"BEAUTY IS THE PROMISE OF HAPPINESS"?

Daniel S. Hamermesh
Jason Abrevaya

"Beauty Is the Promise of Happiness"?
Daniel S. Hamermesh and Jason Abrevaya
NBER Working Paper No. 17327
August 2011
JEL No. C20,I30,J10

ABSTRACT

We measure the impact of individuals' looks on life satisfaction/happiness. Using five data sets, from the U.S., Canada, the U.K., and Germany, we construct beauty measures in different ways that allow placing lower bounds on the effects of beauty. Beauty raises happiness: A one standard-deviation change in beauty generates about 0.10 standard deviations of additional satisfaction/happiness among men, 0.12 among women. Accounting for a wide variety of covariates, particularly effects in the labor and marriage markets, including those that might be affected by differences in beauty, the impact among men is more than halved, among women slightly less than halved.

Daniel S. Hamermesh
Department of Economics
University of Texas
Austin, TX 78712-1173
and NBER
hamermes@eco.utexas.edu

Jason Abrevaya
Department of Economics
University of Texas
Speedway Avenue
Austin, TX 78712-1173 USA
abrevaya@eco.utexas.edu

I. Introduction

While economists have studied happiness for several generations (Easterlin, 2010; Scitovsky, 1976), interest in it has burgeoned in the last 15 years. The Frey and Stutzer (2002) survey captured part of the literature, but there has been a continuing outpouring of research on happiness from an economic viewpoint (e.g., Clark *et al*, 2008; Stevenson and Wolfers, 2008; Deaton and Kahneman, 2010; Oswald and Wu, 2011). Much of the analysis focuses on measuring the short- and long-run effects of changes in income on happiness, but the relation of happiness to other outcomes that are at least partly economically determined (divorce, fertility and others) has also been subject to discussion.

At the same time a smaller, but also burgeoning literature on the effects of beauty on various outcomes has been created (e.g., Hamermesh and Biddle, 1994; Möbius and Rosenblat, 2006; Mocan and Tekin, 2010; Hitsch *et al*, 2010). In these studies the economic focus is on such topics as how beauty is traded for income, how it alters occupational choice, and how it affects marital bargaining. The general issue is how human beauty determines outcomes in various markets and shifts the distribution of surpluses among participants in those markets.

Here we put the two literatures together, examining how happiness is affected by beauty. Some psychologists have correlated subjects' happiness and their self-assessed beauty, but that approach seems flawed. Others have compared the happiness of college students to observers' ratings of their looks (Mathes and Kahn, 1975; Diener *et al*, 1995); examined simple averages of several measures of happiness among a random sample of people whose beauty was rated by interviewers (using one of the data sets we use, Umberson and Hughes, 1987); and offered partial correlations of happiness measures and survey respondents' waist-to-hip ratio, used as a proxy for beauty (Plaut *et al*, 2009).

The analysis should not merely reflect the idiosyncrasies of measuring the subjective concepts of happiness and human beauty. For that reason we use five sets of surveys from four different countries, which we discuss in Section II. The measures of satisfaction/happiness differ across surveys, and even within a given survey different measures are generated. Given evidence of the sensitivity of responses to questions about happiness to the framing and scaling of the questions (Conti and Pudney, 2011), using

these variously constructed measures should minimize concerns over survey-based idiosyncrasies in eliciting responses about happiness. Finally, while happiness is obviously self-rated by the respondents, none of the beauty measures is—we are not relating a person's subjective assessment of one aspect of life to his/her assessment of another (Hamermesh, 2004).

Were it possible to find a convincing instrument for beauty or to undertake some experiment that altered beauty, we would take one of those analytical approaches to studying happiness. The former is not available, and the latter would not seem ethical. Absent them, we take advantage of the fact that the surveys use four different approaches to measuring beauty and in Section III delineate the types of measurement errors implicit in each approach and their implications for inferring the impact of beauty on happiness. These considerations enable us to place a lower bound on this impact. In the end, the validity of our results, which we present in Sections IV and V, depends on their robustness to differing approaches to measuring beauty and to eliciting people's expressions of satisfaction/happiness.

There are two main general mechanisms through which people's looks could affect their satisfaction/happiness. The first is through the many channels that have been shown in the beauty literature to offer routes by which beauty affects economic outcomes. These **indirect effects** may be at least as important as the **direct effects** of beauty on satisfaction/happiness—the halo that good looks might impart to a person independent of the effects of beauty on any market-related outcomes. In the main economic exercise in this study, presented in Section VI, we decompose any measured impact of beauty on satisfaction/happiness into its direct and indirect components and thus examine the extent to which any beauty-happiness relation works through markets.

II. Data Sources and Descriptive Statistics

The five sets of data that we use are especially diverse in terms of their methods of assessing beauty. The first consists of the two cross sections of the Quality of American Life (QAL) surveys, undertaken in 1971 and 1978 as random samples of the U.S. population age 18 plus. At the end of the interview in each of these surveys, the interviewer assessed the interviewee's looks on a five-to-one scale, with 5 being strikingly handsome or beautiful, 1 being homely. The complete list of descriptions

2

associated with each rating of beauty is shown in the first column of the first panel of Table 1. This measure has been used in a variety of studies linking beauty to economic outcomes (e.g., Hamermesh and Biddle, 1994; Leigh and Susilo, 2009), although typically with the top two categories combined into an indicator "good looks" and the bottom two combined into "bad looks," because of the paucity of respondents rated 5 or 1.

Both QAL surveys provide the same measures of happiness, each on a three-to-one scale, as the description in column (3) of Table 1 shows. The surveys also provide direct measures of life satisfaction, focused on the current moment and on the person's total experience, measures that are standard in the literature. Henceforth we distinguish between the determinants of life satisfaction and those of happiness. The analysis of these surveys is thus data-driven, so that we are not inquiring into the various aspects of satisfaction/happiness that have been identified by psychologists (e.g., Seligman, 2004), but merely using general expressions of happiness, as has become standard in economics.

The Quality of Life (QOL) survey was a longitudinal study conducted in Canada biennially from 1977 through 1981, randomly sampling Canadians ages 18 plus in 1977. In each of the three waves a wide array of subjective information was obtained. As in the QAL, at the end of each interview the interviewer rated the subject's looks using the same five-point rating system. Interviewers differed across the years, so that for those participants who remained in the study for all three years we have three independent measures of their beauty. The satisfaction measures use different wording and a different scale from those in the QAL, while the happiness measure is similar.

The German contribution to the General Social Survey program in 2008, the ALLBUS (*Allgemeine Bevölkerungsumfrage der Sozialwissenschaften*), included measures of beauty and happiness. As in the QAL and QOL the interviewer rated the subject's looks at the end of the interview (in this case, on an eleven-point scale); but the interviewer also provided a rating (on the same scale) at the very start of the interview—upon first contact with the subject. The survey also obtained a four-point rating of happiness, more backward-looking than the happiness measures in the QAL and QOL. We use data on all respondents ages 18 plus for whom information is available on the crucial variables.

The Wisconsin Longitudinal Survey (WLS), a study of a cohort of high-school graduates from 1957, also contains measures of beauty and happiness. Unlike the previous three sets of studies, beauty in the WLS was based on assessments of high-school graduation photographs of the participants. In 2004 each respondent's picture was rated by 12 individuals (6 men and 6 women) aged 63 to 91. The ratings were unit-normalized within a given rater and averaged over raters. Each respondent was interviewed in 1992 and 2004 (at ages 53 and 65) and asked how many days last week s/he was happy, how many days s/he enjoyed life, and how many days s/he was sad. The happiness measures were thus obtained 35 and 47 years after the photographs from which the respondents' beauty was rated were taken.[1]

The fifth data set is the British National Child Development Study (NCDS), a longitudinal examination of all Britons born March 3-9, 1958. At age 7, and again at age 11, each student's teacher assessed his/her attractiveness during the school year, along a scale shown in column (1) of Table 1. We aggregated these into the three categories, good-looking, average-looking and unattractive, similar to previous work relating these ratings to subsequent earnings (Harper, 2000). In various later waves of the survey, including 1991, 1999, 2004 and 2009 (at ages 33, 41, 46 and 51), the remaining respondents were asked questions designed to elicit their happiness/life satisfaction, some which have been studied before using these data (e.g., Blanchflower and Oswald, 2008). In the three most recent waves life satisfaction was elicited in a question (column (2) of Table 1) focusing on the respondent's entire life experience. Happiness at age 51 was also measured in a backward-looking manner, while happiness at age 33 was measured with reference to the respondent's current situation only.

In Appendix Tables 1a-1e we present descriptive statistics for the five sets of surveys. For each of the interview studies (the QAL, QOL and ALLBUS) here and in subsequent sections all the results are calculated using sample weights. Consider first the QAL. As is usual in assessing beauty, more people are rated in the top two categories than in the bottom two; and the majority are rated as average-looking. Also as is usual, women are rated more extremely than men (Hamermesh, 2011, Chapter 2). Consistent

---

[1]Other studies have assessed beauty from school pictures taken nearly two decades before the outcome linked to the assessments (Biddle and Hamermesh, 1998), and one study even showed a high correlation between the assessments of pictures of 10-year-olds and those of the same individuals at age 50 (Hatfield and Sprecher, 1986, p. 283).

with the previous satisfaction/happiness literature, most people are fairly happy and satisfied. The beauty measures in the QOL, shown in Appendix Table 1b, look remarkably similar qualitatively to those in the QAL, except for a lesser Canadian willingness to classify subjects as below-average looking. Also, the gender differences are reversed from those in the QAL. For the ALLBUS, shown in Appendix Table 1c, the crucial thing to note is that the ratings of both men's and women's beauty are higher at the end of the interview than at the start, although men are on average rated lower at both times.

