From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Economics

2015

Essays on Testing Non-Monotone Relationships

Kostyshak, Scott

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp01p5547t776

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

ESSAYS ON TESTING
NON-MONOTONE RELATIONSHIPS

SCOTT KOSTYSHAK

A DISSERTATION

PRESENTED TO THE FACULTY

OF PRINCETON UNIVERSITY

IN CANDIDACY FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE

BY THE DEPARTMENT OF ECONOMICS

ADVISOR: BO HONORÉ

SEPTEMBER 2015

© Copyright by Scott Kostyshak, 2015.

All rights reserved.

Abstract:

These three chapters discuss problems and propose solutions to exploring potentially non-monotone regression functions.

Chapter 1 focuses on a specific class of non-monotone regression functions, U-shapes. I propose a non-parametric test of U-shaped regression functions based on critical bandwidth (CB), first introduced by Silverman (1981). The tests in this chapter are extended to work in generalized additive models, which allows investigation of whether there is an inherent U-shape between two variables or if the relationship is actually 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.

Chapter 2 combines three new elements to take a closer look at whether there is an inherent U-shape between subjective well-being and age. First, the latest waves of data from the surveys are used; second, semi-parametric methods are used in addition to the commonly used OLS and ordered logit quadratic specifications; and third, financial satisfaction is added as a control variable which changes the relationship between life satisfaction and age from a U-shape to monotonically decreasing. The U-shape of financial satisfaction explains the U-shape of life satisfaction, which suggests evidence against the theory that the U-shape of life satisfaction is due to a midlife crisis that is hard-coded in DNA.

Chapter 3 provides an intuitive way to explore non-monotone relationships in a multivariate context. It is well-known that OLS is sensitive to a minority's strong positive effect drowning out a majority's weak negative effect. This chapter proposes estimation of a new parameter, the Partial Monotonicity Parameter (PMP), which improves on the common practice of testing predictions of monotonicity with OLS, and provides a single number that summarizes the degree to which two random

variables have a non-monotone relationship. By combining OLS which is sensitive to magnitude, with PMP which is not sensitive to magnitude, a rich summary of the regression function can be succinctly reported. A plug-in estimator based on local polynomial regression is proposed.

Acknowledgments:

I would first like to thank my advisor, Bo Honoré. I was lucky enough to have him as a teacher of econometrics in my first year at Princeton. His skill for communicating the important aspects of econometrics and intuition lead me to understand how econometrics is more than just math. His intuition combined with the gift and effort of explaining his insights is a rare combination. As an advisor, Bo has given me the freedom to be the type of researcher I am, which is somewhere in-between an econometrician and an applied economist. Whenever I came to a river with no crossing Bo was there to advise me on how to build a bridge, or to suggest I turn around and run. I would also like to thank Angus Deaton, who was my main advisor before I moved more towards econometrics. He was always available to push me in the right direction. The most important lesson I have learned from Angus is the appreciation of simple methods. I am also grateful that Angus has provided guidance on studying subjective well-being. I would like to thank Tom Vogl, who is wise beyond his years. I got to know Tom as his teaching assistant for an applied econometrics class in the Woodrow Wilson School. His ability to balance a fascinating research agenda with a sincere and effective effort towards teaching and advising, combined with an always pleasant attitude, is an inspiration for myself as I begin my academic career.

The Center for Health and Wellbeing at Princeton and the Research Program in Development Studies provided an excellent atmosphere for research and was always welcoming. I am especially thankful for the comments from applied researchers which kept me grounded while I was focusing on econometrics. I would particularly like to thank Ilyana Kuziemko for her insightful comments. I would like to thank Anne Case for her comments on my work, and also for her dedication to Princeton's field of economic development. In my second year, development would not have been offered

as a field if it were not for Anne's generous dedication. I would like to thank Hannes Schwandt for our conversations on subjective well-being.

I would not be at Princeton in the first place if it were not for my UCSD mentors Julie Cullen and James Rauch. They believed in me more than I believed in myself. I am very thankful to Ian Abramson for getting me excited about statistics, and planting the seed of non-parametrics that would grow over time at a rate almost as fast as $\sqrt{t}$.

My friend Vladimir Asriyan has been in some sense an advisor to me for the last decade of my life. I highly appreciate his kindness and friendship and all the good times we have shared. Along with Santiago Halty and Zack Zimbalist, I was fortunate enough to find a group of friends at UCSD who shared similar values and who motivated and challenged me to think.

At Princeton, I have had incredible classmates who made for a wonderful experience. I would not have made it past the first couple of years without the support in particular of Judd Cramer, Gabriele la Spada, and Neel Sukhatme. Thank you to Judd for his assists on and off the court. Thank you to Mariano Somale and Natalia Cámpora for making Princeton feel like home; and to Thomas Winberry and Andrés Maggi for all the great memories.

Thank you to Lillian Anderson and Denise Zapecza, for making Wallace a wonderful place to be a part of, and to Laura Hedden for her help though the years. Her sincere care towards each graduate student is exemplary.

Thank you to my mother and father, Betty Semple and Geoff Kostyshak, for their support and for teaching me to love education. Thank you to my brother, Todd Kostyshak, whom I followed to San Diego and then to the East Coast because I knew I had someone close by who would be there for me. I thank my nephew, Jacob Kostyshak, for literally tearing up my dissertation and thus reminding me what is important in life. Thanks to my cousin, Amira Adassen, whose curiosity in world

issues and ambition to help others inspires me and reminds me of why I became interested in economics.

Thank you to Perihan Saygın, for her comments on my research and for making the job market experience more enjoyable than it was supposed to be. Her kindness, wisdom, and passion for research inspires me. She has given me direction in life and a reason to have expressions.

For Grandpa Ted

Contents

Abstract iii

Acknowledgments v

List of Tables xii

List of Figures xiii

1 Non-Parametric Testing of U-Shaped Relationships **1**

1.1 Introduction[1] 2

1.2 Conventional Tests of a U-shape 4

1.2.1 Implicit weighting of OLS on thick marginal density 5

1.2.2 Inconsistency of multivariate OLS under misspecification . . . 12

1.2.3 Simulations 14

1.2.4 Lind and Mehlum (2010) 17

1.3 Critical Bandwidth Tests of U-shapes 19

1.3.1 Non-parametric testing of a U-shape 19

1.3.2 A critical bandwidth approach 20

1.3.3 Consistency of h_{stat} 22

1.3.4 Bootstrapping the null distribution 26

1.3.5 The flatness problem 27

1.3.6 Interpreting the non-parametric test results 29

[1] A previous version of this chapter was presented at the Simposio de la Asociación Española de Economía.

1.4 Semi-parametric Testing of Multivariate Hypotheses 30
1.4.1 Simulations 33
1.5 Conclusion 36
References 37
1.A Denominator of OLS Limit in Section 1.2.2 40

2 Robustness of Subjective Well-Being U-Shapes in Age 41

2.1 Introduction[2] 41
2.2 SWB Model Specification and Identification Issues 44
2.2.1 The identification problem of age, period, and cohort effects . 44
2.2.2 What if the happiest live longer? 50
2.2.3 Is the meaning of "satisfaction" universal? 51
2.2.4 Variable selection 52
2.3 Parametric Data Analysis 54
2.3.1 Description of data 54
2.3.2 Replication of Blanchflower and Oswald (2008) 55
2.3.3 Tests of a U-shape 59
2.4 Semi-parametric Analysis of the Data 62
2.4.1 Applying a semi-parametric test of a U-shape 62
2.4.2 Within-country results 62
2.4.3 OLS results using a quadratic specification 66
2.4.4 Cross-country results 69
2.5 Conclusion 76
References 76
2.A Identification of Second Differences of Age, Period, and Cohort Effect Profiles 79

[2]A previous version of this chapter was presented at the Simposio de la Asociación Española de Economía.

2.B Table 5 of Blanchflower and Oswald (2008) 82
2.C Defining a U-shape 82
2.D Cubic Specification as Improvement to Quadratic 84
2.E Whether to Condition on Health 85
2.F Variables in the World Values Survey 87
2.G Ordered Logit Dummy Plots 88

3 The Partial Monotonicity Parameter 90
3.1 Introduction 90
3.2 Motivation 92
3.2.1 The partial monotonicity parameter 92
3.2.2 Related literature 93
3.2.3 PMP as a generalization of monotonicity 93
3.2.4 Reduced form 95
3.2.5 Ease of understanding 95
3.2.6 Related parameters of interest 95
3.3 Estimation 97
3.3.1 Discussion of analytical finite distribution of $\hat{\delta}_n^{PMP}$ 97
3.3.2 Bootstrapping the finite distribution of $\hat{\delta}_n^{PMP}$ 98
3.3.3 A critical bandwidth test of δ^{PMP} 104
3.4 Multivariate Hypotheses 105
3.5 An Application in Subjective Well-Being 106
3.6 Conclusion 108
References 108

List of Tables

1.1 Inconsistency of Multivariate OLS Test for U-Shape 15

1.2 Inconsistency of Multivariate OLS Test for U-Shape Compared with Semi-Parametric U-Shape Test 18

1.3 Semi-Parametric Test with Two Covariates 33

1.4 Semi-Parametric Test with Five Covariates 35

2.1 Argmins of SWB-Age Relationship, by Country Category and Sex . . 56

2.2 Tests for U-Shape of Life Satisfaction in Age 62

2.3 Tests for U-Shape of Financial Satisfaction in Age 64

2.4 Argmins of SWB-Age Relationship, by Country Category and Sex (Table 5 in Blanchflower and Oswald, 2008) 82

3.1 Simulations of Estimating the PMP (Bootstrapped SEs) 99

3.2 Simulations of Estimating the PMP (Bootstrapped SEs) 101

3.3 OLS and PMP Estimation of SWB on Age and Financial Satisfaction 107

List of Figures

1.1 Quadratic Fits (OLS and Quantile) 6

1.2 Improvements to Quadratic Fits 9

1.3 True Regression Functions 32

2.1 Cohort Effects, No Period Effects 46

2.2 Controlling for Cohort Effects but Not for Period Effects 48

2.3 Trying to Control for All Three 50

2.4 Life Satisfaction by Health 53

2.5 Income vs. Age in France 54

2.6 Developing Countries 58

2.7 Western Countries 58

2.8 Eastern European 59

2.9 Life Satisfaction vs. Age in Iraq 61

2.10 OLS Results for Life Satisfaction 66

2.11 OLS Results for Life Satisfaction with More Control Variables 67

2.12 OLS Results for Financial Satisfaction 68

2.13 OLS Results for Financial Satisfaction with More Control Variables . 69

2.14 Life and Financial Satisfactions Against Age 71

2.15 Life and Financial Satisfactions Against Age (FE) 72

2.16 Life and Financial Satisfactions Against Age (FE Interacted) 73

2.17 Distribution of Life and Financial Satisfactions 74

2.18 Health vs. Age 86

3.1 True Regression Functions (with PMP) 103

3.2 Graphical Depiction of Non-Monotone Relationship 108

Chapter 1

Non-Parametric Testing of U-Shaped Relationships

1.1 Introduction[1]

Theories predicting U-shapes and inverted U-shapes of one variable in another are common in the social sciences, including U-shapes of female labor force participation in development (Goldin, 1995) and life satisfaction in age (Blanchflower and Oswald, 2008); and inverted U-shapes of inequality in development (Kuznets, 1955), of union membership in age (Blanchflower, 2006), and of innovation in competition (Aghion et al., 2005). 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 et al., 1981; Shaper et al., 1988; Marmot and Brunner, 1991), can be tested as a U-shape hypothesis. Any theory that predicts a single change of sign in the derivative of the regression function can be tested as a U-shape theory. For more examples of U-shapes in economics, see the survey in Lind and Mehlum (2010).

Despite the prevalence of theories predicting U-shapes, there are few tools for testing them. These theories are most commonly tested with OLS using a squared term or by observing whether grouped dummies are consistent with a U-shape. In this paper I show some problems with this parametric type of testing and how it can lead to incorrect inference. The goal of this paper is to keep the advantages of a statistical test without sacrificing the advantages of non-parametrics.

Developing a non-parametric test for a theory predicting a U-shaped regression function is especially complicated when the turning point is not specified in the null hypothesis. By reframing the null hypothesis of a U-shaped (inverted U-shaped) regression function as a regression function with a single valley (peak) and zero peaks (valleys), the problem of testing such theories can be viewed in a more general

[1] A previous version of this chapter was presented at the Simposio de la Asociación Española de Economía.

framework of testing the number of modes of a regression function. I propose a non-parametric test based on critical bandwidth, first introduced by Silverman (1981) to test the number of modes of a density. The test statistic is the smallest bandwidth such that a non-parametric regression (e.g. local polynomial regression) is U-shaped or monotone. The essence of the test is an extension of Bowman et al. (1998), which considers a test of monotonicity of the regression function, based on critical bandwidth. To my knowledge, this paper is the first to use critical bandwidth tests to explore U-shapes.

Critical bandwidth tests have been known to be inconsistent for certain types of regression functions.[2] This paper considers an improvement based on adaptive bandwidth that leads to consistency in such cases. This contribution applies more generally to testing the number of bumps for both densities and regression functions. Two papers have proposed critical bandwidth approaches to testing numbers of bumps of regression functions, Harezlak and Heckman (2001) and Harezlak et al. (2007). They do not discuss the inconsistency issue noted above and deal only with the bivariate case, with no control variables.

In this paper I extend the critical bandwidth framework in a practical way to multivariate hypotheses, using partial linear models and generalized additive models, which allows investigation of whether there is an inherent U-shape between two variables or if the relationship is actually 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. These theoretical contributions apply more generally to any critical bandwidth test.

[2]Hall and Heckman (2000) showed that the test can fail to reject if there is a long flat part of the regression function in one part of the x-axis combined with a sharp dip in another part. Hartigan and Hartigan (1985) showed a similar problem for critical bandwidth tests of modes of densities. See Section 1.3.5 for more discussion.

The rest of this paper is organized as follows. In Section 2.3.3 I present existing tests of U-shaped regression functions based on OLS, discuss several inherent problems and corresponding improvements, and give necessary conditions for such tests to be consistent. In Section 1.3 I introduce the non-parametric test that forms the core of this paper and I prove consistency of the test statistic. Finally, I extend the framework to multivariate hypotheses and semi-parametrics in Section 1.4.

1.2 Conventional Tests of a U-shape

In order to examine theories of U-shaped regression functions, researchers most 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 disadvantages to first the quadratic test and then to grouped dummies, and how they can lead to incorrect inference when used to test for a U-shape, especially in multivariate specifications.

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

$$y_i = \alpha + \beta x_i + \gamma f(x_i) + \theta' \mathbf{z}_i + \epsilon_i, \quad i = 1, \ldots, n \tag{1.1}$$

where x_i is the explanatory variable of interest, ϵ_i is an error term, $\mathbf{z}_i$ is a vector of control variables, and f is a function chosen by the researcher, perhaps based on theory. $f(x)$ is usually chosen to be x^2. The null hypothesis of {positive quadratic} is 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 minimum of the fit is inside an acceptable range. An immediate concern of the quadratic test is the parametric restriction: OLS will find evidence of a U-shape if a positive quadratic fits the data better than a

negative quadratic and if the minimum of the fit falls in a reasonable range. We will see how the induced null and alternative sets are often not what we would prefer.

Such tests are attractive because they are easy to carry out and to interpret. Further, obtaining a point estimate of the minimum is simple and the corresponding standard error is a straightforward application of the delta method. Because these tests are so common and they are the main alternative to the non-parametric test proposed in Section 1.3, I spend much of this section exploring their properties. Lind and Mehlum (2010) analyze the seven articles between 2001 and 2009 in the *American Economic Review* whose focus was testing U-shaped regression functions. All of these papers contain a test with a quadratic specification. As Lind and Mehlum show, some of the papers carry out the test incorrectly, such as failing to include the linear term, in which case monotone regression functions are no longer in the alternative set; or using a joint test of the linear and squared terms (only the squared term should be tested). Other papers do not test that the minimum is in an acceptable range. Thus, although the OLS test is easy to apply and interpret, it appears that mistakes are still common which can yield misleading conclusions.

Of the seven papers, Imbs and Wacziarg (2003) is the only one to use non-parametric techniques. However, the thorough non-parametric techniques based on LOWESS do not allow the authors 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). In addition to the non-parametric investigation, the authors use an OLS quadratic specification.

1.2.1 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. The preceding statement holds

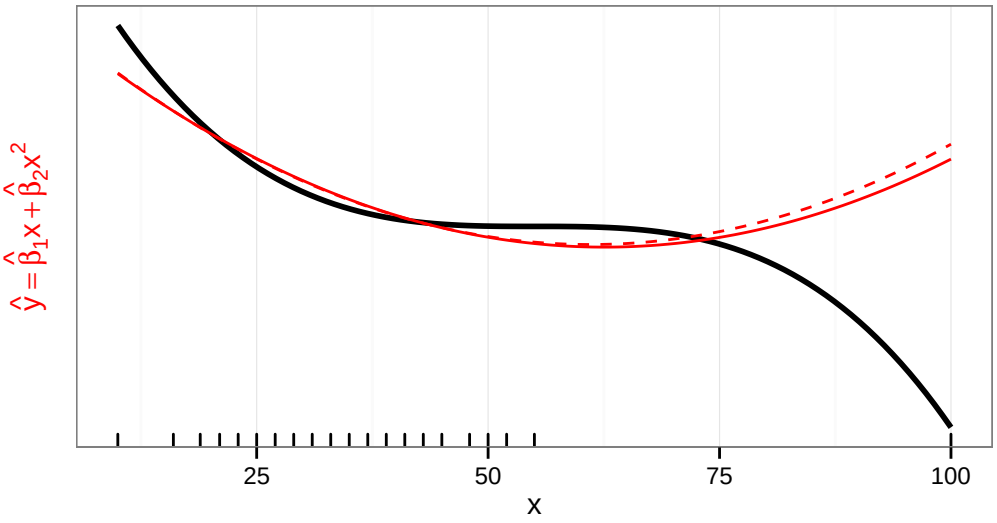


Figure 1.1: Quadratic Fits (OLS and Quantile)

This figure shows two estimated regressions, one using OLS (solid line) and one using median regression (dotted line), of the true regression function which is the black line. The first stick poking up from the x-axis is the minimum of the realizations of X and the last stick is the maximum. Each stick in-between represents 5% increments of the number of observations.

true even (perhaps especially) if we have a large random sample. The reason is that OLS's goal of minimizing squared residuals does not always coincide with the goal of estimating the regression function. Note, however, that if the model is correctly specified, then there is no problem.

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

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

where Y is the variable of interest, X is the variable in which Y is theorized to be U-shaped, and $\mathbf{Z}$ is a vector of covariates. The OLS estimator with a large sample, even under misspecification (e.g. choosing $f(x)$ in equation (1.1) to be x^2 when the

true function is $|x|^{2.5}$), will give us a reasonable (BLUE) answer to minimizing

$$E\left[\left(m\left(X,Z\right)-\hat{m}\left(X,\mathbf{Z}\right)\right)^{2}\right] \tag{1.2}$$

over functions $\hat{m}$. Under misspecification, however, OLS will not, in general, give a reasonable minimization of

$$E\left[\left(m\left(X,\mathbf{Z}\right)-\hat{m}\left(X,\mathbf{Z}\right)\right)^{2}\middle|X=x,\mathbf{Z}=\mathbf{z}\right] \tag{1.3}$$

for all values of x and $\mathbf{z}$. It will likely only give a good answer to minimizing (1.3) for sets in the support of X where the density is thick. To see this, note that in (1.2) the expectation is over randomness in X and $\mathbf{Z}$, and the expectation yields a single real number. In (1.3) the expectation is not over the distribution of X—it is conditional on $X=x$. If we use OLS to minimize (1.2) with the end goal of minimizing (1.3) then we are using a statistic that depends asymptotically 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 $m\left(x,\mathbf{z}\right)$.

In many cases, (1.3) is not of interest. For example, policy makers might not care about what happens to 10 percent of the population; they might only care about where the marginal distribution is thick, and thus a random sample would give them the answer. 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. It applies to any estimation strategy that minimizes the sum of functions of the absolute values of the residuals

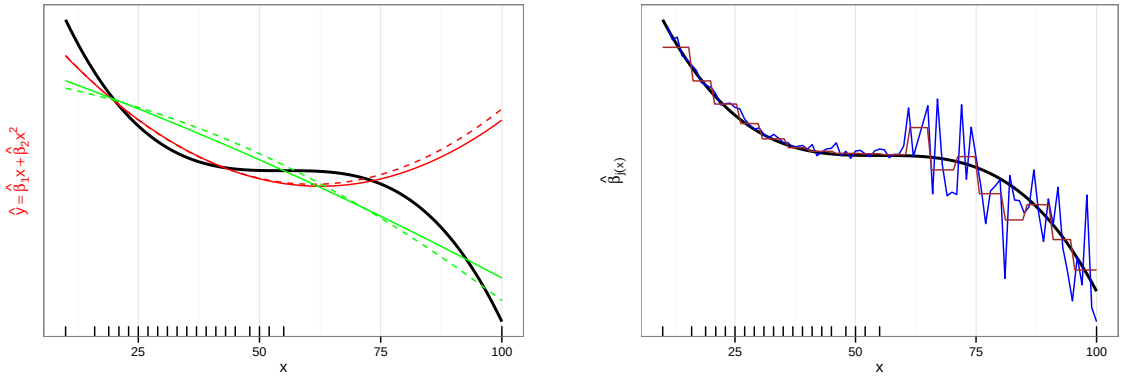
in a parametric fit. Median regression will likely be worse: The property that OLS is sensitive to outliers could help correct some of the problem, whereas a median regression is more able to ignore areas where the fit is very poor.

