Non-Parametric Testing of U-Shaped Relationships*

Scott Kostyshak†

July 20, 2017

Abstract

Many theories in economics predict U-shaped relationships between variables. However, satisfactory tools to examine U-shapes are lacking. After explaining the limitations of the commonly employed quadratic specification, I propose a non-parametric test of U-shaped regression functions based on critical bandwidth, and give sufficient conditions for consistency of the test statistic. The test allows one to determine whether an inherent U-shape exists between two variables or the relationship is instead caused by correlation with other variables. I apply the test to the commonly observed U-shape of life satisfaction in age, and find that much of the U-shape can be explained by the increase in financial satisfaction that typically occurs later in life. This novel insight into a long-studied puzzle is not revealed by using a quadratic specification. A user-friendly and efficient R package is provided.

Keywords: U-shape, Monotonicity, Nonlinear model, Non-parametric regression, Critical bandwidth, Life satisfaction, Well-being, Life cycle

JEL Codes: C12, C14, C18, I31

*This paper is based on material from the 2015 PhD dissertation chapter "Non-Parametric Testing of U-Shaped Relationships," available at `http://arks.princeton.edu/ark:/88435/dsp01p5547t776`. I thank Bo Honoré and Angus Deaton for many helpful conversations. I thank also Chunrong Ai, Vladimir Asriyan, Julie Berry Cullen, Elias Dinopoulos, Jonathan H. Hamilton, Ye Luo, Áureo de Paula, James E. Rauch, Richard E. Romano, David Sappington, Perihan Saygin, and Alan Secor for helpful comments. I am grateful for support from the Center for Health and Wellbeing and the Research Program in Development Studies at Princeton University.

†University of Florida, Department of Economics. Email: skostyshak@ufl.edu

1 Introduction

Theories that predict a single change of sign in the derivative of the regression function are common in economics, and are often referred to as "U-shapes" in the literature.[1] For example, U-shaped relationships have been identified between female labor force participation and economic development (Goldin, 1995), life satisfaction and age (Blanchflower and Oswald, 2008), and, in theory, between average total cost and quantity of output (Walters, 1963). Inverted U-shapes have been identified between inequality and development (Kuznets, 1955), union membership and age (Blanchflower, 2006), and innovation and competition (Aghion et al., 2005). During the past decade, U-shaped relationships have been mentioned in 80% of the issues of the *Review of Economic Studies*, and in 83% of the issues of the *American Economic Review*.[2] These statistics reflect the importance of accounting accurately for non-monotonic economic relationships.

U-shaped relationships occur naturally as the result of two competing effects, such as in Goldin (1995), where an income effect accounts for a decline in female labor supply as economic development improves, until the substitution effect dominates and bends the labor supply upward for higher levels of economic development. Some popular hypotheses are not immediately thought of as U-shape hypotheses, but in fact can be treated as such. For example, the finding that a moderate amount of alcohol consumption is associated with a lower incidence of cardiovascular disease, compared to no consumption or heavy consumption (Marmot and Brunner, 1991), can be tested as a U-shape hypothesis. Similarly, the analysis of job polarization (Goos and Manning, 2007) is an analysis of U-shapes.

Despite many theories predicting U-shapes, few tools are available to test them; this

[1] Without loss of generality, throughout this paper I make statements about U-shapes with the understanding that similar considerations apply to inverted U-shapes.

[2] This statistic is based only on issues that contain the terms "U-shape" or "hump-shape," and thus represents a lower bound, as many other terms are used to refer to U-shaped relationships. Such terms include "quadratic," "valley-shaped," "trough-shaped," "hill-shaped," "unimodal," "single-peaked," and "bell-shaped," as well as more imaginative phrases such as "the Goldilocks principle." There is no inherent reason why "U-shape" is a better term than the others, but since it is the most common, it is the term used in this paper.

limits the insights that can be gained from data analysis. There is no framework designed specifically to answer questions such as, "What features of an industry cause the average total cost curve to be U-shaped?" and "Is cardiovascular disease inherently U-shaped in alcohol consumption, or is there a third variable that drives the relationship?" Questions like these have been examined using ad hoc methods (e.g., manually inspecting graphs), and with tools not designed to answer them, such as ordinary least squares (OLS).

U-shape theories are most commonly tested with OLS using a squared regressor or by observing whether grouped dummies are consistent with a U-shape. In this paper I show how this parametric type of testing can lead to incorrect inference when the functional form is misspecified for either the variable of interest or a control variable. Suppose that $Y = f(X) + \epsilon$, and that we are interested in whether f is U-shaped. Even under misspecification,[3] OLS gives a good estimate of the regression function where the marginal density of X is thick (i.e., where there are many observations); but this property contrasts with the goal of having a good *global* estimate of the regression function. For example, if a data set has observations for age mostly between 20 and 40 years, using the OLS quadratic fit to infer the effect of age on life satisfaction across the entire life cycle would be misleading, because such a regression cannot be expected to give good predictions of life satisfaction for ages over 40.

A separate issue with OLS and semi-parametric techniques that control for variables in a restricted manner (e.g., partially linear models) is that there is a risk of finding a spurious U-shape of Y in X, when in fact Y is only U-shaped in one of the control variables, Z, that is correlated with X. Suppose $Y = f(X) + g(Z) + \epsilon$. If $g(Z)$ is not correctly modeled, the effect of Z on Y can be picked up by the estimate of $f(X)$. In Section 2, I provide necessary and sufficient conditions for OLS to give correct inference when testing for a U-shape of Y in X and controlling for Z.

Given that the common parametric tests of a U-shape require assumptions unlikely to hold in practice, it is natural to consider non-parametric models in order to avoid

[3]Suppose that $f(x)$ is modeled as $g(x)$ (a common choice is $g(x)=x^2$). The term "misspecification" here means that $g(x) \neq f(x)$.

misspecification and the corresponding consequences. Non-parametric regression is often used to explore potential U-shapes graphically, but without a statistical test. In addition, non-parametric techniques are usually abandoned when conditioning on other variables. Even if graphical techniques were used in multivariate analysis, processing and reporting results for several models and subgroups would be impractical.[4] The goal of this paper is to maintain the advantages of a statistical test and the ease of conditioning on other variables, without sacrificing the advantages of non- and semi-parametrics.

Developing a non-parametric test for a theory that predicts a U-shaped relationship seems especially complicated when the turning point is not specified in the null hypothesis. However, by reframing the null hypothesis of a U-shaped regression function as a regression function with a single valley and zero peaks, the problem of testing such theories can be viewed within a more general framework of testing the number of peaks and valleys of a regression function. I propose a non-parametric test based on critical bandwidth, which was first introduced by Silverman (1981) for a different purpose, to test the number of modes of a density. The test statistic in this paper is the smallest bandwidth such that a non-parametric regression (e.g., local polynomial regression) is quasi-convex, which in $\mathbb{R}$ is equivalent to being either U-shaped or monotone. The key insight is that if the underlying regression function has one valley and zero peaks, then choosing a bandwidth that forces the kernel estimator to have one valley and zero peaks should not be a strong restriction.

The essence of the test used in this paper is an extension of Bowman et al. (1998), who tests monotonicity of the regression function, and Harezlak and Heckman (2001) and Harezlak et al. (2007), who consider the general case of testing for an arbitrary number of bumps in the regression function and its derivatives. To my knowledge, the current paper is the first to connect the U-shape literature to the literature on testing the number of regression peaks and valleys, and specifically to use critical bandwidth tests to explore U-shapes.

[4] Imagine that one would like to examine data within 100 countries, for 10 different model specifications (e.g., different control variables), for 2 different data sets, and separately for males and females. To report these results, the number of graphs required would be $100 \times 10 \times 2 \times 2 = 4{,}000$.

After presenting the non-parametric test of bivariate relationships, I extend the critical bandwidth framework in a practical way to multivariate hypotheses using generalized additive models, which allows investigation of whether an inherent U-shape exists between two variables or the relationship is instead caused by correlation with other variables. I give sufficient conditions for consistency of the test statistic, and show that the rate of convergence under the null is at least as fast as any bandwidth sequence leading to pointwise-consistent estimates of the regression function. The test is implemented using a bootstrap procedure that leads to conservative inference. A user-friendly and efficient R package is provided for practitioners to use the tools developed in this paper. The algorithm is fast, even with many covariates and large data sets, and can be used in situations in which OLS is used in practice.

After showing that the tests developed in this paper perform well in simulations, I apply them to the U-shaped relationship of life satisfaction in age that is well established by economists but not well understood.[5] This paper uses repeated cross-sections from the World Values Survey and the European Values Survey to explore the relationship of life satisfaction and age in ninety-eight countries. The main results are that (1) there is evidence of a U-shape of financial satisfaction in age in more countries than there is of a U-shape of life satisfaction in age; and that (2) the U-shape of financial satisfaction explains the U-shape of life satisfaction. These results hold for both within-country and cross-country analyses, and are robust to several baseline specifications. Results from using OLS with a quadratic specification are provided and do not lead to the same inference because the parametric restrictions make it difficult to determine whether life satisfaction or financial satisfaction is the underlying driver of the U-shaped relationship with age. The tests in this paper thus offer a new insight into the life-satisfaction puzzle that would have been missed by using conventional methods.

The rest of the paper is organized as follows. In Section 2, I present existing tests of U-shaped regression functions based on OLS, discuss several inherent problems, and

[5]See Blanchflower and Oswald (2008), Deaton (2008), Stone et al. (2010), Wunder et al. (2013), Schwandt (2016). I provide a literature review in Section 5.

give sufficient conditions for such tests to be consistent. In Section 3, I introduce the non-parametric test that forms the core of the paper, and prove consistency of the test statistic. In Section 4, I extend the framework to include multivariate hypotheses and semi-parametrics. Section 5 applies the tests to the U-shape of life satisfaction. Section 6 concludes, and proofs of the results are given in Online Appendix A.

2 Traditional Tests of a U-Shape

To examine theories of U-shaped regression functions, researchers commonly use either OLS regression with linear and quadratic terms, or carry out analysis based on OLS with a set of grouped dummies. In this section I discuss the disadvantages of, first, the quadratic test and then of grouped dummies, and explain how both techniques can lead to incorrect inference when used to test for a U-shape. I then discuss how controlling for variables in a restricted way (e.g., linearly with OLS or partially linear models) can lead to spurious findings of U-shapes.

The general specification for the OLS regression that tests for a U-shaped regression function is

$$Y = \alpha + \beta X + \gamma f(X) + \theta' \mathbf{Z} + \epsilon, \tag{1}$$

where X is the explanatory variable of interest, $\mathbf{Z}$ is a vector of control variables, ϵ is an error term with $E(\epsilon) = 0$, and f is a function chosen by the researcher, perhaps based on theory. $f(x)$ is usually chosen to be $f(x) = x^2$. The null hypothesis of {positive quadratic} is thus tested against a composite alternative of {monotone or negative quadratic}. The correct way to carry out the test is to reject that the regression function is *not* U-shaped in X if $\hat{\gamma}$ is significant and if the x-value that minimizes the fit is inside an acceptable range. An immediate concern of the quadratic test is the parametric restriction: OLS could find evidence of a U-shape if a positive quadratic term fits the data better than a negative quadratic term, even if both are poor approximations of the true relationship.

Such tests are attractive because they are easy to carry out and to interpret. Further, obtaining a point estimate of the location of the extremum is simple. However, because of the parametric restriction, the point estimate is often found to be biased (e.g., Millimet et al., 2003). In addition to poor properties of the estimates themselves, Hirschberg and Lye (2005) show that standard errors for this type of estimate can vary dramatically depending on the method used, and are often inaccurate; they also note that the commonly used delta method is especially inaccurate in the examples they explore. Similarly, Lind and Mehlum (2010) find that the delta method is only reliable with large sample sizes.

Because tests based on the quadratic specification are so common and are the main alternative to the non-parametric test proposed in Section 3, this section is dedicated to exploring their properties. Lind and Mehlum (2010) analyze seven articles from the *American Economic Review* that focus on testing U-shaped regression functions; all of the papers conduct a test with a quadratic specification. As Lind and Mehlum show, some of the studies perform the test incorrectly. For example, some fail to include the linear term, in which case monotone regression functions are no longer in the alternative set; others employ a joint test of the linear and squared terms when only the squared term should be tested; and still others do not test that the location of the extremum is in an acceptable range. Thus, although the OLS test is easy to apply and interpret, common mistakes yield misleading conclusions.

Of the seven papers, only Imbs and Wacziarg (2003) use non-parametric techniques in addition to an OLS quadratic specification. However, the thorough non-parametric techniques employed by the authors, based on LOWESS,[6] do not allow them to statistically test whether the regression function is U-shaped. Further, the only way to condition on variables is to use the bivariate technique on subgroups (for example, running LOWESS within each category of a categorical variable).

Lind and Mehlum (2010) propose a test of a U-shape that improves on the vanilla quadratic test. Because the test is based on the coefficients from an OLS quadratic regression, the test has the same properties demonstrated in this section, and thus separate

[6]LOWESS is a non-parametric regression technique that is similar to local polynomial regression and thus estimates $E(Y|X = x)$ using the data points that are close to x. For more information, see Cleveland (1979).

discussion of the test is omitted. The test is presented in Online Appendix B.1 for reference.

2.1 Implicit weighting of OLS on thick marginal density

Estimating a parametric regression and then making inferences about the global shape of the population regression function can be deceptive. This is the case even (perhaps especially) if we have a large random sample. The reason is that OLS's goal of minimizing squared residuals does not, in general, coincide with the goal of estimating the regression function, unless the model is correctly specified.

Although OLS has good properties for predicting the response variable given a random draw of independent variables, these properties do not directly translate to estimating the global shape of a regression function when the distribution of the independent variable is not uniform. The parameter of interest is the regression function

$$m\left(x,z\right)=E\left[Y|X=x,\mathbf{Z}=\mathbf{z}\right],$$

where Y is the outcome variable, X is the variable in which Y is theorized to be U-shaped, and $\mathbf{Z}$ is a vector of covariates. Even under misspecification (e.g., choosing $f\left(x\right)$ in equation (1) to be x^2 when the true function is $|x|^{2.5}$), in a large sample the OLS estimator will give us a reasonable answer to minimizing

$$E\left[\left(\hat{m}\left(X,\mathbf{Z}\right)-m\left(X,\mathbf{Z}\right)\right)^2\right] \tag{2a}$$

over functions $\hat{m}$. However, OLS will not give a reasonable minimization of

$$E\left[\left(\hat{m}\left(X,\mathbf{Z}\right)-m\left(X,\mathbf{Z}\right)\right)^2\middle|X=x,\mathbf{Z}=\mathbf{z}\right] \tag{2b}$$

for all values of x and $\mathbf{z}$, unless there is no misspecification. It will likely only give a good answer to minimizing (2b) for sets in the support of X where the density is thick. To see this, note that in (2a) the expectation is over randomness in X and $\mathbf{Z}$, and the expectation yields a single real number. In (2b) the expectation is not over the distribution of X, but rather conditional on $X = x$. If we use OLS to minimize (2a) with the end goal of minimizing (2b), then we are using a statistic that depends on the density of X to estimate a parameter that does not depend on the density of X. This asymmetry can lead to poor estimators of $m(x, \mathbf{z})$.