Because the beauty measures in the WLS were normed, we do not list them in Appendix Table 1d. In these data, people report being happy on most days (on average, between 5 and 6 days) in the week before the survey. The number of days reported as being sad is typically 20 percent or fewer than the number of happy days. With one exception—number of days reported happy in the 1992 wave of the survey—male respondents are happier than females. Appendix Table 1e shows that in the NCDS females' looks (in this case at age 11) were rated more extremely than males'. Perhaps, however, because of their close acquaintance with their charges, the teachers who rated the students' attractiveness included more students in the attractive (good-looking) category than in the excluded category (children viewed as neither attractive nor unattractive). Most of the respondents were fairly happy or satisfied at ages 33-51.

There is no consistent gender difference in average satisfaction/happiness across the sets of surveys. In the QAL the comparisons are mixed; in the QOL and the ALLBUS women are more satisfied/happier, while the opposite is true in the WLS and NCDS. The differences in the nature of the measures across the surveys make them non-comparable along this dimension; but considering them together underscores the benefits of using various different measures of satisfaction/happiness to prevent incorrect generalizations based on only one sample or one type of question.

III. Measuring Beauty in Relation to Happiness

To understand the nature of the measurement difficulties in these data sets, Figure 1 presents the timing of the assessment of the respondent's beauty in relation to the elicitation of his/her satisfaction/happiness. A negative denotes that beauty is assessed after the respondent answers question(s) about his/her satisfaction/happiness; and the widths of the bars are in proportion to the square

root of the number of people rating the subject's beauty. Obviously there is no universal measure of human beauty—it is in the eye of the beholder. But a huge literature (summarized in Hamermesh, 2011, Chapter 2) shows that there is substantial agreement by people about each other's looks. The best possible measure would average ratings by large numbers of individuals who have no physical contact with a set of subjects who are dressed the same way and have the same standard facial expression. Since that kind of measure has not been obtained in any study we know of, we are thrown back on thinking about how the measures in these data sets generate errors in inferring the impacts of beauty.

To focus only on the beauty rating, consider the following simple linear regression model, with H_t (satisfaction/happiness) as the dependent variable and true (latent) beauty B_t^* as the sole explanatory variable:[2]

$$H_t = \alpha + \beta B_t^* + \varepsilon, \ E(\varepsilon) = 0, \ Var(\varepsilon) = \sigma_\varepsilon^2. \tag{1}$$

The subscript t indicates the time at which the happiness measure is observed.

We consider three possible types of difficulty in measuring beauty in relation to happiness:

(1) *Classical measurement error in the beauty rating*: The beauty rating used in the actual regression is an imperfect measure of B_t^*, because of the small number of raters.

(2) *Attenuation in the accuracy in the beauty rating*: Since beauty changes, albeit slowly, over time, the inherent noise in the beauty rating will be larger the more that the rating pre-dates the satisfaction/happiness measure. The variance will be an increasing function of the time interval between observation of the beauty rating and observation of the happiness measure.

(3) *Bias in the beauty rating*: If the beauty rating is elicited after the rater has spent time interacting with the subject, we would expect a positive correlation between the beauty rating and the unobservable component ε of the happiness outcome. For instance, an interviewer might have a

---

[2]To simplify notation we assume homoskedasticity throughout this section and therefore omit conditioning on B_t^*.

better opinion of a subject's beauty if the subject projects self- confidence in the interview, which

might occur if the subject is happier.

The following stylized model for the observed beauty rating incorporates each of these three possible

sources of difficulty:

$$B_s = B_t^* + v_{t-s} + \eta, \tag{2}$$

where $s \leq t$ is the time at which the beauty rating B_s is obtained. The attenuation component of the

measurement error is v_{t-s}, which has a variance assumed to be linear in the time interval $(t-s)$:

$$Var(v_{t-s}) = (t-s)\sigma_v^2, \ Cov(v_{t-s}, \varepsilon) = 0. \tag{3}$$

The other component of the error, denoted η, is similar to a classical measurement error, except that we

allow it to be correlated with the happiness residual ε:

$$Var(\eta) = \sigma_\eta^2, \ Cov(\eta, \varepsilon) = \sigma_{\eta\varepsilon}, \ Cov(\eta, v_{t-s}) = 0. \tag{4}$$

For this general model, the inconsistency of the least-squares estimator is given by the probability limit of

the slope estimate:

$$plim \ \hat{\beta} = \frac{Cov(B_s, H_t)}{Var(B_s)} = \frac{Cov(B_t^* + v_{t-s} + \eta, \alpha + \beta B_t^* + \varepsilon)}{Var(B_t^* + v_{t-s} + \eta)}$$

$$= \beta \left(\frac{\sigma_{B^*}^2 + \frac{\sigma_{\eta\varepsilon}}{\beta}}{\sigma_{B^*}^2 + (t-s)\sigma_v^2 + \sigma_\eta^2} \right), \tag{5}$$

where $\sigma_{B^*}^2$ denotes the variance of B_t^*. (Note that the textbook case of classical-measurement error is a

special case of this formula corresponding to $\sigma_{\eta\varepsilon} = 0$ (no rating bias) and $s = t$ (no depreciation

effect), for which $plim \ \hat{\beta} = \beta \left(\frac{\sigma_{B^*}^2}{\sigma_{B^*}^2 + \sigma_\eta^2} \right)$.)

For the QAL, QOL and the ALLBUS(end) data, the beauty rating is provided by the interviewer

slightly after the satisfaction/happiness measures are elicited. (Below the listing for each data set in Table

1 we present the type of measurement error contained in the beauty rating.) There is no depreciation

effect, since $s = t$, but the interview format leads to the possibility of a bias in the beauty rating ($\sigma_{\eta\varepsilon} > 0$). In this case the probability limit of the slope estimate simplifies to

$$plim\ \hat{\beta} = \beta \left(\frac{\sigma_{B*}^2 + \frac{\sigma_{\eta\varepsilon}}{\beta}}{\sigma_{B*}^2 + \sigma_\eta^2} \right). \tag{6}$$

If β is positive, the usual inconsistency associated with classical measurement error is opposite the inconsistency associated with the beauty-rating bias. The overall direction of the inconsistency depends on whether $\frac{\sigma_{\eta\varepsilon}}{\beta} > \sigma_\eta^2$ (upward inconsistency) or $\frac{\sigma_{\eta\varepsilon}}{\beta} < \sigma_\eta^2$ (downward inconsistency). In regressions based on these data sets the estimated β contains measurement errors of Types 1 and 3.[3]

Estimates based on the ALLBUS(start), in which the interviewer's rating of the respondent's attractiveness is obtained at the very start of the interview, avoid Type 3 measurement error. In this case the classical errors-in-variables result that $plim\ \hat{\beta} = \beta \left(\frac{\sigma_{B*}^2}{\sigma_{B*}^2 + \sigma_\eta^2} \right)$ is all that remains; and it occurs because only one person (the interviewer) assesses the subject's looks.

In the WLS data the beauty rating is based upon a subject's high-school picture, and the happiness measure is elicited during late adulthood. There will be an attenuation effect, since $s < t$, and the lack of interaction between the rater and the subject eliminates concerns about beauty-rating bias. Entering $\sigma_{\eta\varepsilon} = 0$ into the general formula yields:

$$plim\ \hat{\beta} = \beta \left(\frac{\sigma_{B*}^2}{\sigma_{B*}^2 + (t-s)\sigma_v^2 + \sigma_\eta^2} \right). \tag{7}$$

Inconsistency is expected here due to mis-measurement and attenuation, although the Type 1 measurement error is minimized by the large numbers of raters of the photographs. The estimated beauty slope from the WLS regressions should therefore be considered as "too low" (a lower bound to the true effect).

---

[3]In all of these studies the interviewer assessed the subject's looks very near the end of the interview, and substantially after the subject's satisfaction/happiness was elicited. It is very unlikely that the subject's specific response to happiness questions directly affected the beauty rating.

Finally, for the NCDS dataset, all three difficulties could arise, since the beauty ratings were assessed in childhood ($s < t$) by only two teachers, both of whom were very familiar with the subject ($\sigma_{\eta\varepsilon} > 0$). As a result, the general probability-limit formula in (4) would apply. As with the QAL, QOL and ALLBUS(end) data, the beauty-rating bias acts in the opposite direction from the measurement error. The impact of the other errors here, however, would be expected to be much larger than in those data sets, since beauty is assessed in the NCDS decades before the expression of happiness is elicited (as captured by the attenuation bias $(t - s)\sigma_v^2$). Although it is difficult to say how the sizes of $\sigma_{\eta\varepsilon}$ would compare between the NCDS and those interview-based data sets, the large difference in the variance of measurement error between the two studies suggests that the probability limits for the NCDS estimates would be lower than the true effects.

IV. Basic Results

Here we estimate linear models relating life satisfaction/happiness to beauty in each of the five data sets. For each we first include as regressors only the beauty measure(s) and, in the QAL, QOL and ALLBUS, a quadratic in age and a measure of race/ethnicity, which might affect happiness but which cannot be caused by differences in beauty. Then we add a number of covariates that have been shown to affect happiness but may not mediate the effect of beauty on satisfaction/happiness. In the next section we report on a large number of robustness checks that include varieties of additional controls, alternative beauty measures and more complex estimation procedures.