This implicit weighting issue can be seen for a particular case using simulations in Figure 1.1. According to a naive OLS or median regression, the draws are consistent with being from a U-shaped regression function. However, the black line shows that the true regression function is monotonically decreasing and is not convex. The .95 quantile is just before X equals 60. Both OLS and median regression fit the true regression curve poorly at higher values. Because OLS minimizes the total sum of squared errors, minimizing the squared errors for $X < 60$ is given priority over minimizing the squared errors for observations between values 60 and 100 because there are more errors 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. The next section proposes a way to fix this particular problem if the researcher is intent on imposing a parametric form for the regression function.

Finally, it should be noted that 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. This particular issue should be kept in mind but is not discussed further in this paper.

Addressing density weighting with inverse weights

In order to estimate the regression function in a way that is valid globally across the support of X, we should prevent the fitting algorithm (e.g. OLS) from ignoring



(a) Quadratic fits with and without weighting

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

Figure 1.2: Improvements to Quadratic Fits

Figure 1.2a shows the same graph as Figure 1.1 with two additional fits. The dotted lines are weighted with inverse frequency weights, as described in Section 1.2.1. Figure 1.2b shows fits to the same data using single and 5-year 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).

points where the marginal distribution of the independent variables is thin. One way to do this is to artificially redistribute 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 where the vector of covariates, $\mathbf{z}$, is empty. OLS minimizes

$$\int (m(x) - \hat{m}(x))^2 \, dF_X(x) \,. \tag{1.4}$$

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, something more sensible would be to minimize

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

which is the special case of (1.4) 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 original weights in (1.4). Note that we do not have the same problems that, for example, feasible weighted least squares can suffer from, because the ECDF itself gives the ideal weights; we do not use it as an estimator of the CDF of X.

Figure 1.2a shows these proposed weights applied to the OLS and quantile regressions with the quadratic specification. The weighted regressions are the dotted lines, which give a better global fit to the regression function and which correctly reject[3] a U-shape.

Inference based on dummies

A common attempt to address the parametric inflexibility is to make dummies for intervals partitioning x and regressing y on this set of dummies instead of $f(x)$. This approach is often referred to in the applied literature as "non-parametric." Compared to a parametric fit, using a dummies approach can reduce bias when fitting, as shown in Figure 1.2b. 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 indicating 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 has theoretical issues where the turning point is unspecified. One could in practice take the minimum dummy coefficient as a turning point.

[3]The technically correct statement is "fail to reject that the regression function is not U-shaped."

In practice, this test suffers from the property that dummies close to the bottom of the U are very close to each other around the turning point 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 Tables 1.1 and 1.2. If the researcher is intent on using such a test in practice, she should emphasize that the significance away from the turning point is necessary but that the significance tests around the turning point are known to suffer from low power. On a separate note regarding power, this test would also be affected by the "flatness problem" that will be discussed in Section 1.3.5 because of low power from a small derivative.

Note also from a theoretical perspective that the researcher would be making restrictions that are often known to be unlikely. For example, when exploring the effect of alcohol on mortality in Marmot et al. (1981), by using dummies the researchers effectively restrict the effect on mortality of drinking on average 34 grams of alcohol a day to be the same as drinking on average 1 gram of alcohol a day. The fact that the estimate of mortality for drinkers of 34 grams is more affected by observations with a value of 1 gram of alcohol a day than by observations drinking 34.1 grams is a disadvantage of such a model.

Relying on grouped dummies can lead to incorrect inference in a multivariate regression. Unless X_1 is categorical and the dummies are saturated, this test is generally inconsistent because the dummies approach restricts the fit to be stepwise and the residual non-linearity can still be spuriously attributed to X_2. 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 1.2.2. Further, this test is included in the simulations shown in Tables 1.1 and 1.2.

If we were interested in how the regression function in Figure 1.1 behaves for X values between 60 and 100, we could estimate a separate regression involving only those observations. Similarly, we could use spline regressions or other non-parametric methods designed for estimating regression functions. However, we would lose the benefits of a statistical test. This paper aims to keep the advantages of a statistical test without losing the flexibility of non-parametrics.

1.2.2 Inconsistency of multivariate OLS under misspecification

Suppose the researcher is worried 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. In this section I will show that existing tests of U-shapes, based on OLS with a quadratic or grouped dummies[4], can find spurious evidence of a U-shape in X_1, controlling for X_2. Non-linear correlation can often be spuriously attributed to variables under misspecification. Suppose that the researcher is interested in testing the theory that Y is U-shaped in X_1, conditional on X_2, and that

$$Y = f(X_2) + \epsilon_Y, \tag{1.5}$$

where $\epsilon \sim N(0,1)$. Note that X_1 does not enter the equation so that the truth is that Y is not U-shaped in X_1, conditional on X_2. It is common for researchers to test for a U-shape of Y in X_1, controlling for X_2 by regressing Y on X_1, X_1^2 and X_2, and testing the significance of the coefficient on X_1^2. This is generally inconsistent 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.

[4]The issues discussed in this section also extend to dummies and any estimation procedure that imposes restrictions.

An improvement might 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 way. However, unless f is correctly specified, this test is still generally inconsistent. For simplicity, we consider regressing Y on only the non-linear terms (simulations in Section 1.2.3 will show results from including linear terms). We thus consider the following regression:

$$y = \alpha + \beta_1 g_1(x_1) + \beta_2 g_2(x_2) + \epsilon,$$

where g_1 and g_2 are functions chosen by the researcher. The most common choice of g_1 and g_2 is $g(x) = x^2$. We assume that

Assumption 1. *$g_1(X_1)$ and $g_2(X_2)$ are not perfectly correlated and $var(g_1(X_1)) > 0$, $var(g_2(X_2)) > 0$.*

Assumption 2. *$cov[g_1(X_1), \epsilon_Y]$ and $cov[g_2(X_2), \epsilon_Y]$ are both zero.*

In this section, stating that the regression corresponding to equation (1.5) 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.

Let $\hat{\beta}_1$ be the estimated OLS coefficient on $g_1(X_1)$. If the regression is consistent, $\hat{\beta}_1$ should tend to 0 because X_1 does not enter equation (1.5). We now explore under which conditions this is true. 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[5]

[5]Note that it is not required for g_1 or g_2 to be continuous, although the first and second moments of $g_1(X_1)$ and $g_2(X_2)$ must exist.

as

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

The denominator is 0 only when there is perfect correlation between $g_1\left(X_1\right)$ and $g_2\left(X_2\right)$ (proved in Appendix 1.A) so by assumption 1 $\beta_1 \xrightarrow{p} 0$ only when the numerator is 0. By assumption (2), for consistency to hold we must then have that

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

or

$$corr\left[f\left(X_2\right), g_1\left(X_1\right)\right] = corr\left[g_2\left(X_2\right), g_1\left(X_1\right)\right] corr\left[f\left(X_2\right), g_2\left(X_2\right)\right]. \tag{1.6}$$

In the case where f is correctly guessed and so $g_1 = f$, equation (1.6) holds and consistency is achieved if the researcher also sets $g_2 = f$. In a separate case, if $g_1\left(X_1\right)$ is uncorrelated with both $f\left(X_2\right)$ and $g_2\left(X_2\right)$, then consistency also holds.

However, if $f\left(x\right)$ is not correctly guessed and $g_1\left(X_1\right)$ is correlated with either $f\left(X_1\right)$ or $g_2\left(X_2\right)$, $\hat{\beta}_1$ will generally be inconsistent. Intuitively, equation (1.6) shows that in the case of misspecification of f, $\hat{\beta}_1$ can pick up part of what g_2 cannot approximate of f.

1.2.3 Simulations

Table 1.1 shows simulations of an OLS quadratic test. An OLS regression is run based on the following specification:

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_{X_1^2} X_1^2 + \beta_3 X_2 + \beta_{X_2^2} X_2^2 + \epsilon$$

Table 1.1: Inconsistency of Multivariate OLS Test for U-Shape

exp.	test	100	200	500	1000	10000	100000
1.5	$\beta_{X_1^2} > 0$	0.05	0.06	0.06	0.05	0.05	0.06
	$\beta_{X_2^2} > 0$	1.00	1.00	1.00	1.00	1.00	1.00
	$\alpha_{X_1,[5]}$ are U-shaped	0.00	0.00	0.00	0.00	0.00	0.77
	$\alpha_{X_2,[5]}$ are U-shaped	0.00	0.00	0.13	0.54	0.96	1.00
2	$\beta_{X_1^2} > 0$	0.05	0.05	0.05	0.05	0.05	0.05
	$\beta_{X_2^2} > 0$	1.00	1.00	1.00	1.00	1.00	1.00
	$\alpha_{X_1,[5]}$ are U-shaped	0.00	0.00	0.00	0.00	0.04	0.91
	$\alpha_{X_2,[5]}$ are U-shaped	0.00	0.00	0.16	0.59	0.96	1.00
3	$\beta_{X_1^2} > 0$	0.15	0.19	0.21	0.24	0.28	0.29
	$\beta_{X_2^2} > 0$	1.00	1.00	1.00	1.00	1.00	1.00
	$\alpha_{X_1,[5]}$ are U-shaped	0.00	0.00	0.00	0.00	0.27	0.95
	$\alpha_{X_2,[5]}$ are U-shaped	0.00	0.00	0.01	0.29	0.94	1.00
4	$\beta_{X_1^2} > 0$	0.27	0.32	0.36	0.39	0.44	0.46
	$\beta_{X_2^2} > 0$	1.00	1.00	1.00	1.00	1.00	1.00
	$\alpha_{X_1,[5]}$ are U-shaped	0.00	0.00	0.00	0.00	0.00	0.92
	$\alpha_{X_2,[5]}$ are U-shaped	0.00	0.00	0.00	0.00	0.81	0.99

This table shows the proportion of times each test rejects the null hypothesis. The OLS-implemented test $\beta_{X_k^2} > 0$ 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. The data generating process is given in equation (1.7). Proportions are based on 10000 simulations.

and the test is simply whether $\hat{\beta}_2$ is significantly positive. The data are generated as follows:

$$\begin{pmatrix} X_1 \\ X_2 \\ \epsilon \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & .8 & 0 \\ .8 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right]$$

$$Y = |X_2|^a + \epsilon, \tag{1.7}$$

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 $\hat{\beta}_{X_2^2}$ is significant about 5% of the time. However, when a is 3 or 4, the rejection rate is considerably larger than 5% and thus spuriously suggests that there is a U-shape of Y in X_1, conditional on X_2. These simulations thus support the inconsistency we found in equation (1.6).

Although the semi-parametric test has not yet been introduced, I provide results here, in Table 1.2, for easy comparison with the quadratic test. 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 is that Y is U-shaped in X_1, where for the quadratic test, the null hypothesis is that Y is *not* U-shaped in X_1. However, as the sample increases to where 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, where ϵ_k are independent standard normals:

$$\begin{aligned} X_2 &\sim \epsilon_1 \\ X_1 &= X_2 + X_2 \cdot \epsilon_2 \\ Y_2 &= |X_2|^a + .05 \cdot \sin(15 \cdot X_1) + \epsilon_3, \end{aligned} \tag{1.8}$$

Ideally, a test with a null hypothesis of "not U-shaped" should find even less evidence supportive of a U-shape in X_1: instead of the test statistic being insignificant (as it should be in the data generating process shown in equation (1.7)) it should be significant in the opposite direction. Table 1.2 shows the same result as Table 1.1: 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 $\hat{\beta}_{X_1^2}$ is significant more often than 0.05. The semi-parametric tests are consistent and correctly reject that Y_2 is U-shaped in X_1, conditional on X_2.

1.2.4 Lind and Mehlum (2010)

The only test (parametric or non-parametric) of a U-shape in addition to the vanilla OLS quadratic specification is Lind and Mehlum (2010). The test is implemented on top of the coefficients and standard errors obtained from estimating equation (1.1) with OLS. The function f must be chosen to have one extreme point. Then, the relationship is one of inverse U-shaped, 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:

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

Table 1.2: Inconsistency of Multivariate OLS Test for U-Shape Compared with Semi-Parametric U-Shape Test

exp.	test	100	200	500	1000	10000
1.5	$\beta_{X_1^2} > 0$	0.10	0.12	0.13	0.14	0.16
	$\beta_{X_2^2} > 0$	1.00	1.00	1.00	1.00	1.00
	$\alpha_{X_1,[5]}$ are U-shaped	0.00	0.00	0.00	0.00	0.13
	$\alpha_{X_2,[5]}$ are U-shaped	0.00	0.00	0.00	0.00	0.00
	$f_1(x_1)$ is quasi-convex	0.13	0.16	0.17	0.18	0.16
	$f_1(x_1)$ is monotone	0.79	0.96	0.99	1.00	1.00
	$f_2(x_2)$ is quasi-convex	0.01	0.00	0.00	0.00	0.00
	$f_2(x_2)$ is monotone	0.84	0.86	0.93	0.95	1.00
2	$\beta_{X_1^2} > 0$	0.05	0.05	0.05	0.05	0.05
	$\beta_{X_2^2} > 0$	1.00	1.00	1.00	1.00	1.00
	$\alpha_{X_1,[5]}$ are U-shaped	0.00	0.00	0.00	0.00	0.26
	$\alpha_{X_2,[5]}$ are U-shaped	0.00	0.00	0.00	0.00	0.00
	$f_1(x_1)$ is quasi-convex	0.24	0.29	0.32	0.34	0.51
	$f_1(x_1)$ is monotone	0.94	0.99	1.00	1.00	1.00
	$f_2(x_2)$ is quasi-convex	0.01	0.01	0.01	0.00	0.00
	$f_2(x_2)$ is monotone	0.93	0.94	0.97	0.98	1.00
3	$\beta_{X_1^2} > 0$	0.43	0.51	0.54	0.61	0.65
	$\beta_{X_2^2} > 0$	1.00	1.00	1.00	1.00	1.00
	$\alpha_{X_1,[5]}$ are U-shaped	0.00	0.00	0.00	0.00	0.17
	$\alpha_{X_2,[5]}$ are U-shaped	0.00	0.00	0.00	0.00	0.00
	$f_1(x_1)$ is quasi-convex	0.60	0.73	0.88	0.97	1.00
	$f_1(x_1)$ is monotone	0.94	0.98	1.00	1.00	1.00
	$f_2(x_2)$ is quasi-convex	0.01	0.00	0.00	0.00	0.00
	$f_2(x_2)$ is monotone	0.98	0.98	0.99	0.99	1.00
4	$\beta_{X_1^2} > 0$	0.54	0.61	0.64	0.67	0.73
	$\beta_{X_2^2} > 0$	1.00	1.00	1.00	1.00	1.00
	$\alpha_{X_1,[5]}$ are U-shaped	0.00	0.00	0.00	0.00	0.01
	$\alpha_{X_2,[5]}$ are U-shaped	0.00	0.00	0.00	0.00	0.00
	$f_1(x_1)$ is quasi-convex	0.68	0.77	0.91	0.98	1.00
	$f_1(x_1)$ is monotone	0.87	0.93	0.99	1.00	1.00
	$f_2(x_2)$ is quasi-convex	0.02	0.01	0.00	0.00	0.00
	$f_2(x_2)$ is monotone	0.97	0.98	0.98	0.99	1.00

This table shows the proportion of times each test rejects the null hypothesis. The OLS-implemented test $\beta_{X_k^2} > 0$ 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. The quasi-convex and monotone tests are the semi-parametric tests described in Section 1.4. Note that the null hypotheses are different depending on the test and thus they are not easily compared. 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) at the 5% level. For the semi-parametric test of quasi-convexity (monotonicity), the test is that the relationship between X_k and Y is quasi-convex (monotone). The data generating process is given in equation (1.8). Proportions are based on 1000 simulations.

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 suffers from the same problems as the vanilla quadratic specification.

1.3 Critical Bandwidth Tests of U-shapes

1.3.1 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. I refer to a U-shape as a regression function that is downward sloped to a (typically unknown) turning point and upward sloped thereafter. Let $S(X)$ be the support of X and

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

where we assume that m' exists[6]. For a specified set $A_0 \subset S(X)$, we test

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

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

versus

$$H_A \colon \forall a \in A_0,\, \exists x^{'} \in S\left(X\right) \text{ st}$$

$$m^{'}\left(x\right)\left(x-a\right) < 0$$

A non-parametric test in this context is a test that is consistent for any m of a large class of functions (e.g. $\mathbb{C}\left(S\left(X\right)\right)$).

1.3.2 A critical bandwidth approach

Bowman et al. (1998) extend the idea of the critical bandwidth of a density estimate developed by Silverman (1981) to non-parametric regression. A test of single-peakedness can be implemented similarly and is closer to Silverman (1981) than a test of monotonicity in the sense of looking for the smallest bandwidth such that a function has a single maximum. A difference of single-peakedness in regression functions is that a regression function is not restricted by laws of probability to eventually decrease to zero, as a density is.

The non-parametric regression technique that forms the basis of the test statistic in this section is local linear regression. For advantageous properties of this smoother, see Fan (1992), and Fan and Gijbels (1996). Also, note that boundary conditions are important in this context because if a bias causes a fit to curl down that could provide evidence against a U-shape.

The critical bandwidth test statistic is the smallest h such that a 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, with critical bandwidth is consistent with Silverman (1981) and Harezlak and Heckman (2001). Thus,

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

where H is the set of possible smoothing values, parameterized such that smaller values correspond to a less smooth fit; and $\hat{m}_h(x)$ is the estimate of the regression function, using the smoothing parameter h. For a 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.[7] We will see in Section 1.4 that the concept of a critical bandwidth can be generalized to a critical smoothing parameter, where h can be a smoothing parameter in many other models.

The intuition for using h_{stat} is that under the null 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 MSE. Restricting the smooth to be U-shaped is only a matter of smoothing out sampling noise. If the true regression function is U-shaped, as the sample increases, less smoothing is needed because the mean is measured more precisely.

A nice property of h_{stat} that can immediately be seen is that it does not depend on the researcher specifying a bandwidth (e.g. through cross-validation). However, as we will see in Section 1.3.4, evaluating the significance of h_{stat} with the bootstrap method does rely on a bandwidth choice in order to estimate the null distribution (although simulations suggest that the p-value is not very sensitive to the bandwidth selection method).

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 have not yet been explored by the literature and are examined in the next section.

[7] The set H of course depends on the parameterization used by the software implementation. The examples here are for implementations that define the smoothing parameter to be the percentage of data points used in the calculation of each point estimate.

1.3.3 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 in an interval $[a, b]$, say $r_1, \ldots, r_J$ where $r_k = a + (b - a)\frac{(k-1)}{J-1}$ and $r_k \in S(X)$. Note that $r_1 = a$ and $r_J = b$. Let

$$\gamma = r_k - r_{k-1} = \frac{b - a}{J - 1}, \tag{1.9}$$

which defines the longest violation of quasi-convexity that might not be caught. γ will play an important role under the alternative, as shown below in assumption 7 and Theorem 2. 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. Define also

$$\Delta m(r_k) = m(r_k) - m(r_{k-1})$$
$$\Delta \hat{m}_{h_n}(r_k) = \hat{m}_{h_n}(r_k) - \hat{m}_{h_n}(r_{k-1}).$$

The proof below shows consistency of h_{stat} as it would be implemented in software. For example, instead of using $m'(r_k) \neq 0$ in assumption (6) below, I use $\Delta m(r_k) \neq 0$; and instead of requiring pointwise convergence of $\hat{m}'_{h_n}(r_k)$ I require only pointwise convergence of $\hat{m}_{h_n}(r_k)$.[8] I make the following assumptions:

1. m(x) has a bounded second derivative.

[8]This allows for use of smoothers that do not estimate derivatives.

2. There exists a neighborhood around each r_k within which $f_X(\cdot)$ satisfies $|f(x) - f(y)| \leq c|x-y|^{\alpha}$, for $0 < \alpha < 1$.

3. The conditional variance $\sigma^2(x) = Var(Y|X=x)$ is bounded and continuous.

4. The kernel $K(\cdot)$ is a bounded and continuous density function satisfying

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

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

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

Note that (4) could be relaxed but is in any case satisfied by the normal kernel, which is used throughout in this paper. Assumption (5) implies that the following mean integrated squared error converges to 0:

$$E\int_{-\infty}^{\infty} (\hat{m}(x) - m(x))^2 w(x)\,dx$$

Intuitively, assumption (5) 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. This assumption and result are not needed if J stays fixed (or is bounded) as a function of n. In this case we could use a weaker assumption:

5 (alt). $f_X(r_j) > 0$ for all j

I discuss and relax assumption (6) in Section 1.3.5.

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

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

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

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 $\hat{m}_{h_n}\left(r_k\right) \to m\left(r_k\right)$ pointwise so for $1 < k \leq J, \exists N_2$ st for $n > N_2$,

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

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

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

Note that under the null, the event $\left\{sign\left(\Delta \hat{m}_{h_n}\left(r_k\right)\right) = sign\left(\Delta m\left(r_k\right)\right), \text{ for } 1 < k \leq J\right\}$ 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 true that $h_{stat}^{(n)}$ converges in probability to 0 faster than $n^{-\beta}$ for any $0 < \beta < 1$. □

For consistency under the alternative, I put some structure on the type of violations that the test implementation can practically detect by adding another assumption:

7. Under H_A, $\forall a \in A_0$, $\exists\,[c,d]$ with $(c-a)(d-a) > 0$ and $d - c > \gamma$ where for $x \in [c,d] \cap S(X), m'(x)(x-a) < 0$, where γ was defined in equation (1.9).