In many cases, (2b) is not of interest. For example, policy makers might not care about what happens to 10% of the population; they might only care about where the marginal distribution is thick (e.g., where the most potential voters are), and thus inference based on minimizing (2a) from a random sample would answer an interesting question. However, in this paper we are interested in more than unconditional prediction; we are making inference on the *shape* of the population regression function. For example, if we are exploring the U-shape of life satisfaction in age, we might not want the inference to depend on the marginal distribution of age. Thus, we should be aware that dense areas of the independent variables can bully the parametric estimator into focusing on them.

Note that this problem is not particular to OLS; rather, it applies to any estimation strategy that minimizes the sum of functions of the absolute values of the residuals in a parametric fit.[7] Median regression will likely be worse: The property by which OLS is sensitive to outliers could help correct some of the problem, whereas a median regression is more able to ignore areas in which the fit is very poor.

This implicit weighting issue is illustrated for a particular case using simulations in Figure 1. According to a naive OLS or median regression, the simulated observations are consistent with coming from a U-shaped regression function. However, the black line shows that the true regression function is monotonically decreasing and is not convex. The 0.95 quantile of X is just before 60. An example of an X variable whose density has a thin right tail is age in many developing countries. Both OLS and median regression fit the true regression curve poorly at higher values. Because OLS minimizes the sum of squared residuals, minimizing the squared residuals for observations with X less than 60

[7]The density-weighting problem can strike in more subtle ways and is not just limited to parametric methods. For example, when estimating a non-parametric regression with a global bandwidth where the bandwidth is chosen from cross-validation, the bandwidth will be chosen based on the best smoothing properties for predicting parts of the regression function where the marginal density is thick.

is given priority over minimizing the squared residuals for observations with X between values 60 and 100 because there are more residuals to minimize in the first part. Thus, we should not infer much from the fit in the upper half of the support of X, unless we accept the consequences of relying on extrapolation. Online Appendix B.2 proposes a way to fix this particular problem if the researcher is intent on imposing a parametric form for the regression function.

2.2 Inference based on dummies

A common attempt to address parametric inflexibility is to make dummies for intervals partitioning x, and then to regress y on this set of dummies instead of on $f(x)$. This approach is often referred to in the applied literature as being "non-parametric."[8] Compared to the previously discussed quadratic fit, using a dummies approach can reduce bias when fitting. However, in this section I discuss several problems that can occur when relying on a test based on grouped dummies.

Testing a U-shape using dummies is not as straightforward as it might seem. When this technique is used in practice, often a table is shown with stars that indicate whether the dummies are individually significant. To test for a U-shape, however, one would need to test whether each dummy is significantly different from the previous.

This test suffers from the consequence that there are dummies close to each other around the turning point of the U because of a small first derivative, if the null is true. For the test to find a significant U-shape, the dummies must be far enough away from each other to be significant, but around the turning point should be close under the null. These competing properties lead to a test that requires a large sample, as shown in simulations in Online Appendix B.3. If such a test is used, the researcher should emphasize that the significance tests around the turning point are known to suffer from low power.

Arbitrary decisions are often made when using grouped dummies, such as the locations and widths of the intervals for the groups. It is thus not surprising that restrictions are

[8]When I refer to "non-parametric" techniques in this paper, I do not refer to OLS with grouped dummies.

often imposed that are known to be unlikely. For example, when exploring a U-shape of mortality in alcohol consumption, Marmot et al. (1981) use the following categories for grams of alcohol consumption: $0, (0, 9], (9, 34], (34, \infty)$. The researchers effectively restrict the effect on mortality of drinking an average of 34 grams of alcohol per day to be the same as drinking an average of 9.1 grams of alcohol per day. The property that the estimate of mortality for drinkers of 34 grams is more influenced by observations with a value of 9.1 grams of alcohol per day than by observations with a value of 34.1 grams is a disadvantage of such a model.

Relying on grouped dummies can lead to incorrect inference in a multivariate regression. Unless the control variables are categorical and the dummies for them are saturated, this test is generally inconsistent because the dummies approach restricts the fit to be stepwise, and the residual nonlinearity can still be spuriously attributed to the variable of interest.[9] No formal demonstration of inconsistency is given here, but the concern is parallel to the inconsistency of OLS with a quadratic under misspecification, which is explored in depth in Section 2.3. In addition, a test based on dummies is included in the simulations shown in Online Appendix B.3.

2.3 Inconsistency of multivariate OLS under misspecification

Suppose that Y is not actually U-shaped in X_1, and that the only reason we observe a bivariate U-shape between Y and X_1 is because X_1 is correlated with X_2, which is the real driver of the U-shaped relationship. For example, in the application part of this paper (Section 5), we explore whether the U-shape between life satisfaction and age can be explained by a third variable, financial satisfaction.

In this section I will show that existing tests of U-shapes, based on OLS with a quadratic or grouped dummies, can find spurious evidence of a U-shape in X_1, even when controlling for X_2. Nonlinear correlation can often be spuriously attributed to variables under misspecification. Suppose that the researcher is interested in testing the theory that Y

[9] This problem is worse in practice, because the researcher often controls for variables linearly instead of using dummies for them as well, thus making it easier for the effect of a control variable on y to be attributed to the more flexibly modeled variable of interest.

is U-shaped in X_1, conditional on X_2, and that (unknown to the researcher) the data generating process is

$$Y = f(X_2) + \epsilon. \tag{3}$$

Note that X_1 does not enter the equation, so the truth is that Y is not U-shaped in X_1 conditional on X_2. If a test asymptotically leads to inference suggesting that Y is U-shaped in X_1 conditional on X_2, it is referred to as an "inconsistent" test in this section.

It is common for researchers testing for a U-shape of Y in X_1 to control for X_2 by regressing Y on X_1, X_1^2, and X_2, and to then test the significance of the coefficient on X_1^2. This is generally an inconsistent test, for the obvious reason that if X_1 is correlated with X_2, in a large sample the coefficient on X_1^2 will be significant because X_2^2 was not included.

An improvement would be to regress Y on X_1, X_1^2, X_2, and X_2^2. This makes intuitive sense because we are now giving the fit of X_2 as many degrees of freedom as the fit of X_1, and in the same parametric manner. However, unless the functional form is correctly specified, this test is still generally inconsistent. For simplicity, we take the case in which Y is regressed on only the nonlinear terms (simulations in Online Appendix B.3 show results from including linear terms). We thus consider a general OLS regression specification,

$$Y = \alpha + \beta_1 g_1(X_1) + \beta_2 g_2(X_2) + \epsilon, \tag{4}$$

where g_1 and g_2 are functions chosen by the researcher. The most common choices of g_1 and g_2 are $g_1(x) = g_2(x) = x^2$. In this section, stating that the regression corresponding to equation (4) is correctly specified means that $g_1 = f$. It is not required that $g_2 = g_1$, but we will see that the optimal choice, in a certain sense, is to set them equal. We make the following assumptions:

Assumption 1. *$g_1(X_1)$ and $g_2(X_2)$ are not perfectly correlated and $0 < var(g_1(X_1)) < \infty$, $0 < var(g_2(X_2)) < \infty, 0 < var(f(X_2)) < \infty$.*

Assumption 2. *$cov[g_1(X_1), \epsilon] = cov[g_2(X_2), \epsilon] = 0$.*

Let $\hat{\beta}_1$ be the estimated OLS coefficient on $g_1(X_1)$. If the test is consistent, $\hat{\beta}_1$ should converge to 0 in probability, because X_1 does not enter equation (3). The following theorem gives the necessary and sufficient conditions under which this is true.

Theorem 1. *Under assumptions (1)–(2), if the data generating process of Y is given by equation (3), and equation (4) is the OLS specification chosen, then $\hat{\beta}_1 \xrightarrow{p} 0$ if and only if*

$$corr\left[f(X_2), g_1(X_1)\right] = corr\left[g_2(X_2), g_1(X_1)\right] corr\left[f(X_2), g_2(X_2)\right]. \tag{5}$$

Proof. Proved in Online Appendix A.1. □

For the case in which f is guessed correctly (so $g_1 = f$), if the researcher also sets $g_2 = f$ then equation (5) holds and consistency is achieved. In a separate case, if $g_1(X_1)$ is uncorrelated with both $f(X_2)$ and $g_2(X_2)$, then consistency also holds. However, if f is not guessed correctly and if $g_1(X_1)$ is correlated with $f(X_1)$ or $g_2(X_2)$, $\hat{\beta}_1$ will in general be inconsistent. Intuitively, equation (5) shows that in the case of misspecification of f, $\hat{\beta}_1$ can pick up part of what g_2 cannot approximate of f.

It should be clear that the problem discussed in this section applies to any estimation model that controls for variables in a restricted way. For example, partially linear models do not suffer from the problems of a restricted fit of the variable of interest that were discussed in Section 2.1. However, they do suffer from the same problem as OLS of finding spurious U-shapes in multivariate specifications because they free the variable of interest to pick up the residual nonlinearity of control variables.

For simulations that demonstrate the concepts discussed above, see Online Appendix B.3.

3 Critical Bandwidth Tests of a U-Shape

3.1 Non-parametric testing of a U-shape

In this section I define a U-shape and the null and alternative hypotheses associated with a non-parametric test of a U-shape. Let $S(X)$ be the support of X and

$$Y = m(X) + \epsilon,$$

where $E(\epsilon) = 0$. Throughout this section I assume there are no confounding variables, in order to focus on the statistical intuition of the test. The test will be extended to allow for confounding variables in Section 4. We assume that the first derivative of m exists and denote it as m'.[10]

If a regression function decreases across most of $S(X)$ and turns upward just at the end, the shape is usually not referred to as a U-shape. The framework below allows the researcher to specify what they consider to be a U-shape by choosing the interval, denoted by A_0, of the turning point.[11] For a specified set $A_0 \subset S(X)$, we are interested in testing

$$H_0\colon \exists a \in A_0 \text{ st } \forall x \in S(X)$$
$$m'(x)(x-a) \geq 0$$

versus

$$H_A\colon \forall a \in A_0, \exists x \in S(X) \text{ st}$$
$$m'(x)(x-a) < 0$$

The null hypothesis corresponds to a regression function that is downward sloping to a (typically unknown) turning point, a, and upward sloping thereafter. A non-parametric test, in this context, is a test that is consistent for any m of a large class of functions (e.g.,

[10]We could define a U-shape more generally, without requiring the existence of m', but this assumption is required by most non-parametric smoothers anyway, which are the backbone of the test that will be proposed.

[11]Choosing $A_0 = S(X)$ corresponds exactly to a test of quasi-convexity, but usually we are interested in setting A_0 to a subset of $S(X)$.

$\mathbb{C}(S(X)))$. For example, in Section 5 we test for a U-shape of life satisfaction in age from age 20 to 70, where the turning point is between ages 30 and 60. Using the framework above, this corresponds to $S(X) = [20, 70] \cap \mathbb{Z}$ and $A_0 = [30, 60] \cap \mathbb{Z}$.

3.2 A critical bandwidth approach

Bowman et al. (1998) extend the idea of the critical bandwidth of a density estimate, developed by Silverman (1981), to test monotonicity of the regression function. The current paper is the first to use critical bandwidth to test that the regression function is U-shaped.

Let $\hat{m}(x)$ represent a general estimator of $m(x)$ and let $\hat{m}_h(x)$ represent a non-parametric estimator of $m(x)$ using the smoothing parameter (e.g., bandwidth) h, which controls the trade-off between variance and bias of the estimator. Small values of h correspond to low bias of $\hat{m}_h(x)$ but large variance, leading to a jittery curve when graphed against x. Conversely, large values of h correspond to low variance of $\hat{m}_h(x)$ and a smooth curve. As the sample size, n, increases, the variance of $\hat{m}_h(x)$ for a fixed h decreases. Thus, h is usually specified as a function of n (by writing h_n), that decreases as n increases in order to reduce the bias of $\hat{m}_h(x)$. The specific non-parametric regression estimator used in this section is local polynomial regression.[12]

The critical bandwidth test statistic is equal to the smallest bandwidth, h, such that the fit has 0 peaks and at most 1 valley, which is equivalent to the function being quasi-convex. Testing the composite null of k or fewer bumps, instead of exactly k bumps, is consistent with Silverman (1981) and Harezlak and Heckman (2001). In Section 3.5, we will discuss how to disentangle a monotone regression function from a U-shaped regression function when interpreting a test of quasi-convexity. The test statistic is thus

$$h_{stat} = \min \{h \mid h \in H, \hat{m}_h(x) \text{ is quasi-convex}\},$$

[12] For advantageous properties, see Fan (1992) and Fan and Gijbels (1996). Properties on the boundary are important in this context, because if bias causes a fit to curl down, a test would interpret this as evidence against a U-shape. This is one reason why a first-degree polynomial is preferred over a zero-degree in local polynomial regression.

where H is the set of possible values for the smoothing parameter. For local polynomial regression, $H = R^{+}$ is the set of possible bandwidth values. An example of a different H is $H = (0, 1]$ for both generalized additive models and LOWESS.[13]

The intuition for using h_{stat} as the test statistic is that under the null hypothesis, we would expect h_{stat} to become small as $n \to \infty$ because as the sample increases, a smaller amount of smoothing is needed for a small mean squared error. Restricting the fit to be quasi-convex is only a matter of smoothing out sampling noise. If the true regression function is quasi-convex, less smoothing is needed as the sample increases because the mean is measured more precisely.

A nice property of h_{stat} that can immediately be seen is that its value does not depend on the choice of a bandwidth (e.g., through cross-validation). Practical implementations of a critical bandwidth statistic (e.g., Bowman et al., 1998) use a grid over the space of smoothing parameter values. The theoretical properties of this implementation of critical bandwidth for regression functions have not yet been explored in the literature and are examined in the next section.

3.3 Consistency of h_{stat}

Consider the regression function,

$$m(x) = E(Y|X = x),$$

with the goal of estimating $m(x)$ at a fixed number J of equally spaced points $r_1, \ldots, r_J$ in an interval $[a, b]$, where $r_k = a + (b - a)\frac{(k-1)}{J-1}$ and $r_k \in S(X)$. Note that $r_1 = a$ and $r_J = b$. If using smoothers that do not have good properties near the boundary, r_1 and r_J can be excluded from the grid of points to test. This section focuses on local polynomial regression of degree one, although in practice many smoothers could be used.

[13] The set H depends on the parameterization of the software. The examples here are for implementations that define the smoothing parameter as the percentage of data points used in the calculation of each point estimate.

Define the operator Δ such that

$$\Delta m\left(r_{k}\right)=m\left(r_{k}\right)-m\left(r_{k-1}\right)$$
$$\Delta \hat{m}_{h_{n}}\left(r_{k}\right)=\hat{m}_{h_{n}}\left(r_{k}\right)-\hat{m}_{h_{n}}\left(r_{k-1}\right).$$

I make the following assumptions:

Assumption 1. *$m(x)$ has a bounded second derivative.*

Assumption 2. *The kernel $K(y)$ is a bounded and continuous density function satisfying*

$$\begin{aligned}\int_{-\infty}^{\infty} K(y)\, dy &= 1\\ \int_{-\infty}^{\infty} yK(y)\, dy &= 0\\ \int_{-\infty}^{\infty} y^{2}K(y)\, dy &\neq 0\\ \int_{-\infty}^{\infty} y^{2l}K(y)\, dy &< \infty \text{ for } l=1,2,\ldots\end{aligned}$$

Assumption 3. *There exists a neighborhood for each r_k within which the density of X, $f_X(x)$, satisfies*

$$\left|f_{X}(x)-f_{X}(y)\right| \leq c\left|x-y\right|^{\alpha},$$

for $0<\alpha<1$.