Table 2a presents the estimates based on the two QAL surveys. Among women all the coefficients have the expected signs—positive on the indicator for good looks (above-average or beautiful—the upper third of looks), negative on the indicator for bad looks (below-average or homely—the bottom eighth of looks). This is true whether or not we control for age, education, race, number of children and marital status.[4] Indeed, the addition of the vector of controls hardly alters the point estimates

---

[4]Whether we should be controlling for marital status here is unclear. There is substantial evidence that married people are happier (e.g., Blanchflower and Oswald, 2004; Oswald and Wu, 2011); but one's gains from marriage are affected by one's looks (Hamermesh and Biddle, 1994). In the data sets used here, however, there is mixed

of the coefficients among women; and nearly all the estimates are statistically significantly different from

zero.   Among men almost all of the point estimates have the expected sign, and they are generally

statistically significant in the 1978 data. As with women, adding the vector of controls does not greatly

alter the point estimates.   The effects of differences in beauty on life satisfaction or happiness are not

small, at least in the 1978 data. Using the estimates from Specification 2, going from the bottom eighth

of women's (men's) looks (those rated below-average) to the top third (those rated above-average) raises

satisfaction with life by 0.45 (0.48) standard deviations; the effects on happiness of this difference in

beauty are 0.38 (0.48) standard deviations. The impacts of differences in beauty in the 1971 data are

smaller, but still average about 0.20 standard deviations.

The QOL results, shown in in Table 2b, are qualitatively similar to those of the QAL. Almost all

the effects are in the expected directions, and the negative impact on satisfaction/happiness of being

among the small fraction of Canadians classified as being below-average in looks is substantial. There is

no obvious gender difference in the impacts of beauty. Adding indicators of education, marital status and

number of children has little effect on the estimates. The estimated effects are even larger than those in

the QAL: Going from the bottom twelfth of women's (men's) looks to the top third raises life satisfaction

by 0.36 (0.45) standard deviations, and raises happiness by 0.64 (0.75) standard deviations.

Table 2c presents the estimates of the impacts of beauty—the ALLBUS(start) and ALLBUS(end)

measures—on happiness. Specification 2 adds indicators of educational attainment, of marital status and

of partnered status.[5] Increases in the eleven-point beauty rating have significant positive effects on

happiness in all cases, although adding the covariates typically reduces the impacts by about one-third.

The effects are slightly smaller among men than among women, but the gender differences are not

significant statistically. Most interestingly, and in line with the discussion of measurement error, the

_____

evidence on the relationship between beauty and the probability of being married, with better-looking people being
significantly more likely to be married in the NCDS, but with little evidence of any effect in the other studies.

[5]The education categories are other, "*mittlere Reife*," and "*Hochschul,*" each accounting for about one-third of each
sample. We include a separate indicator of life partner here but nowhere else, because nearly 10 percent of the
sample reported being unmarried but having a life partner.

effects are smaller when we use the ALLBUS(start) ratings, with the difference being greater among men. Picking the same percentile points as in the distribution of looks in the QAL and moving from the equivalent of the median below-average looking woman (man) to the median above-average looking woman (man) produces an increase in happiness of 0.31 (0.23) standard deviations based on ALLBUS(start), and 0.34 (0.31) standard deviations based on ALLBUS(end). The former is somewhat smaller than in the QAL or QOL, perhaps because using ALLBUS(start) vitiates what we have denoted as Type 3 measurement error.[6]

The results from the WLS, with number of days happy, enjoyed and/or sad, are presented in Table 2d. The upper part of the table contains results from equations including only the unit-normal measure of beauty, while Specification 2 in the bottom part adds years of education, marital status, number of children, BMI observed at high-school graduation, and current BMI. As with the results for the interview data sets, adding this vector of covariates hardly alters the estimated impacts of attractiveness on the measures of satisfaction/happiness. There is no significant impact of attractiveness on happiness among men at either of the two ages at which these adults are observed. Among women, however, in all the equations the more attractive respondents are significantly happier at age 53 than less attractive respondents. The impacts are smaller relative to the standard deviations of satisfaction/happiness than in the interview data sets, suggesting the importance of measurement error arising from attenuation.

We can explain the disappearance of the results for women as they age by the possibility that the correlation of attractiveness at age 18 with attractiveness at age 53 may be greater than that with attractiveness at age 65. The absence of any relation between attractiveness and happiness among men is harder to explain, especially in light of the fact that the labor-market effects of beauty are at least as large among men as among women. One possibility is that there is inherently more measurement error in the

---

[6]The eleven-point scale in these beauty ratings allows us to compare how well different pairs of indicators of good- and bad looks describe happiness. While no pair performs as well as the eleven-point measure, defining good looks as the top third of respondents on the eleven-point scale, and bad looks as the bottom sixth, roughly the divisions shown in the other interview data, describes happiness in these samples better (in terms of adjusted R-squared) than any other pair of indicators.

ratings (assigned over 40 years after the pictures were taken) of men's high-school graduation pictures than of women's.

Table 2e shows the results of relating measures of happiness and satisfaction in adulthood in the NCDS sample to attractiveness as assessed by a child's teacher at age 11. The first part of the table includes only indicators for being rated as attractive or as unattractive (with a middle category excluded). All of the estimated impacts that are statistically significant are of the expected sign, and there is no obvious difference in the size or significance of the effects between men and women. The second part of Table 2e reports the estimates of the impacts of the beauty measures when indicators for educational attainment, marital status, number of children, BMI at age 11 and current BMI are added to the equations.[7] The estimated effects of attractiveness are typically somewhat attenuated when the control variables are added, although the overall conclusions remain the same: Where significantly nonzero, the beauty measures have the expected effects; and, as in the upper part of the table, the impacts of beauty are roughly the same by gender.[8]

There are a large number of estimates of the impact of looks here—30 coefficients for each gender for each of Specifications 1 and 2. Among men in all the samples taken together, in Specification 1(2) 27(27) of the 30 estimated coefficients have the expected signs, of which 14(9) are significantly nonzero. Among women, in Specification 1(2) the comparable summaries are 28(26) and (17)14. None of the few "incorrectly" signed parameter estimates is statistically different from zero. The data strongly support the notion that better looks produce a gross positive effect on life satisfaction/happiness.

While these comparisons clearly suggest a positive answer to the titular question of this study, we would like to compare the estimates across the samples, given the differences in the potential biases. To

---

[7]The categories represented by the vector of education indicators are: cse or equivalent, O-level or equivalent, A-level or equivalent, higher qualification, or university degree or higher, with no qualification as the excluded category.

[8]Current BMI and BMI at age 11 have mixed effects on happiness/satisfaction, but their inclusion hardly alters the impacts of beauty on happiness. This absence of any effect is not surprising, in light of the demonstrated insignificant correlations between beauty and BMI over all but the very extreme upper tail of BMI (e.g., Thornhill and Grammer, 1999).

do so we calculate the effect of being at different percentiles of the distribution of beauty on the level of satisfaction/happiness measured in standard deviations. Thus, for example, we assume that the average male among the 12.5 percent rated as below-average in the QAL 1971 is at the 6th percentile of the distribution of looks and is thus 1.53 standard deviations below the mean beauty of men. We use this type of approximation for the QAL, QOL and NCDS. For the ALLBUS we find the percentile points of the distributions of the eleven-point scale corresponding to percentiles in the averages of the QAL, QOL and NCDS, and for the WLS we do the same thing at the percentiles of the unit normal deviates.

The results of these calculations are shown in Figures 2a and 2b, with each of the 30 points in a Figure representing the fractional change in standard deviations of satisfaction/happiness generated by a movement from the mean beauty to some point in the distribution below or above the mean. Among women (men) the average good-looking respondent is 0.79 (0.89) standard deviations above the mean of beauty, while the average bad-looking respondent is 1.62 (1.67) standard deviations below the mean. On average, in Specification 2 among women (men) the gain from being this good-looking is 0.053 (0.048) standard deviations along the satisfaction/happiness index compared to the average male (female), while the loss from being this bad-looking is 0.157 (0.176) standard deviations of satisfaction/happiness. Assuming, as these calculations must, that the effects are linear within the categories above-average and average, or attractive and unattractive, the results in the expanded specifications imply that a one standard-deviation increase in beauty raises satisfaction/happiness by 0.087 (0.088). These are not large, far smaller than the impact of income on happiness in a cross-section (computed from Frey and Stutzer, 2002, Table 1), although that calculation is based on decile averages rather than individual observations. In comparison to standard-deviation impacts of the crucial "experimental" variables that are reported in related literatures, however, including those on education and health, they are not small.

The relative sizes of the estimates shown in Figures 2a and 2b generally accord with the discussion of measurement error. They are largest in the regressions based on the QAL, QOL and ALLBUS(end), where the assessment of beauty late in a long interview might have created upward biases due to Type 3 measurement error. They are somewhat smaller using the ALLBUS(start); and they are

smallest, and certainly negatively biased, in the estimates based on the WLS, where changes in beauty will have led to Type 2 measurement error that has grown over time. The direction of the bias in the estimates based on the NCDS is unclear, since the errors induce opposite-signed biases, but the estimates are generally below those from the interview studies.

V. Robustness Checks and Methodological Extensions

Although we were concerned about measurement issues in discussing the results in Section IV, in none of the estimation did we consider alternative measures and specifications, nor did we use alternative approaches to estimation. We do that here, in each case basing the estimates on the expanded specifications with control variables (Specification 2) in Tables 2a-2e.