In other words, assumption (7) defines the width of violations of quasi-convexity we are looking for and disallows the situation where the only violations are small and over a potential turning point. If we want a strict test of quasi-convexity asymptotically, we can let γ tend to 0 (by letting J go to infinity) as the sample increases.

Theorem 2. *Under assumptions (1)—(7) and if $m(x)$ is not quasi-convex, h_{stat} does not converge in probability to zero.*

Proof. Suppose H_0, as defined in Section 1.3.1, is not true. Then, by assumption (7), $\forall a \in A_0$ there exists an interval $[c,d]$ with $d - c > \gamma$ such that $m'(x)(x-a) < 0$ for $x \in [c,d] \cap S(X)$. Then, $\exists j \leq J$ st $r_j \in [c,d]$ and $\Delta m(r_j)(x-a) < 0$. Take any bandwidth sequence, h_n, that satisfies

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

Because $\hat{m}_{h_n}(r_j) \to m(r_j)$, $\exists N$ such that for $n > N$,

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

and thus using assumption (6)

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

Fix such an N and thus an h_N. Note that $h_N > 0$. For any $h < h_N, h \neq h_{min}$ with probability 1. □

1.3.4 Bootstrapping the null distribution

In order to get a distribution for the test statistic h_{stat}, we would ideally take draws from $f_{XY}(x, y)$ of the same sample size, and calculate h_{stat} for each bootstrap sample. However, we cannot observe f_{XY}. Further, we also cannot assume that the null is true. If we bootstrap from a $\hat{f}_{XY}$ that does not yield a U-shaped regression function, we will be generating a distribution of h_{stat} for that particular function, which is not part of the null set. Thus, we restrict $\hat{m}(x)$ to yield a U-shaped regression function by using $\hat{m}_{hstat}(x)$, which gives us the fit that is the least biased of all fits while still being consistent with the null hypothesis. The intuition is consistent with Efron and Tibshirani (1994, p. 230).

In order to have 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 both under the null and under the alternative. This method is consistent with Bowman et al. (1998). This is the only place where 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 with local linear regression*

1. *find* h_{stat}
2. *estimate a fit using a cross-validated smooth 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* $h_{stat}\left(x, y^b\right)$

1.3.5 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) is violated). Even if (6) is true, there could be power issues in finite samples if there is a stretch on the x-axis with $\Delta m\left(r_k\right)$ small in absolute value. Just as Hartigan and Hartigan (1985) shows regarding Silverman (1981), Hall and Heckman (2000) shows that the Bowman et al. (1998) critical bandwidth test of monotonicity is not consistent and that there are cases where the power does not tend to one. In particular, they give an example of a regression function that is not monotone but that is not rejected asymptotically (we will define this function in Section 1.4.1 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 ($m'\left(x\right) = 0$ for $x \in [a, b]$), h_{stat} will be large because a large amount of smoothing is required before the smooth is flat[9]. Because a global bandwidth is used, a large bandwidth can smooth out non-monotone segments in other parts of the underlying regression function.

[9]Here, whether a smooth is flat or not depends on the decimal precision that is being used in the computation.

The flatness problem is relevant to all critical bandwidth tests and can occur when both of the following are true:

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

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

This problem is not just a theoretical problem and could show up in practical problems. Below, I give a few practical suggestions for addressing this issue.

1. Use an adaptive smoothing parameter.

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

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

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

In this paper I plan to explore 1. At first this seems incompatible with a critical bandwidth test because there is no single h_{stat} with an adaptive bandwidth. However, we define a new test statistic

$$\alpha = \min \{a \in R | \hat{m}(x; a \cdot h(x)) \text{ is quasi-convex}\},$$

where $\hat{m}(x; h(x))$ is a fit using an adaptive bandwidth. The idea is that instead of increasing the bandwidth uniformly, we can view a as the critical smoothing parameter, which scales the set of bandwidths. This allows each section on the x-axis to become smoother at different rates. 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 an MSE sense) tradeoff between bias and variance. In areas with a large density of x, a smaller bandwidth can be used to reduce bias without a significant sacrifice of variance of $\hat{m}(x)$. In areas where there is a lot of

curvature, a small bandwidth is used because oversmoothing can lead to large costs of biasedness.

In our case, we do not want the second property, because an example similar to the one given in Hall and Heckman (2000) could be constructed. For example, imagine a case where 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.

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

Note that in the case of testing for monotonicity, the problem could be improved by breaking the test into smaller pieces so that a flat part in one area will not affect the smooth 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. This approach is not explored in this paper.

1.3.6 Interpreting the non-parametric test results

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

1. Test for quasi-convexity

2. Test for monotonicity

If (1) rejects, then there is no need to proceed to (2) because the null in (2) is a subset of the null in (1). Both tests failing to reject is consistent with a monotone regression function and a U-shaped regression function. Finally, 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 and monotonicity.

One might think that the p-value of the test of quasi-convexity should always be lower than the p-value of the test for monotonicity. Often this is the case, but there are cases when this is not true, and understanding why can help with understanding 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 before it becomes U-shaped. A bootstrapped test statistic, h^b_{stat}, is more likely to be smaller for the test of quasi-convexity because $\hat{m}_{hstat_b}(x)$ 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. As for interpretation, if h_{stat} is the same for both tests, then this suggests that the true regression function is not U-shaped. As for which p-value should be used for inference about monotonicity, the p-value for monotonicity should be used.

1.4 Semi-parametric Testing of Multivariate Hypotheses

We now extend the discussion started in Section 1.2.2 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 where a non-parametric test

would not have enough precision to be useful. The idea behind the critical bandwidth test introduced in Section 1.3 can be generalized in an intuitive way. In addition to a bandwidth, smoothing parameters from other models can be used as test statistics. This approach is also used in Harezlak and Heckman (2001) and in fact they find that the method of smoothing does not affect inference.

Generalized Additive Models (GAMs) (Hastie and Tibshirani, 1986) provide a reasonable way to flexibly estimate a multivariate regression function. We model y as

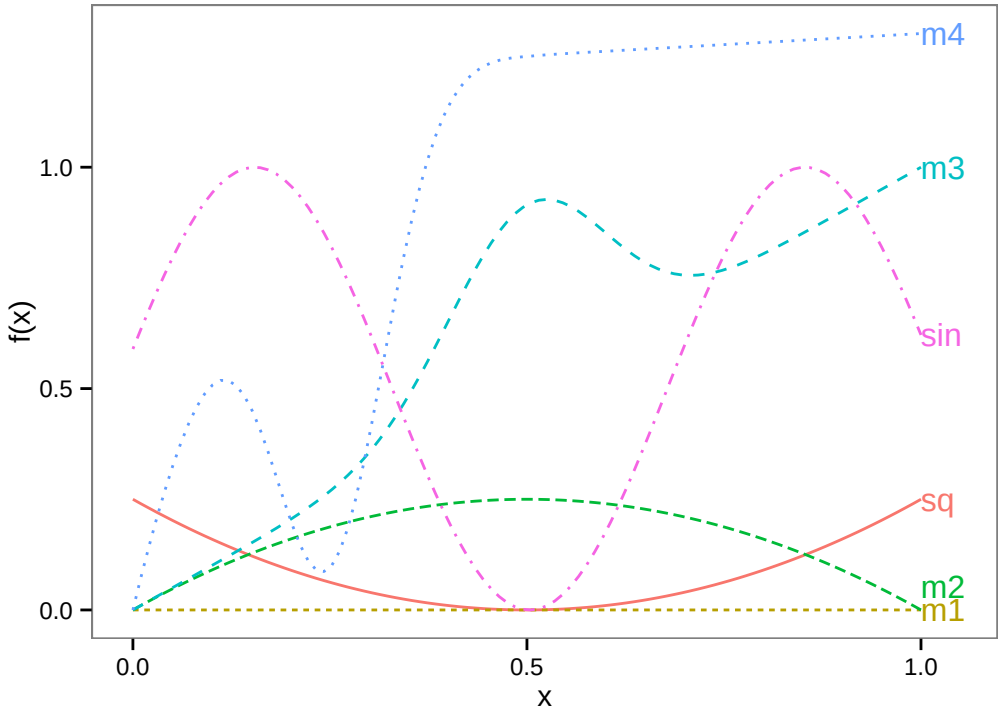
$$y = \sum f_k\left(X_k\right) + \epsilon$$

and estimate $\hat{f}_k$ without having to choose a functional form as in OLS. Testing for a U-shape with GAMs is a straightforward extension to the non-parametric test implemented in Section 1.3. Note that there is a smoothing parameter for each k. The algorithm for implementing a critical smoothing parameter test using GAMs is similar to the algorithm for critical bandwidth and is shown below:

Algorithm 2. *Critical bandwidth test with GAMs*

1. *find h_{stat}, scaling all RHS vars by same smooth parameters*
2. *fit a GAM with cross-validated smooth 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 $h_{stat}\left(x, y^b\right)$*

Figure 1.3: True Regression Functions



These functions are used in simulations throughout this paper. For their definitions, see Section 1.4.1.

Table 1.3: Semi-Parametric Test with Two Covariates

test	100	500	1000	5000	10000
$M(X_{sq}) + s(X_{m3})$	0.05	0.11	0.14	0.41	0.57
$M(X_{m3}) + s(X_{sq})$	0.03	0.02	0.05	0.14	0.48
$M(X_{sq}) + s(X_{sin})$	0.08	0.10	0.13	0.41	0.57
$M(X_{sin}) + s(X_{sq})$	0.14	0.38	0.56	0.99	1.00
$M(X_{m3}) + s(X_{sin})$	0.04	0.03	0.06	0.17	0.51
$M(X_{sin}) + s(X_{m3})$	0.17	0.39	0.64	0.98	1.00
$Q(X_{sq}) + s(X_{m3})$	0.04	0.01	0.00	0.00	0.01
$Q(X_{m3}) + s(X_{sq})$	0.04	0.01	0.06	0.11	0.47
$Q(X_{sq}) + s(X_{sin})$	0.04	0.01	0.00	0.01	0.01
$Q(X_{sin}) + s(X_{sq})$	0.04	0.17	0.65	1.00	1.00
$Q(X_{m3}) + s(X_{sin})$	0.03	0.03	0.05	0.17	0.48
$Q(X_{sin}) + s(X_{m3})$	0.04	0.27	0.73	1.00	1.00

This table shows the proportion of times each test rejects the null hypothesis. For the test name (column 1), "$M(Z)$" means the null hypothesis is regression monotonicity in Z; "$Q(Z)$" means a null of regression quasi-convexity; and $s(W)$ indicates that W is controlled for semi-parametrically (additively without specifying a functional form). All tests use GAMs. For the shapes of the regression functions used, see Figure 1.3. Proportions are based on 1000 simulations.

1.4.1 Simulations

We no longer consider OLS-based tests for a U-shape in these simulations because as shown in Tables 1.1 and 1.2 of Section 1.2.2, such tests are inconsistent under misspecification of functional forms. We start by looking at whether there is a U-shape in one variable, conditional on one other and will later consider the more general case of conditioning on multiple covariates. The true regression functions that we will

consider here are shown in Figure 1.3, and are defined as follows:

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

The m_n functions were explored by Ghosal et al. (2000), and some of them as well in other papers testing 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 a function used by Hall and Heckman (2000) to illustrate the flatness problem discussed in Section 1.3.5. Notice the presence of the two features contributing to the flatness problem: a dip around 0.25, and a relatively small derivative after 0.50.

Table 1.3 shows semi-parametric regressions of Y on X, conditioning on one other variable. Tests of quasi-convexity and monotonicity are both shown, as discussed in Section 1.3.6. The regressors are all uniformly distributed and the dependent variable is constructed additively, with a normally distributed error. For example, for the test

Table 1.4: Semi-Parametric Test with Five Covariates

test	100	1000	10000	100000
$M(X_{sq}) + ...$	0.05	0.16	0.67	1.00
$M(X_{m1}) + ...$	0.06	0.06	0.07	0.01
$M(X_{m2}) + ...$	0.10	0.18	0.57	1.00
$M(X_{m3}) + ...$	0.02	0.02	0.54	1.00
$M(X_{m4}) + ...$	0.07	0.16	0.40	0.74
$Q(X_{sq}) + ...$	0.03	0.00	0.01	0.01
$Q(X_{m1}) + ...$	0.06	0.06	0.05	0.03
$Q(X_{m2}) + ...$	0.15	0.27	0.82	1.00
$Q(X_{m3}) + ...$	0.03	0.03	0.53	1.00
$Q(X_{m4}) + ...$	0.07	0.17	0.39	0.74

This table shows the proportion of times each test rejects the null hypothesis. For the test name (column 1), "$M(Z)$" means the null hypothesis is regression monotonicity in Z, and "$Q(Z)$" means a null of regression quasi-convexity. Each test conditions on all the other variables. For example, test "$Q(X_{sq}) + \ldots$" tests for quasi-convexity in X_{sq}, conditioning on X_{m1}, X_{m2}, X_{m3}, and X_{m4}. All tests use GAMs. For the shapes of the regression functions used, see Figure 1.3. Proportions are based on 1000 simulations.

"$Q(X_{sq}) + s(X_{m3})$", the data are constructed as follows:

$$
\begin{aligned}
U_{sq} &\sim unif\,(0,1) \\
U_{m_3} &\sim unif\,(0,1) \\
\epsilon &\sim N(0,1) \\
X_{sq} &= sq\,(U_{sq}) \\
X_{m_3} &= m_3\,(U_{m_3}) \\
Y &= X_{sq} + X_{m_3} + \epsilon,
\end{aligned}
$$

with 0.25 correlation between the uniforms. Then, algorithm (2) is carried out to get a p-value for testing whether Y is quasi-convex in X_{sq}, conditional on X_{m_3}.

Table 1.4 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 is generated as $X_f = f(U_f)$, where $U_f \sim unif(0,1)$ with .25 correlation between them. Then, a test is performed on each of the X variables, conditioning on all the others. For example, for the row "$Q(X_{m2}) + \ldots$", 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}.

1.5 Conclusion

Economists have been exploring U-shaped relationships in the data since at least (Kuznets, 1955), but econometric tools focusing on testing U-shapes have not been developed. Using other tools to explore U-shapes, such as OLS, can lead to unwanted consequences, such as testing for "more quadratic than inverse-quadratic." This paper proposes to explore U-shapes in a more natural way by testing exactly what the definition is, not a restricted approximation of it.

Further, controlling for other variables in a non-parametric way ensures that the effect of those variables is not being picked up by the variable of interest only because of parametric restrictions. This paper contributes to the trend that non- and semi-parametric techniques are becoming more important in economics as 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 (see e.g. Varian (2014)).

In addition to using critical bandwidth tests for testing U-shapes, the contributions in this paper apply more generally to the regression "bump hunting" literature by proving consistency of the test statistic for critical bandwidth-type tests and by

specifying a lower-bound for the rate of convergence. Further, I proposed a solution to the well-known flatness problem through the use of adaptive bandwidth. Theoretical results were confirmed with simulations using a variety of data generating processes. This paper extends the tests to multivariate analysis through the use of GAMs, making them more attractive to use in practice.

References

Aghion, P., Bloom, N., Blundell, R., Griffith, R. and Howitt, P.: 2005, Competition and innovation: an inverted-U relationship, *The 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.

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

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, *The 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.

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* p. 61.

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, *The Annals of Statistics* pp. 70–84.

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

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

Kuznets, S.: 1955, Economic growth and income inequality, *The American economic review* pp. 1–28.

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.

Marmot, M., Shipley, M., Rose, G. and Thomas, B.: 1981, Alcohol and mortality: a U-shaped curve, *The Lancet* **317**(8220), 580–583.

Sasabuchi, S.: 1980, A test of a multivariate normal mean with composite hypotheses determined by linear inequalities, *Biometrika* **67**(2), 429–439.

Shaper, A. G., Wannamethee, G. and Walker, M.: 1988, Alcohol and mortality in British men: explaining the U-shaped curve, *The Lancet* **332**(8623), 1267–1273.

Silverman, B. W.: 1981, Using kernel density estimates to investigate multimodality, *Journal of the Royal Statistical Society* **43**, 97–99.

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

Appendix

1.A Denominator of OLS Limit in Section 1.2.2

Theorem 3. *If* $var\left(g_1\left(X_1\right)\right) > 0$ *and* $var\left(g_2\left(X_2\right)\right) > 0$*, then*

$$var\left(g_1\left(X_1\right)\right) var\left(g_2\left(X_2\right)\right) - cov\left(g_1\left(X_1\right), g_2\left(X_2\right)\right)^2 = 0$$

if and only if there is perfect correlation (correlation is 1 or -1) between $g_1\left(X_1\right)$ *and* $g_2\left(X_2\right)$

Proof. Without loss of generality, suppose that $cov\left(g_1\left(X_2\right), g_1\left(X_1\right)\right) \geq 0$. We start with the Pearson correlation inequality (easily proven with the Cauchy-Schwartz inequality):

$$\begin{aligned}
\frac{cov\left(g_2\left(X_2\right), g_1\left(X_1\right)\right)}{\sqrt{var\left(g_1\left(X_1\right)\right)}\sqrt{var\left(g_2\left(X_2\right)\right)}} &\leq 1 \\
cov\left(g_2\left(X_2\right), g_1\left(X_1\right)\right)^2 &\leq var\left(g_1\left(X_1\right)\right) var\left(g_2\left(X_2\right)\right) \\
0 &\leq var\left(g_1\left(X_1\right)\right) var\left(g_2\left(X_2\right)\right) - cov\left(g_1\left(X_2\right), g_1\left(X_1\right)\right)^2
\end{aligned}$$

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. □

Chapter 2

Robustness of Subjective Well-Being U-Shapes in Age

2.1 Introduction[1]

The attention given to subjective well-being (SWB) as an outcome variable of interest in the field of economics has been increasing (see Kahneman and Krueger, 2006, or for a more extreme position, Layard, 2011). This paper focuses on the effect of age on SWB. Although age is not a choice variable for any individual, understanding the SWB dynamics across ages could be important when making comparisons of SWB across time or groups. Further, policy decisions might affect the *way* in which age affects SWB (Landeghem, 2012). Understanding age effects also gives us a better foundation for understanding issues regarding period and cohort effects, such as trends of subjective well-being over time.

[1] A previous version of this chapter was presented at the Simposio de la Asociación Española de Economía.

Several authors have found a U-shaped relationship between SWB and age. Blanchflower and Oswald (2008), for example, find that such a relationship holds in 72 countries. This finding has even been claimed to hold in apes (Weiss et al., 2012), suggesting perhaps that there are biological foundations for the U-shaped relationship. 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 Europe and developing countries. The U-shape has been found to be robust to including individual fixed effects, in British panel data (Clark, 2007).

This paper aims to understand whether there is a U-shaped relationship, and where and why it occurs. Solving the puzzle of why SWB would increase in later ages, despite worsening health and declining wage-earning ability, in some sense gets at the basic question of "What makes us happy?" Further, if a U-shaped relationship holds, the next question, which would have policy implications, is whether we could find some way to learn from what makes people of older ages happy and apply that knowledge to increase SWB in middle ages. Alternatively, it might be that people in middle-ages willingly sacrifice SWB in the short-run either for increased SWB in later ages or for something else, such as increased SWB of children (which might not necessarily show up in all measures of SWB of the parents).

This paper uses repeated cross-sections from the World Values Survey and the European Values Survey and finds that in places where the life satisfaction U-shape does exist, financial satisfaction is the main driver. When conditioning on financial satisfaction, in places where there is a U-shape between life satisfaction and age, the U-shape becomes monotonically decreasing in many countries and in pooled regressions. In fact, it appears that there is a U-shaped relationship between financial satisfaction and age, and that the positive correlation of financial satisfaction and life satisfaction is the cause of the U-shape of life satisfaction and age. These results are

evidence against theories that the U-shape of life satisfaction is found because of a hard-coded midlife crisis in our DNA.

The results discussed above are consistent across a variety of statistical tools, including parametric tools such as OLS and ordered logits. Semi-parametric tools such as generalized additive models are also used, which allow for a flexible way of conditioning on variables. These flexible models provide graphical results, as well as forming the basis of a semi-parametric critical bandwidth test of a U-shape, which has not yet been used to look at the U-shape of SWB. I confirm the findings of Ferrer-i Carbonell and Frijters (2004) that making assumptions about cardinality of SWB outcome variables does not qualitatively change results.

In addition to using several statistical tools, many regression specifications are considered. Unconditional results are reported, as well as regressions conditional on control variables that are commonly used in the SWB literature. Quadratic specifications are considered, both unweighted and weighted by age in order to get a global estimate of the regression function that is not sensitive to the marginal distribution of age. Grouped dummy specifications are also used. Different ways of addressing age-period-cohort collinearity are examined. Tests are run at the country level, country category level (western countries, developing countries, and Eastern Europe), and world-wide.