Assumption 4. *The conditional variance $\sigma^{2}(x)=Var(Y|X=x)$ is bounded and continuous.*

Assumption 5. *$f_X(x)$ is bounded away from 0 on the interval $[a,b]$.*

Assumption (5) implies that the mean integrated squared error of $\hat{m}(x)-m(x)$ converges to 0. Intuitively, the assumption ensures that the density is not decreasing at a rate faster than the regression estimator can gain precision. This result gives us uniform convergence across the support of x, but is not needed if J stays fixed (or is bounded) as a function of n. In this case, we could use the weaker assumption that $f_X(r_j)>0$ for all j.

Assumption 6. $\Delta m(r_k) \neq 0$ *for all* $1 < k \leq J$.

Theorem 2. *Suppose assumptions (1)–(6) hold and* $m(x)$ *is quasi-convex. Then,* h_{stat} *converges in probability to zero, and is* $o_p\left(n^{-\beta}\right)$ *for any* $\beta < 1$.

Proof. Proved in Online Appendix A.2. □

The following definition introduces the concept of a function being quasi-convex on a (possibly non-convex) subset of the domain. This definition is used in the next assumption.

Definition 1. A function $f : S \to \mathbb{R}$ is *quasi-convex on a subset* $W \subset S$ if for all $x, y, z \in W$, and $y \in [x, z]$,

$$f(y) \leq \max\{f(x), f(z)\}.$$

For consistency under the alternative, I impose some structure on the type of violations that the test implementation can practically detect:

Assumption 7. *Under* H_A, m *is not quasi-convex on* $\{r_j\}_{1 \leq j \leq J}$.

Assumption (7) requires that the violation of quasi-convexity can be detected by precise estimation of $m(x)$ only on the grid points r_j. If we want a strict test of quasi-convexity asymptotically, we could make the grid finer by letting J be a function of n that goes to infinity at a reasonable rate.

Finally, to rule out the theoretical possibility of $\hat{m}_h$ giving a spurious quasi-convex fit from high variance due to a small h, we make the following assumption:

Assumption 8. *The total number of changes of derivatives in* $\{\hat{m}_h(r_j)\}_{1 \leq j \leq J}$ *is monotone decreasing in* h.

The "monotonicity property" of assumption (8) leads to attractive theoretical properties of critical bandwidth tests. Whether this assumption holds depends on various factors, such as the grid size, the starting value of h when searching for h_{stat}, and the smoother that is used. For example, local polynomial regression does not satisfy this assumption in general for $\hat{m}_h(x)$ (Harezlak and Heckman, 2001). However, similar to Harezlak and Heckman (2001) and Harezlak et al. (2007), I have found that with both simulated and real data, in practice it makes no difference as long as a reasonable grid size is used.

Theorem 3. *Suppose assumptions (1)–(8) hold and $m(x)$ is not quasi-convex. Then, h_{stat} does not converge in probability to zero.*

Proof. Proved in Online Appendix A.3. □

3.4 Bootstrapping the null distribution for conservative inference

If we bootstrapped directly from the (X_i, Y_i) data to estimate the distribution of h_{stat}, we would not be imposing the null hypothesis. Thus, we restrict $\hat{m}_h(x)$ to yield a U-shaped regression function by using $\hat{m}_{h_{stat}}(x)$, which gives us the fit that is the least biased of all fits while still being consistent with the null hypothesis. Then, we instead bootstrap residuals and add those to the restricted fit to produce bootstrapped data points that give bootstrapped h_{stat}. The intuition is consistent with Efron and Tibshirani (1994, p. 230).

To obtain a bootstrapped distribution of the residuals, we use as the sample population the residuals from the best fit, which is taken as the fit according to the smoothing parameter chosen from cross-validation. Note that the distribution of these residuals will be valid under both the null and the alternative. This method is consistent with Bowman et al. (1998). This is the only step in carrying out the test in which selection of a smoothing parameter is used and the choice of method does not seem to affect results in practice. The algorithm for generating critical values of h_{stat} is as follows:

Algorithm 1. *Critical bandwidth test*

1. *find h_{stat} by scaling smoothing parameter up*
2. *estimate a fit using a cross-validated smoothing parameter*
3. *calculate residuals e from (2)*
4. *draw bootstrapped residuals e^b from e in (3)*
5. *calculate y^b from fitted values in (1) and residuals in (4)*
6. *generate a distribution of h^b_{stat} based on $h_{stat}\left(x, y^b\right)$*

The test is conservative because the test statistic leads to an undersmoothed fit under the null. To see this, suppose that the true regression function is quasi-convex. Because of statistical noise, there likely exists a different quasi-convex function that is more consistent with the data than the true regression function. Estimation of the regression function with the smoothing parameter set to the test statistic thus leads to overfitting and a less smooth fit than the true regression function. Because the fit is less smooth, after the bootstrapped data are constructed (in step 5), more smoothing on average is needed to obtain a quasi-convex fit than was needed for the original data, so the bootstrapped test statistics are on average larger. The p-value thus has a distribution skewed away from zero under the null, leading to a conservative test. This property can be seen in the results from simulations in Section 4.1, and is consistent with other tests based on critical bandwidth (Silverman, 1981; Bowman et al., 1998; Harezlak and Heckman, 2001; Harezlak et al., 2007).

3.5 Interpretation of the non-parametric test results

The test for a U-shaped regression function is implemented as a combination of (1) a test of quasi-convexity (2) a test of a monotonicity. A quasi-convex function can either be single-dipped or monotone. Results from a test of monotonicity can thus distinguish between the two. Consider the following order of testing:

1. Test for quasi-convexity.

2. Test for monotonicity.

If (1) rejects, there is no need to proceed to (2) because the null in (2) is a subset of the null in (1). Failing to reject in (1) and rejecting in (2) is consistent with a U-shape and not consistent with monotonicity. If both tests fail to reject, it is difficult to distinguish between low power, a quasi-convex regression function, and a non-quasi-convex regression function. Using simulation for further power analysis could be useful in this case.

Observe that the p-value of the test of quasi-convexity need not be lower than the p-value of the test of monotonicity, and understanding why this is true can give intuition for how to interpret the results. It is possible for h_{stat} to be the same for both tests. This happens if, as h is increased, the fit becomes monotone without first passing through a U-shaped fit. A bootstrapped test statistic, h^b_{stat}, is more likely to be smaller for the test of quasi-convexity, because $\hat{m}_{h^b_{stat}}(x)$ is (weakly) more likely to be quasi-convex than it is to be monotone. Thus $P\left(h^b_{stat} < h_{stat}\right)$ is larger for the test of quasi-convexity than the test of monotonicity. If h_{stat} is the same for both tests, this suggests that the true regression function is not U-shaped.

For more practical discussion on interpreting results, see the application in Section 5.2.

4 Semi-Parametric Tests of Multivariate Hypotheses

We now extend the discussion started in Section 2.3 of testing for a U-shape conditional on other variables. A non-parametric test for Y being U-shaped in X_1 conditional on X_2 could be implemented by testing that for any value of X_1, Y is U-shaped in X_2. However, the focus of this paper is on a practical test, and the curse of dimensionality leads to many common situations in which a non-parametric test would not have enough precision to be useful. The idea behind the critical bandwidth test introduced in Section 3 can be generalized in an intuitive way. As discussed before, in addition to a bandwidth, smoothing parameters from other non-parametric regression techniques can be used as test statistics.[14]

Generalized additive models (GAMs) (Hastie and Tibshirani, 1986) provide a reasonable way to flexibly estimate a multivariate regression function. We model Y as

$$Y = \sum f_k\left(X_k\right) + \epsilon,$$

and estimate f_k without having to guess a functional form as in OLS.[15] This paper is the first to extend the concept of a critical bandwidth to GAMs in order to test multivariate hypotheses in a semi-parametric way. When compared to the bivariate non-parametric test in Section 3, an important difference with GAMs is that there are multiple smoothing

[14] This approach is also used by Harezlak and Heckman (2001), who find that "the method of smoothing is not crucial" (p. 714).

[15] The locations of the f_k are not identified because of the additivity, but we are only interested in the shape.

parameters, one for each k. The algorithm for implementing a critical smoothing parameter test using GAMs is similar to the bivariate case (Algorithm 1) and is shown below.

Algorithm 2. *Critical bandwidth test with GAMs*

1. *find* h_{stat} *by scaling up smoothing parameter(s)*
2. *fit a GAM with cross-validated smoothing parameters*
3. *calculate residuals* e *from (2)*
4. *draw bootstrapped residuals* e^b *from* e *in (3)*
5. *calculate* y^b *from fitted values in (1) and residuals in (4)*
6. *generate a distribution of* h^b_{stat} *based on* $h_{stat}\left(x, y^b\right)$

Step (1) can be implemented in two ways. The first method is to fix the smoothing parameters of the control variables equal to their cross-validated values, while increasing the smoothing parameter of the variable of interest until the fit becomes quasi-convex. The second method is to scale up all smoothing parameters at the same time, again until the fit of interest becomes quasi-convex. The potential advantage of the first method is that the fit for the variable of interest should not be confounded by poor fits of the other variables. The potential advantages of the second method is that the degrees of freedom given to each fit are more similar for all smoothing values, and there is no reliance on a bandwidth selector. A third variation, which combines elements from each of the first two, is to scale up all bandwidths proportional to their cross-validation values. Which method is better depends on the implementation and properties of the smoother used, but in practice the choice does not seem to make a difference.

4.1 Simulations

We start by looking at whether there is a U-shape in one variable conditional on another, and will later consider the more general case of conditioning on multiple covariates. The true regression functions that we will consider are shown in Figure 2, and are defined as follows:

$$m_1(x) = 0$$
$$m_2(x) = x(1-x)$$
$$m_3(x) = x + 0.415e^{-50(x-0.5)^2}$$
$$m_4(x) = \begin{cases} 10(x-0.5)^3 - e^{-100(x-0.25)^2} & \text{if } x < 0.5, \\ 0.1(x-0.5) - e^{-100(x-0.25)^2} & \text{otherwise} \end{cases}$$
$$sin(x) = 0.5 \cdot \sin(9 \cdot (x + 0.02))$$
$$sq(x) = (x-0.5)^2$$

The m_n functions were explored by Ghosal et al. (2000), and some were also studied in other papers that tested for monotonicity of regression functions. m_1 is used to examine the level of the test. m_2 is a simple function that is inverse-U-shaped. m_3 was considered in both Bowman et al. (1998) and Hall and Heckman (2000). m_4 was used by Hall and Heckman (2000) to illustrate the flatness problem, which affects tests based on critical bandwidth and is presented in Online Appendix C.

Table 1 shows results from semi-parametric regressions of Y on X, conditioning on another variable. Tests of quasi-convexity and monotonicity are both shown, as discussed in Section 3.5. The regressors are all uniformly distributed and the dependent variable is constructed additively, with a normally distributed error. For example, for the test "$Q(X_{sq}) + s(X_{m_3})$" the data are constructed as follows:

$$U_{sq} \sim \text{unif}(0,1), \quad U_{m_3} \sim \text{unif}(0,1), \quad \epsilon \sim N(0,1)$$
$$X_{sq} = sq(U_{sq}), \quad X_{m_3} = m_3(U_{m_3})$$
$$Y = X_{sq} + X_{m_3} + \epsilon,$$

with correlation 0.25 between the uniforms. Then, Algorithm 2 is carried out to obtain a p-value for testing whether Y is quasi-convex in X_{sq}, conditional on X_{m_3}. The simulations in the table are consistent with the test being asymptotically valid. Not surprisingly, the power is lower for relationships that do not deviate as much from monotonicity, such as m_3, compared to sin. The test is conservative, as discussed in Section 3.4: Under the null, the test carried out at the nominal level $\alpha = 0.05$ rejects less than 5% of the time. This can be seen in the rows with H_0 containing $Q(X_{sq})$.

Table 2 shows semi-parametric tests of a U-shape conditioning on multiple other variables. In all cases, the data are generated as follows:

$$Y = X_{sq} + X_{m_1} + X_{m_2} + X_{m_3} + X_{m_4} + \epsilon,$$

where $\epsilon \sim N(0, 1)$ and $X_f = f(U_f)$, where $U_f \sim \text{unif}(0, 1)$ with 0.25 correlation between them. Then, a test is performed on each of the X variables conditioning on all of the others. For example, for the row "$Q(X_{m_2}) + \ldots$" the null hypothesis is that Y is quasi-convex in X_{m_2}, conditional on X_{sq}, X_{m_1}, X_{m_3}, X_{m_4}. The symptoms of low power from the flatness problem (discussed in Online Appendix C) can be seen in the table by looking at the slow convergence toward 1 of both the monotonicity and quasi-convexity tests of X_{m_3}.

Comparison of the results across Tables 1 and 2 suggests that power is not noticeably affected by controlling for more variables. This nice property is due to using GAMs (as opposed to a fully non-parametric model), and provides for practical multivariate analysis of U-shapes by applied researchers: If conditioning on a variable causes a pattern to disappear, it is likely because the conditional relationship is different and not because there is lower precision from conditioning on another variable.

5 The U-Shape of Life Satisfaction in Age

A U-shape of life satisfaction in age has been found by many economists. Blanchflower and Oswald (2008), for example, find that such a relationship holds in 72 countries. There is more agreement that the U-shape holds in certain sets of countries, notably English-speaking countries (Deaton, 2008), than in other sets of countries, such as Eastern European and developing countries. Using data from Gallup for the United States, after controlling for gender, unemployed, with a partner, and living with a child at home, Stone et al. (2010) find that not only does the U-shape still exist, but its shape is almost unchanged. The U-shape has been found to be robust to the inclusion of individual fixed effects with British panel data (Clark, 2007) and to various panel data methods dealing with age-period-cohort issues in German panel data (Landeghem, 2012; Schwandt, 2016). The U-shape findings in both of these panel data sets are generally confirmed using partially linear models by Wunder et al. (2013). Several studies using different types of data are consistent with the life satisfaction U-shape, such as an inverted U-shape of antidepressants in age (Blanchflower and Oswald, 2016), and a U-shape of a measure of happiness for apes in age (Weiss et al., 2012).

Despite the popularity of the topic, no clear theory has emerged that definitively explains the channel through which the U-shape emerges. Solving the puzzle of why life satisfaction increases in later ages, despite worsening health and declining wage-earning ability, would provide insight into the fundamental question of "What makes us happy?" Further, although age is not a choice variable, policy decisions might affect the *way* in which age affects life satisfaction (Landeghem, 2012). Understanding age effects also offers a foundation for understanding issues regarding period and cohort effects, such as trends of subjective well-being over time.

Wunder et al. (2013) propose an explanation for the initial decline in life satisfaction by which time passes more quickly as one ages, which causes individuals to perceive that in recent times they experienced fewer pleasurable events. A separate approach suggests that the U-shape can be explained by initial decreasing satisfaction due to unmet aspirations, followed by an increase due to acceptance of those unmet aspirations as one ages; This has been found to be consistent with panel data from Germany (Schwandt, 2016). Finally, after observing a related U-shape in apes over their life cycles, Weiss et al. (2012) propose that there are biological foundations for the U-shaped relationship in humans.

In this section I apply the semi-parametric tests developed in Section 4 to the life satisfaction U-shape. I show that the U-shape goes away when conditioning on financial satisfaction, for several baseline specifications. Using the data explained in Section 5.1, both within-country (Section 5.2) and cross-country (Online Appendix D.4) results suggest the same conclusion: that the U-shape of financial satisfaction in age explains the U-shape of life satisfaction in age. Section 5.3 explains potential channels through which financial satisfaction is U-shaped in age.