A.   Re-specifying Proxies for Beauty and Considering Confounding Variables

No sensible reformulations of the beauty ratings in the QAL or ALLBUS surveys can be done to check their robustness, but we can use alternative measures from the other three data sets.[9] In the QOL we substituted measures from the other two years in which the respondent's satisfaction/happiness was elicited for the measure noted at the end of the particular interview. These might be viewed as instruments for current beauty. This approach will reduce Type 1 measurement error and also reduce Type 3 measurement error (but not eliminate it, assuming there is some correlation in happiness across the biennia), but it will introduce some Type 2 (attenuation) error. We present the results of this re-estimation in Table 3. The estimated effects are generally larger and more significant statistically than the comparable estimates shown in the bottom panel of Table 2b. Implicitly the reduction in Type 1 measurement error has a larger effect than do the introduction of some Type 2 measurement error and the reduction in Type 3 error.

In the WLS we re-estimated the expanded specifications using first the normalized beauty ratings given by female raters to pictures of female respondents, and by male raters to male subjects. We then

---

[9]In the QAL 1971, for example, only 59 of the respondents are rated as strikingly handsome or beautiful, and only 44 are rated as homely. When we re-estimated the models with measures encompassing each of the five beauty ratings, unsurprisingly, given the cell sizes at the extremes, this extension hardly altered the conclusions. The number in the lowest category in the QOL panel is even less.

switched and re-estimated the equations using opposite-sex ratings. Most of the estimates are attenuated slightly, just as expected assuming that there is more measurement error when fewer raters are used; but all of those that were statistically significant (women in 1992) remain so.

The NCDS respondents' appearance was assessed by their teachers at age 7 and by their teachers at age 11 (the measure used in Section III). To the extent that measurement error in the variable we used arose from random errors in individual teachers' assessments of the child's appearance, averaging the teachers' ratings at ages 7 and 11 will reduce Type 1 error. Accordingly, we average the indicator variables for appearance at 11 with identically defined variables describing appearance at age 7. These average measures replace the age-11 measures in the estimating equations, and the age-11 BMI is replaced with the average of BMI at ages 7 and 11. [10] In a few cases some previously insignificant parameter estimates in Table 3b become marginally significant, but otherwise there is no change. Implicitly, whatever measurement error exists in the age-11 proxies is highly positively correlated with that in the age-7 data and is not much reduced by averaging. [11]

Another concern is that different assessors rate beauty differently, and that their idiosyncrasies may be correlated with the subjects' happiness. With most teachers in the NCDS assessing only one subject's appearance, this issue cannot be examined in those data; and we cannot identify the raters in the WLS. In the interview surveys, however, we know which raters assessed each subject's beauty. Accordingly, we re-estimate the equations in Tables 2a-2c adding interviewer fixed effects. With one exception (the impact of bad looks among women in the QAL 1978 data) none of the significant impacts shown in Tables 2a-2c became statistically insignificant, nor did any of the estimated effects of looks on satisfaction/happiness reverse sign.

---

[10]As Lubotsky and Wittenberg (2006) show, the appropriate method in this case and in the re-estimation for the QOL is to introduce the information from the other two years separately rather than as averages. We present and discuss only the estimates based on averages to save space, as the sums of the coefficients, and the pairs' statistical significance, are always almost identical to those of the averages for both data sets.

[11]We also used the age-7 measures alone—looks and BMI—in place of the age-11 measures. Perhaps unsurprisingly, the results were slightly weaker than with the age-11 measures, and thus somewhat weaker still than the specifications based on the age-7 and age-11 averages.

Another potential difficulty with the results is that there may be location-specific determinants of beauty that may also directly affect people's perceived satisfaction/happiness. For example, perhaps living in Los Angeles with its proximity to mountains and ocean makes people happier and also attracts good-looking people. In the two data sets containing long-duration longitudinal data there is strong evidence, consistent with Gautier *et al* (2010), that changes in location are related to beauty.[12] To investigate this issue in the QAL and QOL surveys, we add state or province fixed effects to the basic equations in Tables 2a and 2b.[13] As in the other re-specifications, this extension hardly altered the results. Although the vectors of state and province effects were themselves statistically significant, and their inclusion did increase slightly the absolute values of the estimated effects of the beauty indicators, no qualitative changes resulted from their addition.

In the WLS the only locational information is whether the high-school graduate still resides in Wisconsin at the time of the interview. Adding this location variable to the specifications for the three outcomes in Table 2d produces only minute changes in the estimated impacts of beauty. No signs change, and the impacts remain statistically significant only for women observed in 1992.

In the NCDS we can account for regional effects in both the assessments of beauty during childhood and the effect of childhood beauty on adult happiness. There may be regional differences in beauty standards, which using region in childhood as a control would account for; and there may be regional differences in the relationship between beauty and satisfaction/happiness, which using location

---

[12]For example, in the WLS we can compare the beauty of the Wisconsin high-school graduates who remained there to those who were not living there at age 53. The average beauty of those still residing in Wisconsin was 0.004 (s.e.= 0.045), while the beauty of those who had left was 0.098 (s.e.= 0.030). Those who remained are significantly worse-looking than those who left. Because the NCDS was national, it allows us to compare the beauty of those who entered, those who stayed and those who left an area. Because the definitions of the British regions were not the same in all the NCDS waves, we cannot examine mobility and beauty for all areas; but southeast England, and Scotland and Wales, are consistently identified at ages 11 and 33. (Since most geographic mobility in the sample occurs between these ages, this is the most useful single comparison.) 0.572 (s.e. = 0.013) of those who moved to the Southeast were good-looking at age 11, but only 0.545 (s.e. = 0.013) of those who stayed were, and only 0.512 (s.e.= 0.032) of those who left were good-looking. The Southeast attracted good-looking people, while less good-looking people moved elsewhere in the U.K. While not statistically significant, the differences in Scotland and Wales are exactly opposite those in the Southeast: 0.571 (s.e.= 0.040) of those who entered were good-looking; 0.587 (s.e.= 0.013) of those who stayed were; and 0.598 (s.e.= 0.031) of those who left were.

[13]While we used indicators for all states in the QAL, in the QOL we divided Canada into the Atlantic provinces, Quebec, Ontario and the West.

as an adult will account for. We thus re-specify the equations in Table 2e to include vectors of regional indicators at age 11 and at the time the respondent's happiness/satisfaction is reported. In no case did either of these vectors of fixed effects approach statistical significance; nor did their inclusion qualitatively alter the impacts of beauty on satisfaction/happiness. In these data, at least, regional differences in childhood and adulthood are unimportant in affecting the estimated relationships between beauty and satisfaction/happiness.

Before examining additional covariates that we have ignored, consider a conundrum in the WLS results: The effects of beauty on happiness are apparent (in Table 2d) in 1992 (at age 53) but not in 2004 (at age 65). One reason might be that beauty effects on happiness generally diminish as one advances past middle age. To explore this possibility the only way that these data sets allow, we re-estimate Specification 2 for the interview surveys, adding interactions of the quadratic in age with the variables measuring beauty assessments. Only in one of the re-specifications for the QAL was the vector of interactions statistically significant; and in none of the re-specifications was the average beauty effect altered by accounting for possible nonlinear effects in age.

Might the results differ if we had good objective measures of intelligence and non-cognitive abilities? The interview studies suffer from an absence of objective measures of these characteristics. The NCDS, however, provides test scores measuring the respondent's general ability while a child, which we add to the estimated equations. While those respondents who tested as more able as children are typically more satisfied/happy in these re-estimates, the inclusion of this additional measure has tiny and erratic impacts on the estimated impacts of beauty on satisfaction/happiness. Also, Scholz and Sicinski (2011) show in the WLS that IQ and a wide array of non-cognitive characteristics are not correlated with raters' assessments of the photographs of the subjects. The answer to the initial question of this paragraph would seem to be negative.[14]

---

[14]Only in the QOL can we construct a proxy for self-esteem, obviously self-reported and thus quite problematic in describing happiness. Nonetheless, we do so, combining answers to a set of four questions. In each case respondents who score higher on the self-esteem scale are happier, but including the self-esteem measures has essentially no impacts on the effect of beauty on happiness. No doubt there are other, perhaps genetically determined

Other studies (Stevenson and Wolfers, 2008, Blanchflower and Oswald, 2008) have found health status to be an important determinant of happiness. The difficulty here is that in all the data sets except in the NCDS at age 51 the measures of health are subjective, self-assessed, so that they are very likely to be determined by some of the same factors that determine happiness/satisfaction. Nonetheless, we add health measures to the second specifications in each data set. In the QAL surveys subjective health is based on the response to a question about whether the respondent has health problems (with about 30 percent responding yes). All the other sets of data report self-rated health on five- or four-point scales that typically range from "excellent" to "poor." Indicators of these responses, measured as whether the person states that s/he is in one of the top two categories of self-reported health, are strongly positively correlated with the satisfaction/happiness measures in all the specifications. The estimated impacts of the beauty measures, however, change only slightly, with only small decreases in their statistical significance and no changes in their signs.