The rest of this paper is organized as follows. In Section 2.2 I discuss model specification and identification issues that are important for SWB topics such as which variables should be controlled for in SWB regressions on age, and the well-known problem of identifying all effects in an additive age, period, cohort model. As a starting point for data analysis, in Section 2.3 a well-known study (Blanchflower and Oswald, 2008) is replicated with additional data and with several ways of controlling for period and cohort effects. In Section 2.4 semi-parametric analysis is used to focus on how controlling for financial satisfaction changes the relationship between life

satisfaction and age. Both within-country results (Section 2.4.2) and cross-country results (Section 2.4.4) provide the same conclusion, which is that much of the life satisfaction U-shape can be explained by financial satisfaction increasing later in life. Finally, Section 3.6 concludes.

2.2 SWB Model Specification and Identification Issues

2.2.1 The identification problem of age, period, and cohort effects

When analyzing age effects, it often makes sense to consider period and cohort effects. This section introduces a framework for modeling age, period, and cohort effects and shows how there are problems for identifying all 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, that 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 $t_{n+m} = t_n + m$, that is, if the observation years are consecutive.[2] 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 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 the data in this paper because the observation years are not consecutive.

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

Thus, I refer to the cohort born in year b as cohort b. For individual i in cohort b, of age a in 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 $t = c + a$. Plugging in, this gives us $y_{i,b,a,b+a}$. We drop the unnecessary last subscript and model $y_{i,b,a}$ as

$$y_{i,b,a} = \alpha_b + \beta_a + \gamma_{b+a} + \epsilon_{i,b,a}, \tag{2.1}$$

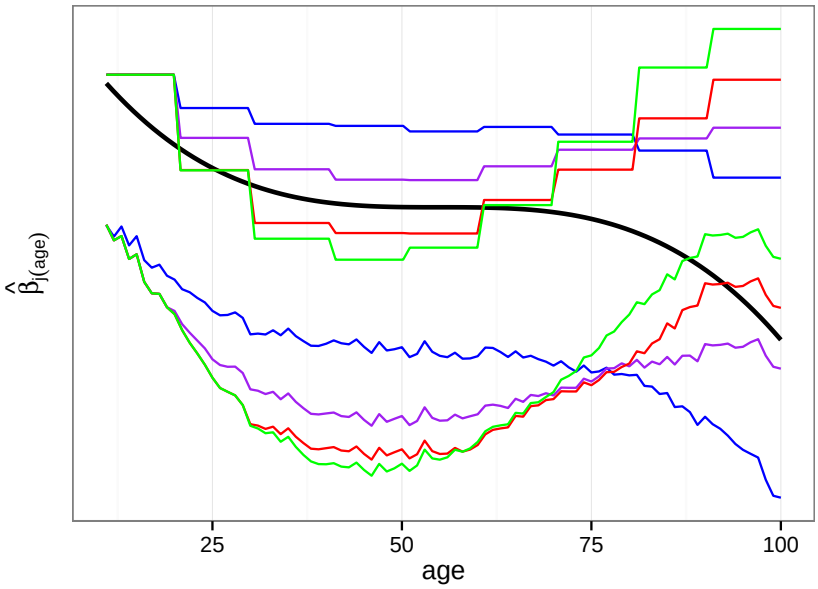
where α_b is the effect of being in the cohort with birth year b, β_a is the effect of having age a and γ_{b+a} is the effect of being observed in year $b + a$. This additive structure is not enough to get identification of any of the effects. To see this, consider that we could add a time trend to the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

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

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 (2.1) we can identify second differences of the effect profiles. I extend this strategy in Appendix 2.A to allow for data from non-consecutive years.

In a recent paper, Landeghem (2012) uses the identification framework of McKenzie (2006) to test the convexity of well-being in age. He uses panel data from the

Figure 2.1: Cohort Effects, No Period Effects



This figure plots single and grouped age dummy estimates when controlling for cohort effects using different sized cohort dummy groups. $\hat{\beta}_{j(\text{age})}$ is the coefficient for the interval that age belongs to (or the coefficient for age itself in the case of single dummies for age). Blue: 1-year cohort dummies; purple: 10-year cohort dummies; red: 20-year cohort dummies; green: no cohort dummies. Note that as the dummy groups of the cohorts get larger, the estimated age effects converge towards the green line, which does not control for cohort effects at all.

German Socio-Economic Panel in which individuals were observed in consecutive years from 1984 to 2007. He 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. I mostly follow the setup of McKenzie (2006) except that I do not require that we observe the cohorts every year.

Simulations

The simulations in this section provide a demonstration of what goes wrong when attempting to identify the age effects by controlling for age, period, and cohort effects in a certain way. I give intuition, backed up by situations, to suggest that the most conservative way to get an estimate of the age effects is to distribute the most degrees

of freedom to controlling for the period and cohort effects. That is, the age effect dummies should be the most coarsely grouped.

I keep the sample size and distribution of the error the same throughout all simulations so that properties of the estimators such as precision can be compared across different simulations. The birth years are uniformly distributed from 1890 to 1970. Observation years are uniformly distributed from 1980 to 1990. There are no deaths. The age effect function is

$$\beta\left(\text{age}\right) = \frac{-\left(\text{age} - 55\right)^{3}}{100}.$$

$\beta\left(\text{age}\right)$ is our parameter of interest. It is always the black line in the simulation graphs. It might appear stretched or condensed according to the range of the estimates, but the true age effect function is constant throughout. The cohort effect function is

$$\alpha\left(\text{birth year}\right) = \left(\text{birth year} - 1940\right)^{2}$$

and the period effect function is

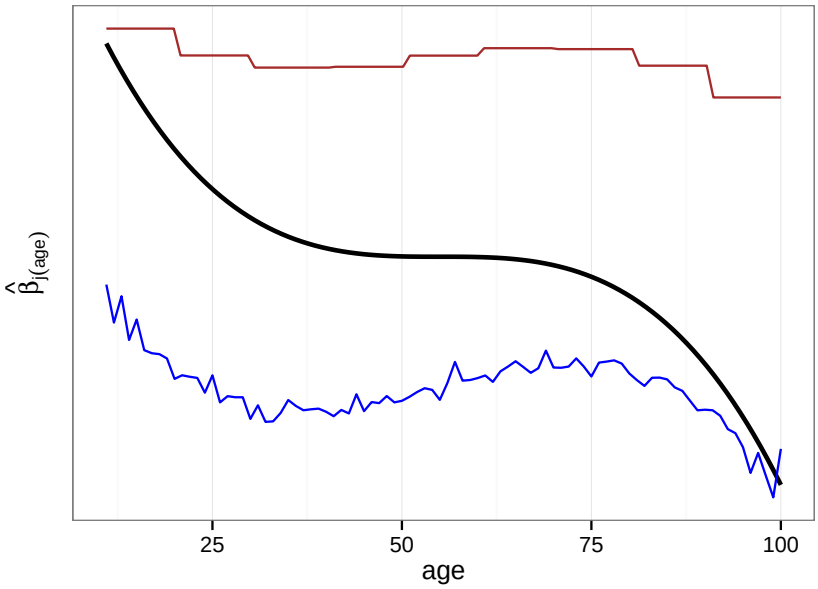
$$\gamma\left(\text{age} + \text{birth year}\right) = \gamma\left(\text{year}\right) = \left(x - 1980\right)^{2}.$$

The cohort effect function and period effect function are only present in some of the simulations. Unless stated, $\alpha\left(\text{birth year}\right) = \gamma\left(\text{year}\right) = 0$. Finally, a noise vector is generated as $\epsilon \sim N(0, 1000^{2})$. The vector of life satisfaction outcomes is then

$$LS = \alpha\left(\text{birth year}\right) + \beta\left(\text{age}\right) + \gamma\left(\text{year}\right) + \epsilon.$$

All the simulations have a sample size of 100,000.

Figure 2.2: Controlling for Cohort Effects but Not for Period Effects



This figure shows estimates of age effects controlling for cohort effects but not for period effects when all three effects are present. $\hat{\beta}_{j(\text{age})}$ is the coefficient for the interval that age belongs to (or the coefficient for age itself in the case of single dummies for age). Even though single age dummies and single birth-year dummies are included, not controlling for period effects leads to a spurious U-shape.

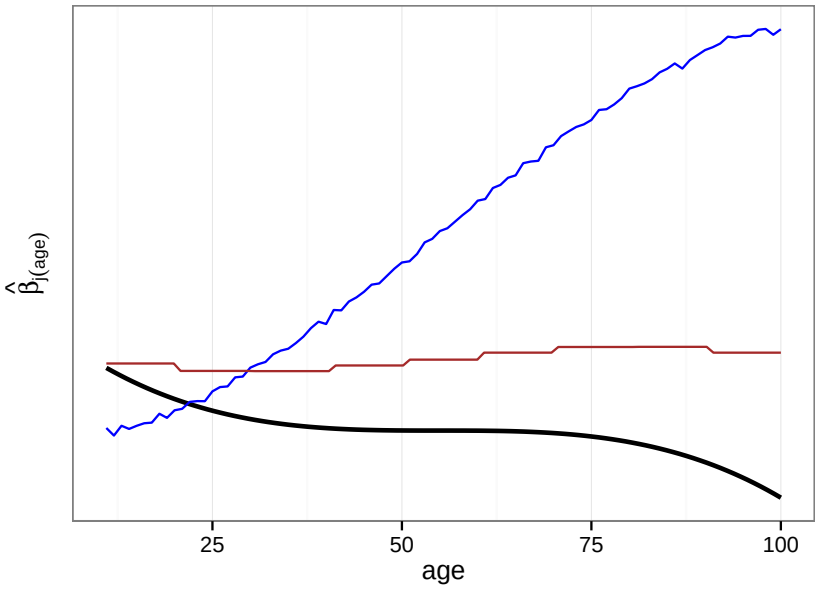
Simulation results

Figure 2.1 shows a simulation with age and cohort effects (but no period effects). The different lines are estimates of age effects controlling for cohort effects using differently sized intervals for birth year. As the interval size increases, the estimates of age effects converges towards the estimates that do not control for cohort effects at all. This figure shows how incorrectly controlling for cohort effects can lead to a spurious U-shape finding. In this case, if the researcher is limited because of observations, he should use as many degrees of freedom as possible to reduce the size of the cohort dummy intervals. Leaving the age dummies grouped does not lead to bias. We lose finer details, but the general shape is a good approximation of the true age effect profile. It seems that large cohort group dummies commonly bias age effect estimates in misleading directions whereas large age group dummies have the less damaging property of just squishing the age effect estimates compared to the true age effect profile.

When there are age, period, and cohort effects in the simulation, the estimates of the age effects are worse when trying to control for all effects (excluding dummies as necessary to avoid multicollinearity) in Figure 2.3 than when not including any period dummies, as shown in 2.2. Thus, it appears that in some cases it is better to group dummies than try to force a saturated specification.

The McKenzie (2006) estimator (derived in Appendix 2.A) correctly and precisely estimates the true second differences of the age effects in all the simulations. If we believe there to be strong age, cohort, and period effects, as in Figure 2.3, and we are not sure which effects we can ignore or group, then estimating second differences is a robust strategy that should be considered.

Figure 2.3: Trying to Control for All Three



In this figure, as many period dummies as possible are added (that is, avoiding multicollinearity) in addition to age and cohort dummies. $\hat{\beta}_{j(\text{age})}$ is the coefficient for the interval that age belongs to (or the coefficient for age itself in the case of single dummies for age).

Proxies for period effects

Instead of dealing with the collinearity problem outlined above, if we had an idea of which variables are driving the period effects, we could put those variables directly into the specification and not have to rely on catch-all dummies (Heckman and Robb, 1985). For example, if we think that the period effect is being driven by the inflation rate or whether the country is at war, including the underlying variables directly would capture most of the period effect. I do not explore this strategy further in this paper.

2.2.2 What if the happiest live longer?

If the happiest live longer, then there could be a spurious finding of a U-shape even if the true age profile is decreasing, because the decrease in SWB due to age might be dominated by the increase in the individual happiness effect from selection. In other

words, we might not be able to generalize a finding of a U-shape across individuals to a U-shape within individuals. Controlling for variables correlated with the individual happiness effect alleviates some of the concern regarding this possibility. In order to see if the theory could potentially explain the magnitude of the increase of SWB in old age, data could be used from the Human Mortality Database (University of California, Berkeley (USA) and Max Planck Institute for Demographic Research (Germany), 2012).

A separate identification issue is that even if there were biological reasons driving a U-shaped relationship between SWB and age, the relationship might be closer between SWB and *biological* age, which is a measure that takes into account that people with the same chronological age might have different physiological deterioration across organ systems. See, for example, Belsky et al. (2015). Differences in biological and chronological age could potentially explain why the U-shape appears in some countries and not in others. Macro level health variables might be able to explain these differences. For example, we might try transforming the x-axis so that instead of being age, it is some normalization of age with respect to life expectancy.

2.2.3 Is the meaning of "satisfaction" universal?

In addition to econometric identification, there is a higher-level identification issue at the heart of the subject of most SWB studies. If the SWB questions are interpreted in a fundamentally different way by different age groups, then it is not clear if comparing SWB measures between those groups gives us something of use.[3] We would not know if it is the underlying SWB parameter that we have in mind which is causing observed variation or if it is the personal interpretation of SWB which is changing as one ages.

[3] A 20-year-old should not be expected to assess her SWB in exactly the same way as a 70-year-old. What is important is that the variables that do affect SWB affect it in the same way (or at least to the same first-order approximation) for both.

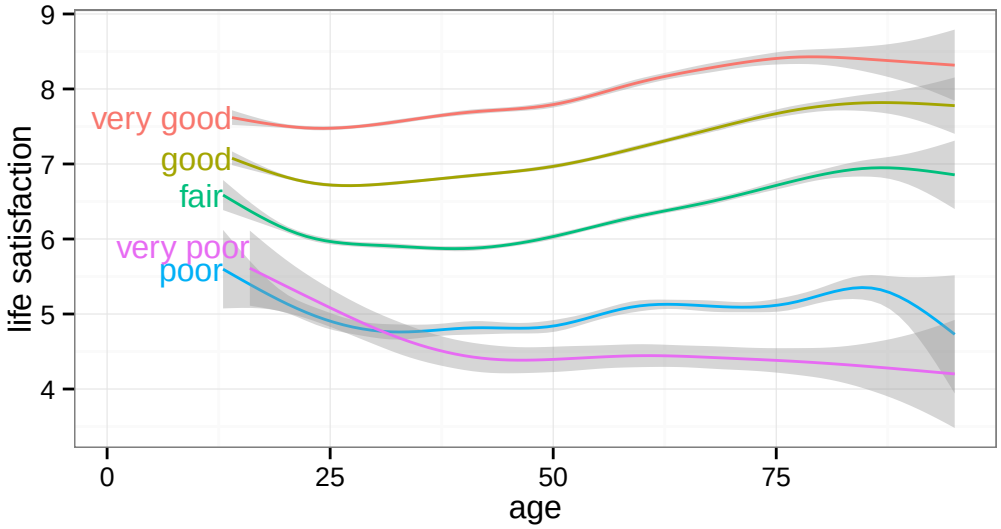
Similar to the problem above is the question of which groups of people are comparable. That is, which groups can be pooled together. Considerations for the level of grouping in the SWB context can be placed into two categories: groupings should contain (1) people who interpret and report SWB questions similarly, and (2) people for which age affects SWB through similar mechanisms. A concern falling under the first category is whether the term "life satisfaction" in French means the same thing as it does in German. If it does not, then pooling French-speaking Swiss and German-speaking Swiss might not make sense. The second category contains concerns such as whether age affects the SWB of women in the same way as it affects men. A common way to address this concern is to run separate regressions for men and women, show that the coefficients of interest are not statistically different, and then pool them together based on this justification.

Most papers dealing with SWB run regressions at the country level. Blanchflower and Oswald (2008) run some regressions that pool all people in the sample, some at the country level, and some using categories that they define for countries (Western, Eastern Europe, and Developing Countries). Sometimes they run separate regressions for males and females and sometimes, when the sample sizes are small, they pool them together. I will follow their setup, but note that it is not clear without a theoretical framework which groups should be combined.

2.2.4 Variable selection

It is not clear which conditional relationship between SWB and age should be most interested in. It is generally agreed that a more interesting relationship excludes health. In Appendix 2.E, I discuss how it might make sense to control for the variable in theory, but that in practice we often do not have the particular health variable that is really of interest. Further even if we did have the variable, excluding it still allows for (conservative) inference on that conditional relationship.

Figure 2.4: Life Satisfaction by Health

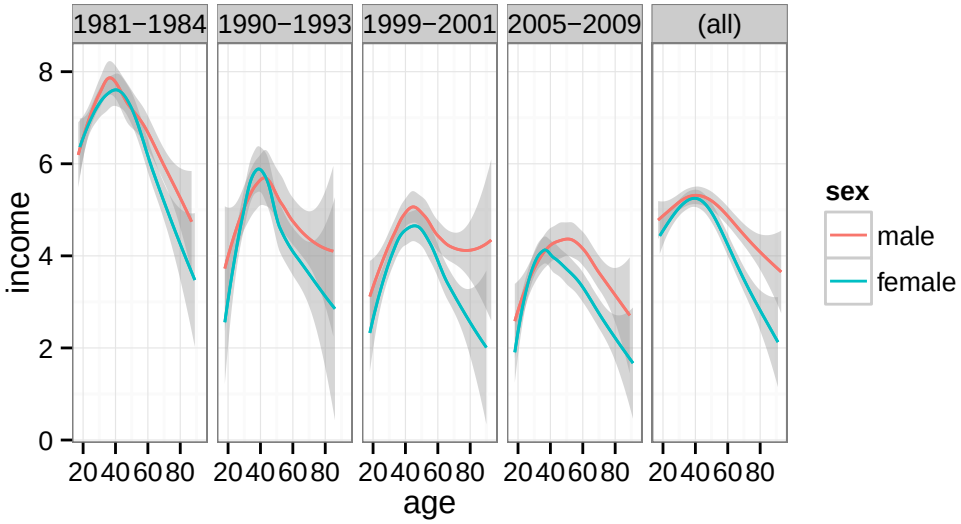


The question (a009) and response codings for self-reported health were "All in all, how would you describe your state of health these days? Would you say it is... 1 'Very good' 2 'Good' 3 'Fair' 4 'Poor' 5 'Very poor'." See Appendix 2.F for the life satisfaction question.

Income is controlled for by most authors who have studied the U-shape. The income variable available in the data that is described in the next section is presumably household income. However, it is not clear that the income data reflects that intent. The respondent might not have understood that she was being asked for household income or might have replied with inaccurate household income. For example, a 20-year-old and a 50-year-old in the same household might give (systematically) different answers. Figure 2.5 shows LOWESS curves of income against age in France for each survey wave. These data could be compared to tax data for individual and household income in order to see if the survey responses can be interpreted as a weighted average of the two. In the figure, the distributions seem pretty consistent except for the first wave.

The income variable could be viewed as problematic whether it represents individual income, household income, or what is most likely, some noisy weighted average of

Figure 2.5: Income vs. Age in France



The income question (x047) and response codings were "Here is a scale of incomes. We would like to know in what group your household is, counting all wages, salaries, pensions and other incomes that come in. Just give the letter of the group your household falls into, before taxes and other deductions. "Lower step", "Second step", ..., "Tenth step"."

the two. Similar to the health variable discussion, if this variable represents individual income, then comparing a 16-year-old to a 40-year-old, holding income constant, could produce strange estimates. If the variable represents household income, then is that really what we would ideally control for when analyzing personal life satisfaction?

2.3 Parametric Data Analysis

2.3.1 Description of data

I use data from the World Values Survey (WVS) manually merged[4] 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,

[4]The agreement between WVS and EVS which lead to the creation of a merged dataset, covering the period 1981-2004, was not renewed. See `http://www.wvsevsdb.com/wvs/WVSFaqs.jsp`.

2010–2014 with the EVS waves 1981–1983, 1989–1993, 1999–2001, 2008–2010. 98 countries are used and the median sample size for a country (pooling across survey waves) is 2431. For comparison purposes, I sometimes restrict the dataset in ways consistent with Blanchflower and Oswald (2008).[5] For descriptive graphs such as a bivariate non-parametric regression, I use the whole dataset. For problems with using the WVS, see Deaton (2008). Also, see Figure 2.5 for evidence that the populations represented by different sampling waves are different: the survey wave carried out in 1981–1984 has, for each age, a mean income that is 2 income deciles larger than the 1989–1993 wave and 3 income deciles larger than the 1999–2004 wave. Further, there is surely selection bias on at least two levels—selection into the WVS-EVS and selection by regressions based on missing values. The regressions in Blanchflower and Oswald (2008) and my replication in this paper only use about 67% of the WVS-EVS observations.

2.3.2 Replication of Blanchflower and Oswald (2008)

In this section I attempt to replicate the results of Blanchflower and Oswald (2008), which used the WVS-EVS in their paper. I use the same specification but use the updated WVS-EVS data set. The specification in Blanchflower and Oswald (2008) for the WVS-EVS is an ordered logit with life satisfaction as the response variable and age and age squared as the regressors of interest. In the regression they include dummies for income deciles, marital status, education, employment, and three year-dummies.

Blanchflower and Oswald (2008) present one table based on the WVS-EVS. They show six regressions, which result from breaking their three country categories down by sex. For a list of which countries belong to each category and for the original WVS-EVS results of Blanchflower and Oswald (2008), see Table 2.4 in Appendix

[5]Thus, I drop Azerbaijan, Armenia, El Salvador, Georgia, Israel, South Korea, New Zealand, and Pakistan, and I restrict the dataset to only observations containing data on all the variables they use.