5.1 Description of the data

I use data from the World Values Survey (WVS) manually merged[16] with the European Values Survey (EVS) (I refer to the merged dataset as WVS-EVS). The WVS-EVS merges the WVS waves 1981–1984, 1990–1994, 1995–1998, 1999–2004, 2005–2009, and 2010–2014 with the EVS waves 1981–1983, 1989–1993, 1999–2001, and 2008–2010. Ninety-eight countries are used for most parts of the analysis, and the median sample size for a country (pooling across survey waves) is 2,431. See Online Appendix D.1 for the survey questions that correspond to the WVS-EVS variables used in the data analysis.

The semi-parametric methods used in this paper ignore the ordinal nature of the outcome variable. Findings by Ferrer-i-Carbonell and Frijters (2004) suggest that analysis of subjective well-being data is not sensitive to the use of methods that do take into account the ordinal nature of the variable. Similar studies (e.g., Wunder et al., 2013) take the same approach. Regressions are not run separately by gender, which is consistent with the within-country regressions in Blanchflower and Oswald (2008); similar relationships have been found between males and females for life satisfaction, as well as for a variety of hedonic well-being measures across age (Steptoe et al., 2015).

5.2 Within-country results

Table 3 shows columns that summarize the results of the semi-parametric test of a U-shape of life satisfaction in age for various specifications, using the WVS-EVS data. The hypothesis explored is that there is a U-shape between life satisfaction and age, over the age range 20 to 70, with a turning point between 30 and 60. Because the semi-parametric test has a

[16] The agreement between WVS and EVS that lead to an official merged dataset, covering the period 1981–2004, was not renewed.

composite null hypothesis of monotone and U-shaped, interpretation requires care in order to differentiate between the two classes of functions, as discussed in Section 3.5. Column names for Table 3 are defined as follows:

U There is evidence consistent with a U-shape: The test of quasi-convexity fails to reject; the test of quasi-concavity rejects; the turning point associated with the test statistic is inside the specified range of turning points.

U-out The same criteria as the U column, except that the turning point is *outside* the specified range of turning points (there must, however, be a turning point).

mono. ↓ There is evidence consistent with a monotone regression function: Both of the quasi- tests fail to reject; the turning points (if they exist) for the quasi-tests do not fall inside the range; the fit associated with the monotone test statistic is decreasing.

reject all There is evidence against a U-shape, inverted U-shape, and monotonicity: Both quasi- tests reject.

low power Both quasi- tests fail to reject, and the criteria for MONO. ↓ and MONO. ↑ are not met.

The INV-U columns and MONO. ↑ column have descriptions parallel to those of the U columns and MONO. ↓ column. In the U column, it is not enough to fail to reject quasi-convexity; there could be low power, or the true regression function could be monotone. To address both of these possibilities, a criterion for the U column is that quasi-concavity is rejected (recall that monotone functions are in the null sets of both quasi- tests). A reasonable way to separate monotone increasing from monotone decreasing is to look at the slope of the fit associated with the test statistic for monotonicity, as done in the MONO. columns. There are not many entries in the U-OUT and INV-U-OUT columns, because it is not common for a country to have a U-shape with the turning point close to one of the ends and, at the same time, have enough power to reject monotonicity.[17]

[17] As the level of the test tends to zero, the table entries will move toward the MONO. columns and the LOW POWER column. Thus, when interpreting the MONO. columns, it is

The results in Table 3 suggest that for the unconditional relationship, there is more evidence of a monotone decreasing relationship (36 countries) than of a U-shaped relationship (15 countries). Including the control variables income, education, employment status, and marital status is consistent with the within-country regressions of Blanchflower and Oswald (2008). When controlling for these variables, the results flip compared to the unconditional results: There is evidence of a U-shape in 37 countries and of a monotone decreasing relationship in 19 countries.

The table also shows that financial satisfaction can explain much of the U-shaped relationship between life satisfaction and age, where it exists. No matter which of the three baseline specifications is used, after conditioning on financial satisfaction, the number of countries in the U column is more than halved. With the second (rows three and four) and third (rows five and six) specifications, many of the countries move from the U column to the MONO. $\downarrow$ column. Conditioning on another variable might decrease the power of the test, which could be a reason countries move out of the U column (because of failing to reject quasi-concavity). However, most of the countries migrating out of the U column into the MONO. $\downarrow$ column do so because the turning point (if it exists) is no longer in the specified range of turning points, which is not a consequence of low power. In addition, the simulations in Section 4.1 suggest that adding control variables does not noticeably decrease power, as it might in a fully non-parametric framework.

There is reason to be concerned with regressing one subjective measure on another, because such a regression could suffer from correlated measurement error. For example, if a survey respondent is in a good mood, they could respond with a high response to both subjective measures. We might see a high correlation between the two variables when in fact the correlation is caused by a (hidden) third variable. One robustness check to address this possibility is to run the same tests as above, but switch the positions of life satisfaction and financial satisfaction. By running these parallel tests, we can get an idea of which of

important to recognize that there is no criterion of a statistical test rejecting, only tests failing to reject. Hence, if entries move from the OUT columns to the MONO. columns (e.g., when adding a control variable or using a subset of the data), it is possible that this is purely the result of reduced power.

the highly correlated variables is responsible for the U-shape with age.

Table 4 shows results from testing for a U-shape of financial satisfaction in age. There is evidence of a U-shaped relationship in more countries than in Table 3, which suggests that financial satisfaction could be the underlying driver of the observed U-shaped relationship between life satisfaction and age. Recall that when financial satisfaction was added as a control variable in the unconditional life satisfaction test (second row of Table 3), the number of countries consistent with a U-shape dropped from 15 to 5. Conversely, when life satisfaction is added as a control in the unconditional financial satisfaction test (second row of Table 4), the number stays constant at 24. Similar patterns hold for the other baseline specifications shown in the table. This suggests that financial satisfaction explains the U-shaped relationship between life satisfaction and age more than the other way around.

A potential reason for why financial satisfaction might have more explanatory power than life satisfaction could be that financial satisfaction has more variance, and thus inherently contains more information than life satisfaction. Both satisfaction variables should be viewed as collapsed approximations of underlying continuous variables, so if the financial satisfaction variable were collapsed less than life satisfaction, financial satisfaction would mechanically have more explanatory power. However, this does not seem to be the case: As Online Appendix D.2 shows, the two distributions are similar. Thus, it does not seem that the differences in marginal distribution are the reason for financial satisfaction's higher explanatory power of the U-shape.

As shown in Online Appendix D.3, the OLS results do not suggest the same conclusion as the semi-parametric results: that financial satisfaction explains the U-shape of life satisfaction. Instead, the results are consistent with the common situation that when a variable that is highly correlated with the dependent variable is controlled for, there are larger standard errors. As demonstrated in Section 2, because OLS does not capture the effect of one variable on another in a flexible way, it is more difficult to discern which variable is the underlying driver of an observed relationship.

Online Appendix D.4 provides estimations of several different pooled models. The results are consistent with the within-country results, and include graphs of semi-parametric estimates to demonstrate that the tests of a U-shape agree with what one would conclude from visual inference.

5.3 Discussion of results

The results in the previous section suggest that financial satisfaction can explain much of the U-shaped relationship between life satisfaction and age, and that financial satisfaction itself has a U-shaped relationship with age. In this section, some potential reasons are discussed for why the above patterns are found in the data, in order to encourage future research to quantitatively determine why financial satisfaction is U-shaped in age.

It is not surprising that financial satisfaction is positively related to life satisfaction. However, financial satisfaction is also positively related to age later in life in most countries. This result might initially be surprising, given that income decreases in old age. Taking the subset with ages between 40 and 70, although age and income are positively correlated in only 8 of the 102 countries, age and financial satisfaction are positively correlated in 59 countries. The most immediate explanation to reconcile the divergent directions of income and financial satisfaction is that although income decreases in old age, expenses also decrease. The increase in financial satisfaction at later ages was also listed as one reason for increasing life satisfaction by Wunder et al. (2013).

Financial obligations and sources of financial stress typically are most pronounced during the middle years of a person's life. For instance, raising children increases both present financial expenditures and stress about meeting future expenditures, and the parents of middle-aged individuals are usually no longer in the workforce and thus less able to provide support. Further, in many cultures, children are expected to care for their elderly parents. In some countries, it is common to have large debts in the middle years of life, such as from mortgages. Finally, the middle-aged are more likely to hold risky financial assets (Guiso et al., 2000). These sources of financial stress appear to outweigh the high wage-earning ability of the middle-aged.

In addition to fewer expenditures in old age, stress about one's financial situation might be lower as the result of reduced uncertainty about the future, because the future is shorter.

Many 70-year-olds do not have the stress of planning for retirement because that planning has already been done. Similarly, there is less need to save as insurance against events such as losing one's job. Also, in contrast to the situation of middle-aged individuals, in many countries there are welfare systems or cultural expectations about being cared for by children that ensure that the elderly will not have financial concerns.

6 Conclusion

Economists have been exploring U-shaped relationships in the data since at least Kuznets (1955), but econometric tools that focus on testing U-shapes have not been developed. Using traditional tools (e.g., OLS) to explore U-shapes can lead to unwanted consequences, such as testing for the regression function being more positive-quadratic than negative-quadratic. This paper provides a framework for analyzing U-shapes, and proposes to test them exactly how they are defined, rather than by a restricted approximation. By bridging the literatures that test U-shapes and that test bumps in regression functions, this framework opens the door for further contributions of tools that analyze this common type of relationship between variables. In addition to being the first to use critical bandwidth for testing U-shapes, the methodologies developed in this paper apply more generally to all tests based on critical bandwidth by extending them to multivariate analysis through the use of GAMs, which renders them more attractive for use in practice.

OLS will always be a useful tool for data analysis because of its convenient properties; the advanced techniques developed in this paper are intended to supplement the use of OLS, not replace it. Controlling for other variables in a flexible way ensures that those variables' effects are not picked up by the variable of interest only because of parametric restrictions. This paper contributes to the trend for non- and semi-parametric techniques to become more important in economics; the requirements for their advantages to stand out are more easily met, as a result of both computational power increasing and size of data increasing (e.g., Varian, 2014). A user-friendly R package accompanies this paper and allows for easy application of the methods described.

The theoretical tools developed were applied to the open question of why life satisfaction is commonly found to be U-shaped in age. Results from the semi-parametric tests within countries and in several pooled models suggest that the U-shape of financial satisfaction in age explains the U-shape of life satisfaction in age. This insight is not found when using the classic OLS quadratic specification. The discovery that financial dissatisfaction causes the midlife dip in life satisfaction leads to important policy considerations: Many countries have policies that encourage saving by the middle-aged for retirement, but the results in this paper suggest that if the social planner's goal is to smooth well-being across the life cycle, then perhaps more attention should be given to alleviating financial stress during the middle years of life, rather than later. The optimal policies can be designed only by identifying the sources of financial dissatisfaction.

References

Aghion, P., Bloom, N., Blundell, R., Griffith, R. and Howitt, P.: 2005, Competition and Innovation: An Inverted-U Relationship, *Quarterly Journal of Economics* **120**(2), 701–728.

Blanchflower, D. G.: 2006, A Cross-Country Study of Union Membership, *Technical report*, IZA Discussion Papers.

Blanchflower, D. G. and Oswald, A. J.: 2008, Is Well-Being U-Shaped Over the Life Cycle?, *Social Science & Medicine* **66**(8), 1733–1749.

Blanchflower, D. G. and Oswald, A. J.: 2016, Antidepressants and Age: A New Form of Evidence for U-Shaped Well-Being through Life, *Journal of Economic Behavior & Organization* **127**, 46–58.

Bowman, A., Jones, M. and Gijbels, I.: 1998, Testing Monotonicity of Regression, *Journal of Computational and Graphical Statistics* **7**(4), 489–500.

Clark, A. E.: 2007, Born to Be Mild? Cohort Effects Don't (Fully) Explain Why Well-Being Is U-Shaped in Age, *Technical report*, IZA Discussion Papers.

Cleveland, W. S.: 1979, Robust Locally Weighted Regression and Smoothing Scatterplots, *Journal of the American Statistical Association* **74**(368), 829–836.

Deaton, A.: 2008, Income, Health, and Well-Being around the World: Evidence from the Gallup World Poll, *Journal of Economic Perspectives* **22**(2), 53–72.

Efron, B. and Tibshirani, R. J.: 1994, *An Introduction to the Bootstrap*, Vol. 57, CRC press.

Fan, J.: 1992, Design-Adaptive Nonparametric Regression, *Journal of the American Statistical Association* **87**(420), 998–1004.

Fan, J.: 1993, Local Linear Regression Smoothers and Their Minimax Efficiencies, *Annals of Statistics* pp. 196–216.

Fan, J. and Gijbels, I.: 1996, *Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability*, Vol. 66, CRC Press.

Ferrer-i-Carbonell, A. and Frijters, P.: 2004, How Important is Methodology for the Estimates of the Determinants of Happiness?, *Economic Journal* **114**(497), 641–659.

Ghosal, S., Sen, A. and Van Der Vaart, A. W.: 2000, Testing Monotonicity of Regression, *Annals of Statistics* pp. 1054–1082.

Goldin, C.: 1995, The U-Shaped Female Labor Force Function in Economic Development and Economic History, *Investment in Women's Human Capital* pp. 61–90.

Goos, M. and Manning, A.: 2007, Lousy and Lovely Jobs: The Rising Polarization of Work in Britain, *Review of Economics and Statistics* **89**(1), 118–133.

Guiso, L., Jappelli, T. and Haliassos, M.: 2000, Household Portfolios: An International Comparison, Available at SSRN: `https://ssrn.com/abstract=245805`.

Hall, P. and Heckman, N. E.: 2000, Testing for Monotonicity of a Regression Mean by Calibrating for Linear Functions, *Annals of Statistics* pp. 20–39.

Harezlak, J. and Heckman, N. E.: 2001, CriSP: A Tool for Bump Hunting, *Journal of Computational and Graphical Statistics* **10**(4), 713–729.

Harezlak, J., Naumova, E. and Laird, N. M.: 2007, LongCriSP: A Test for Bump Hunting in Longitudinal Data, *Statistics in Medicine* **26**(6), 1383–1397.

Hartigan, J. A. and Hartigan, P.: 1985, The Dip Test of Unimodality, *Annals of Statistics* pp. 70–84.

Hastie, T. and Tibshirani, R.: 1986, Generalized Additive Models, *Statistical Science* pp. 297–310.

Hirschberg, J. and Lye, J. N.: 2005, Inferences for the Extremum of Quadratic Regression Models, Available at SSRN: `https://ssrn.com/abstract=918626`.

Imbs, J. and Wacziarg, R.: 2003, Stages of Diversification, *American Economic Review* **93**(1), 63–86.

Kuznets, S.: 1955, Economic Growth and Income Inequality, *American Economic Review* pp. 1–28.

Landeghem, B. V.: 2012, A Test for the Convexity of Human Well-Being over the Life Cycle: Longitudinal Evidence from a 20-year Panel, *Journal of Economic Behavior & Organization* **81**(2), 571–582.

Lind, J. T. and Mehlum, H.: 2010, With or Without U? The Appropriate Test for a U-Shaped Relationship, *Oxford Bulletin of Economics and Statistics* **72**(1), 109–118.