An additional set of checks includes covariates related to the respondent's parents. To the extent that there are family background effects in happiness (Hartog and Oosterbeek, 1997), including these measures could help to isolate the effect of beauty independent of any correlation that it might have with unmeasured characteristics in the respondent's family background. In the NCDS we know whether the respondent's parents are alive at his/her age 41, 46 and 51. Since parental death might affect well-being, for these latter two waves we thus form an indicator equaling 1 if a respondent's parent died in the five years preceding the interview (which occurred for 16 percent of the respondents at age 46 and 14 percent at age 51).[15] In each of the six equations describing satisfaction/happiness at ages 46 and 51 a parental death in the quinquennium does have a negative effect; unsurprisingly, given that there is no reason to

characteristics that are correlated with the genetic characteristic beauty and that also affect happiness. Absent longitudinal data in which beauty changes significantly—essentially, some almost certainly unethical experiment— it is impossible to isolate the impact of other possibly genetic determinants of happiness from those of beauty. Suffice it to note, as argued here and earlier, that the correlations of beauty with many other inherent characteristics seem to be low enough so that we arguably have good estimates of its effects.

[15]Five- and even two-year retrospectives may be too long to observe an effect on current happiness, but those are the limitations of the data.

expect the assessment of beauty at age 11 to be correlated with parental mortality three decades later, this additional variable leaves the estimated beauty effects essentially unchanged.[16] The NCDS also provides data on the social status of the respondent's parents when s/he is age 11. Adding indicators of parental status to the equations describing adult satisfaction/happiness also produces only tiny changes in the estimated impacts of beauty.[17]

A final set of possibilities relates to changes in the respondent's marital status. In the estimates based on the QOL and the NCDS we added indicators for whether the respondent got married since the previous interview or whether her/his marriage had ended since then. In no case did the inclusion of these indicators produce any but minute changes in the estimated effects of beauty.

B. Methodological Extensions

All of the specifications reported in Tables 2a-2e were estimated by least squares—although in most cases the expressions of satisfaction/happiness take only a small number of values. To examine whether a discrete-variable modeling approach would change the results, we re-estimate the models (except for satisfaction in the QAL 1978, where the questionnaire allowed a large range of discrete responses). Except for the WLS this means estimating ordered probits over these specifications; in the WLS, since the measures are of numbers of days (ranging from 0 to 7), we re-estimated the model using a count-data method (Poisson estimation). These various estimation methods yielded qualitatively very similar results to those reported in Section IV, with all previously significant effects remaining.

In the WLS data the respondents' underlying happiness is measured in three ways—by the days s/he identifies as being happy, as sad, or as having been enjoyed in the past week. Each of these can be viewed as a noisy measure of the respondent's underlying mental state. To remove some of the noise we first subtract days sad from days happy and re-estimate the specifications that were presented in the bottom half of Table 2d. Then, since we do not know what the appropriate weights on these particular

---

[16]The effects do not change greatly in any of the data sets when all the additional variables are included at the same time. This is not surprising, as within each data set these additional variables are typically nearly orthogonal.

[17]Roughly one-fourth of the sample was coded as top-class, and one-half as middle class.

expressions of happiness might be, we re-estimate the equations using the first eigenvector (first principal factor) to weight the three expressions of happiness—days happy, days sad and days enjoyed   These alternatives do not add much to the basic results. Among men the effects of beauty are positive but statistically insignificant; among women they are positive and highly significant in 1992, insignificantly negative in 2004.

VI. Inferring the Direct and Indirect Effects of Beauty

The main **economic** question in this study is whether the effect of beauty on satisfaction/happiness works through markets: How much of the effect is direct—with people who are otherwise identical in every respect being happier/more satisfied than their less good-looking peers? How much is due to the fact that beauty enhances one's outcomes in various markets, including the labor and marriage markets?[18]

We can interpret the estimate of β in equation (1) as the gross effect of beauty on satisfaction/happiness. Adding covariates X_t, both those included in the expanded specifications in Tables 2 and also self-reported health and measures of earnings and spouse quality, we obtain:

$$H_t = \alpha' + \beta'B_t^* + X_t\gamma + \varepsilon'. \tag{8}$$

We can then define the direct effect of beauty on satisfaction/happiness as β', and the indirect effect as the difference $\beta - \beta'$.

In the QAL the only additional covariates are self-reported health and the available measures of individual income. No doubt this paucity of additional covariates will generate an additional upward bias in the estimated direct effect beyond the possible overall bias that we already noted. In the QOL and the ALLBUS we add self-reported health, personal income and family income (thus presumably proxying the income of a spouse if one is present) to the covariates in Specification 2. In the WLS for each of the two years we add health status and variables measuring the respondent's own income and the household's income. Finally, in the NCDS estimates at ages 33, 41 and 51 we add health and own and

---

[18]The effect of a different personal endowment, height, on happiness was decomposed into these components by Deaton and Arora (2009), with the adjustment limited to accounting for the impact of height on earnings.

spouse/partner's weekly earnings, while at   age 46 we replace spouse/partner's earnings with family income (spouse/partner's weekly earnings being unavailable).

Rather than presenting the estimates of these expanded specifications, the first statistic in each triplet in Table 4 shows the average effects of moving from the mean beauty to being good-looking or bad-looking, measured in standard-deviation units of satisfaction/happiness. (The statistics listed in Table 4 for Specification 2 are based on the averages of those shown in Figures 2.) Taking the results at face value suggests that roughly half of the gross effect of looks works through the marriage and labor markets; but while this further expansion of the estimates accounts for some of the indirect effects of beauty on satisfaction/happiness, the available data do not allow us to account for impacts in other markets. As just one example, there is growing evidence that beauty confers benefits on good-looking borrowers in lending markets (Hamermesh, 2011, Chapter 7). Moreover, our proxies for the outcomes in the labor and marriage markets that are affected by beauty are far from perfect. It thus seems fair to conclude that the expanded estimates suggest that the direct effect of beauty is at most one-half of the total effect, and perhaps much less. The majority of the impact of beauty on satisfaction/happiness appears to be economic—through its effects on outcomes in various markets.[19]

The second statistic in each triplet in Table 4 is the average of the estimated effects over all the interview studies that used a beauty rating assessed near the end of the interview (and thus contained Type 3 measurement errors). To take advantage of the exclusion of Type 3 measurement error from the ALLBUS(start) assessment of beauty, we assume that the biases in the QAL and the QOL ratings induced by that error component are proportionate to those in the ALLBUS. We then re-calculate the estimates from those data for each gender by prorating them by the difference between the estimates in the

---

[19]One might object that the surveys key the respondents into thinking about economic issues when they respond to questions about their life satisfaction/happiness, so that the relative importance of the indirect effects is overstated. We do not believe that this is a problem in these data sets. In the ALLBUS and the NCDS questions about pay and income long precede those on satisfaction/happiness, in the QOL the opposite is the case, while in the QAL own pay is elicited long before satisfaction/happiness, while family income is elicited long after. Moreover, to the extent that there are reporting errors in own and spouse's pay (or incomes), they will lead us to underestimate the indirect effects. In sum, survey-induced biases seem to be minimal or, indeed, will lead us to understate the relative importance of indirect effects.

ALLBUS using the ALLBUS(start) and ALLBUS(end) measures. Thus we scale down the estimates from the other interview studies based on the proportionalities between the estimates from the ALLBUS in each of the three specifications. These re-estimates and those from the ALLBUS based on beauty rated at the start of the interviews offer a lower bound on the impact of beauty on satisfaction/happiness in the interview-based studies.[20]

Averaging these re-estimates together across all three interview-based studies (QAL, QOL and the ALLBUS) yields the lower bounds, the third measure in each triplet in Table 4. In some cases these lower bounds exceed the simple averages over all studies (the first statistic in the triplet), not surprisingly since the latter include results from the WLS and NCDS, in which the beauty ratings suffer from Type 2 measurement errors. Overall, however, these lower bounds yield the same conclusions as do the simple averages—at least half of the impact of beauty on satisfaction/happiness is indirect.

The results in Table 4 show only slightly greater gross effects of beauty among women than among men. The direct effects, however, are larger among women, with relatively less of the impact of beauty on women's satisfaction/happiness working through markets. This gender difference may explain why, although most of the studies measuring the impact of beauty on earnings find larger effects among men than among women, laypeople believe that looks matter more to women.[21]

VII. Conclusions and Extensions

We have examined the relationship between people's life satisfaction/happiness and their beauty. Both are subjective, although in each of our empirical examples the agent describing his/her satisfaction differs from the agent(s) describing his/her beauty. While the beauty measures introduce difficulties into the inference of the true effect of beauty on happiness, those difficulties, which differ across our data sets,

---

[20]Regressing the end-of-interview beauty rating on the rating at the start and the interview duration, the estimated effect of start-of-interview beauty is positive but statistically less than one, suggesting regression to the mean. Among men, conditional on the initial rating, the length of the interview (of contact between interviewer and interviewee) had no effect on the rating at the end; among women, the effect was positive, statistically significant, but tiny compared to the dispersion in the beauty rating.

[21]"Women face greater discrimination when it comes to looks," *Wall Street Journal*, Friday-Saturday, November 26-27, 1993, page 1, quoting indirectly Naomi Wolf, author of *The Beauty Myth*.

do not result because we make the simple mistake of essentially relating happiness to a proxy for happiness. The difficulties with the beauty measures are more subtle in our context, but the use of a large number of different types of assessments of beauty allows us to circumvent them and place a lower bound on the magnitudes of the true impacts of beauty on happiness.

The results suggest that a person's beauty does increase his/her satisfaction/happiness, with effects that are not tiny. Among both men and women at least half of the increase in satisfaction/happiness generated by beauty is indirect, resulting because better-looking people achieve more desirable outcomes in the labor market (higher earnings) and the marriage market (higher-income spouses). That relatively more of the impact among women is direct, not mediated through the effects of beauty on market outcomes, might help to explain gender differences in people's attitudes about their own looks.

