	0.000	0.000	0.000	0.000	0.000	0.000
The U.K. and the U.S.						
$UK_{income} \succeq_s US_{income}$	n.a.	0.000	0.000	n.a.	0.005	0.000
$US_{income} \succeq_s UK_{income}$	n.a.	0.005	0.035	n.a.	0.235	0.410
$UK_{leisure} \succeq_s US_{leisure}$	n.a.	1.000	1.000	n.a.	1.000	1.000
$US_{leisure} \succeq_s UK_{leisure}$	n.a.	0.000	0.000	n.a.	0.000	0.000
$UK \succeq_s US$	n.a.	0.275	0.340	n.a.	1.000	1.000
$US \succeq_s UK$	n.a.	0.000	0.000	n.a.	0.000	0.000

Table 11– Estimated p-values for tests of restricted first-order stochastic dominance for singles 25 years of age and older

	1983	**1990**	**2000**
Income cut-off	25000	32000	40000
Leisure cut-off	7000	7000	7000
H_0		**P-value**	
$Ger \succeq_1 UK$	n.a.	0.965	0.955
$UK \succeq_1 Ger$	n.a.	0.000	0.000
$Ger \succeq_1 US$	1.000	1.000	1.000
$US \succeq_1 Ger$	0.000	0.000	0.000
$UK \succeq_1 US$	n.a.	0.930	1.000
$US \succeq_1 UK$	n.a.	0.010	0.015