Marmot, M. and Brunner, E.: 1991, Alcohol and Cardiovascular Disease: The Status of the U Shaped Curve, *BMJ: British Medical Journal* **303**(6802), 565–568.

Marmot, M., Shipley, M., Rose, G. and Thomas, B.: 1981, Alcohol and Mortality: A U-Shaped Curve, *Lancet* **317**(8220), 580–583.

McKenzie, D. J.: 2006, Disentangling Age, Cohort and Time Effects in the Additive Model, *Oxford Bulletin of Economics and Statistics* **68**(4), 473–495.

Millimet, D. L., List, J. A. and Stengos, T.: 2003, The Environmental Kuznets Curve: Real Progress or Misspecified Models?, *Review of Economics and Statistics* **85**(4), 1038–1047.

Sasabuchi, S.: 1980, A Test of a Multivariate Normal Mean with Composite Hypotheses Determined by Linear Inequalities, *Biometrika* **67**(2), 429–439.

Schwandt, H.: 2016, Unmet Aspirations as an Explanation for the Age U-Shape in Well-being, *Journal of Economic Behavior & Organization* **122**, 75–87.

Silverman, B. W.: 1981, Using Kernel Density Estimates to Investigate Multimodality, *Journal of the Royal Statistical Society* **43**, 97–99.

Steptoe, A., Deaton, A. and Stone, A. A.: 2015, Subjective Wellbeing, Health, and Ageing, *Lancet* **385**(9968), 640–648.

Stone, A. A., Schwartz, J. E., Broderick, J. E. and Deaton, A.: 2010, A Snapshot of the Age Distribution of Psychological Well-Being in the United States, *Proceedings of the National Academy of Sciences* **107**(22), 9985–9990.

Varian, H. R.: 2014, Big Data: New Tricks for Econometrics, *Journal of Economic Perspectives* **28**(2), 3–27.

Walters, A. A.: 1963, Production and Cost Functions: An Econometric Survey, *Econometrica* pp. 1–66.

Weiss, A., King, J. E., Inoue-Murayama, M., Matsuzawa, T. and Oswald, A. J.: 2012, Evidence for a Midlife Crisis in Great Apes Consistent with the U-Shape in Human Well-Being, *Proceedings of the National Academy of Sciences* **109**(49), 19949–19952.

Wunder, C., Wiencierz, A., Schwarze, J. and Küchenhoff, H.: 2013, Well-Being Over the Life Span: Semiparametric Evidence from British and German Longitudinal Data, *Review of Economics and Statistics* **95**(1), 154–167.

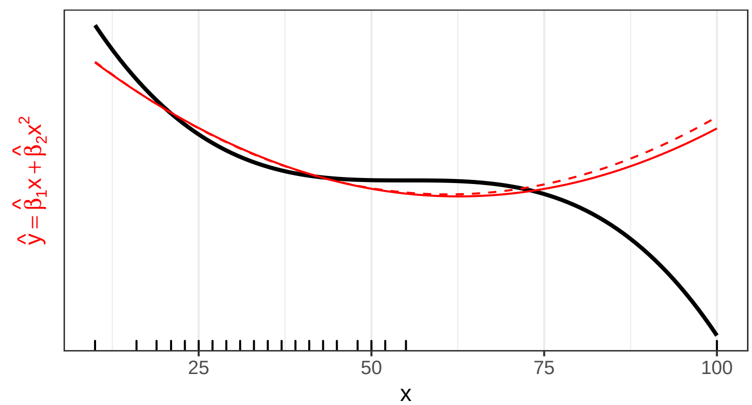


Figure 1: Quadratic Fits (OLS and Quantile)

This figure shows two estimated regressions, one using OLS (the solid red curve) and one using median regression (the dotted red curve), of the true regression function (the thick black curve). The first stick on the x-axis is the minimum of the realizations of X, at 10, and the last stick is the maximum, at 100. Each stick in between represents 5% increments of the number of observations.

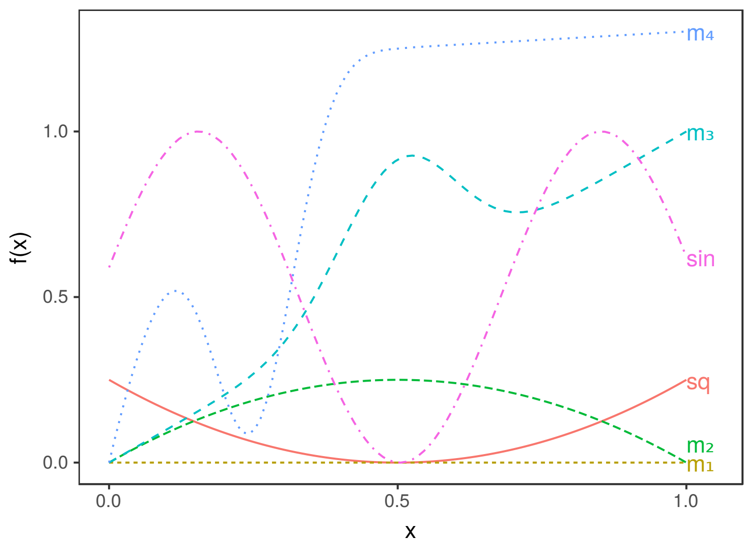


Figure 2: True Regression Functions

The functions shown in this figure are used in simulations throughout the paper. For their definitions, see Section 4.1.

Table 1: Semi-Parametric Test with Two Covariates

	sample size					
H_0	100	500	1,000	5,000	10,000	100,000
$M(X_{m_3}) + s(X_{sq})$	0.02	0.03	0.03	0.15	0.45	1.00
$Q(X_{m_3}) + s(X_{sq})$	0.02	0.03	0.02	0.15	0.45	1.00
$M(X_{sin}) + s(X_{sq})$	0.14	0.36	0.59	0.98	1.00	1.00
$Q(X_{sin}) + s(X_{sq})$	0.02	0.22	0.71	1.00	1.00	1.00
$M(X_{m_3}) + s(X_{sin})$	0.02	0.04	0.02	0.17	0.51	1.00
$Q(X_{m_3}) + s(X_{sin})$	0.02	0.03	0.03	0.17	0.50	1.00
$M(X_{sin}) + s(X_{m_3})$	0.15	0.38	0.64	0.98	1.00	1.00
$Q(X_{sin}) + s(X_{m_3})$	0.03	0.27	0.77	1.00	1.00	1.00
$M(X_{sq}) + s(X_{m_3})$	0.07	0.09	0.13	0.34	0.57	1.00
$\mathbb{Q}(X_{sq}) + s(X_{m_3})$	0.04	0.01	0.00	0.01	0.01	0.01
$M(X_{sq}) + s(X_{sin})$	0.08	0.08	0.13	0.36	0.59	1.00
$\mathbb{Q}(X_{sq}) + s(X_{sin})$	0.03	0.01	0.00	0.01	0.01	0.02

This table shows the proportion of times each test rejects the null hypothesis at the 0.05 level. For the test name (column 1), "$M(Z)$" means a null hypothesis of regression monotonicity in Z; "$Q(Z)$" means a null of regression quasi-convexity in Z; and $s(W)$ indicates that W is controlled for semi-parametrically. All tests use GAMs. The solid horizontal line separates the sets of tests in which the null hypotheses are false from sets in which at least one is true (indicated in the test name using a different font, $\mathbb{Q}$ and $\mathbb{M}$). For the shapes of the regression functions, see Figure 2. Proportions are based on 10,000 simulations.

Table 2: Semi-Parametric Test with Five Covariates

H_0	100	500	1,000	5,000	10,000	100,000
			sample size			
$M(X_{m_2}) + ...$	0.07	0.10	0.13	0.37	0.64	1.00
$Q(X_{m_2}) + ...$	0.10	0.17	0.23	0.58	0.83	1.00
$M(X_{m_3}) + ...$	0.03	0.02	0.02	0.19	0.52	1.00
$Q(X_{m_3}) + ...$	0.03	0.02	0.02	0.19	0.51	1.00
$M(X_{m_4}) + ...$	0.05	0.12	0.18	0.30	0.45	0.77
$Q(X_{m_4}) + ...$	0.05	0.13	0.18	0.31	0.46	0.77
$M(X_{sq}) + ...$	0.07	0.11	0.14	0.39	0.62	1.00
$\mathbb{Q}(X_{sq}) + ...$	0.04	0.01	0.01	0.01	0.01	0.01
$\mathbb{M}(X_{m_1}) + ...$	0.06	0.06	0.05	0.05	0.05	0.05
$\mathbb{Q}(X_{m_1}) + ...$	0.06	0.06	0.05	0.04	0.04	0.05

This table shows the proportion of times each test rejects the null hypothesis at the 0.05 level. For the test name (column 1), "$M(Z)$" means a null hypothesis of regression monotonicity in Z, and "$Q(Z)$" means a null of regression quasi-convexity. Each test conditions on all other variables. For example, test "$Q(X_{sq}) + \ldots$" tests for quasi-convexity in X_{sq}, conditioning on X_{m_1}, X_{m_2}, X_{m_3}, and X_{m_4}. All tests use GAMs. The solid horizontal line separates the sets of tests in which the null hypotheses are false from sets in which at least one is true (indicated in the test name using a different font, $\mathbb{Q}$ and $\mathbb{M}$). For the shapes of the regression functions, see Figure 2. Proportions are based on 10,000 simulations.

Table 3: Tests for U-Shape of Life Satisfaction in Age

controls	U	U out	Inv-U	Inv-U out	mono. ↓	mono. ↑	reject all	low power
unconditional	15	3	3	1	36	10	2	28
FS	5	1	6	0	38	11	0	37
inc	28	4	6	0	22	10	2	26
inc FS	7	2	8	0	31	11	2	37
inc ed emp marital	37	2	1	0	19	8	1	30
inc ed emp marital FS	18	4	3	0	36	4	1	32

This table summarizes results from various country-level specifications for a test of a U-shape of life satisfaction in age. Each numeric cell of the table is the number of countries that fall into the category indicated in the column, for the specification listed in the row. The first column indicates the variables that were conditioned on for the test. Variable abbreviations are as follows: "FS" is financial satisfaction, "inc" is income, "ed" is education, "emp" is employment status, and "marital" is marital status. See Section 5.2 for descriptions of the columns. The level of significance used is 0.1.

Table 4: Tests for U-Shape of Financial Satisfaction in Age

controls	U	U out	Inv-U	Inv-U out	mono. ↓	mono. ↑	reject all	low power
unconditional	24	6	4	1	28	12	1	22
LS	24	5	4	0	26	13	0	26
inc	39	7	3	0	17	10	1	21
inc LS	42	5	3	0	10	12	0	26
inc ed emp marital	47	0	1	0	15	11	0	24
inc ed emp marital LS	32	2	3	0	13	21	0	27

This table summarizes results from various country-level specifications for a test of a U-shape of financial satisfaction in age. Each numeric cell of the table is the number of countries that fall into the category indicated in the column, for the specification listed in the row. The first column indicates the variables that were conditioned on for the test. Variable abbreviations are as follows: "LS" is life satisfaction, "inc" is income, "ed" is education, "emp" is employment status, and "marital" is marital status. See Section 5.2 for descriptions of the columns. The level of significance used is 0.1.

Online Appendix

A Proofs of Results in the Main Text

Lemma 1. *Suppose that $var(U) > 0$ and $var(V) > 0$. Then*

$$var(U)\,var(V) - cov(U,V)^2 = 0$$

if and only if there is perfect correlation (correlation is 1 or -1) between U and V.

Proof. Without loss of generality, suppose that $cov(V,U) \geq 0$. We start with the Pearson correlation inequality (easily proved with the Cauchy-Schwartz inequality):

$$\begin{aligned}\frac{cov(U,V)}{\sqrt{var(U)}\sqrt{var(V)}} &\leq 1 \\ cov(U,V)^2 &\leq var(U)\,var(V) \\ 0 &\leq var(U)\,var(V) - cov(U,V)^2\end{aligned}$$

Note that if the top inequality is strict, the following two are as well, and if the top inequality is an equality, the following two are as well. □

A.1 Proof of Theorem 1

Proof. It is easily shown that

$$\hat{\beta}_1 = \frac{\widehat{cov}[Y, g_1(X_1)]\,\widehat{var}[g_2(X_2)] - \widehat{cov}[g_2(X_2), g_1(X_1)]\,\widehat{cov}[Y, g_2(X_2)]}{\widehat{var}[g_1(X_1)]\,\widehat{var}[g_2(X_2)] - \widehat{cov}[g_2(X_2), g_1(X_1)]^2},$$

where $\widehat{cov}$ and $\widehat{var}$ are sample covariance and sample variance. The probability limit of $\hat{\beta}_1$ is then given by the law of large numbers and the continuous mapping theorem as

$$\hat{\beta}_1 \xrightarrow{p} \frac{cov[Y, g_1(X_1)]\,var[g_2(X_2)] - cov[g_2(X_2), g_1(X_1)]\,cov[Y, g_2(X_2)]}{var[g_1(X_1)]\,var[g_2(X_2)] - cov[g_2(X_2), g_1(X_1)]^2}.$$

By Lemma 1, the denominator is 0 if and only if there is perfect correlation between $g_1(X_1)$ and $g_2(X_2)$, so by assumption (1), $\hat{\beta}_1 \xrightarrow{p} 0$ only when the numerator is 0. By assumption (2), for consistency to hold, we must then have that

$$cov\left[f\left(X_2\right), g_1\left(X_1\right)\right] var\left[g_2\left(X_2\right)\right] = cov\left[g_2\left(X_2\right), g_1\left(X_1\right)\right] cov\left[f\left(X_2\right), g_2\left(X_2\right)\right].$$

Dividing both sides by $var\left[g_2\left(X_2\right)\right]\sqrt{var\left[f\left(X_2\right)\right] var\left[g_1\left(X_1\right)\right]}$ gives the equivalent and more intuitive condition that

$$corr\left[f\left(X_2\right), g_1\left(X_1\right)\right] = corr\left[g_2\left(X_2\right), g_1\left(X_1\right)\right] corr\left[f\left(X_2\right), g_2\left(X_2\right)\right].$$

□

A.2 Proof of Theorem 2

Proof. Fix $\epsilon > 0$, $\delta > 0$, $\beta < 1$. We want to show that $\exists N$ st for $n > N$,

$$P\left[n^{\beta} h_{stat}^{(n)} < \epsilon\right] > 1 - \delta.$$

Take any bandwidth sequence, h_n, that satisfies

$$h_n = dn^{-\beta}, \quad 0 < \beta < 1, \quad d > 0.$$

Take N_1 such that $h_{N_1} < \epsilon$. Let

$$M = \min_k \left|\left(\Delta m\left(r_k\right)\right)\right|.$$

$M > 0$ by assumption (6). Under assumptions (1)–(5), using the results of Fan (1993) we have that $\hat{m}_{h_n}(r_k) \xrightarrow{p} m(r_k)$ pointwise so for $1 < k \leq J$, $\exists N_2$ st for $n > N_2$,

$$\begin{aligned} &P\left[sign\left(\Delta\hat{m}_{h_n}(r_k)\right) = sign\left(\Delta m(r_k)\right)\right] \\ &\geq P\left[\left|\Delta\hat{m}_{h_n}(r_k) - \Delta m(r_k)\right| < M\right] \\ &> 1 - \frac{\delta}{J}. \end{aligned}$$