Calls from political leaders and renowned economic theorists for combining happiness with GDP to obtain new measures of economic welfare may have some popular appeal.[22] Substantial evidence (see Hamermesh, 2011, Chapter 2), however, makes it clear that even radical measures to alter one's looks have fairly small effects. Coupling that observation with our findings implies that much of the differences in happiness that exist in a society arise from characteristics that are beyond one's control. While there are many other good reasons to avoid combining GDP measures with measures of subjective well-being, our discussion showing the importance of this one, essentially immutable determinant of happiness suggests that focusing on creating a happier society may not be fruitful.

---

[22]This is argued by the *Report by the Commission on the Measurement of Economic Performance and Social Progress* to President Sarkozy of France, who publicized it http://www.businessinsider.com/sarkozy-happiness . For an alternative view, see the report of a U.S. National Academy of Sciences Panel (Abraham and Mackie, 2005).

REFERENCES

Abraham, Katharine, and Christopher Mackie, *Beyond the Market*. Washington: National Academies Press, 2005.

Biddle, Jeff, and Daniel Hamermesh, "Beauty, Productivity and Discrimination: Lawyers' Looks and Lucre," *Journal of Labor Economics*, 16 (Jan. 1998): 172-201.

Blanchflower, David, and Andrew Oswald, "Well-Being over Time in Britain and the USA," *Journal of Public Economics*, 88 (July 2004): 1359-86.

-------------------------, and ----------------------, "Hypertension and Happiness Across Nations," *Journal of Health Economics*, 27 (March 2008): 218-33.

Clark, Andrew, Ed Diener, Yannis Georgellis and Richard Lucas, "Lags and Leads in Life Satisfaction: A Test of the Baseline Hypothesis," *Economic Journal*, 118 (June 2008): F222-43.

Conti, Gabriella, and Stephen Pudney, "Survey Design and the Analysis of Satisfaction," *Review of Economics and Statistics*, 93 (2011): 1087-93.

Deaton, Angus, and Raksha Arora, "Life at the Top: The Benefits of Height," *Economics and Human Biology*, 7 (July 2009): 133-6.

Diener, Ed, Frank Fujita and Brian Wolsic, "Physical Attractiveness and Subjective Well-Being," *Journal of Personality and Social Psychology*, 69 (July 1995): 120-129.

Easterlin, Richard. *Happiness, Growth, and the Life Cycle*. New York: Oxford University Press, 2010.

Frey, Bruno, and Alois Stutzer, "What Can Economists Learn from Happiness Research?" *Journal of Economic Literature*, 40 (June 2002): 402-435.

Gautier, Pieter, Michael Svarer and Coen Teulings, "Marriage and the City: Search Frictions and Sorting of Singles," *Journal of Urban Economics*, 67 (March 2010): 206-18.

Hamermesh, Daniel, "Subjective Outcomes in Economics," *Southern Economic Journal*, 71 (July 2004): 2-11.

----------------------. *Beauty Pays*. Princeton, NJ: Princeton University Press, 2011.

---------------------- and Jeff Biddle, " Beauty and the Labor Market," *American Economic Review*, 84 (Dec. 1994): 1174-94.

Harper, Barry, "Beauty, Stature and the Labour Market: A British Cohort Study," *Oxford Bulletin of Economics and Statistics,* 62 (Dec. 2000): 771-800.

Hartog, Joop, and Hessel Oosterbeek, "Health, Wealth and Happiness: Why Pursue a Higher Education?" *Economics of Education Review*, 17 (June 1998): 245-56.

Hatfield, Elaine, and Susan Sprecher, *Mirror, Mirror…: The Importance of Looks in Everyday Life*. New York: SUNY Press, 1986.

Hitsch, Günter, Ali Hortaçsu and Dan Ariely, "Matching and Sorting in Online Dating," *American Economic Review*, 100 (March 2010): 130-163.

Kahneman, Daniel, and Angus Deaton, "High Income Improves Evaluation of Life but not Emotional Well-Being," *Proceedings of the National Academy of Science*, (Aug. 2010):

Leigh, Andrew, and Tirta Susilo, "Is Voting Skin-deep? Estimating the Effect of Candidate Ballot Photographs on Election Outcomes," *Journal of Economic Psychology*, 30 (Feb. 2009): 61-70.

Lubotsky, Darren, and Martin Wittenberg, "Interpretation of Regressions with Multiple Proxies," *Review of Economics and Statistics*, 88 (Aug. 2006): 549-62.

Mathes, Eugene, and Arnold Kahn, "Physical Attractiveness, Happiness, Neuroticism and Self-Esteem," *Journal of Psychology: Interdisciplinary and Applied*, 90 (May 1975): 27-30.

Mocan, H. Naci, and Erdal Tekin, "Ugly Criminals," *Review of Economics and Statistics*, 92 (Feb. 2010): 15-30.

Möbius, Markus, and Tanya Rosenblat, "Why Beauty Matters," *American Economic Review*, 96 (March 2006): 222-35.

Oswald, Andrew, and Stephen Wu, "Well-Being Across America: Evidence from a Random Sample of One Million Americans," *Review of Economics and Statistics*, 93 (2011), forthcoming.

Plaut, Victoria, Glenn Adams and Stephanie Anderson, "Does Attractiveness Buy Happiness? 'It Depends on Where You're From," *Personal Relationships*, 16 (Dec. 2009): 619-30.

Scholz, John Karl, and Kamil Sicinski, "Facial Attractiveness and Lifetime Earnings: Evidence from a Cohort Study," Unpublished Paper, University of Wisconsin—Madison, 2011.

Scitovsky, Tibor, *The Joyless Economy*. New York: Oxford University Press, 1976.

Seligman, Martin, *Authentic Happiness*. New York: Free Press, 2004

Stevenson, Betsey, and Justin Wolfers, "Economic Growth and Subjective Well-Being: Reassessing the Easterlin Paradox," *Brookings Papers on Economic Activity* (Spring 2008): 1-87.

Thornhill, Randy, and Karl Grammer, "The Body and Face of Woman: One Ornament that Signals Quality?" *Evolution and Human Behavior*, 20 (March 1999): 105-20.

Umberson, Debra, and Michael Hughes, "The Impact of Physical Attractiveness on Achievement and Psychological Well-Being," *Social Psychology Quarterly*, 50 (Sept. 1987): 227-36.

Table 1. Descriptions of Beauty, Happiness and Satisfaction Measures, Five Data Sets

	Beauty	Satisfaction	Happiness
	(Measurement errors)		
QAL 1971, 1978 US	5-point rating by interviewer at end of interview: Strikingly handsome or beautiful Good-looking (above average for age and sex) Average looks for age and sex Quite plain (below average for age and sex) Homely (1,3)	1971: How satisfied are you with your life as a whole these days? (7 to 1 scale) 1978: How satisfied are you with your life as a whole (100 point scale)?	Taking all things together, how would you say things are these days --- would you say you're very happy, pretty happy or not too happy these days? (3 to 1 scale)
QOL 1977, 1979, 1981 CDN	Same as QAL 1971 and 1978 (1,3)	All things considered, how satisfied would you say you are? (11 to 1 scale)	Generally speaking, how happy are you with your life as a whole? (Very, fairly, not too).
ALLBUS 2008 DE	11-point scale, attractive to unattractive (1); (1,3)		If you look at your entire life, would you say you are: very happy, rather happy, not very happy, not happy at all?

Table 1, cont.

WLS **Wisc.**	Constructed from ratings on an 11-point scale, with endpoints labeled as "not at all attractive" (1) and "extremely attractive" (11), based upon an individual's high-school yearbook photo (in 1957); each photo was rated by six men and six women, and the constructed measure is an average of the z-scores across raters (1,2)		Age 53: On how many days during the past week did you feel happy? (sad?) (values 0 through 7) Age 65: Same
NCDS **UK**	Teachers' ratings of the student's appearance at age 7 and at age 11. Which best describes the student? Attractive; unattractive; looks underfed; abnormal feature; scruffy and dirty. "Looks underfed" and "scruffy and dirty" were coded as missing, "attractive" as good-looking, "unattractive" and "abnormal feature" as bad-looking, others as neither. (1,2,3).	Age 41: How satisfied are you with the way your life has turned out so far? (10 to 0 scale, from completely satisfied to completely unsatisfied) Age 51: Same	Age 33: All things considered, how happy are you? (4 to 1 scale) Age 51: On balance I look back on life with a sense of happiness. Often; sometimes; not.

Table 2a. Results from Regressions of Life Satisfaction and Happiness and Beauty Ratings, QAL 1971 and 1978*

| | Men | | Women | |
	Good looks	Bad looks	Good looks	Bad looks
Specification 1: LS, beauty, age quadratic, race, 1971				
Life Satisfaction	-0.034	**-0.255**	0.116	**-0.203**
	(0.096)	(0.129)	(0.084)	(0.100)
Happiness	0.041	-0.093	**0.097**	**-0.124**
	(0.047)	(0.062)	(0.039)	(0.046)
Specification 1: LS, beauty, age quadratic, race, 1978				
Life Satisfaction	**2.673**	**-2.290**	**1.423**	**-3.279**
	(0.745)	(1.155)	(0.703)	(1.016)
Happiness	**0.108**	-0.068	**0.120**	**-0.096**
	(0.031)	(0.048)	(0.028)	(0.041)
Specification 2: Add, education indicators, number of children, married, 1971				
Life Satisfaction	-0.088	-0.177	0.113	-0.122
	(0.096)	(0.130)	(0.083)	(0.099)
Happiness	0.019	-0.058	**0.082**	-0.077
	(0.047)	(0.064)	(0.039)	(0.046)
Specification 2: Add education indicators, number of children, married, 1978				
Life Satisfaction	**2.653**	**-2.405**	**1.323**	**-3.421**
	(0.075)	(1.159)	(0.707)	(1.016)
Happiness	**0.089**	-0.044	**0.101**	**-0.084**
	(0.031)	(0.048)	(0.028)	(0.041)

*Estimates that are significantly non-zero, one-sided 5-percent level, in **bold**.