Table 2.1: Argmins of SWB-Age Relationship, by Country Category and Sex

country category	sex	OLS	oLogit	B&O	N
Developing	M	41.6	41.2	42.6	89352
		(0.9)	(0.9)		
Developing	F	42.4	42.2	44.3	94111
		(0.9)	(0.8)		
Eastern European	M	50.6	50.5	46.5	22582
		(1.0)	(1.0)		
Eastern European	F	54.3	54.6	48.2	25247
		(1.5)	(1.5)		
Western	M	46.0	45.6	45.2	48375
		(0.7)	(0.6)		
Western	F	45.5	45.2	47	51679
		(0.8)	(0.8)		

Column "oLogit" (ordered logit, quadratic model) shows my parametric estimates using the same specification as Blanchflower and Oswald (2008), whose table of WVS-EVS results is reproduced in 2.4 in Appendix 2.B. See that table for details. Column "OLS" (ordinary least squares, quadratic model) uses the same regressors, but converts life satisfaction to a numeric variable. Standard errors based on the delta method are in parentheses. Adding a cubed term is discussed in Appendix 2.D.

2.B. My attempt to replicate these estimates of the minima is in column "oLogit" of Table 2.1. When restricted (not shown) to the survey waves that were used by Blanchflower and Oswald the sample sizes of each regression are the same, which suggests that the variables are indeed the same (most variables in the dataset have differing amounts of missing observations). The estimates of the minima with the updated data are similar to the estimates of Blanchflower and Oswald. In some sense, this consistency in the estimates can serve as out-of-sample evidence that the regression results from the specification used by Blanchflower and Oswald are stable over newer survey waves. I present standard errors based on the delta method below each estimate, in parentheses.

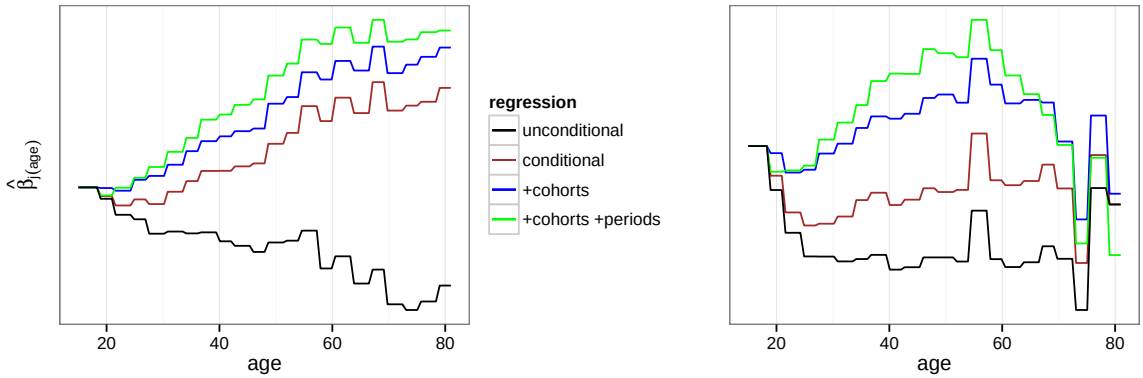
Column "OLS" (OLS, quadratic model) of Table 2.1 uses the same regressors as the "oLogit" column, but converts life satisfaction to a numeric variable. Consistent with other papers (e.g. Ferrer-i Carbonell and Frijters 2004), the differences in the

inference resulting from the OLS estimates and OL estimates are small. Adding a cubed term is discussed in Appendix 2.D.

As discussed in Section 2.2.1, there is no way to estimate the age effects while controlling for arbitrary period and cohort effects. Further, the simulations give examples of how it is possible to arrive at wrong conclusions using grouped dummies to attempt to control for period and cohort effects. Nonetheless, because such approaches are consistent with the literature, results from several ways of controlling for cohort and period effects are compared here. I present estimates of coefficients in an ordered logit regression of life satisfaction on the same controls as Blanchflower and Oswald (2008), except that I use period dummies instead of survey wave dummies. In these plots, it does not seem to make a large difference in most cases (exceptions are noted) whether one controls for cohort or cohort and period effects. The only discordant line in each plot is the unconditional regression, which regresses life satisfaction on age dummies and nothing else.

In Figure 2.6, the unconditional female regression estimates hint at a U-shape from age 20 to age 55, at which it decreases. For males, there also seems to be a U-shape for the unconditional and conditional specifications, before 60. The spike at age 55 is present for both females and males. Going from the red line (conditional), where one could claim that a U-shape exists, to the blue line (+cohorts) and green line (+cohorts +periods) where one could claim an inverted U-shape exists, shows how fragile conclusions based on these specifications can be.

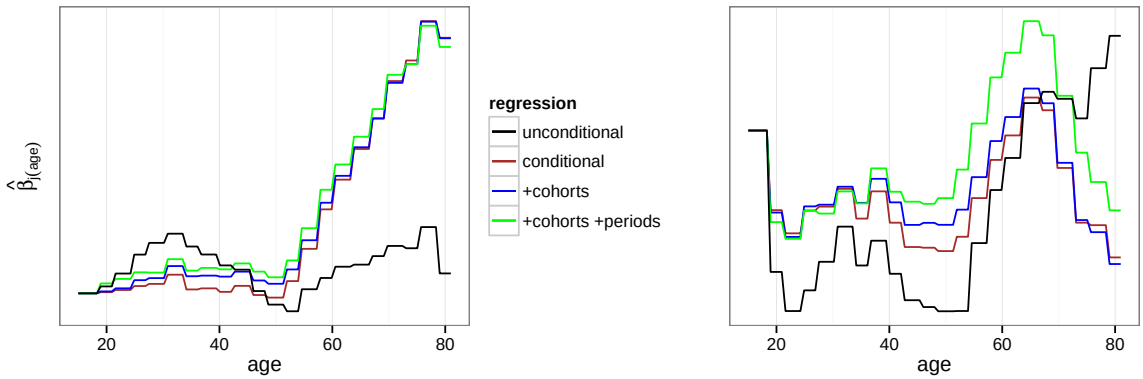
Figure 2.7 shows similar plots for the Western category. Neither the male nor female age profile suggests a strict U-shape. For females, the life satisfaction at later ages seems to strongly dominate younger and middle ages for all of the conditional specifications. For males, by age seventy life satisfaction decreases, where for females it keeps increasing. It might be that health effects take over early for males than for females.



(a) female (b) male

Figure 2.6: Developing Countries

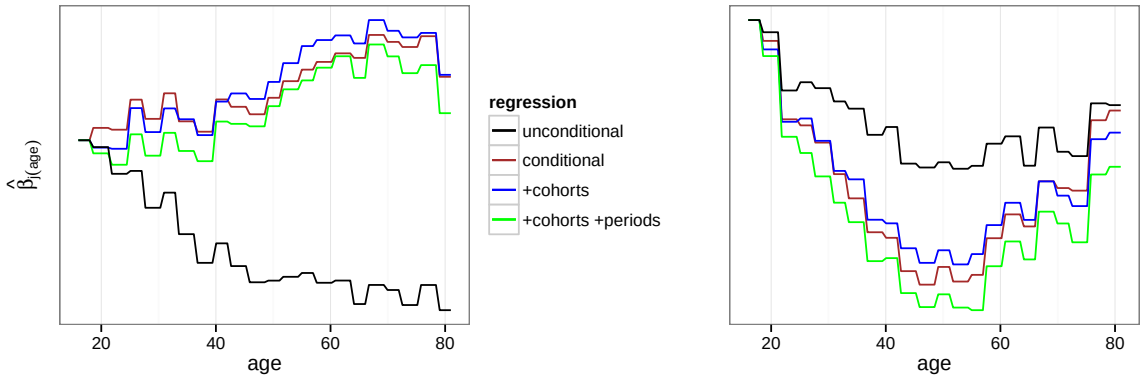
$\hat{\beta}_{j(\text{age})}$ is the ordered logit coefficient for the interval that age belongs to. The unconditional regression is of life satisfaction on age. The conditional regression uses the same controls as Blanchflower and Oswald (2008) (except for cohort and wave dummies). "+cohorts" extends the "conditional" regression by adding decade of birth dummies and "+cohorts +periods" additionally controls for single periods. No y-axis labels are drawn because coefficients in an ordered logit cannot be compared across regressions.



(a) female (b) male

Figure 2.7: Western Countries

$\hat{\beta}_{j(\text{age})}$ is the ordered logit coefficient for the interval that age belongs to. The unconditional regression is of life satisfaction on age. The conditional regression uses the same controls as Blanchflower and Oswald (2008) (except for cohort and wave dummies). "+cohorts" extends the "conditional" regression by adding decade of birth dummies and "+cohorts +periods" additionally controls for single periods. No y-axis labels are drawn because coefficients in an ordered logit cannot be compared across regressions.



(a) female (b) male

Figure 2.8: Eastern European

$\hat{\beta}_{j(\text{age})}$ is the ordered logit coefficient for the interval that age belongs to. The unconditional regression is of life satisfaction on age. The conditional regression uses the same controls as Blanchflower and Oswald (2008) (except for cohort and wave dummies). "+cohorts" extends the "conditional" regression by adding decade of birth dummies and "+cohorts +periods" additionally controls for single periods. No y-axis labels are drawn because coefficients in an ordered logit cannot be compared across regressions.

Figure 2.8 shows the results for Eastern European countries. The age effect profile for Eastern European men is the closest to a U-shape of the figures, with the trough being around 55. This sharply contrasts with the spike at 55 in developing countries. For females, controlling for the additional variables changes the estimated profile from monotonically decreasing to monotonically increasing.

2.3.3 Tests of a U-shape

Quadratic functions have two properties in particular that are restrictive in an SWB context: a quadratic functions have at most one inflection point and quadratic functions are symmetric. By allowing for a better fit we would reduce the problems that occur from misspecification and we could increase power. Imposing a quadratic specification is a strong restriction in general when making inference based on OLS, but it can be especially misleading when estimating the shape across the entire support of the X-variable, which is what we would like to do in order to understand how age affects SWB across the whole life cycle. I first show this disadvantage

of parametric specification and then suggest weighting as an improvement. The majority of this paper will be focused on using non-parametric solutions, which are more flexible.

There are several ways of testing for a U-shape, but the most common method is to use a quadratic specification, either implemented through OLS or an ordered logit. There are variations on the simple quadratic test. For example, Lind and Mehlum (2010) provide a refinement for which simulations suggest provides an improvement over the simple quadratic specification. Further the test has the nice property of defining a U-shape in an intuitive way. Appendix 2.C of this paper explores some problems of defining a U-shape and of constructing a test for one. I present a cubic specification in Appendix 2.D which is more flexible and allows for better null and alternative sets of the test. Further, the cubic specification is more in line with theory, which suggests that at some point poor health drives life satisfaction downward and thus calls for another inflection point (which may or may not be in the range of ages). However, I only consider the cubic specification as an improvement to the quadratic specification because the functional form is still restrictive in approximating the true relationship. If sample size allows and a rigorous test is not required, using age dummies is the most flexible estimation strategy.

Non-parametric methods

As mentioned at the start of this section, there are advantages to a fully parametric approach to exploring for a U-shape. However, non-parametric methods allow for a less assuming exploration of a U-shape. This flexibility comes with the cost that it is more difficult to implement a rigorous statistical test with non-parametric methods and more difficult to interpret the results.

If a positive quadratic fits the data better than a negative quadratic, and the minimum of the fit falls in the accepted range, the quadratic specification could

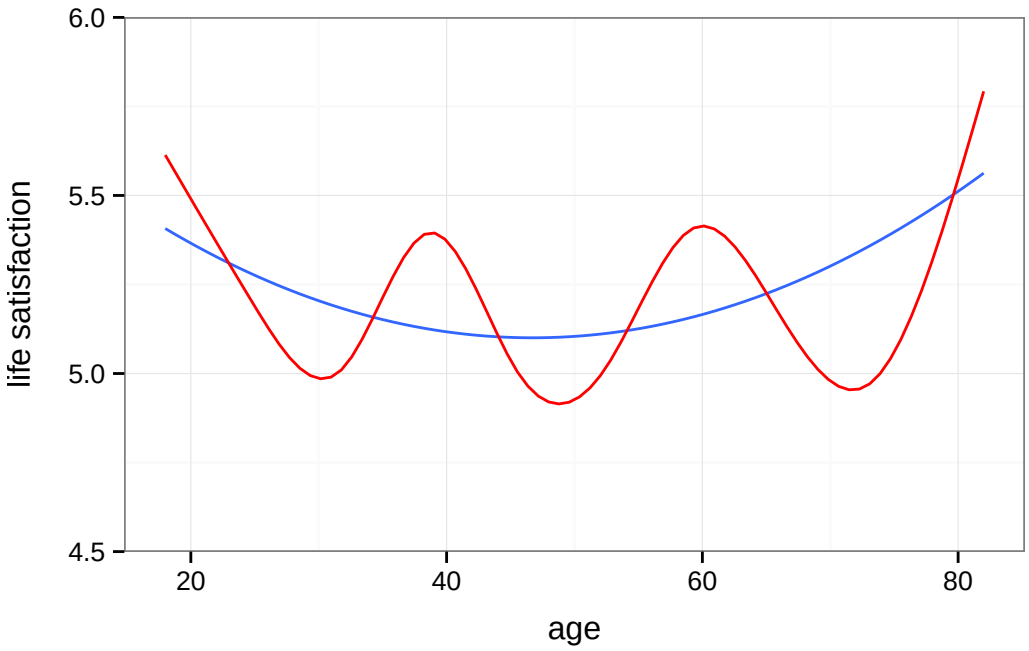


Figure 2.9: Life Satisfaction vs. Age in Iraq

This figure shows the unconditional relationship between life satisfaction and age in Iraq. The blue line shows an OLS quadratic fit, which would suggest a U-shape. The red line is a non-parametric fit using splines and generalized cross-validation to choose the smoothing parameter.

suggest a U-shape even when the function it is approximating would not commonly be described as U-shaped. An example of this in practice is shown in Figure 2.9. There is no theory that predicts a sinusoidal-like effect of age on life satisfaction, but having a test that rejects that it is U-shaped helps in realizing that there are likely confounding effects (e.g. period and cohort effects).

The standard way to estimate effects in a flexible way is to use dummies in an OLS or other linear specification. This approach is commonly referred to as "non-parametric." Adding in a full set of age, period, and cohort dummies—removing one from each set as normal—still leads to perfect multicollinearity. We must remove one additional dummy. The dummies are commonly grouped in order to solve the extra multicollinearity problem and to deal with small sample sizes. Grouping dummies is in essence a restriction that usually incorrect and not desired. For example, if we use OLS to regress y on a dummy which groups categories 2 and 3, age_{2-3}, instead of on

Table 2.2: 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	2	3	1	40	12	1	24
FS	5	1	5	0	39	11	0	37
inc	30	4	6	0	24	12	2	20
inc FS	8	3	7	0	34	11	0	35
inc ed emp marital	38	1	1	0	19	8	2	29
inc ed emp marital FS	18	2	3	0	36	4	1	34

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 sum 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. The 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 the text for descriptions of the columns. The level of significance used was 0.1.

both age_2 and age_3, then it is similar to running restricted least squares, requiring that the coefficients on age_3 and age_2 are equal. This restriction has secondary effects in that the variation caused by the individual components might be attributed to coefficients of covariates of those variables.

2.4 Semi-parametric Analysis of the Data

2.4.1 Applying a semi-parametric test of a U-shape

The non-parametric methods used in this section ignore the ordinal nature of the outcome variable, which is reasonable because OLS results were comparable to ordered logit results. The findings of Ferrer-i Carbonell and Frijters (2004) suggest that SWB results are not sensitive to using methods that do take into account the ordinal nature of the variable.

2.4.2 Within-country results

Table 2.2 shows columns summarizing the results of using the semi-parametric test of a U-shape of life satisfaction in age for various specifications, using the WVS-EVS

data. The null hypothesis is that there is a U-shape between life satisfaction and age, over the age range 20 to 70, with a turning point anywhere between 30 and 60. Because of the composite null hypothesis of monotone and U-shaped, interpretation requires care. The column entries 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 in an acceptable range.

U-outside The same criteria as the column U, except for the criterion that the turning point be in a certain range (there must, however, be a turning point).

mono. $\downarrow$ There is evidence consistent with a monotone regression function: both the test of quasi-convexity and quasi-concavity fail to reject; the turning point (if it exists) for both the test of quasi-convexity and quasi-concavity tests does not fall inside the range; the fit associated with the monotone test statistic is monotone decreasing (and not increasing).

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

low power Both quasi- tests fail to reject and the criteria for MONO. $\downarrow$ and MONO. $\uparrow$ are not met.

The INV-U columns and MONO. $\uparrow$ column have descriptions parallel to those of the U columns and MONO. $\downarrow$ 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, to be in the U column it is required that the test of quasi-concavity rejects. It is important to remember that monotone regression functions are in the null sets of both the quasi-convexity test and the quasi-concavity test. Further, note that all monotone regression functions (both

Table 2.3: 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	29	11	1	22
LS	24	5	4	0	27	14	0	24
inc	40	6	3	0	16	9	0	24
inc LS	39	5	3	0	10	13	0	28
inc ed emp marital	43	0	1	0	15	12	0	27
inc ed emp marital LS	31	3	2	0	13	22	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 sum 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. The 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 the text for descriptions of the columns. The level of significance used was 0.1.

increasing and decreasing) are in the nulls of the monotonicity test and the quasi-tests. The reason for this is the same as why we do not test for a U-shape directly using a critical smoothing parameter—so that the test statistic is well-defined in all cases. 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 is done in the MONO. columns. There are not many entries in the U-OUTSIDE and INV-U-OUTSIDE 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.[6]

Including the control variables income, education, employment status, and marital status is consistent with Blanchflower and Oswald (2008) for their within-country regressions. Regressions were not run separately by gender, which is also consistent with the within-country regressions in Blanchflower and Oswald (2008). Further, similar relationships have been found between males and females for life satisfaction and a variety of hedonic well-being measures across age (Steptoe et al., 2014).

[6]As the level of the test tends to zero, the table entries will move towards the MONO. columns and the LOW POWER column. Thus, when interpreting the MONO. columns, it is important to recognize that there is no criterion of a statistical test rejecting, only tests failing to reject. Hence, if entries move from the OUTSIDE columns to the MONO. columns (e.g. when adding a control variable or using a subsample), it is possible that it is purely a result of reduced power.

The results in Table 2.2 suggest that for the unconditional relationship, there is more evidence of a monotone decreasing relationship (40 countries) than for a U-shaped relationship (15 countries). When conditioning on income, education, employment, and marital status, the opposite is true—there is evidence of a U-shape in 38 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. After conditioning on financial satisfaction, several countries move from the U column to the MONO. $\downarrow$ column. Conditioning on another variable might lower the power of the test, which could be a reason for countries moving out of the U column (because of failing to reject quasi-concavity). However, most of the countries migrating out of the U column go into the MONO. $\downarrow$ column because the turning point (if it exists) is no longer in the U-shape range, which is not a consequence of low power.

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 be surprising, given that income expectedly decreases in old age. Taking the subset of 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. Further research is needed to understand these divergent relations. Both questions are household questions. Income might be going down not only because the old themselves earn less but because the old live in households with fewer other wage earners. Similarly, perhaps once they move out of such households they shed themselves of financial burdens. Differences in composition of households could explain why the U-shape is more commonly found in high-income, English speaking countries. A separate explanation is that although income is decreasing, expenses are also decreasing (e.g. costs of raising children).

	Inv-U*	Inv-U$^{\not\star}$	mono. $\downarrow$	mono. $\uparrow$	U*	U$^{\not\star}$
Developing	6	11	18	5	11	12
E. European	0	0	6	0	6	1
Western	3	3	3	5	6	6
Sum	9	14	27	10	23	19

(a) Unconditional

	Inv-U*	Inv-U$^{\not\star}$	mono. $\downarrow$	mono. $\uparrow$	U*	U$^{\not\star}$
Developing	4	17	13	8	7	14
E. European	0	0	5	0	7	1
Western	3	6	5	2	5	5
Sum	7	23	23	10	19	20

(b) Controlling for financial satisfaction

Figure 2.10: OLS Results for Life Satisfaction

This table provides OLS results for testing a U-shape of life satisfaction with respect to age. Figure (a) has control variables for education, marital status, and employment status. Figure (b) includes the same control variables as in (a) and additionally controls for financial satisfaction. $\star$ means "significant" and $\not\star$ means "not significant" (at the 0.05 level).

Table 2.3 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 there was in Table 2.2, which suggests that financial satisfaction could be an underlying driver of the observed U-shaped relationship between SWB and age. When financial satisfaction is added as a control variable in the life satisfaction test (second row of the table), the number of countries consistent with a U-shape drops from 15 to 5. Conversely, when life satisfaction is added as a control in the financial satisfaction test, the number stays constant at 24. This suggests that financial satisfaction affects the U-shaped relationship between life satisfaction and age more than the other way around.