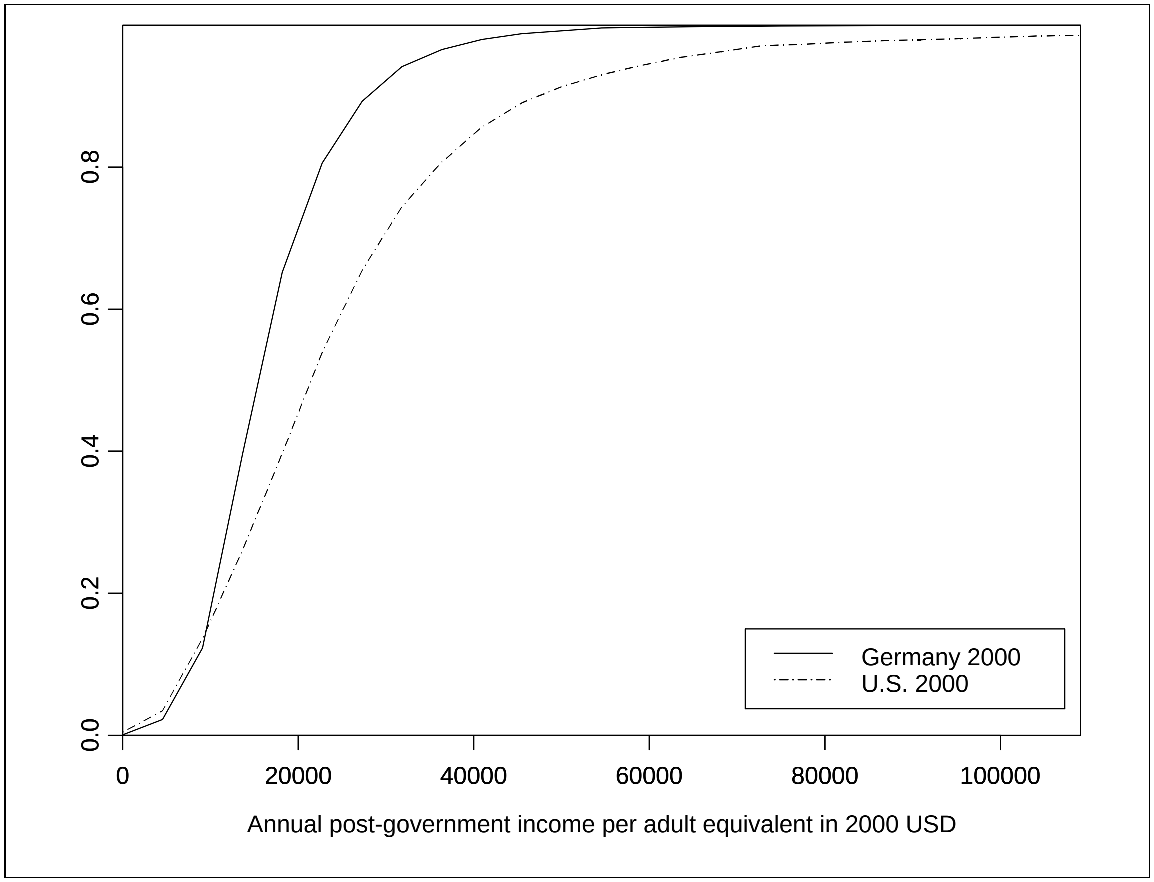

Figure 1 - Empirical income distributions for individuals 25 years of age and older in Germany and the U.S., 2000

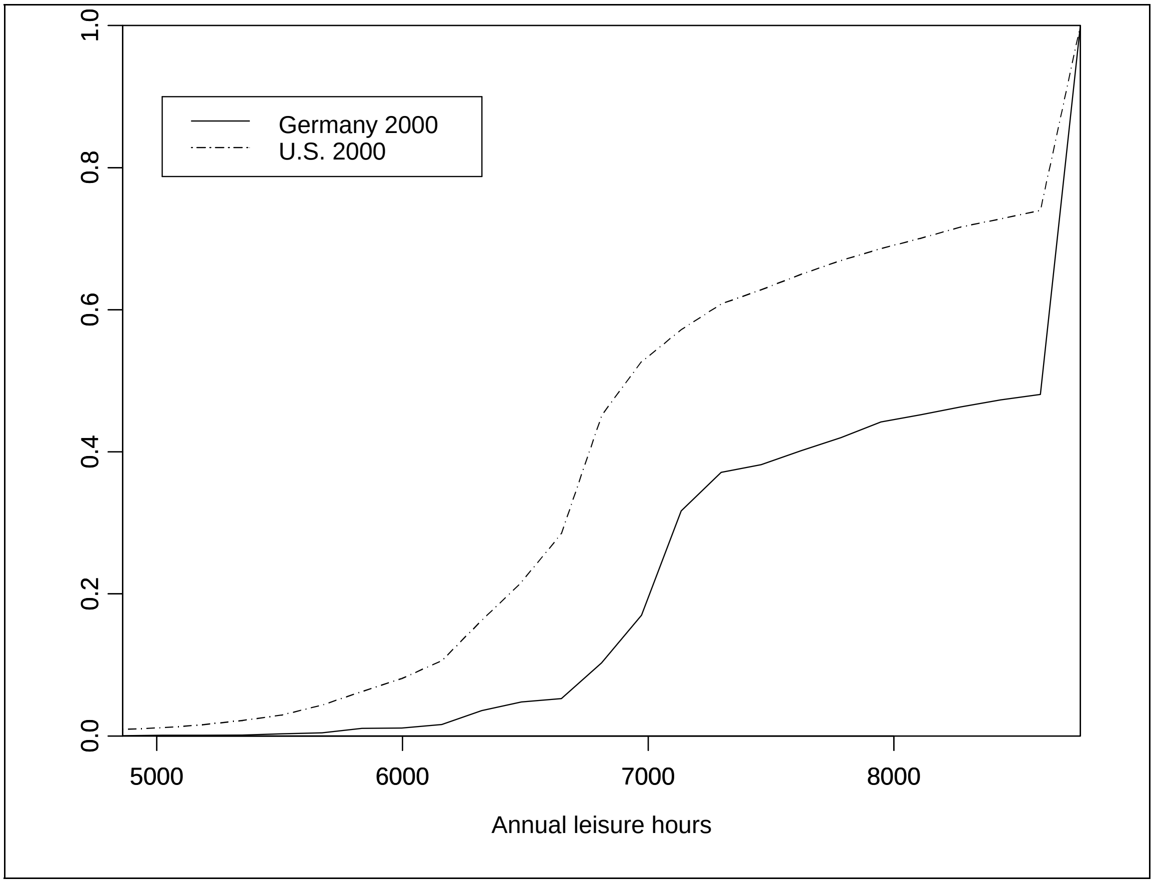

Figure 2 - Empirical leisure time distributions for individuals 25 years of age and older in Germany and the U.S., 2000