Table 2b. Results from Regressions of Life Satisfaction and Happiness and Beauty Ratings, QOL, 1977-81*

	Men		Women	
	Good looks	**Bad looks**	**Good looks**	**Bad looks**
Specification 1: LS, beauty, age quadratic, language, year indicators.				
Life Satisfaction	-0.095	**-0.451**	0.160	-0.299
	(0.140)	(0.232)	(0.134)	(0.246)
Happiness	0.031	**-0.225**	0.057	**-0.200**
	(0.053)	(0.088)	(0.040)	(0.076)
Specification 2: Add, education indicators, number of children, married.				
Life Satisfaction	-0.075	**-0.448**	0.156	-0.259
	(0.146)	(0.234)	(0.130)	(0.243)
Happiness	0.033	**-0.219**	0.053	**-0.168**
	(0.053)	(0.086)	(0.040)	(0.073)

*Estimates that are significantly non-zero, one-sided 5-percent level, in **bold**. Standard errors are clustered on individuals.

Table 2c. Regressions of Happiness on Beauty Ratings, ALLBUS Germany, 2008*

	Men		Women	
Beauty rated at:	**Start**	**End**	**Start**	**End**
Specification 1: LS, beauty, age quadratic, German				
Beauty rating	**0.047**	**0.056**	**0.062**	**0.066**
	(0.008)	(0.008)	(0.008)	(0.008)
Specification 2: Add education indicators, married, partnered				
Beauty rating	**0.033**	**0.042**	**0.048**	**0.052**
	(0.008)	(0.009)	(0.008)	(0.009)

 *Estimates that are significantly non-zero, one-sided 5-percent level, in **bold**.

Table 2d. Regressions of Days Happy, Enjoyed or Sad on Beauty Rating, WLS Ages 53 and 65

| | Men | | Women | |
	1992	2004	1992	2004
Specification 1: LS, beauty only				
# days happy	-0.012	0.016	**0.128**	-0.026
	(0.056)	(0.044)	(0.048)	(0.044)
# days enjoyed	0.069	0.008	**0.142**	0.006
	(0.056)	(0.045)	(0.050)	(0.049)
# days sad	-0.030	-0.014	**-0.100**	-0.040
	(0.033)	(0.028)	(0.042)	(0.039)
Specification 2: Add completed education, married, number of children, HS BMI, current BMI				
# days happy	-0.028	0.025	**0.118**	-0.052
	(0.056)	(0.046)	(0.049)	(0.045)
# days enjoyed	0.058	0.010	**0.116**	-0.026
	(0.056)	(0.047)	(0.051)	(0.047)
# days sad	-0.035	-0.012	**-0.096**	-0.032
	(0.032)	(0.030)	(0.043)	(0.041)

*Estimates that are significantly non-zero, one-sided 5-percent level, in **bold**.

Table 2e. Results from Regressions of Life Satisfaction and Happiness on Beauty Ratings, NCDS Ages 33, 41, 46 and 51

	Men		Women	
	Attractive Age 11	**Unattractive Age 11**	**Attractive Age 11**	**Unattractive Age 11**
Specification 1: LS, beauty only				
Age 33				
Happiness	**0.036**	-0.023	0.036	-0.034
	(0.020)	(0.035)	(0.022)	(0.036)
Age 41				
Life Satisfaction	**0.267**	-0.053	**0.146**	-0.043
	(0.058)	(0.104)	(0.068)	(0.114)
Age 46				
Life Satisfaction	**0.118**	**-0.247**	-0.005	**-0.227**
	(0.049)	(0.088)	(0.055)	(0.092)
Age 51				
Life Satisfaction	**0.132**	-0.177	**0.203**	**-0.367**
	(0.060)	(0.107)	(0.069)	(0.118)
Happiness	**0.106**	-0.120	0.057	-0.060
	(0.042)	(0.076)	(0.044)	(0.077)

Table 2e, cont.

	Men			Women	
	Attractive **Age 11**	**Unattractive** **Age 11**		**Attractive** **Age 11**	**Unattractive** **Age 11**
Specification 2: Add education indicators, number of children, BMI 11, BMI current, married/partnered					
Age 33					
Happiness	0.029	-0.021		0.026	-0.013
	(0.020)	(0.034)		(0.022)	(0.036)
Age 41					
Life Satisfaction	**0.228**	-0.006		0.025	-0.073
	(0.062)	(0.110)		(0.073)	(0.123)
Age 46					
Life Satisfaction	0.070	-0.120		-0.060	**-0.176**
	(0.052)	(0.093)		(0.058)	(0.098)
Age 51					
Life Satisfaction	0.053	-0.091		**0.137**	**-0.267**
	(0.064)	(0.114)		(0.073)	(0.126)
Happiness	**0.099**	-0.054		0.028	0.049
	(0.046)	(0.082)		(0.048)	(0.084)

*Estimates that are significantly non-zero, one-sided 5-percent level, in **bold**.

Table 3. Estimates from Regressions of Life Satisfaction and Happiness Using Average of Other Years' Beauty Ratings, QOL, 1977-81*

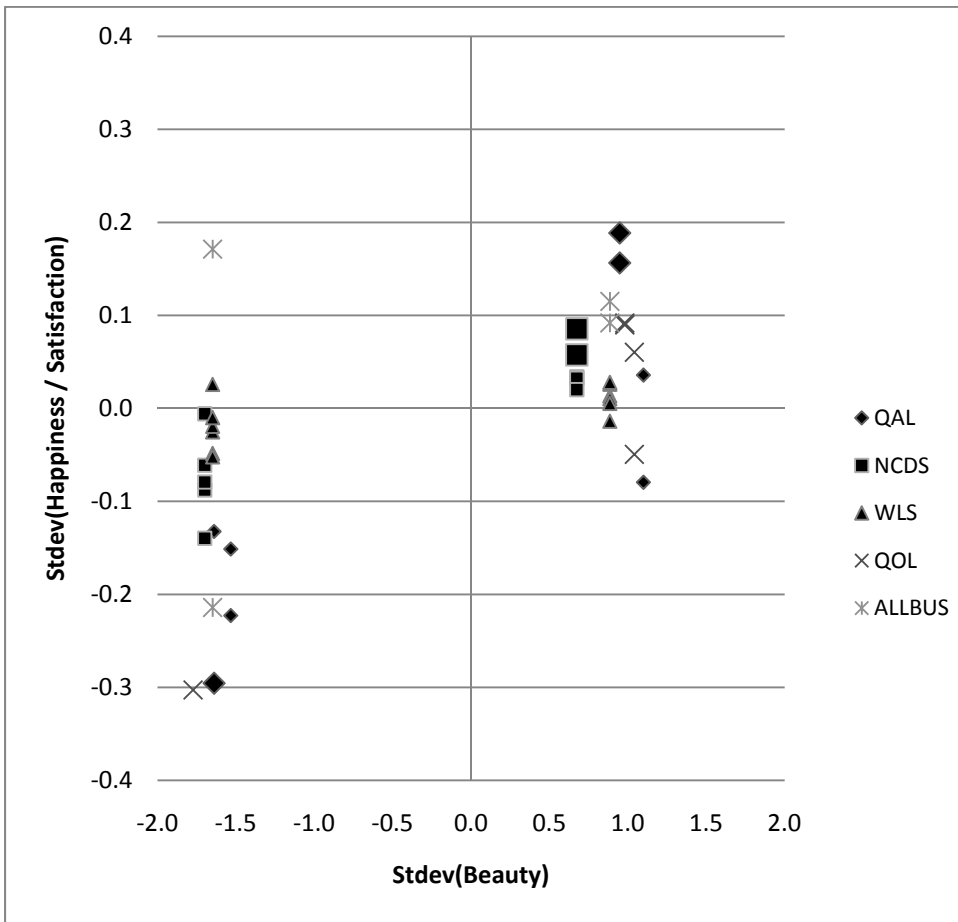|	Men		Women	
	Good looks	**Bad looks**	**Good looks**	**Bad looks**
Specification 1: LS, beauty, age quadratic, language, year indicators.				
Life Satisfaction	0.237	**-0.773**	**0.323**	-0.330
	(0.184)	(0.394)	(0.187)	(0.343)
Happiness	**0.143**	**-0.255**	0.068	**-0.191**
	(0.070)	(0.114)	(0.055)	(0.097)
Specification 2: Add, education indicators, number of children, married.				
Life Satisfaction	0.309	**-0.811**	**0.335**	-0.264
	(0.193)	(0.392)	(0.181)	(0.349)
Happiness	**0.162**	**-0.253**	0.069	**-0.161**
	(0.073)	(0.113)	(0.054)	(0.095)

*Estimates that are significantly non-zero, one-sided 5-percent levels, in bold.
Standard errors are clustered on individuals.