2.4.3 OLS results using a quadratic specification

Tables use the quadratic test to explore the U-shape of SWB. Although this method of testing is less flexible than the semi-parametric results presented, it is the most common method used in the literature and the simplicity of it makes it attractive. The OLS-based test is carried out for each country. A country is categorized as

	Inv-U$^{\bigstar}$	Inv-U$^{\not\bigstar}$	mono. $\downarrow$	mono. $\uparrow$	U$^{\bigstar}$	U$^{\not\bigstar}$
Developing	4	6	11	2	22	18
E. European	0	0	2	0	11	0
Western	1	1	3	1	15	5
Sum	5	7	16	3	48	23

(a) Control variables from Blanchflower and Oswald (2008)

	Inv-U$^{\bigstar}$	Inv-U$^{\not\bigstar}$	mono. $\downarrow$	mono. $\uparrow$	U$^{\bigstar}$	U$^{\not\bigstar}$
Developing	4	8	13	5	12	21
E. European	0	0	3	0	7	3
Western	1	2	5	1	9	8
Sum	5	10	21	6	28	32

(b) Additionally controlling for financial satisfaction

Figure 2.11: OLS Results for Life Satisfaction with More Control Variables

This table provides OLS results for testing a U-shape of life satisfaction with respect to age. Figure (a) has control variables for education, marital status, and employment status. Figure (b) includes the same control variables as in (a) and additionally controls for financial satisfaction. $\bigstar$ means "significant" and $\not\bigstar$ means "not significant" (at the 0.05 level).

having a U-shape if for the subset of 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 corresponding with the bottom of the fit is within the range $[30, 60]$. Similarly, a country is categorized as having an inverse-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 corresponding with 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. There are improvements that could be made to the test described above. For example, it could be tested that the turning point is *significantly* within the range. See Lind and Mehlum (2010, Section III) for a discussion.

The columns $U^{\not\bigstar}$ and Inv-$U^{\not\bigstar}$ indicate that the signs of the coefficient were of the right sign and the turning point was within the required range, but the significance tests did no pass. They are useful for concerns that a country leaves the U category only because of power (e.g. because of added control variables).

	Inv-U*	Inv-U$^{\not\star}$	mono. $\downarrow$	mono. $\uparrow$	U*	U$^{\not\star}$
Developing	7	5	22	3	15	11
E. European	0	0	5	0	5	3
Western	2	0	2	9	7	6
Sum	9	5	29	12	27	20

(a) Unconditional

	Inv-U*	Inv-U$^{\not\star}$	mono. $\downarrow$	mono. $\uparrow$	U*	U$^{\not\star}$
Developing	3	9	18	3	15	15
E. European	0	0	5	3	3	2
Western	1	0	3	14	5	3
Sum	4	9	26	20	23	20

(b) Controlling for life satisfaction

Figure 2.12: OLS Results for Financial Satisfaction

This table provides OLS results for testing a U-shape of financial satisfaction with respect to age. Figure (a) has control variables for education, marital status, and employment status. Figure (b) includes the same control variables as in (a) and additionally controls for life satisfaction. $\star$ means "significant" and $\not\star$ means "not significant" (at the 0.05 level).

The "mono." columns are meant to give an idea of why a country is not in one of the U-shape categories and not as evidence that a country has a monotone life satisfaction relationship with age. A country falls into a "mono." category if both coefficients have the same sign or if the turning point of the U-shape or inverse-U-shape is not in the range $[30, 60]$ (that is, it is in 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 different specifications are used as the base specification—an unconditional specification, and a specification using the same control variables as in Blanchflower and Oswald (2008).

When testing for a U-shape of financial satisfaction with respect to age, and controlling for life satisfaction in the same way as financial satisfaction was controlled for above, the results are similar. That is, there are fewer U-shapes when the additional satisfaction variable is controlled for, but not that many fewer. The OLS results do not suggest the same conclusion as the semi-parametric results, which was that financial satisfaction explains the U-shape of life satisfaction. Instead,

	Inv-U*	Inv-U$^{\not\star}$	mono. $\downarrow$	mono. $\uparrow$	U*	U$^{\not\star}$
Developing	2	5	8	6	28	14
E. European	0	0	1	0	10	2
Western	0	1	1	5	16	3
Sum	2	6	10	11	54	19

(a) Control variables from Blanchflower and Oswald (2008)

	Inv-U*	Inv-U$^{\not\star}$	mono. $\downarrow$	mono. $\uparrow$	U*	U$^{\not\star}$
Developing	3	6	8	7	14	25
E. European	0	0	2	3	4	4
Western	0	0	1	11	11	3
Sum	3	6	11	21	29	32

(b) Additionally controlling for life satisfaction

Figure 2.13: OLS Results for Financial Satisfaction with More Control Variables
This table provides OLS results for testing a U-shape of financial satisfaction with respect to age. Figure (a) has control variables for education, marital status, and employment status. Figure (b) includes the same control variables as in (a) and additionally controls for life satisfaction. $\star$ means "significant" and $\not\star$ means "not significant" (at the 0.05 level).

they support the conclusion that when a variable that is highly correlated with the dependent variable is controlled for, there are fewer significant coefficients. One possible explanation for the difference in OLS results and semi-parametric results is that OLS does not capture the effect of one variable on another in a flexible way and thus it is more difficult to figure out which variable is the underlying driver of an observed relationship.

2.4.4 Cross-country results

In addition to running regressions within country, we now look at pooled regressions, motivated by cross-country models in this section. These models allow for greater power and allow for another test of the theory that there are biological foundations to the U-shape. If indeed the theory holds, there should be a common relationship that holds across countries.

Pooling the data could also 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 of the test. If those violations occurred because of cohort effects that were specific to a country, then by averaging across several countries, if the cohort effects are not correlated, they would be smoothed away. To see this more formally, let us take the general age, period, cohort framework that was presented in Section 2.2.1,

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

For simplification, let us assume that period effects are zero and further that cohort effects depend on the country, but that averaging the cohort effects over countries yields a constant. Formally, we assume the model

$$y_{i,b,a,c} = \alpha_{bc} + \beta_a + \epsilon_{i,b,a},$$

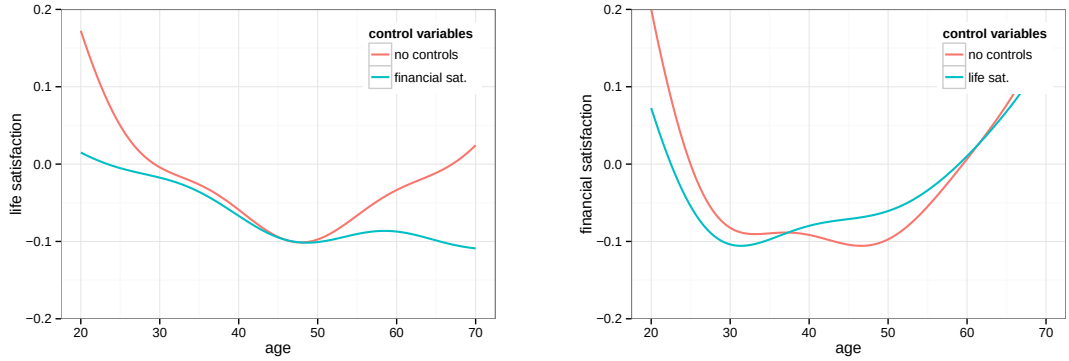
where $\sum_c \alpha_{bc} = k$ where k does not depend on birth year, b. This assumption would be violated if cohort effects were correlated across countries, for example during World War 2 or world-wide economic crises that scarred people of certain ages more than others. If the above assumptions do hold, GAM estimations of age effects would be biased for the within-country regressions but not for the pooled regressions.[7]

The red line in Figure 2.14a gives a first look at the pooled data relationship between life satisfaction and age. The panel reproduces the U-shaped relationship that is commonly found in similar regressions. In the same panel, the blue line shows that after controlling for financial satisfaction, the effect of age on life satisfaction changes—the right arm of the U mostly flattens out. The two lines correspond to estimates of $f(age)$ in the following models:

$$LS = f(age) + \epsilon \tag{2.2}$$

$$LS = f(age) + g\,(FS) + \epsilon \tag{2.3}$$

[7]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.



(a) Specifications (2.2) and (2.3)

(b) Specifications (2.4) and (2.5)

Figure 2.14: Life and Financial Satisfactions Against Age

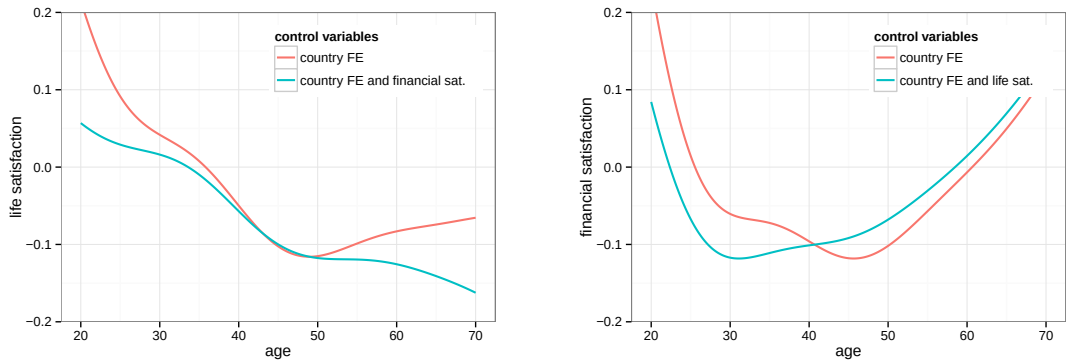
This figure shows estimates of the age effects on life satisfaction and financial satisfaction using GAMs, for the specifications referenced below each plot.For estimates of the same models using ordered logits and grouped dummies, see Appendix 2.G. The estimates are similar.

There is reason to be concerned with the concept of regressing one subjective measure on another. Such a regression could suffer from correlated measurement error. For example, if a survey respondent is in a good mood, she could respond with a high response to both subjective measures. We might see a high correlation between the two when in fact the correlation is caused by a (hidden) type variable. One robustness check to address this possibility is to run the same regressions as above, but switching the positions of LS and FS. Specifically, the following specifications are additionally estimated:

$$FS = f(age) + \epsilon \tag{2.4}$$

$$FS = f(age) + g\left(LS\right) + \epsilon \tag{2.5}$$

These regressions also allow us to answer the question of how it is possible that financial satisfaction can explain so much of the U-shaped relationship between life satisfaction and age. A possibility is that financial satisfaction is U-shaped in age (and financial satisfaction positively affects life satisfaction). This hypothesis is positively supported by the estimation of the above two specifications, shown in Figure 2.14b.



(a) Specifications (2.6) and (2.7)

(b) Specifications (2.8) and (2.9)

Figure 2.15: Life and Financial Satisfactions Against Age (FE)

This figure shows estimates of the age effects on life satisfaction and financial satisfaction using GAMs, for the specifications referenced below each plot.

In fact, financial satisfaction seems to have a stronger U-shaped relationship with age than life satisfaction does, and the shape shifts only slightly when including life satisfaction as a control variable.

If the same countries that have older populations are also the countries that generally have higher life and financial satisfactions, 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 2.4.2 should address this concern, but we can also address the concern with the pooled estimations by including country fixed effects. Figure 2.16 shows estimations of the following specifications:

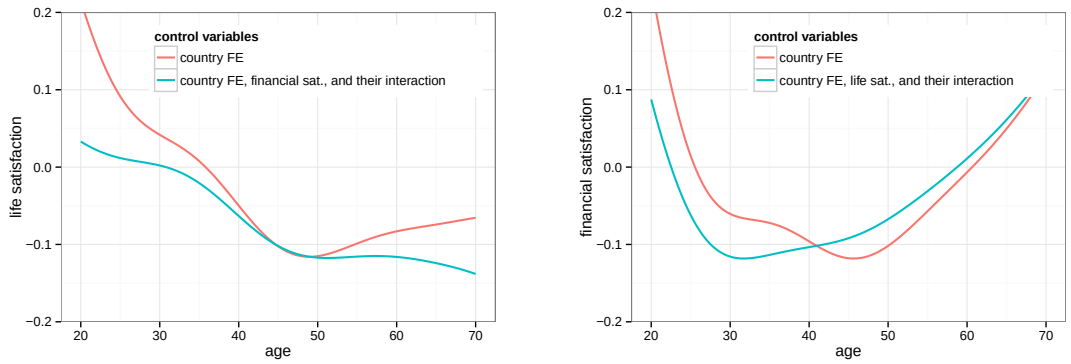
$$LS = f(age) + \sigma_c + \epsilon \tag{2.6}$$

$$LS = f(age) + g(FS) + \sigma_c + \epsilon \tag{2.7}$$

$$FS = f(age) + \sigma_c + \epsilon \tag{2.8}$$

$$FS = f(age) + g(LS) + \sigma_c + \epsilon \tag{2.9}$$

Although there is some prior theory for introducing the effect of age on life satisfaction as a function that is shared across countries, there is less reason to require that the



(a) Specifications (2.6) and (2.10)

(b) Specifications (2.8) and (2.11)

Figure 2.16: Life and Financial Satisfactions Against Age (FE Interacted)
This figure shows estimates of the age effects on life satisfaction and financial satisfaction using GAMs, for the specifications referenced below each plot.

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 their 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 one's life depends on the welfare system in the country. The following specifications thus allow for the relationship between financial satisfaction and life satisfaction to differ by country:

$$LS = f(age) + g_c(FS) + \sigma_c + \epsilon \tag{2.10}$$

$$FS = f(age) + g_c(LS) + \sigma_c + \epsilon \tag{2.11}$$

The results, shown as the blue lines in the two panels of Figure 2.16. Note that in 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.[8] As the distributions in Figure 2.17 show, the support of the two variables

[8]The location of the y-axis is not, however, comparable, or interpretable and is thus centered around 0.

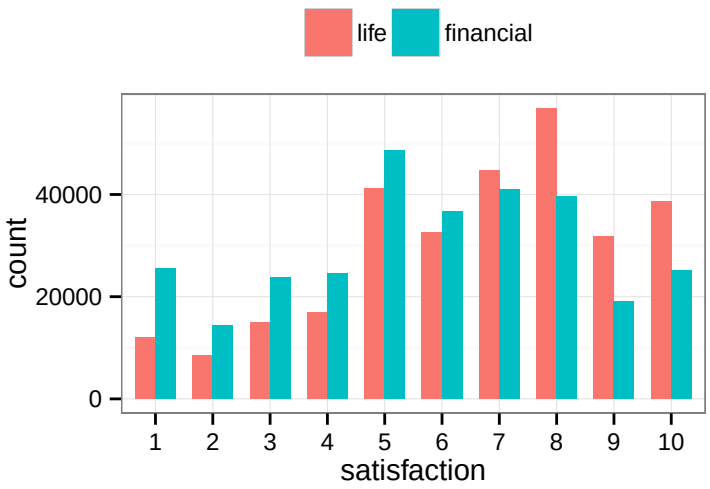


Figure 2.17: Distribution of Life and Financial Satisfactions
This figure shows the marginal distributions of life satisfaction (red) and financial satisfaction (blue).

is the same, and the distributions are similar. The main difference between the two distributions is that in the bottom three categories there is more mass for financial satisfaction than for life satisfaction, and the opposite for the top three categories. Thus, it does not seem that the differences in marginal distribution are the reason for financial satisfaction's higher explanatory power of the U.

Why financial satisfaction explains the U-shape

The results in the previous sections suggest that financial satisfaction can explain much of the U-shaped relationship between life satisfaction and age, and that financial satisfaction itself shows a U-shaped relationship with age. In this section some reasons are discussed for why the above patterns are found in the data.

Financial obligations are the greatest in middle ages. Raising children increases current financial expenditures and stress about meeting future expenditures. The parents of middle-aged individuals are usually no longer in the workforce and thus less able to provide support. Further, in many cultures there are expectations to take

care of parents, which is another cause of financial stress. In some countries, it is common to have large debts in the middle ages, such as from mortgages.

In addition to fewer expenditure in old ages, stress about one's financial situation might be lower because of reduced uncertainty about the future (because the future is shorter). Because there are fewer years left to live, there is less planning to do. At 70 years old, one does not have the stress of planning for retirement. Similarly, there is less need to save as insurance against events such as losing one's job. And conversely to the situation of middle ages, in many countries there are welfare systems or cultural expectations about being cared for by children that ensure that the old do not have financial concerns.

If financial satisfaction does indeed explain the U-shape between life satisfaction and age, it becomes important to consider how to use the link between financial satisfaction and life satisfaction to improve life satisfaction, which seems like a reasonable target of maximization by social planners. A simple way to increase life satisfaction through financial satisfaction is to improve the financial situation of the individual, for example by increased income. A second possible way in which to use the channel of financial satisfaction to improve life satisfaction would be to change the mechanism that translates financial situation to financial satisfaction—for example, being more satisfied with what one has. A third way would be to change the mechanism through which financial satisfaction affects life satisfaction. This latter idea is likely something difficult to change and could be part of the culture. In fact, the cultural differences of how financial satisfaction translates to life satisfaction could explain why the U-shape of life satisfaction in age appears in some countries and not in others.

2.5 Conclusion

Both the quadratic specifications (using OLS and ordered logits) and semi-parametric specifications suggest that financial satisfaction is U-shaped in age in more countries than life satisfaction is. Further, results from the semi-parametric analysis, both within countries and in several pooled models, suggest that financial satisfaction explains the U-shape of life satisfaction in age. These results suggest that the U-shape of life satisfaction is not due to a midlife crisis that is hard-coded in our DNA. Because the U-shape of life satisfaction is affected by financial satisfaction, we can thus think of questions such as: Should we have policies that help us smooth our life satisfaction? If so, one way to do this might be through affecting financial satisfaction. More research is needed to address the identification issues inherent in age, period, and cohort models and to disentangle reverse causality between life satisfaction and financial satisfaction.

References

Belsky, D. W., Caspi, A., Houts, R., Cohen, H. J., Corcoran, D. L., Danese, A., Harrington, H., Israel, S., Levine, M. E., Schaefer, J. D. et al.: 2015, Quantification of biological aging in young adults, *Proceedings of the National Academy of Sciences* p. 201506264.

Blanchflower, D. G. and Oswald, A. J.: 2008, Is well-being U-shaped over the life cycle?, *Social Science & Medicine* **66**(8), 1733 – 1749.

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.

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

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

Heckman, J. and Robb, R.: 1985, Using longitudinal data to estimate age, period and cohort effects in earnings equations, *Cohort analysis in social research*, Springer, pp. 137–150.

Kahneman, D. and Krueger, A. B.: 2006, Developments in the Measurement of Subjective Well-Being, *The Journal of Economic Perspectives* **20**(1), pp. 3–24.

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.

Layard, R.: 2011, *Happiness: Lessons from a New Science*, second edn, The Penguin Press, New York.

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.

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

Steptoe, A., Deaton, A. and Stone, A. A.: 2014, Subjective wellbeing, health, and ageing, *The Lancet* .

University of California, Berkeley (USA) and Max Planck Institute for Demographic Research (Germany): 2012, Human Mortality Database.

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.

Appendix

2.A Identification of Second Differences of Age, Period, and Cohort Effect Profiles

In this section I generalize the framework of McKenzie (2006) to allow for not having data of consecutive years. Denote the sample size of cohort c_b observed at age a as $n_{b,a}$. Averaging over each (cohort, age) pair yields

$$\overline{y}_{b,a} = \frac{1}{n_{b,a}} \sum_{i=1}^{n_{b,a}} y_{i(b,a),b,a} = \alpha_b + \beta_a + \gamma_{b+a} + \overline{\epsilon}_{b,a}$$

where

$$\overline{\epsilon}_{b,a} = \frac{1}{n_{b,a}} \sum_{i=1}^{n_{b,a}} \epsilon_{i(b,a),b,a}.$$

Note that because the data used is a pseudo-panel and not a panel, the same i most likely refers to a draw of a different person depending on the year. That is, $y_{i,b,a}$ does not refer to the same person as $y_{i,b,a-1}$ and in fact does not even exist if $i > n_{b,a}$. This distinction is highlighted by having the individual index be a function of the birth year and age. Then, define a difference operator Δ_a to get that

$$\Delta_a \overline{y}_{b,a} = \overline{y}_{b,a} - \overline{y}_{b,a-1} = (\beta_a - \beta_{a-1}) + (\gamma_{b+a} - \gamma_{b+a-1}) + (\overline{\epsilon}_{b,a} - \overline{\epsilon}_{b,a-1}) \tag{2.12}$$

The cohort effects were differenced out. Note that equation (2.12) might be interesting in its own right. Combined with theory this equation could provide some identification. For example, suppose that we expect the age effect to change considerably more (in absolute) from age 18 to age 19 than the period effect changes from one period to the next—perhaps we think that graduating from high school leads to a much larger change in life satisfaction than the period change for a certain period. In this case, the difference in age effects would dominate the difference in period effects and we could get an idea of the first difference $\beta_{19} - \beta_{18}$.