Let $N^* = \max(N_1, N_2)$. It follows from Boole's inequality that

$$P\left[sign\left(\Delta\hat{m}_{h_n}(r_k)\right) = sign\left(\Delta m(r_k)\right), \text{ for } 1 < k \leq J\right] > 1 - \delta.$$

Note that under the null, the event $\left\{sign\left(\Delta\hat{m}_{h_n}(r_k)\right) = sign\left(\Delta m(r_k)\right), \text{ for } 1 < k \leq J\right\}$ implies the event $\left\{h_{stat}^{(n)} \leq h_n\right\}$. We then have that for $n > N^*$,

$$P\left[h_{stat}^{(n)} \leq h_n\right] > 1 - \delta$$

and so

$$P\left[h_{stat}^{(n)} < \epsilon\right] > 1 - \delta.$$

Because h_n above can be any sequence such that $h_n = dn^{-\beta},\ 0 < \beta < 1,\ d > 0$, and because

$$P\left[h_{stat}^{(n)} \leq h_n = dn^{-\beta}\right] \to 1$$

for all $\beta < 1$, it must be that $h_{stat}^{(n)}$ converges in probability to 0 faster than $n^{-\beta}$ for any $0 < \beta < 1$. □

A.3 Proof of Theorem 3

Proof. Suppose H_0, as defined in Section 3.1, is not true. Then, by assumption (7), $\forall a \in A_0$ $\exists j \leq J$ such that $\Delta m(r_j)(x - a) < 0$. Take any bandwidth sequence h_n that satisfies

$$h_n \to 0,\ \ nh_n \to \infty.$$

Because $\hat{m}_{h_n}(r_j) \xrightarrow{p} m(r_j)$ (Fan, 1993) and by assumption (6), $\exists N$ such that for $n > N$,

$$P\left[\left|\Delta\hat{m}_{h_n}(r_j) - \Delta m(r_j)\right| < \left|\Delta m(r_j)\right|\right] > 1 - \delta$$

and

$$P\left[sign\left(\Delta\hat{m}_{h_n}(r_j)\right) = sign\left(\Delta m(r_j)\right)\right] > 1 - \delta.$$

Fix such an N and thus an h_N. Then,

$$P\left[sign\left(\Delta\hat{m}_{h_N}(r_j)\right) = sign\left(\Delta m(r_j)\right)\right] > 1 - \delta$$

for all $n > N$ because if we fix the bandwidth, as the sample increases, precision can only be gained. In other words, if h_N were used as the bandwidth for $n > N$, the bias of $\hat{m}_{h_N}(r_j)$ would not increase, and the variance would decrease. Thus, the probability that $\hat{m}_h$ is quasi-convex is less than δ for any h smaller than h_N and greater than h_n, for $n > N$.

There is still the possibility that $\hat{m}_h$ could be quasi-convex for h less than h_n. However, this case is ruled out by assumption (8). To see this, suppose that because of high variance caused by a small h, the probability that $\hat{m}_h$ is quasi-convex for some $h < h_n$ is greater than δ. Because the total number of changes of the sign of the derivative of $\hat{m}_h$ is monotone in h, this would imply that the probability that $\hat{m}_{h_N}$ is quasi-convex is also greater than δ, which contradicts the result above. Thus, $P\left[h_{stat} > h_N\right] > 1 - \delta$ for $n > N$. □

B Supplements for OLS Discussions

B.1 The Lind and Mehlum test

This section describes the test by Lind and Mehlum (2010), which has been the only test (parametric or non-parametric) of a U-shape in addition to the vanilla OLS quadratic specification. The test is implemented on top of the coefficients and standard errors obtained from estimating equation (1) with OLS. The function f must be chosen to have one extremum. Then, the relationship is U-shaped, inverse-U-shaped, or monotone, depending on β and γ. A U-shape exists if and only if the slope is negative at the start and positive at the end of a chosen interval $[x_l, x_h]$ (e.g. $[x_l, x_h] = [\min(x_i), \max(x_i)]$), in which case the following inequalities hold:

$$\beta + \gamma f'(x_l) < 0 < \beta + \gamma f'(x_h)\,.$$

To test that the inequalities hold, Lind and Mehlum propose a likelihood ratio test, based on the framework of Sasabuchi (1980). Simulations in the paper suggest that the Lind and Mehlum test provides an improvement over the vanilla test of the coefficient on the quadratic term. Further, it has the nice property of testing for a U-shape in an intuitive way. Nonetheless, because the test relies on the choice of f in the OLS specification, it shares the same parametric properties as the vanilla quadratic specification discussed in Section 2.

B.2 Inverse frequency weighting

Section 2.1 discusses how using OLS can lead to incorrect inference about the shape of the regression function because of its implicit weighting on the density of the X variable. To estimate the regression function in a way that has better global properties, we should prevent the fitting algorithm (e.g., OLS) from ignoring points at which the marginal distribution of the independent variable is thin. One way to do this is to artificially redistribute the density of X in the sample until it is uniform. We could resample from the dataset conditional on the values of X where the density is thin. It is on this set of estimates that we would then make inference about a U-shape. This resampling method is equivalent to using generalized least squares, which is easier to implement.

To see the above intuition mathematically, for notational simplicity I take the case in which the vector of covariates $\mathbf{z}$ is empty. OLS minimizes

$$\int (\hat{m}(x) - m(x))^2 \, dF_X(x)\,. \tag{6}$$

In general, we do not want to force weighting the loss function $L(x) = [m(x) - \hat{m}(x)]^2$ by the distribution of X, $dF_X(x)$. When looking at the shape of $m(x)$ over the entire support, it would be more sensible to minimize

$$\int [\hat{m}(x) - m(x)]^2 \, dx,$$

which is the special case of (6) when X is uniformly distributed. This suggests that we could get estimates with better global properties using generalized least squares, where the weight of each observation is the inverse of the empirical cumulative distribution function (ECDF) of X evaluated at each point. We would in essence be weighting in order to cancel the underlying weights in (6). Note that we do not have the same problems that, for example, feasible generalized least squares can suffer from,[18] because the ECDF itself gives the ideal weights; we do not use it as an estimator of the CDF of X.

Figure 3a shows these proposed weights applied to OLS and quantile regressions with the quadratic specification. The weighted regressions are the green lines, which give a better global fit to the regression function and which correctly fail to reject that the regression function is not U-shaped. Alternatively, a dummies approach can reduce bias when fitting, as shown in Figure 3b. However, as discussed in Section 2.2, and shown in Tables 5 and 6, a test based on dummies has poor properties.

B.3 Simulations of OLS inconsistency

Section 2.3 shows analytically how an OLS-implemented test of a U-shape is inconsistent. This appendix provides simulations to explore the properties of the inconsistency for a few specific cases. Table 5 shows simulations of an OLS quadratic test based on the following specification:

$$Y = \beta_0 + \beta_{11} X_1 + \beta_{12} X_1^2 + \beta_{21} X_2 + \beta_{22} X_2^2 + \epsilon$$

[18] Although feasible generalized least squares is asymptotically more efficient than OLS, in finite samples it can actually be less efficient because of the reliance on estimating the variances to use as weights.

The test is simply whether $\hat{\beta}_{12}$ is significantly positive. The coefficient β_{11} on the linear term should not be tested because the support of X_1 is $\mathbb{R}$, as shown below. Table 5 also shows results from a test based on dummies, as discussed in Section 2.2. The data are generated as follows:

$$\begin{pmatrix} X_1 \\ X_2 \\ \epsilon \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & .8 & 0 \\ .8 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right] \tag{7}$$
$$Y = |X_2|^a + \epsilon,$$

where a is a constant (specified in the "exp." column of the table). When a is 2, the model is correctly specified. In this case, the OLS test of a U-shape of Y in X_1, conditional on X_2, is consistent. This can be seen in the table by noting that $\hat{\beta}_{12}$ is significantly positive about 5% of the time. However, when a is 3 or 4, the rejection rate is considerably larger than 5% and it is common to find a spurious U-shape of Y in X_1, conditional on X_2. These simulations are consistent with the theoretical results of the previous section.

A comparison between a quadratic test of a U-shape and a semi-parametric test is not straightforward, because the null and alternative sets are different. In the semi-parametric test, the null hypothesis is that Y is U-shaped in X_1; for the quadratic test, the null is that Y is *not* U-shaped in X_1. However, as the sample increases to the point at which power is "large," we can compare the inference that one would get from each. It makes more sense to include X_1 in the data generating process because if it is absent, the semi-parametric test will expectedly fail to reject. We thus add X_1 such that its relationship to Y is clearly not U-shaped. Here we have the following data generating process, in which the ϵ_k are independent standard normals:

$$\begin{aligned} X_2 &\sim \epsilon_1 \\ X_1 &= X_2(1+\epsilon_2) \\ Y &= |X_2|^a + .05\sin(15X_1) + \epsilon_3, \end{aligned} \tag{8}$$

Ideally, a test with a null hypothesis of "not U-shaped" should find even less evidence of a U-shape in X_1 than when X_1 did not enter the equation of Y: Instead of the test statistic being insignificant (as it should be in the data generating process shown in equation (7)), it should be significant in the opposite direction. Table 6 yields the same conclusion as Table 5: The quadratic specification finds a spurious U-shape with respect to X_1 because of the correlation between X_1 and X_2. In all panels of the table except the $a = 2$ panel, the coefficient $\hat{\beta}_{12}$ is significantly positive in more than 5% of the samples. The semi-parametric tests are consistent and correctly reject that Y is U-shaped in X_1, conditional on X_2.

C The Flatness Problem

Tests based on critical bandwidth can be inconsistent if the regression function has a first derivative that is zero for a long enough interval (that is, if assumption (6) in Section 3.3 is violated). Even if assumption (6) holds, there could be power issues in finite samples if there is a stretch on the x-axis with $\Delta m(r_k)$ small in absolute value. Just as Hartigan and Hartigan (1985) demonstrate regarding Silverman (1981), Hall and Heckman (2000) show that for the critical-bandwidth test of monotonicity of Bowman et al. (1998), there are cases in which the power does not tend to one. In particular, they give an example of a regression function that is not monotone, but is not rejected asymptotically (we will define this function in Section 4.1 as m_4). In fact, the power of the test for this regression function can decrease as the sample size increases. The intuition for why this can happen is that if there is a flat region ($\exists c, d$ st $m'(x) = 0$ for $x \in [c, d]$), h_{stat} will be large because a lot of smoothing is required before the fit becomes flat.[19] Because a global bandwidth is used, a large bandwidth can smooth out non-monotone segments in other parts of the underlying regression function.

The flatness problem is relevant to all critical bandwidth tests and can occur when all of the following three conditions are present:

1. The smoothness of $m(x)$ is not uniform.

[19] Here, whether a smooth is flat depends on the decimal precision used in the computation.

2. There is a long interval where $m'(x) \approx 0$.

3. The sample size is not large enough.

The closer $m'(x)$ is to 0 and the longer the flat interval (condition 2), the larger the sample size must be to overcome the decrease in power. Below, I give a few practical suggestions for addressing this issue.

1. Use an adaptive smoothing parameter for the critical bandwidth.

2. Look at plots of $\hat{m}_h(x)$ using several smoothing parameters.

3. Determine whether $m''(x)$ varies considerably, using a test or exploring graphically as in (2).

4. Be wary of a large h_{stat}.

5. Be wary if $h_{stat} - h_{cv}$ is large, where h_{cv} is the smoothing parameter from cross-validation.

At first, (1) seems incompatible with a critical bandwidth test because there is no single h with an adaptive bandwidth. However, we define a new test statistic

$$\alpha = \min \left\{ a \in R \mid \hat{m}\left(x; ah\left(x\right)\right) \text{ is quasi-convex} \right\},$$

where $\hat{m}(x; h(x))$ is a fit using an adaptive bandwidth. Instead of increasing the bandwidth uniformly, we can view the coefficient a as the smoothing parameter, which scales the set of bandwidths. This allows each section on the x-axis to become smoother at a different rate. Most popular adaptive bandwidths have two properties: They are (1) positively related to a measure of smoothness of the regression function at each point; and they are (2) inversely related to the density. These properties lead to a good (in a mean squared error sense) trade-off between bias and variance. In areas of the x-axis with a large density, a smaller bandwidth can be used to reduce bias without a significant sacrifice of variance of $\hat{m}(x)$. In areas in which there is a lot of curvature, a small bandwidth is used because over-smoothing can lead to a large cost of bias.[20]

In our case we do not want the second property (that the adaptive bandwidth depends on the density), because an example similar to the one given by Hall and Heckman (2000) could be constructed. For example, imagine a case in which the density is thin in the non-monotone interval and thick elsewhere. A large bandwidth would be used in the non-monotone interval and could smooth out a non-stochastic wrinkle in the underlying regression function.

In the case of testing for monotonicity, the flatness problem could be improved by breaking the test into smaller pieces so that a flat part in one area would not affect the fit in other areas. For testing a U-shape, this is not easily done. However, a similar approach could be taken in testing for a U-shape by composing monotone tests.

D Supplements for Life Satisfaction Application

D.1 Variable definitions for the World Values Survey

life satisfaction All things considered, how satisfied are you with your life as a whole these days? 1 (Dissatisfied), 2, ..., 10 (Satisfied).

financial satisfaction How satisfied are you with the financial situation of your household? 1 (Dissatisfied), 2, ..., 10 (Satisfied).

income Scale of household incomes: 1 (Lower step), 2, ..., 11 (Highest step).

marital status Married, Living together as married, Divorced, Separated, Widowed, Single/Never married.

[20]The assumption of smoothness of m is important here, because if the regression function has a very sharp linear kink such that $m''(x) = 0$ except for a couple of points, this might not be detected because typical adaptive bandwidth techniques measure smoothness based on m''.

education At what age did you complete your education? <12 years, 13 years, 14 years, ..., 20 years, 21+ years.

employment status Full time, Part time, Self employed, Retired, Housewife, Student, Unemployed.

number of children Have you had any children? 0, 1, ...,7, 8+.

D.2 Distributions of life and financial satisfactions

Figure 4 compares the marginal distributions of life satisfaction and financial satisfaction in the WVS-EVS. The two distributions are similar. The main difference is that in the bottom three categories, there is more mass for financial satisfaction than for life satisfaction; the opposite is true for the top three categories.

D.3 Results from OLS quadratic specification

Tables 7–10 below show results from using a quadratic specification to explore the life satisfaction U-shape, in order to determine whether inference leads to the same conclusion as obtained with the semi-parametric tests in Section 5.2. The OLS-based test is carried out for each country. A country is categorized as having a U-shape if, for the subset with ages 20 to 70, the coefficient on the age term is significantly negative, the coefficient on the squared age term is significantly positive, and the age that corresponds to the bottom of the fit is within the range $[30, 60]$. Similarly, a country is categorized as having an inverted U-shape if the coefficient on the age term is significantly positive, the coefficient on the squared age term is significantly negative, and the age that corresponds to the peak of the fit is within the range $[30, 60]$. It is not always necessary to test the linear term when implementing a quadratic test of a U-shape, but in this case it is because the support of age is positive.

The distribution of age is not uniform in most countries, and a situation similar to the one shown in Figure 1 is a concern: It might be that life satisfaction monotonically decreases as one ages, but that the relationship is not picked up by the quadratic because of a thin right tail in the density of age. This is especially a concern in developing countries. Thus, in order to get a good global estimate of the regression function, weighting is performed as suggested in Online Appendix B.2. The inference and lack of inference discussed in this section do not change when regressions are run without weighting.