Table 4. Average Effects (Lower Bounds in Interview-Based Studies) of Beauty on Satisfaction/Happiness, SDOutcome/SDLooks, Five Data Sets, Four Countries

	Good Looks	Bad Looks
MEN		
SD Difference from Mean	0.886	-1.672
Specification No.		
1. (Gross effect)	0.062 (0.049)	-0.181 (-0.238)
2.	0.048 (0.033)	-0.176 (-0.124)
3. (Direct effect)	0.022 (0.010)	-0.087 (-0.114)
WOMEN		
SD Difference from Mean	0.792	-1.620
Specification No.		
1. (Gross effect)	0.068 (0.100)	-0.217 (-0.324)
2.	0.053 (0.104)	-0.157 (-0.180)
3. (Direct effect)	0.042 (0.064)	-0.124 (-0.170)
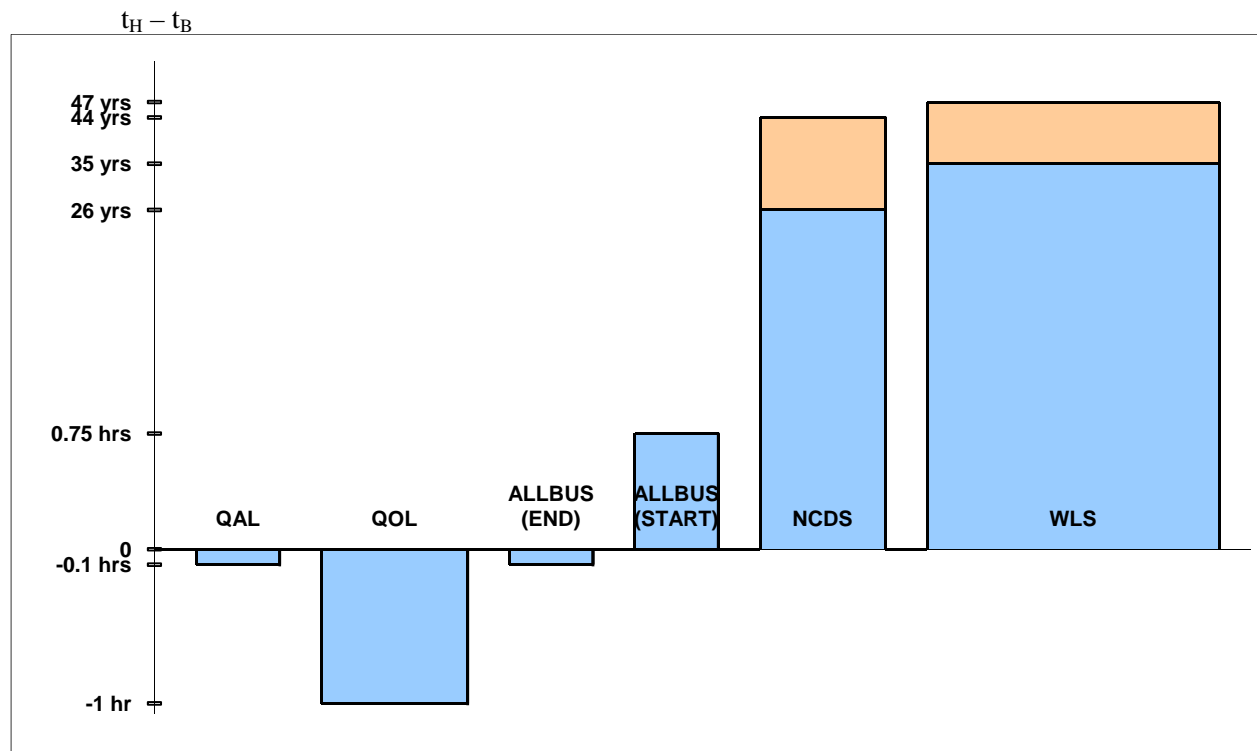
Figure 1. Relative Timing of Beauty and Happiness Measures

Figure 2a. Effects of Beauty on Happiness/Satisfaction, Men, All Data Sets
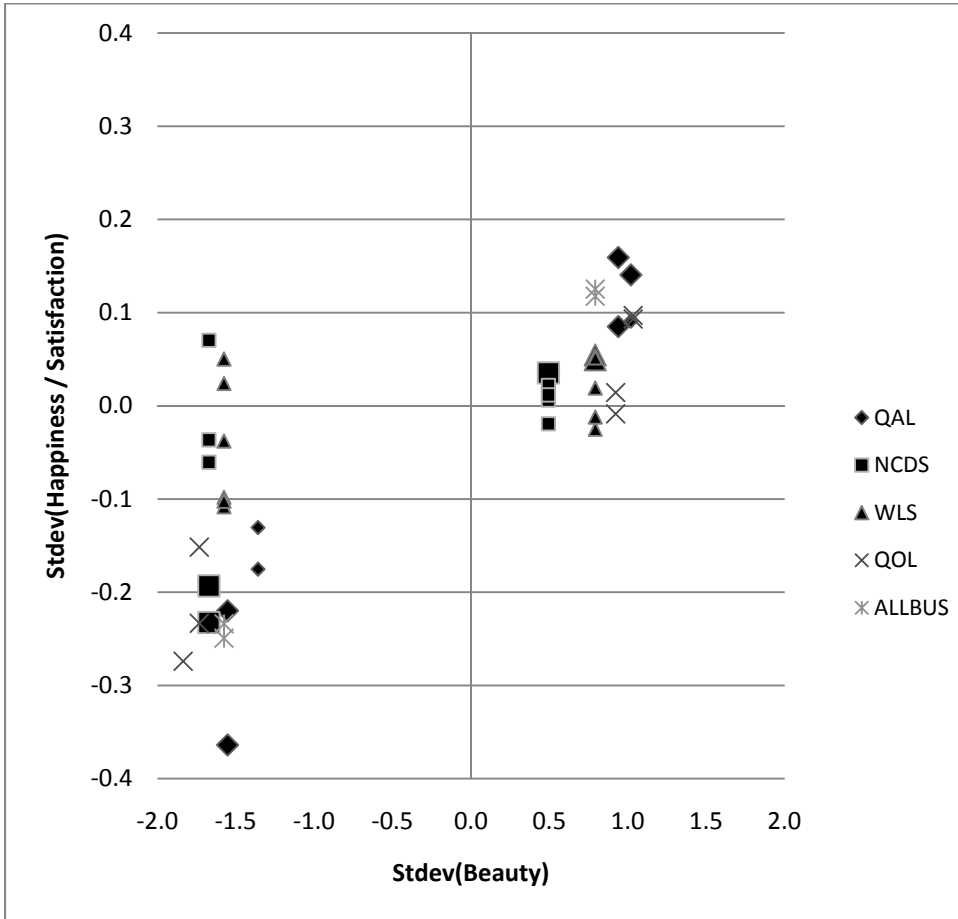
Figure 2b. Effects of Beauty on Happiness/Satisfaction, Women, All Data Sets

Appendix Table 1a. Descriptive Statistics, QAL71 and QAL78, All Observations with Beauty Rating, Satisfaction and Happiness Responses*

| | 1971 | | 1978 | |
	Men	Women	Men	Women
Good looking	0.271	0.308	0.334	0.348
Bad looking	0.125	0.174	0.101	0.120
Life satisfaction	5.556 (0.041)	5.538 (0.036)	82.313 (0.342)	82.096 (0.319)
Happiness	2.187 (0.020)	2.195 (0.017)	2.232 (0.014)	2.217 (0.013)
N	870	1217	1482	2069

*Sample averages with their standard errors for non-binary variables in parentheses.

Appendix Table 1b. Descriptive Statistics, QOL, 1977-81

	Men	Women
Good-looking	0.298	0.301
Bad-looking	0.081	0.066
Happiness	2.371	2.457
	(0.014)	(0.012)
Life Satisfaction	8.680	8.772
	(0.040)	(0.036)
N individuals	517	767

*Sample averages with their standard errors for non-binary variables in parentheses.

Appendix Table 1c. Descriptive Statistics, ALLBUS Germany, 2008, All Usable Observations*

	Men		Women	
	Start	End	Start	End
Beauty rating	7.324	7.462	7.491	7.612
	(0.050)	(0.048)	(0.050)	(0.049)
Happiness	3.047		3.053	
	(0.016)		(0.016)	
N	1554		1623	

*Sample averages with their standard errors for non-binary variables in parentheses.

Appendix Table 1d. Descriptive Statistics, WLS, All Observations with Beauty Rating, Satisfaction and Happiness Responses

	Men		Women	
	1992	2004	1992	2004
# days happy	5.321	5.673	5.466	5.654
	(0.027)	(0.058)	(0.054)	(0.053)
# days enjoyed	5.763	6.040	5.701	5.907
	(0.065)	(0.057)	(0.057)	(0.05b)
# days sad	0.668	0.465	1.115	0.841
	(0.042)	(0.037)	(0.049)	(0.044)
N	801	788	993	952

*Sample averages with their standard errors for non-binary variables in parentheses.

Appendix Table 1e. Descriptive Statistics, NCDS, All Observations with Beauty Rating

	Men	Women
Attractive age 11	0.469	0.592
Unattractive age 11	0.100	0.103
N	7886	7450

	Men (Age 33)	Women	Men (Age 41)	Women
Happiness	3.322 (0.009)	3.388 (0.010)		
Life Satisfaction			7.277 (0.028)	7.359 (0.030)
N	3514	3886	3945	4381

	Men (Age 46)	Women	Men (Age 51)	Women
Happiness			4.315 (0.020)	4.207 (0.020)
Life Satisfaction	7.562 (0.024)	7.657 (0.024)	7.335 (0.029)	7.314 (0.031)
N	3583	3920	3559	3773

*Sample averages with their standard errors for non-binary variables in parentheses.