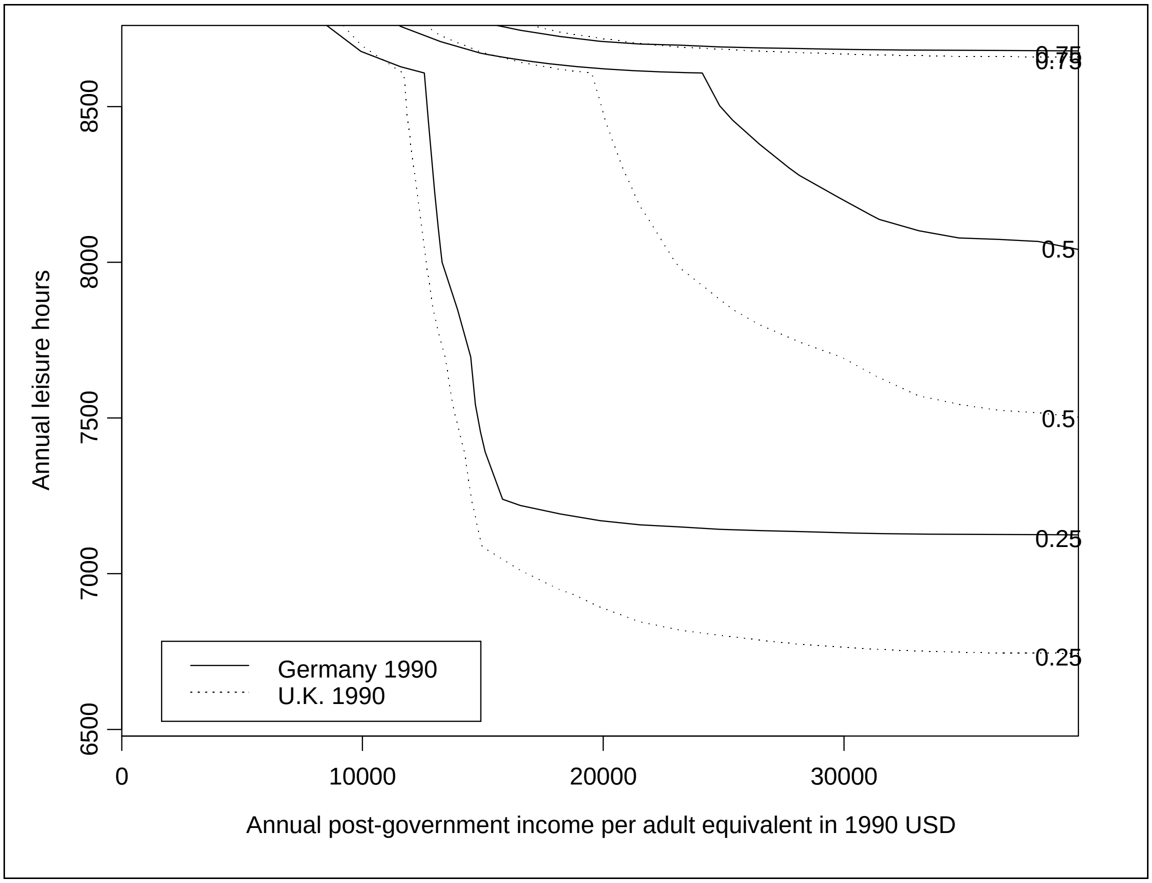

Figure 3 – Contour plots of the empirical CDFs of income and leisure for individuals 25 years of age and older in Germany and the U.K., 1990