The columns U$^{\divideontimes}$ and Inv-U$^{\divideontimes}$ indicate that the signs of the coefficients yield the corresponding shape and the turning point is within the required range, but at least one of the coefficients was insignificant. The columns are useful for concerns that a country leaves the U column only because of decreased power (e.g., because of added control variables). The mono. columns are meant to give an idea of why a country is not in one of the U-shape columns, rather than as evidence that a country has a monotone life-satisfaction relationship with age. A country falls into a mono. column if both coefficients have the same sign or if the turning point of the U-shape or inverted U-shape is not in the range [30,60] (that is, it falls within the first 10 years or the last 10 years). Only the signs of the point estimates are checked. This is just to show that the results are likely not due to lack of power. Finally, two specifications are used as the base specification: an unconditional specification and the set of control variables from Blanchflower and Oswald (2008) that were also used in the semi-parametric analysis.

Table 7 shows that when conditioning on financial satisfaction, the number of countries that have a U-shape falls only slightly compared to the unconditional results. Similarly, when adding control variables in Table 8, although the number of countries with a significant U-shape decreases, most of those countries move to having an insignificant U-shape. When comparing these results to the same models—except for switching life satisfaction and financial satisfaction (Tables 9 and 10)—the results are similar.

D.4 Cross-country results

In this appendix section we approach the data from a different angle than the within-country analysis in Section 5.2. We now consider pooled regressions, motivated by cross-country models in this section. The models necessarily rely on strong assumptions that certain variables affect the dependent variable the same way across all countries. To the extent that these assumptions hold approximately, pooling the data allows to overcome some potential shortcomings of the within-country results regarding cohort effects that are discussed below. This section asks the same question as the within-country analysis, of what drives the U-shape of life satisfaction in age, but from a different perspective, and arrives at the same conclusion: the U-shape is driven by financial satisfaction.

Pooling the data could solve a particular type of issue related to cohort effects: Because in some of the tests a strict test was used, a small violation of a U-shape could lead to a rejection by the test. If those violations occurred because of cohort effects that were specific to a country, and if the cohort effects were not correlated across countries, then by averaging across countries the cohort effects would be averaged away.[21] To see this formally, consider an age, period, and cohort framework in which the life satisfaction of individual i who has age a and was born in year b is

$$y_{i,b,a} = \alpha_b + \beta_a + \gamma_{b+a} + \epsilon_{i,b,a}.$$

For a detailed introduction to this model, see Online Appendix D.5. For simplification, let us assume that the period effects γ_{b+a} are zero, and further that cohort effects depend on the country, but that averaging the cohort effects over countries yields a constant. Formally, for an individual in country q we assume the model

$$y_{i,b,a,q} = \alpha_{bq} + \beta_a + \epsilon_{i,b,a},$$

where $\sum_q \alpha_{bq} = k$ does not depend on birth year, b. This assumption is strong and likely does not hold exactly,[22] but the framework allows us to motivate analysis of the aggregated data. If the above assumption does hold, GAM estimations of age effects could be biased

[21] "Averaging across countries" refers to how if a GAM is estimated on the pooled data set, the estimate of life satisfaction for a given age will be the average of life satisfaction in the data.

[22] For example, this assumption would be violated if cohort effects were correlated across countries during, e.g., World War II or worldwide economic crises that scarred people of certain ages more than others.

for the within-country regressions but not for the pooled regressions.[23] At the very least, as long as there is not perfect correlation, the cohort effects would be smoothed away more than for the within-country tests.

Figure 5a gives a first look at the pooled-data relationship between life satisfaction and age, and how controlling for financial satisfaction affects that relationship. The two curves correspond to estimates of $f(\text{age}_i)$, the effect of age on life satisfaction (LS), before and after controlling for financial satisfaction (FS), as in the following GAM models:

$$\text{LS}_i = f(\text{age}_i) + \epsilon_i \tag{9a}$$

$$\text{LS}_i = f(\text{age}_i) + g(\text{FS}_i) + \epsilon_i \tag{9b}$$

The estimate of $f(\text{age}_i)$ from the unconditional specification in (9a), shown as the red curve in Figure 5a, reproduces the U-shape that is commonly found in similar regressions in previous literature. In the same plot, the blue curve shows that after controlling for financial satisfaction in (9b), the effect of age on life satisfaction changes: The right arm of the U flattens out.[24]

For the same reason as explained in the cross-country regressions, to see if the age effects weaken only because life satisfaction and financial satisfaction are highly correlated, we consider symmetric models with financial satisfaction as the dependent variable:

$$\text{FS}_i = f(\text{age}_i) + \epsilon_i \tag{10a}$$

$$\text{FS}_i = f(\text{age}_i) + g(\text{LS}_i) + \epsilon_i \tag{10b}$$

Figure 5b shows that financial satisfaction has a stronger U-shaped relationship with age than life satisfaction does, and that the shape shifts only slightly when including life

[23]For the cohort effects to average to a constant, we would need to weight the GAM estimations by the inverse of country size. The results do not change much when this is done.

[24]For all results plotted in this section, the y-axis has been fixed to be the same, and that the range of the estimated effect functions is indeed comparable across specifications in GAMs. The location of the y-axis, however, is not comparable or interpretable, and is thus centered around 0.

satisfaction as a control variable. These regressions are consistent with the within-country regressions, and further suggest that the reason life satisfaction is U-shaped in age is because financial satisfaction is also U-shaped in age.

If the same countries that have older populations are also the countries that generally have higher life satisfaction and financial satisfaction, the right arm of the U-shape could be explained simply by the differences in marginal distribution of age. The results from the within-country estimations in Section 5.2 should address this concern, but we can also address it with the pooled estimations by including country fixed effects. For individual i in country c, consider the sets of specifications

$$\mathrm{LS}_{ic} = f(\mathrm{age}_i) + \sigma_c + \epsilon_i \tag{11a}$$

$$\mathrm{LS}_{ic} = f(\mathrm{age}_i) + g\left(\mathrm{FS}_i\right) + \sigma_c + \epsilon_i \tag{11b}$$

and

$$\mathrm{FS}_{ic} = f(\mathrm{age}_i) + \sigma_c + \epsilon_i \tag{12a}$$

$$\mathrm{FS}_{ic} = f(\mathrm{age}_i) + g\left(\mathrm{LS}_i\right) + \sigma_c + \epsilon_i \tag{12b}$$

Figure 6 shows the estimations of $f(\mathrm{age}_i)$. For the regression of life satisfaction, conditioning on financial satisfaction (Figure 6a) leads to the right arm not only flattening out but to it becoming monotonically decreasing. For the regression of financial satisfaction (Figure 6b), conditioning on life satisfaction does not change the U-shaped relationship with age.

Although there is some prior theory for introducing the effect of age on life satisfaction as a function that is shared across countries,[25] there is less reason to require that the effect of financial satisfaction on life satisfaction be shared across countries. Viewing financial satisfaction as one of several components of life satisfaction, it makes sense to think that the effect of financial satisfaction on life satisfaction would differ by country. When assessing

[25] It has been proposed by some authors (e.g., Weiss et al., 2012) that the life satisfaction U-shape is innate in human DNA.

one's life as a whole, a person in one culture might give more weight to financial satisfaction than someone in a different culture. Similarly, how much one's financial woes affect their life depends on the welfare system of the country. The following specifications thus allow for the relationship between financial satisfaction and life satisfaction to differ by country:

$$\text{LS}_{ic} = f(\text{age}_i) + g_c(\text{FS}_i) + \sigma_c + \epsilon_i \quad (13)$$

$$\text{FS}_{ic} = f(\text{age}_i) + g_c(\text{LS}_i) + \sigma_c + \epsilon_i \quad (14)$$

The results, shown as the blue curves in the two panels of Figure 7, support the same conclusion as the previous specifications and as the within-country tests in Section 5.2: The U-shape of life satisfaction is driven by the U-shape of financial satisfaction.

D.5 Identification of age, period, and cohort effects

When analyzing age effects, it often makes sense to also consider period and cohort effects. This section introduces a framework for modeling age, period, and cohort effects and shows how there are inherent problems for identifying all of the effects. Suppose that we have observations on individuals of A age groups $a_1, \ldots, a_A$ observed during T time periods $t_1, \ldots, t_T$. Let C be the number of cohorts that we observe, which is the number of distinct birth years. In McKenzie (2006) and Landeghem (2012), $C = A + T - 1$. This happens if each t_n represents a year and if $t_{n+m} = t_n + m$ (that is, if the observation years are consecutive).[26] In that case, the first time that the cohorts are observed (t_1) there are A new cohorts, and in each following year one new cohort ages into the cohorts observed.

In the McKenzie (2006) setup, the cohort aged a_j observed in time period t_k can be denoted as c_{j-k+1}. This is because the cohort aged a_j observed in time period t_k is the same as the cohort aged a_{j+n} observed in time period t_{k+n}. However, this is not true for data from the WVS-EVS, because the observation years are not consecutive. Thus, I refer to the cohort born in year b as cohort b. For individual i in cohort b of age a in

[26] More generally, $C = A + T - 1$ if the differences in t_n from one observation period to the next are the same differences as in the age categories. Usually these units are years, but they could, for example, be months.

time period t, let the variable of interest (e.g., life satisfaction) be denoted as $y_{i,b,a,t}$. The inconvenience of this notation is that the subscripts do not express the restriction that only certain combinations make sense. For example, it is nonsensical to talk about the life satisfaction of the cohort born in 1940 and aged 50 when surveyed in 1970 (which could be allowed by the notation as $y_{i,1940,50,1970}$). We capture this restriction by noting that $t = c + a$. Substituting in the subscript gives $y_{i,b,a,b+a}$. We drop the unnecessary last subscript and model $y_{i,b,a}$ as

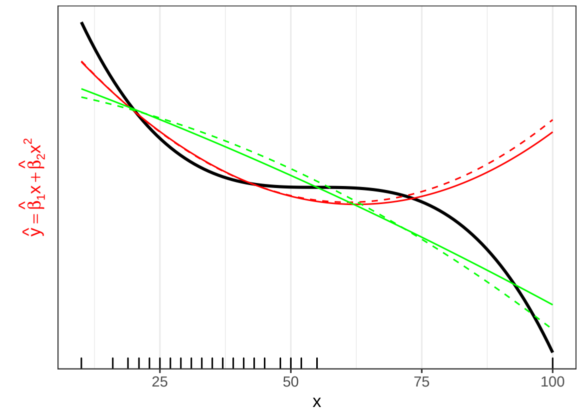
$$y_{i,b,a} = \alpha_b + \beta_a + \gamma_{b+a} + \epsilon_{i,b,a}, \tag{15}$$

where α_b is the effect of being in the cohort with birth year b, β_a is the effect of having age a, and γ_{b+a} is the effect of being observed in year $b + a$. This additive structure is not enough to identify any of the effects. To see this, consider that we could add a time trend to the cohort and age effect profiles and then subtract it from the period effect profile: Define a new set of effect profiles as

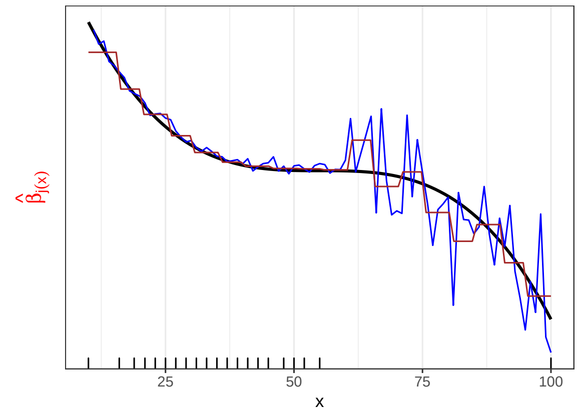
$$\begin{aligned}
\tilde{\alpha}_b &= \alpha_b + sb \\
\tilde{\beta}_a &= \beta_a + sa \\
\tilde{\gamma}_{b+a} &= \gamma_{b+a} - s(b + a)
\end{aligned}$$

where s is a fixed scalar. Then $\alpha_b + \beta_a + \gamma_{b+a} = \tilde{\alpha}_b + \tilde{\beta}_a + \tilde{\gamma}_{b+a}$. Without theory to put more structure on what s would be, for example, we are stuck. However, McKenzie (2006) shows that in model (15) we can identify second differences of the effect profiles.

In a recent paper, Landeghem (2012) uses the identification framework of McKenzie (2006) to test the convexity of well-being in age. Landeghem uses panel data from the German Socio-Economic Panel in which individuals were observed in consecutive years from 1984 to 2007, and adapts McKenzie's framework to allow for true panel data instead of pseudo-panel data. He concludes that well-being is indeed convex in age for the majority of the life cycle.



(a) Quadratic fits with and without weighting



(b) Dummies (1- and 5- grouped)

Figure 3: Improvements to Quadratic Fits

Figure 3a shows the same graph as Figure 1, with two additional fits (the green curves), which are weighted with inverse frequency weights, as described in Section B.2. The OLS regressions are the solid curves and the median regressions are the dotted curves. The thick black curve is the true regression function. Figure 3b shows fits to the same data using single (blue) and 5-year (red) dummy specifications. $\hat{\beta}_{j(\mathrm{x})}$ is the coefficient for the interval that x belongs to (or the coefficient for x itself in the case of single dummies for x).

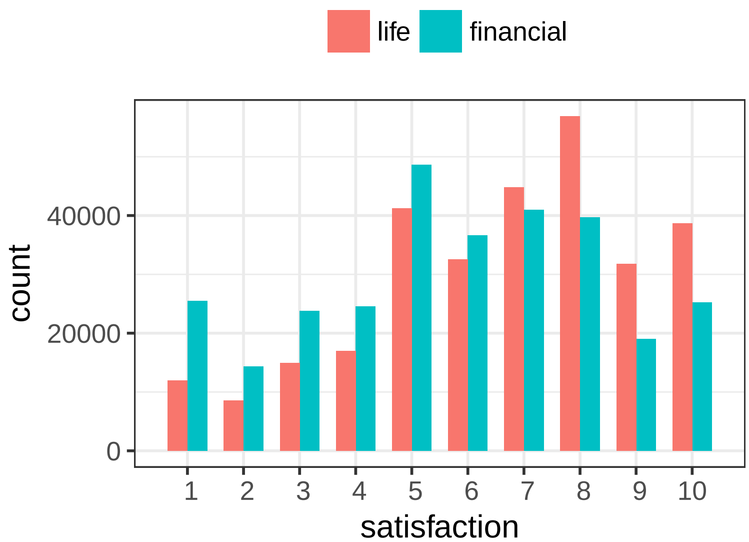
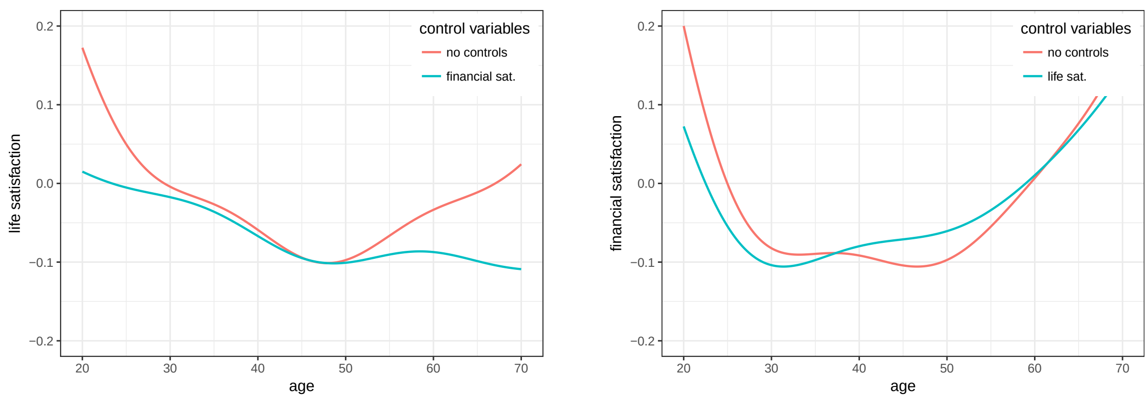


Figure 4: Distributions of life and financial satisfactions

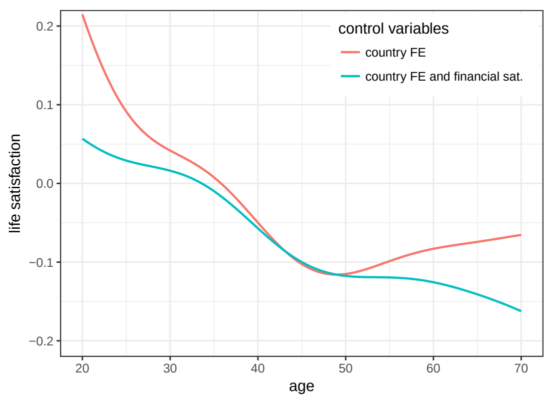
This figure compares the marginal distributions of life satisfaction (red) and financial satisfaction (blue) in the WVS-EVS.