Differencing again for a different cohort,

$$\Delta_a \overline{y}_{b-1,a+1} = (\beta_{a+1} - \beta_a) + (\gamma_{b+a} - \gamma_{b+a-1}) + (\overline{\epsilon}_{b-1,a+1} - \overline{\epsilon}_{b-1,a}) \tag{2.13}$$

Define $\Delta_b \Delta_a$ such that subtraction eliminates the difference in period effects:

$$\Delta_b \Delta_a \overline{y}_{b-1,a+1} = \Delta_a \overline{y}_{b-1,a+1} - \Delta_a \overline{y}_{b,a} = \\ (\beta_{a+1} - \beta_a) - (\beta_a - \beta_{a-1}) + (\overline{\epsilon}_{b-1,a+1} - \overline{\epsilon}_{b-1,a}) - (\overline{\epsilon}_{b,a} - \overline{\epsilon}_{b,a-1}) \tag{2.14}$$

Instead of the cross-difference in differences in equation (2.12) there is now a double difference of the age effects in equation (2.14). For this double difference to give us identification the time must be offset by the same amount as age in equation (2.13) so that in equation (2.14) the same difference is subtracted in cohort effects. In the above differencing, this amount was chosen to be one. Because in the WVS-EVS countries are not observed in consecutive years, $\overline{y}_{b,a}$ and $\overline{y}_{b,a-1}$ will not both exist. A more general framework is needed. Let $\omega_k = t_k - t_{k-1}$ be the difference between observation years. Suppose that cohort b is observed in year t_k and let $a = t_k - b$ so that $\overline{y}_{b,t_k-b}$ is observed. For a meaningful difference we would like to cancel the

cohort effect so we use the same cohort observed ω_k years earlier:

$$\overline{y}_{b,t_k-b} - \overline{y}_{b,t_k-b-\omega_k} = \\ (\beta_{t_k-b} - \beta_{t_k-b-\omega_k}) + (\gamma_{t_k} - \gamma_{t_k-\omega_k}) + (\overline{\epsilon}_{b,t_k-b} - \overline{\epsilon}_{b,t_k-b-\omega_k}) \quad (2.15)$$

Now the goal is to get rid of the time effect. This can be done by using the difference of another cohort to get the same difference in time effects. Theory could be used to suggest a pair of time effects whose difference would be approximately equal to $(\gamma_{t_k} - \gamma_{t_k-\omega_k})$ but the agnostic approach is to use only pairs that give us that same difference. Thus,

$$\overline{y}_{b-\omega_k,t_k-b+\omega_k} - \overline{y}_{b-\omega_k,t_k-b} = \\ (\beta_{t_k-b+\omega_k} - \beta_{t_k-b}) + (\gamma_{t_k} - \gamma_{t_k-\omega_k}) + (\overline{\epsilon}_{b-\omega_k,t_k-b+\omega_k} - \overline{\epsilon}_{b-\omega_k,t_k-b})$$

For this it would be necessary that ω_{k+1}=ω_k so that $t_{k+1} = t_k - b + (t_k - t_{k-1})$ I skip defining a bulky difference operator that would depend on ω_k. The equation of interest is

$$\left(\overline{y}_{b-\omega_k,t_k-b+\omega_k} - \overline{y}_{b-\omega_k,t_k-b}\right) - \left(\overline{y}_{b,t_k-b} - \overline{y}_{b,t_k-b-\omega_k}\right) = \\ (\beta_{t_k-b+\omega_k} - \beta_{t_k-b}) - (\beta_{t_k-b} - \beta_{t_k-b-\omega_k}) + \\ + (\overline{\epsilon}_{b-\omega_k,t_k-b+\omega_k} - \overline{\epsilon}_{b-\omega_k,t_k-b}) - (\overline{\epsilon}_{b,t_k-b} - \overline{\epsilon}_{b,t_k-b-\omega_k}) \quad (2.16)$$

Note that when $\omega_k = 1$ in equation (2.16) the result is equation (2.14).

2.B Table 5 of Blanchflower and Oswald (2008)

Table 2.4 is an exact replication of Table 5 in Blanchflower and Oswald (2008) except that I point out that "Western Europe" should be "Western Countries." This table shows the only results using the WVS-EVS presented in the paper.

Table 2.4: Argmins of SWB-Age Relationship, by Country Category and Sex (Table 5 in Blanchflower and Oswald, 2008)

Table 5
Life satisfaction equations: World Values Survey data 1981−2004

	(1)	(2)	(3)	(4)	(5)	(6)
	Western Europe		Eastern Europe		Developing Countries	
	Men	Women	Men	Women	Men	Women
Age	−0.0570 (13.74)	−0.0371 (9.70)	−0.0772 (10.68)	−0.0593 (9.20)	−0.0414 (8.95)	−0.0360 (8.12)
Age^2	0.0006 (14.24)	0.0004 (9.72)	0.0008 (10.67)	0.0006 (8.83)	0.0005 (9.29)	0.0004 (7.86)
Cut1	−4.9134	−4.4746	−4.0432	−3.7387	−2.5244	−2.1663
Cut2	−4.2999	−3.8588	−3.4915	−3.1577	−1.9276	−1.5718
Cut3	−3.4685	−3.0849	−2.7613	−2.4412	−1.3525	−1.0131
Cut4	−2.8399	−2.4890	−2.2223	−1.9426	−0.9067	−0.5640
Cut5	−1.9622	−1.5927	−1.3586	−1.0357	−0.0194	0.3228
Cut6	−1.3187	−0.9949	−0.8194	−0.4997	0.4750	0.8155
Cut7	−0.4274	−0.2169	−0.1014	0.0982	1.0796	1.3702
Cut8	0.8268	0.9366	0.9300	1.0265	1.8427	2.1340
Cut9	1.7985	1.9007	1.6424	1.7271	2.5311	2.7979
Sample size	33,470	35,448	12,806	14,419	33,631	33,072
Pseudo R^2	0.0324	0.0279	0.0284	0.0274	0.0488	0.0464
Log likelihood ratio	−62,681	−67,801	−27,268	−30,959	−70,942	−69,561
Life satisfaction minima	***45.2***	***47.0***	***46.5***	***48.2***	***42.6***	***44.3***

Notes: The source is: World Values Surveys, 1981−1984; 1989−1993; 1994−1999 and 1999−2004. Controls are 9 income deciles, 5 marital-status dummies, 9 education dummies, 3 year dummies and 7 work-force status dummies. Western countries are Australia; Austria; Belgium; Canada; Denmark; Finland; France; Germany; Greece; Iceland; Ireland; Italy; Japan; Luxembourg; Malta; Netherlands; Norway; Northern Ireland; Portugal; Spain; Sweden; Switzerland; Great Britain; USA; West Germany. Eastern Europe countries are Bosnia and Herzegovina; Bulgaria; Croatia; Czech Republic; Estonia; Hungary; Latvia; Lithuania; Poland; Romania; Slovakia; Slovenia; Macedonia; Serbia and Montenegro. Developing countries are Albania; Algeria; Argentina; Bangladesh; Belarus; Brazil; Chile; China; Colombia; Dominican Republic; Egypt; India; Indonesia; Iran; Iraq; Jordan; Kyrgyzstan; Mexico; Moldova; Morocco; Nigeria; Peru; Philippines; Puerto Rico; Russia; Saudi Arabia; Singapore; South Africa; Taiwan; Tanzania; Turkey; Uganda; Ukraine; Uruguay; Venezuela; Vietnam and Zimbabwe. Question is "All things considered, how satisfied are you with your life as a whole these days?" A. 1 'Dissatisfied' to 10 'Satisfied'. *t*-statistics are in parentheses.

*Note that the column heading "Western Europe" should be "Western Countries."

2.C Defining a U-shape

The hardest part of specifying a potentially U-shaped relationship between two variables is deciding which variables we are interested in controlling for. This topic was discussed in Section 2.E, using health and income as examples. There is also a

technical difficulty in that a U-shape is not easily defined. For example, should we distinguish between a U, a V, and a cursive $\mathcal{U}$?

The first step in creating a test is to decide what the null and alternative sets should be. The contents of the null and alternative sets should be decided by the researcher, before looking at the data, based on what her theory would predict. This is where the particular properties of the problem must be taken into account. For example, must the U be symmetric? In an SWB context, there is no reason to impose symmetry. Some properties seem desirable in all applications. For example, should a smile, $\smile$, be considered a U? Just as magnitude is left out of most tests involving the significance of a regression coefficient, how vertically stretched a U-shape is should not affect whether it is a U-shape, but could affect a separate measure of *how* U-shaped it is. That is, the test should be invariant to monotonically increasing transformations of the y-axis. This point is particularly important in an SWB context, where there is no inherent measure.

One attractive definition of a U-shape is a function that is downward sloping to a point and then upward sloping. But should it be allowed to have some bumps and deviations along its way? It is likely here that theorists and practitioners split. A practitioner would care only that the regression function has a general U-shape. Another issue is whether a checkmark-like function should be included in the U-null. The answer would depend on how close the minimum of the shape is to the beginning or end of the range of X, but this requires arbitrary decision making.

Once the ideal null and alternative sets have been chosen by the researcher, based on theory, the next step to constructing a test is to come up with properties that capture the essence of the null that are not shared by elements of the alternative. Then, based on the goal of the researcher (e.g., public policy), she chooses a loss function. Finally, estimators are developed and compared.

This discussion on abstract development of tests was meant to lay the groundwork for understanding problems with parametric tests that were not tailored to the specifics of an SWB U-shape in age.

2.D Cubic Specification as Improvement to Quadratic

Some of the inflexibility of the quadratic specification comes from its single inflection point, requiring it to be either globally convex or globally concave. In the SWB context, it seems we would want to allow for possibilities such as having a convex shape on one side of the U and a concave shape on the other side. For example, it is possible that SWB decreases in a convex way to a certain point and then increases, but in a concave manner, allowing for the increased SWB that comes with old age to be overcome by the decrease in SWB that comes from worsening health. A cubic specification allows for this.

A cubed term adds flexibility because it allows for fits with an additional inflection point. A more flexible parametric setup, if consistent with theory, will allow for a better fit of the data and will increase the power of tests. The cubic setup shares the advantages of a quadratic setup in that (1) a small sample size can be used, (2) the estimate of the minimum is easy to carry out, and (3) the standard error of that estimate is straightforward to calculate, using the delta method.

I propose a cubic age specification for consideration because it is manageable and allows for a better fit, increasing power and credibility of estimates of the minimum and their standard errors. Compared to a quadratic specification, it is reasonable in the SWB context that we would want convexity on the left side of the potential U-shape and concavity on the right side, allowing for health issues to pull SWB downward at older ages. Higher order polynomials are not suggested by theory and less manageable properties because of their complicated roots.

2.E Whether to Condition on Health

Suppose that happiness were a function of age and health. For example, suppose that happiness of individual i is the sum of a function of his age and a function of the amount of pain he experiences during normal daily activities:

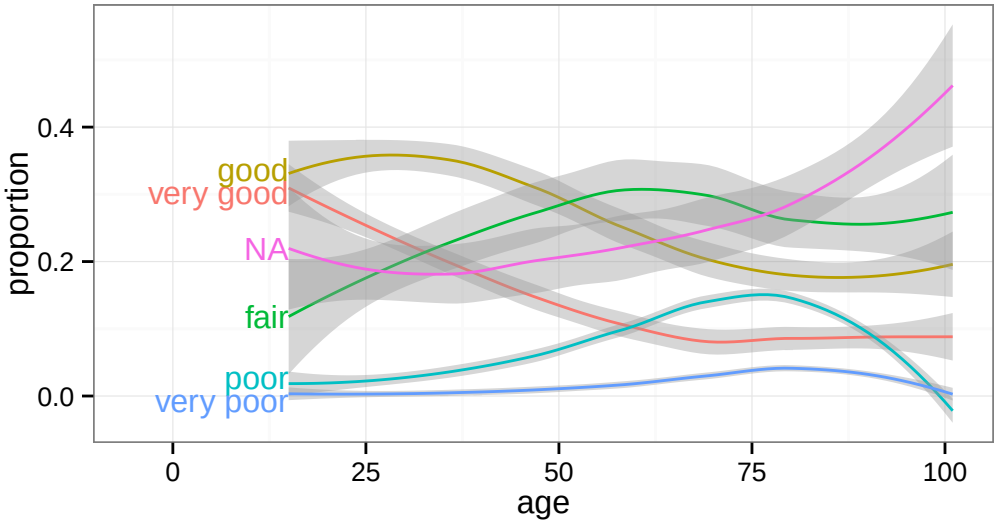
$$H_i = f(\text{age}_i) + g(\text{pain}_i)$$

Because pain_i is likely correlated with age_i, if pain_i is omitted from the regression, then $f(\text{age}_i)$ will be incorrectly estimated. One could claim that because increased pain is part of growing old, we should not control for pain, thus allowing our specification to attribute the effect of pain on happiness to the effect of age on happiness. However, if we believe that the effect on happiness of age (net of pain) is mainly driven by emotional capital, we might want to isolate this effect from the effect of pain.

Isolating the individual effects that age is a proxy for will give us a deeper structural understanding and will thus allow us to better conceptualize counterfactuals. If we keep age and pain aggregated, we are only equipped for wondering what happens, on average, when age changes, whereas if we have an understanding of the disaggregated terms we can examine a finer level of counterfactuals. For example, we could compare the effects on happiness of technology that decreases pain but does not necessarily increase the length of our lives with the effects of technology that increases the length of our lives but does nothing to alleviate old-age pain.

In the empirical part of this paper, I do not have the luxury of deciding whether to include a pain variable in the regressions because there was no question about pain in the WVS-EVS. Further, in testing for a U-shape in age net of pain, omitting pain likely biases our tests towards falsely rejecting a U-shape. Thus evidence in favor of a U-shape is not negated by the concern raised above; in fact, it is strengthened.

Figure 2.18: Health vs. Age



The question (a009) and response codings for self-reported health were "All in all, how would you describe your state of health these days? Would you say it is... 1 (Very good), 2 (Good), 3 (Fair), 4 (Poor), 5 (Very poor)." See Appendix 2.F for the life satisfaction question.

Health is not controlled for in the regressions by Blanchflower and Oswald (2008) for two reasons—"This is partly pragmatic: our data sets have no objective measures and few subjective ones. But the decision is partly substantive: it seems interesting to ask whether people become happier as they age once only demographic and economic variables are held constant." Indeed, it is not clear why we would want to compare an 80-year-old to a 30-year-old, holding health constant. If the identifying power came from such comparisons, what the estimates are actually estimating would be difficult to figure out. Thus, there are understandable reasons for not including health in the regressions.

However, although it might not make sense to hold constant triple jump distances, holding constant whether one can walk without pain or how much time was spent in a hospital in the last year might make sense. Additionally, even the decision of whether to include a general health variable depends on how the question is asked and how it is thought to have been interpreted and answered. If the question were answered with the respondent comparing herself to other people of the same age, then

it would make more sense to include the variable because comparing an 80-year-old claiming to be in relatively good health with a 30-year-old claiming to be in relatively good health would not hold actual health constant, but it might hold constant the portion of health that most affects life satisfaction if life satisfaction is affected more by self-assessment of health rather than health itself. It is interesting to note that health does seem to affect SWB when comparing individuals within the same age, across all ages. This is shown in Figure 2.4 which presents LOWESS curves fitted to the entire WVS-EVS sample.

One way to look at whether people use their peers as a reference group and if that is all that matters when reporting their health is by graphing reported health against age, by health category, and seeing if the fraction of people responding that they are in each category is constant across age. If people compared themselves only to people of the same age, if that comparison were the only thing that mattered in the response, and if these data were randomly sampled, then the lines would be roughly horizontal.

2.F Variables in 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.

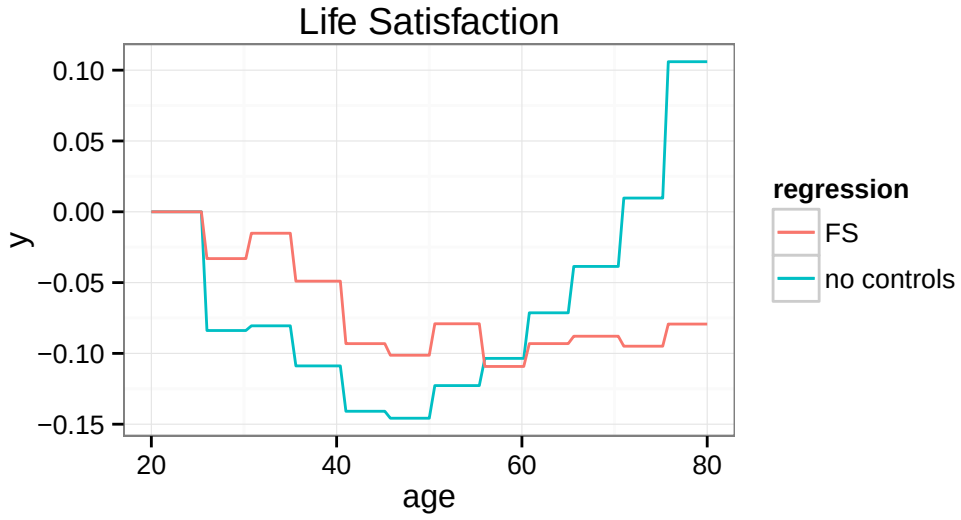
education At what age did you complete your education? <12 years, 13 years, 14 years, ..., 20 years, 21+ years.

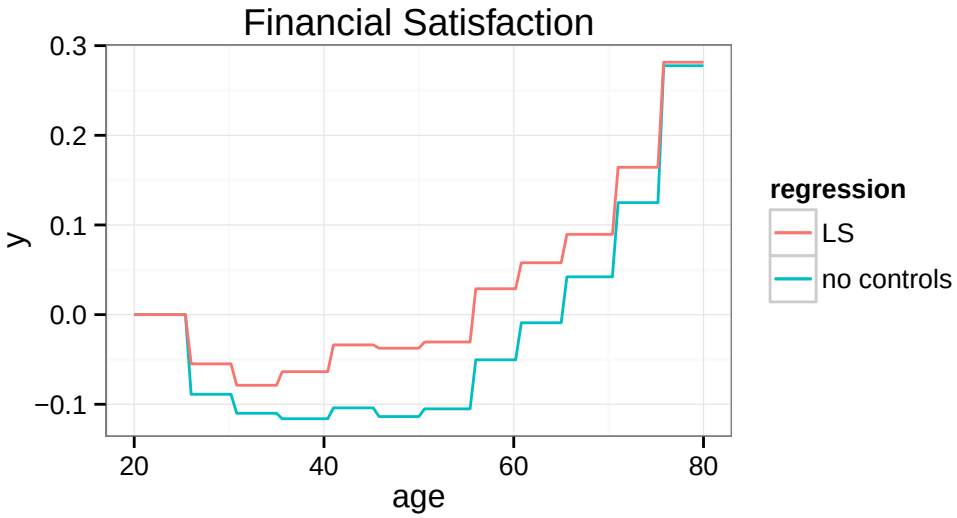
employment status Full time, Part time, Self employed, Retired, Housewife, Student, Unemployed.

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

2.G Ordered Logit Dummy Plots

The following plots use specifications similar to Figures 2.14a and 2.14b, but instead of using GAMs to estimate the age profiles, ordered logits are used with grouped dummies to capture the age effects for life satisfaction in the first plot and financial satisfaction in the second plot. Financial satisfaction dummies are added as control variables in the first plot, and life satisfaction dummies in the second plot. The results are similar to the results in the text that use GAMs.




Financial Satisfaction
0.3
0.2
0.1
0.0
−0.1
y
20
40
60
80
age
regression
LS
no controls

Chapter 3

The Partial Monotonicity Parameter

3.1 Introduction

This paper proposes a new parameter, the partial monotonicity parameter (PMP), which answers the question "For what percentage of the population is a small increase in X associated with an increase in Y?" This question cannot be answered by a simple OLS regression of Y on X. It has long been recognized that OLS estimates are sensitive to a minority's strong positive effect drowning out a majority's weak negative effect. Parameters estimated by OLS are (purposefully) sensitive to magnitude and as a result do not in general pick up information regarding the question asked above. By abstracting away from magnitude, PMP is able to capture information that can be used to complement OLS. Similar to the common practice of reporting the mean and the median of a variable, OLS and PMP together can provide a more complete picture of the regression function being studied than OLS alone.

From a more structural point of view, the PMP allows researchers to improve on the common practice of testing predictions of monotonicity with OLS, by providing a continuous measure that generalizes the notion of monotonicity. Researchers no

longer have to be concerned that a theory that holds for the most part could be rejected by a test of monotonicity solely because of a small violation of monotonicity.

The PMP can be estimated parametrically and non-parametrically. The focus in this paper will be on non- and semi-parametric estimation of the PMP. There is often a tradeoff between non-parametric techniques and parametric techniques when exploring data and reporting results of analysis. Non-parametric techniques make fewer assumptions and often provide a more complete picture of the relationships between variables. However, they also often require large data sets and commonly do not allow for convenient reporting or hypothesis testing, and are not as easily interpreted as parametric techniques.

Even when a large dataset is available, researchers commonly choose to present only OLS results. One reason for this practice is that the results of an OLS regression can be summarized by presenting coefficients and their standard errors. Each regression can be presented in one column of a table, and many columns can be used to show how one regression compares to another. It is common to see tables that show ten regressions, each containing on average ten regressors. Doing fully non-parametric regression with ten independent variables is unmanageable and difficult to report. Even doing semi-parametric regression using generalized additive models, in order to present the full set of results in the same context, the researcher would need to show 100 graphs, taking up considerably more space than a single ten-by-ten table. In the case of regression functions, the difference in ease of reporting is due to the dimension of the parameters space. However, the parameter space of the PMP is the same (it is $[0, 1]$) whether it is estimated parametrically or non-parametrically, and so reporting is just as easy. Similarly, hypothesis tests are approached in the same way. Thus, in the case of the PMP, the argument for focusing on non-parametric estimation methods is strong.