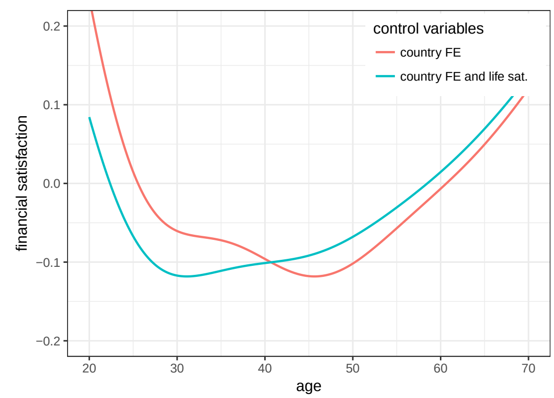
(a) Specifications (9a) and (9b) (b) Specifications (10a) and (10b)

Figure 5: Life Satisfaction and Financial Satisfaction Against Age

This figure shows estimates of the age effects on life satisfaction (left panel) and on financial satisfaction (right panel) using GAMs, for the specifications referenced below each plot. The red line represents the estimates from the baseline specification. The blue line represents the estimates when additionally controlling for financial satisfaction (left panel), or life satisfaction (right panel).



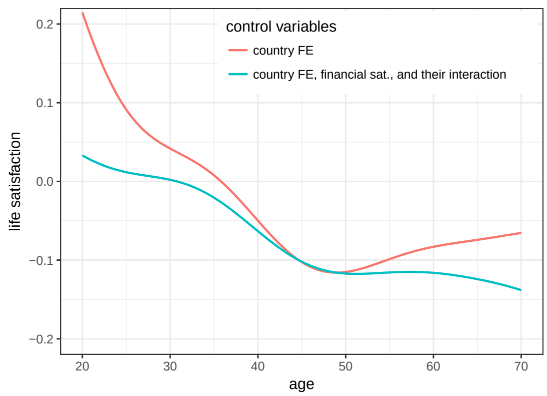
(a) Specifications (11a) and (11b)



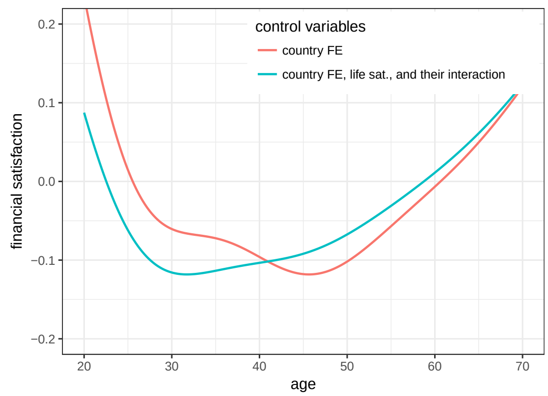
(b) Specifications (12a) and (12b)

Figure 6: Life Satisfaction and Financial Satisfaction Against Age (FE)

This figure shows estimates of the age effects on life satisfaction (left panel) and on financial satisfaction (right panel) using GAMs, for the specifications referenced below each plot. The red line represents the estimates from the baseline specification. The blue line represents the estimates when additionally controlling for financial satisfaction (left panel), or life satisfaction (right panel).



(a) Specifications (11a) and (13)



(b) Specifications (12a) and (14)

Figure 7: Life Satisfaction and Financial Satisfaction Against Age (FE Interacted)

This figure shows estimates of the age effects on life satisfaction (left panel) and on financial satisfaction (right panel) using GAMs, for the specifications referenced below each plot. The red line represents the estimates from the baseline specification. The blue line represents the estimates when additionally controlling for financial satisfaction (left panel), or life satisfaction (right panel).

Table 5: Inconsistency of Multivariate OLS Test for U-Shape

		sample size					
exp.	H_0	100	500	1,000	5,000	10,000	100,000
1.5	$\beta_{12} \leq 0$ (no U-shape)	0.05	0.06	0.05	0.06	0.06	0.06
	$\beta_{22} \leq 0$ (no U-shape)	1.00	1.00	1.00	1.00	1.00	1.00
	$\alpha_{X_1,[5]}$ are not U-shaped	0.00	0.00	0.00	0.00	0.00	0.77
	$\alpha_{X_2,[5]}$ are not U-shaped	0.00	0.14	0.53	0.91	0.96	1.00
2	$\beta_{12} \leq 0$ (no U-shape)	0.05	0.05	0.05	0.05	0.05	0.05
	$\beta_{22} \leq 0$ (no U-shape)	1.00	1.00	1.00	1.00	1.00	1.00
	$\alpha_{X_1,[5]}$ are not U-shaped	0.00	0.00	0.00	0.00	0.03	0.90
	$\alpha_{X_2,[5]}$ are not U-shaped	0.00	0.17	0.58	0.93	0.96	1.00
3	$\beta_{12} \leq 0$ (no U-shape)	0.15	0.22	0.24	0.26	0.28	0.29
	$\beta_{22} \leq 0$ (no U-shape)	1.00	1.00	1.00	1.00	1.00	1.00
	$\alpha_{X_1,[5]}$ are not U-shaped	0.00	0.00	0.00	0.00	0.28	0.95
	$\alpha_{X_2,[5]}$ are not U-shaped	0.00	0.02	0.28	0.88	0.94	0.99
4	$\beta_{12} \leq 0$ (no U-shape)	0.27	0.37	0.38	0.43	0.44	0.45
	$\beta_{22} \leq 0$ (no U-shape)	1.00	1.00	1.00	1.00	1.00	1.00
	$\alpha_{X_1,[5]}$ are not U-shaped	0.00	0.00	0.00	0.00	0.00	0.92
	$\alpha_{X_2,[5]}$ are not U-shaped	0.00	0.00	0.00	0.57	0.81	0.98

This table shows the proportion of times each test rejects the null hypothesis, specified in the H_0 column, at the 0.05 level. The data generating process is given in equation (7). The OLS-implemented test $\beta_{k2} \leq 0$ rejects if the coefficient on the quadratic term for the kth regressor is significantly positive. The OLS-implemented test $\alpha_{X_k,[5]}$ rejects if the 5 dummies that partition X_k are in the shape of a U and if each is significantly different from the previous. Proportions are based on 10,000 simulations.

Table 6: OLS Test for U-Shape Compared to Semi-Parametric Test

		sample size					
exp.	H_0	100	500	1,000	5,000	10,000	100,000
1.5	$\beta_{12} \leq 0$ (no U-shape)	0.09	0.11	0.13	0.14	0.18	0.16
	$\beta_{22} \leq 0$ (no U-shape)	1.00	1.00	1.00	1.00	1.00	1.00
	$\alpha_{X_1,[5]}$ are not U-shaped	0.00	0.00	0.00	0.05	0.11	0.63
	$\alpha_{X_2,[5]}$ are not U-shaped	0.00	0.00	0.00	0.00	0.00	0.00
	$f_1(x_1)$ is quasi-convex	0.13	0.17	0.15	0.15	0.17	0.21
	$f_1(x_1)$ is monotone	0.80	1.00	1.00	1.00	1.00	1.00
	$f_2(x_2)$ is quasi-convex	0.00	0.00	0.00	0.00	0.00	0.00
	$f_2(x_2)$ is monotone	0.82	0.92	0.94	1.00	1.00	1.00
2	$\beta_{12} \leq 0$ (no U-shape)	0.05	0.04	0.04	0.05	0.06	0.05
	$\beta_{22} \leq 0$ (no U-shape)	1.00	1.00	1.00	1.00	1.00	1.00
	$\alpha_{X_1,[5]}$ are not U-shaped	0.00	0.00	0.00	0.11	0.25	0.78
	$\alpha_{X_2,[5]}$ are not U-shaped	0.00	0.00	0.00	0.00	0.00	0.00
	$f_1(x_1)$ is quasi-convex	0.26	0.31	0.33	0.49	0.52	0.55
	$f_1(x_1)$ is monotone	0.93	1.00	1.00	1.00	1.00	1.00
	$f_2(x_2)$ is quasi-convex	0.02	0.00	0.01	0.00	0.00	0.00
	$f_2(x_2)$ is monotone	0.94	0.96	0.98	1.00	1.00	1.00
3	$\beta_{12} \leq 0$ (no U-shape)	0.44	0.56	0.56	0.63	0.66	0.68
	$\beta_{22} \leq 0$ (no U-shape)	1.00	1.00	1.00	1.00	1.00	1.00
	$\alpha_{X_1,[5]}$ are not U-shaped	0.00	0.00	0.00	0.03	0.17	0.68
	$\alpha_{X_2,[5]}$ are not U-shaped	0.00	0.00	0.00	0.00	0.00	0.00
	$f_1(x_1)$ is quasi-convex	0.62	0.88	0.97	1.00	1.00	1.00
	$f_1(x_1)$ is monotone	0.93	1.00	1.00	1.00	1.00	1.00
	$f_2(x_2)$ is quasi-convex	0.01	0.00	0.00	0.00	0.00	0.00
	$f_2(x_2)$ is monotone	0.97	0.98	1.00	1.00	1.00	1.00
4	$\beta_{12} \leq 0$ (no U-shape)	0.56	0.67	0.70	0.75	0.72	0.77
	$\beta_{22} \leq 0$ (no U-shape)	1.00	1.00	1.00	1.00	1.00	1.00
	$\alpha_{X_1,[5]}$ are not U-shaped	0.00	0.00	0.00	0.00	0.01	0.52
	$\alpha_{X_2,[5]}$ are not U-shaped	0.00	0.00	0.00	0.00	0.00	0.00
	$f_1(x_1)$ is quasi-convex	0.69	0.91	0.98	1.00	1.00	1.00
	$f_1(x_1)$ is monotone	0.86	0.99	1.00	1.00	1.00	1.00
	$f_2(x_2)$ is quasi-convex	0.02	0.00	0.01	0.00	0.00	0.00
	$f_2(x_2)$ is monotone	0.98	0.98	0.99	1.00	1.00	1.00

This table shows the proportion of times each test rejects the null hypothesis at the 0.05 level. The data generating process is given in equation (8). In addition to the tests in Table 5, this table includes the semi-parametric tests of Section 4. For the OLS-based tests, the null is that there is no U-shape of the dependent variable in the indicated variable (X_1 or X_2). For the semi-parametric test of quasi-convexity (monotonicity), the null is that the relationship between X_k and Y is quasi-convex (monotone). Proportions are based on 10,000 simulations.

	U☆	U※	Inv-U☆	Inv-U※	mono. ↓	mono. ↑
Developing	11	12	6	11	18	5
E. European	6	1	0	0	6	0
Western	6	6	3	3	3	5
Sum	23	19	9	14	27	10

(a) Unconditional

	U☆	U※	Inv-U☆	Inv-U※	mono. ↓	mono. ↑
Developing	7	14	4	17	13	8
E. European	7	1	0	0	5	0
Western	5	5	3	6	5	2
Sum	19	20	7	23	23	10

(b) Controlling for financial satisfaction

Table 7: OLS Results for Life Satisfaction

This table provides OLS results for testing a U-shape of life satisfaction in age. Subtable (a) has no control variables. Subtable (b) controls for financial satisfaction. ☆ means "significant" and ※ means "not significant" (at the 0.05 level).

	U☆	U⛝	Inv-U☆	Inv-U⛝	mono. ↓	mono. ↑
Developing	22	18	4	6	11	2
E. European	11	0	0	0	2	0
Western	15	5	1	1	3	1
Sum	48	23	5	7	16	3

(a) Controlling for income, education, employment status, and marital status

	U☆	U⛝	Inv-U☆	Inv-U⛝	mono. ↓	mono. ↑
Developing	12	21	4	8	13	5
E. European	7	3	0	0	3	0
Western	9	8	1	2	5	1
Sum	28	32	5	10	21	6

(b) Additionally controlling for financial satisfaction

Table 8: OLS Results for Life Satisfaction with Control Variables

This table provides OLS results for testing a U-shape of life satisfaction in age. Subtable (a) controls for the variables specified, which is consistent with Blanchflower and Oswald (2008). Subtable (b) controls for the same variables as in (a), and additionally for financial satisfaction. ☆ means "significant" and ⛝ means "not significant" (at the 0.05 level).

	U☆	U⛝	Inv-U☆	Inv-U⛝	mono. ↓	mono. ↑
Developing	15	11	7	5	22	3
E. European	5	3	0	0	5	0
Western	7	6	2	0	2	9
Sum	27	20	9	5	29	12

(a) Unconditional

	U☆	U⛝	Inv-U☆	Inv-U⛝	mono. ↓	mono. ↑
Developing	15	15	3	9	18	3
E. European	3	2	0	0	5	3
Western	5	3	1	0	3	14
Sum	23	20	4	9	26	20

(b) Controlling for life satisfaction

Table 9: OLS Results for Financial Satisfaction

This table provides OLS results for testing a U-shape of financial satisfaction in age. Subtable (a) has no control variables. Subtable (b) controls for life satisfaction. ☆ means "significant" and ⛝ means "not significant" (at the 0.05 level).

	U☆	U※	Inv-U☆	Inv-U※	mono. ↓	mono. ↑
Developing	28	14	2	5	8	6
E. European	10	2	0	0	1	0
Western	16	3	0	1	1	5
Sum	54	19	2	6	10	11

(a) Controlling for income, education, employment status, and marital status

	U☆	U※	Inv-U☆	Inv-U※	mono. ↓	mono. ↑
Developing	14	25	3	6	8	7
E. European	4	4	0	0	2	3
Western	11	3	0	0	1	11
Sum	29	32	3	6	11	21

(b) Additionally controlling for life satisfaction

Table 10: OLS Results for Financial Satisfaction with Control Variables

This table provides OLS results for testing a U-shape of financial satisfaction in age. Subtable (a) controls for the variables specified, which is consistent with Blanchflower and Oswald (2008). Subtable (b) controls for the same variables as in (a), and additionally for life satisfaction. ☆ means "significant" and ※ means "not significant" (at the 0.05 level).