The rest of this paper is organized as follows. Section 3.2 defines the PMP and focuses on motivation for its estimation. Because the main purpose of this paper is to propose a new parameter of interest for researchers, the most important part is the motivation for why conventional methods miss information and how the PMP captures this information in a desirable and simple way. Section 3.3 discusses the asymptotic econometric properties of the plug-in estimator of the PMP, and provides results from simulations using the bootstrap to estimate the finite distribution of the PMP. Section 3.4 extends the estimation of PMP in a natural way to multivariate analysis. Section 3.5 applies the estimation of PMP to explore a topic that is researched in the subjective well-being literature. Finally, Section 3.6 concludes.

3.2 Motivation

3.2.1 The partial monotonicity parameter

Let X and Y be random variables and for the sake of terminology let us suppose that a larger Y is always "better." Let the relationship between X and Y be modeled as

$$Y = f(X) + \epsilon, \tag{3.1}$$

where ϵ has mean zero and is independent of X. I define the partial monotonicity parameter (PMP) as

$$\delta^{PMP} = Pr\left\{f'(X) > 0\right\}.$$

The PMP is thus a functional of the distribution of the derivative of the regression function. We can see immediately from the definition that the PMP has an intuitive interpretation: If an observation is taken at random from the population, the PMP is the probability that increasing that observation's X by a small amount is beneficial.

Similarly, the PMP can be described as the fraction of the population such that an increase in X is associated with an increase in Y.

3.2.2 Related literature

Assuming a model similar to (3.1), Powell et al. (1989) and Härdle and Stoker (1989) propose estimating the average derivative of the regression function,

$$E\left\{f'\left(X\right)\right\}.$$

This parameter captures similar information to OLS, and does so in a more flexible way (one does not need to impose a restrictive functional form, as in OLS). However, such parameters are sensitive to magnitude and thus can miss some information about the average *sign* of the derivative of f, as discussed in the context of OLS above.

3.2.3 PMP as a generalization of monotonicity

From a structural perspective, many theories in economics predict monotone relationships between variables. However, monotonicity is rarely tested because it is viewed as too strict: a theory that holds for the most part could be rejected by a minor violation of monotonicity due to data issues, selection, or a violation of the theory for a small subgroup. In such cases, a researcher might want to test that monotonicity holds "to a large extent" and be able to quantify to exactly what extent it holds in the data. The PMP can fulfill this role and thus reconcile the asymmetry between prediction of monotonicity and the common practice of testing a much weaker prediction.

By introducing the notion of "partial monotonicity," estimating the PMP can help to quantify just how much strict monotonicity is violated. For example, estimating the PMP allows us to make statements such as "the relationship is approximately 90% monotonically increasing." Further, hypothesis tests based on the PMP nest two

tests of monotonicity:

$$H_0\colon\ \delta^{PMP} = 1 \iff \text{test of a strictly monotonically increasing } f(x)$$
$$H_0\colon\ \delta^{PMP} = 0 \iff \text{test of a weakly monotonically decreasing } f(x)$$

Similarly, the notion of partial monotonicity could play a role in testing for a U-shape. The most common way to test for a U-shape is to use OLS and include a quadratic term. This practice can in some case be too weak of a test of a U-shape. What really is tested is that a positive quadratic term approximates the true regression function better than a negative quadratic term, not that a positive quadratic approximates the regression function *well* (in which case there would be evidence that the regression function is close to U-shaped). It is straightforward to create a strict test of a U-shape by composing two tests of one-sided monotonicity. Such a test would be too strict to be attractive to practitioners. However, one could compose two tests of 75% monotonicity, which would test that the theory holds for the most part. In addition to hypothesis tests, one could in a similar way create a statistic that would provide a continuous measure of how U-shaped a regression function is. For example, the following is a reasonable measure for the upper bound of a confidence interval of level α for how U-shaped a regression function is:

$$\max\left\{a \in [0,1], b \in \mathbb{I} \subset \mathbb{R} \middle| \text{both PMP tests fail to reject } H_0\colon \gamma^{PMP} \geq a \text{ at level } \alpha\right\},$$

where $\mathbb{I}$ is an interval in which the turning point of a U could exist, and would be set by the researcher depending on what exactly a U-shape means in a given context.

3.2.4 Reduced form

From a more reduced form perspective, there are often violations of monotonicity in the data (because of selection issues, data corruption, type variables correlated with the X variable, etc.). Estimating γ^{PMP} provides an easy and interpretable way to summarize these violations and signal that there are heterogeneous effects. The researcher could then go back to the original specification to model these heterogeneous effects. Estimating the PMP of the regressors provides a quick check of whether polynomials are needed to capture non-monotone relationships, or whether to include an interaction with a type variable.

3.2.5 Ease of understanding

Focusing on a parameter that is easy to interpret is important in conveying information to the general audience, as well as experts who need to base actions on statistical results, such as doctors when recommending a medicine, or lawmakers when considering whether to enact a policy. Further, the simple interpretation of γ^{PMP} is an advantage to economists as well. Soyer and Hogarth (2012) show that economists often misunderstand OLS results, as they are commonly reported.

There are two easy ways to interpret γ^{PMP} (the terminology still assumes that the error term is independent of X):

1. If a person is drawn at random from the population, γ^{PMP} is the probability that increasing that person's X is beneficial.

2. γ^{PMP} is the proportion of the population for which increasing X is beneficial.

3.2.6 Related parameters of interest

In addition to the PMP, we might want to estimate the following related parameters:

1. $Pr_U\{f'(X) > 0\}$, where U is the uniform distribution over the range of X.

   This parameter could be interesting when comparing $f'(X)$ across different populations when those populations have different distributions of X.

2. $g(a) = \Pr\{f'(X) > a\}$, for $a \in \mathbb{R}$.

   This resulting function might be a more intuitive way to interpret the regression function. $g(a)$ is one minus the CDF of the derivative of the regression function. Although $g(a)$ could potentially be inferred from graphing $f(x)$ or $f'(x)$, understanding the uncertainty of $g(a)$ would be more easily done directly.

3. $\delta^{PMP\Delta} = \delta^{PMP} - \delta^{PMP^-} = Pr\{f'(X) > 0\} - Pr\{f'(X) < 0\}$.

   This parameter has attractive properties for estimation. See Section 3.3.2.

4. $h(a, b) = \Pr\{f(b \cdot X + a) - f(X) > 0\}$, for $a, b \in \mathbb{R}$.

   Estimating the function h might be interesting if the regression function is very curvy and we are interested in large changes of X, whose effect is not always well-approximated by looking at the derivative. For example, suppose we wanted to look at the effect of a policy that has a tax cut as well as a lump sum transfer. Note that h is defined for non-continuous f.

5. $Pr_X\left\{f'(X) > \hat{\beta}^{OLS}\right\}$.

   Part of the motivation for estimating the PMP is that it captures information that OLS does not. This parameter gives us an idea of whether the OLS estimate only applies to a small amount of the population.

6. $\alpha = \max\{a \in [0, 1] | \Pr\{f'(X) > a\} > 0.5\}$

   Half of the population experiences an increase in Y of at least $\Delta x \cdot \alpha$ for a small increase Δx of X. Similar to (5), it might be interesting to compare α to $\hat{\beta}^{OLS}$.

7. $\Omega = \{a \in \mathbb{R} | f'(x) > 0\}$

   Often we would like to know the set of values such that an increase in X is associated with an increase in Y.

3.3 Estimation

Many strategies could be used to estimate the PMP, due to the many non-parametric regression techniques available for estimating $f(x)$ as well as different ways of estimating standard errors. In this section I discuss a few estimation strategies. Any technique used to estimate the regression function could potentially be used to estimate the PMP, because the PMP is just a transformation of the regression function. A simple approach, which follows the intuition of many researchers who rely mainly on OLS, is to add polynomials to an OLS regression. Alternatively, using local polynomial estimation has been shown to have good properties (see Fan, 1992 and Fan and Gijbels, 1996). Below a few strategies are discussed for estimating the PMP using non-parametric methods. The estimator used in this paper for simulations is the plug-in estimator using local polynomial regression. Standard errors using the non-parametric percentile bootstrap are presented.

3.3.1 Discussion of analytical finite distribution of $\hat{\delta}_n^{PMP}$

The regression function, $f(x)$, can be estimated using local polynomial regression of degree k, by which at each x a linear regression is run as follows:

$$\hat{f}(x) = \hat{\beta}_0^x + \hat{\beta}_1^x \cdot x + \hat{\beta}_2^x \cdot x^2 + \ldots + \hat{\beta}_k^x \cdot x^k. \tag{3.2}$$

To estimate the pth derivative, it is commonly recommended to regress on a $p+1$ degree polynomial. Thus to estimate the PMP, which depends on the first derivative of f, a second degree polynomial is used. Taking the derivative of (3.2) for the case

of $k = 2$ yields estimates of the derivative constructed for each x as

$$\widehat{f'(x)} = \hat{\beta}_1^x + 2 \cdot \hat{\beta}_2^x \cdot x.$$

Because the coefficients are estimated using OLS, the vector $\left\langle \hat{\beta}_1^x, \hat{\beta}_2^x \right\rangle$ has a multivariate normal distribution asymptotically. The plug-in estimator

$$\hat{\delta}_n^{PMP} = \frac{1}{n} \sum_{i=1}^{n} 1 \left\{ \widehat{f'(X_i)} > 0 \right\}$$

then is an average of Bernoulli random variables. The distribution of $\hat{\delta}_n^{PMP}$ is non-trivial—the Bernoullis are dependent and not identically distributed; further, a weakly dependent central limit theorem cannot be immediately used because although the dependence does decrease towards zero[1] as $|i - j|$ increases, the sequence is not stationary.

for the finite distribution of $\hat{\delta}_n^{PMP}$ is not straight-forward, if the distribution of $\left\langle \hat{\beta}_1^{x_1}, \hat{\beta}_2^{x_1}, \ldots, \hat{\beta}_1^{x_n}, \hat{\beta}_2^{x_n} \right\rangle$ can be approximated by a multivariate normal distribution in large samples, simulation[2] can be used to approximate the distribution of $\hat{\delta}_n^{PMP}$.

3.3.2 Bootstrapping the finite distribution of $\hat{\delta}_n^{PMP}$

An alternative to relying on the asymptotic distribution and to simulation relying on asymptotic normality of $\left\langle \hat{\beta}_1^x, \hat{\beta}_2^x \right\rangle$, is to bootstrap the finite distribution of $\hat{\delta}^{PMP}$. To estimate the standard error of $\hat{\delta}_n^{PMP}$, a bootstrap is used. Hardle and Marron (1991) motivate the use of the bootstrap well in their work on simultaneous error bars for non-parametric regression:

[1] The speed of such convergence depends on the kernel, which is a further complication.

[2] Note that there are quick computer methods to simulate from a large normally distributed vector. See, e.g. Ripley (1987) and for an efficient modern implementation, see the R function `mvrnom` in the package `MASS` (Venables and Ripley, 2002).

Table 3.1: Simulations of Estimating the PMP (Bootstrapped SEs)

test	50	100	150	200	250	500	1000
$H_0\colon \delta_{sq}^{PMP} = 0.5$	0.00	0.00	0.03	0.02	0.05	0.06	0.06
$H_0\colon \delta_{sin}^{PMP} = 0.5$	0.00	0.00	0.03	0.05	0.03	0.04	0.03
$H_0\colon \delta_{m_1}^{PMP} = 0.5$	0.00	0.00	0.04	0.05	0.03	0.07	0.06
$H_0\colon \delta_{m_2}^{PMP} = 0.5$	0.00	0.01	0.05	0.09	0.07	0.02	0.06
$H_0\colon \delta_{m_3}^{PMP} = 0.5$	0.02	0.02	0.87	0.91	0.98	1.00	1.00
$H_0\colon \delta_{m_4}^{PMP} = 0.5$	0.02	0.02	0.93	0.98	1.00	1.00	1.00

This table shows simulations for inference of PMP using the plug-in estimator. For each simulation, a non-parametric percentile bootstrap confidence interval is estimated at the 95% level. Each cell contains the proportion of times the confidence interval does not contain the value 0.5. This is equivalent to the hypothesis test in column 1. For example, $\delta_{sq}^{PMP} = Pr\left\{f'\left(X\right) > 0\right\}$ in the model $Y = f\left(X_k\right) + \epsilon$ where the data for the simulations is generated such that $f\left(x_k\right) = \left(x_k - 0.5\right)^2$. For the shapes of the regression functions used and the true value of each PMP, see Figure 3.1. Proportions are based on 100 simulations.

> One approach to the problem of finding simultaneous error bars would be to work with limiting normal distributions of the estimator at the grid points. However, the joint distribution of the estimator at these gridpoints has substantial positive correlation, which makes the derivation of joint normal theory confidence intervals nontrivial. In fact, they essentially should be done by simulation methods. Since simulation methods are needed anyway, it seems better to use a more direct approach through bootstrapping, as opposed to relying on the normal approximation and also to facing the problems of parameter estimation that such an approach entails.

Table 3.1 shows simulations using the regression functions graphed in Figure 3.1, for inference on PMP using the plug-in estimator and the non-parametric percentile (not bias-corrected) bootstrap. The percentile's range-preserving feature is especially important here, because $\delta^{PMP} \in [0, 1]$. For each simulation, a confidence interval is

estimated at the 95% level. Each cell contains the proportion of times the confidence interval does not contain the value 0.5. If the estimated confidence interval does not contain 0.5, then the hypothesis in column 1 of the table is rejected. For example, δ_{sq}^{PMP} is equal to $Pr\left\{f'\left(X\right)>0\right\}$ in the model $Y=f\left(X_k\right)+\epsilon$ where the data for the simulations is generated such that $f\left(x_k\right)=\left(x_k-0.5\right)^2$. 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 is generated as $X_f=f\left(U_f\right)$, where $U_f\sim unif\left(0,1\right)$.

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

For the regression functions sq, sin, and m_2, the true PMP is 0.5, and thus the null hypothesis is true. Table 3.1 suggests that the size of the implemented test is conservative for small sample sizes and converges to 0.05 as the number of observations grows. For the regression functions m_3 and m_4 the true PMP is not 0.5 and thus the null hypothesis is false. The power of the test in these cases is low for small sample sizes and quickly jumps at 150 observations and increases towards 1 as the number of observations grows.

Table 3.2: Simulations of Estimating the PMP (Bootstrapped SEs)

test	125	250	500	1000	2000	4000
$H_0\colon \delta_{sq}^{PMP} = 0.7$	0.28	0.33	0.41	0.44	0.76	0.90
$H_0\colon \delta_{sin}^{PMP} = 0.7$	0.59	0.84	0.99	1.00	1.00	1.00
$H_0\colon \delta_{m_1}^{PMP} = 0.7$	0.20	0.20	0.13	0.18	0.16	0.16
$H_0\colon \delta_{m_2}^{PMP} = 0.7$	0.22	0.18	0.30	0.43	0.68	0.92
$H_0\colon \delta_{m_3}^{PMP} = 0.7$	0.03	0.13	0.29	0.43	0.68	0.92
$H_0\colon \delta_{m_4}^{PMP} = 0.7$	0.09	0.22	0.35	0.66	0.94	1.00

This table shows simulations for inference of PMP using the plug-in estimator. For each simulation, a non-parametric percentile bootstrap confidence interval is estimated at the 95% level. Each cell contains the proportion of times the confidence interval does not contain the value 0.7. This is equivalent to the hypothesis test in column 1. For example, $\delta_{sq}^{PMP} = Pr\left\{f'\left(X\right) > 0\right\}$ in the model $Y = f\left(X_k\right) + \epsilon$ where the data for the simulations is generated such that $f\left(x_k\right) = \left(x_k - 0.5\right)^2$. For the shapes of the regression functions used and the true value of each PMP, see Figure 3.1. Proportions are based on 100 simulations.

The test fails to reject the hypothesis for m_1, even though as shown in Figure 3.1, the true PMP is 0. The reason for this is that for any given estimation of the regression function, $\hat{f}'\left(x\right)$ will never be exactly to zero. If $f'\left(x\right) = 0$, then because of ϵ we would expect about half of the estimates to be slightly above zero and half to be slightly below zero, and this is indeed what the simulations reflect. In fact, for the test that is implemented, the results suggest that

$$\begin{aligned} \hat{\delta}^{PMP} &\xrightarrow{p} Pr\left\{f'\left(X\right) > 0\right\} + \frac{1}{2} \cdot Pr\left\{f'\left(X\right) = 0\right\} \\ &= \delta^{PMP} + \frac{1}{2} \cdot Pr\left\{f'\left(X\right) = 0\right\}. \end{aligned}$$

This issue could be addressed by taking one of the following approaches:

1. Look at the sets $\left\{a|\hat{f}'(a) > 0\right\}$ across the bootstrapped simulations. The sets probably change a lot in the case of m_1.

2. Estimate the bias, $\frac{1}{2} \cdot Pr\left\{f'\left(X\right) = 0\right\}$, using a different method.

3. Inspect the graph of $\hat{f}'(x)$ manually to look for parts where the derivative fluctuates around zero.

4. Calculate the set $\hat{Z} = \#\{a|\hat{f}'(a) = 0\}$, where $\#$ is the counting measure (the number of elements in the set). If $\hat{Z}$ is large, then there is evidence that the set $Z = \{a|f'(a) = 0\}$ is also large. Similar to (1), $\hat{Z}$ should not change much across the bootstrapped simulations.

Another approach is to define

$$\delta^{PMP^-} = Pr\left\{f'\left(X\right) < 0\right\}$$

and estimate the parameter

$$\delta^{PMP\Delta} = \delta^{PMP} - \delta^{PMP^-} = Pr\left\{f'\left(X\right) > 0\right\} - Pr\left\{f'\left(X\right) < 0\right\}.$$

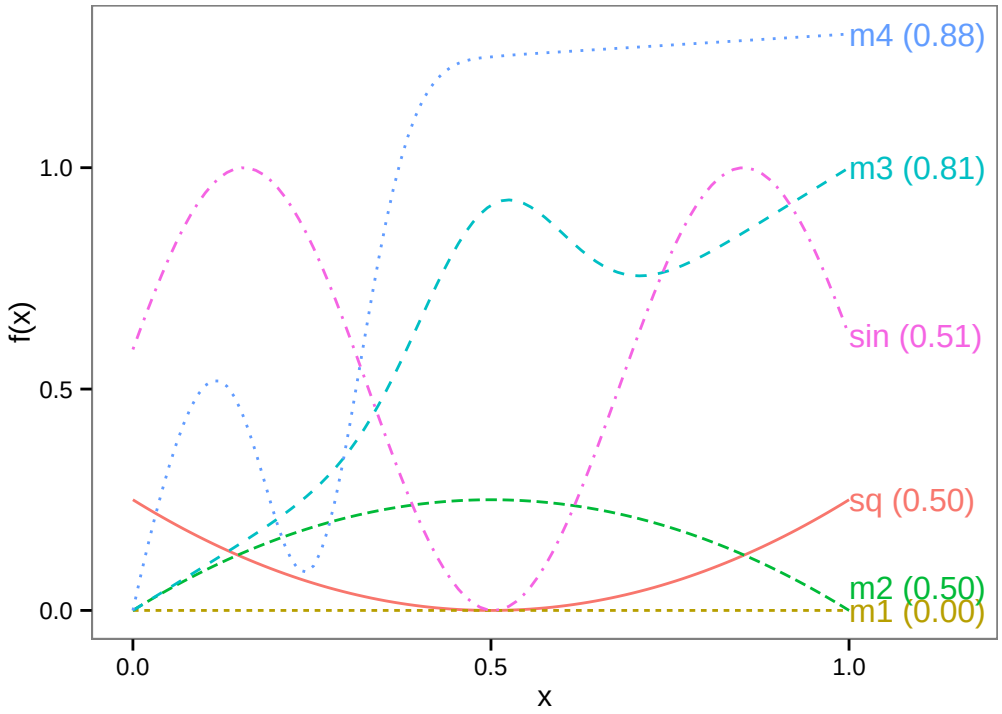
The estimation of this parameter would be consistent because the biases would cancel out:

$$\begin{aligned}
\hat{\delta}^{PMP\Delta} &= \hat{\delta}^{PMP} - \hat{\delta}^{PMP^-} \\
&\xrightarrow{p} Pr\left\{f'\left(X\right) > 0\right\} + \frac{1}{2} \cdot Pr\left\{f'\left(X\right) = 0\right\} - \\
&\quad Pr\left\{f'\left(X\right) < 0\right\} - \frac{1}{2} \cdot Pr\left\{f'\left(X\right) = 0\right\} \\
&= \delta^{PMP} - \delta^{PMP^-}.
\end{aligned}$$

Note that computation of $\hat{\delta}^{PMP^-}$ is easy because assuming the common scenario that $\hat{f}'(x_n) = 0$ for all n,

$$\hat{\delta}^{PMP^-} = 1 - \hat{\delta}^{PMP}.$$

Figure 3.1: True Regression Functions (with PMP)



These functions are used in simulations throughout this paper. The PMP corresponding to a uniformly distributed X is displayed in parentheses after each function's name. For the function definitions, see Section 3.3.2.

3.3.3 A critical bandwidth test of δ^{PMP}

A separate strategy from the plug-in estimator for hypothesis tests of δ^{PMP} is to use the concept of a critical bandwidth, first used by Silverman (1981). To test unimodality, Silverman used as a test statistic the smallest bandwidth such that a density estimate was unimodal. Similarly, in order to test

$$H_0\colon \delta^{PMP} \geq a$$

versus

$$H_A\colon \delta^{PMP} < a$$

the smallest bandwidth, h, such that $\hat{\delta}_h^{PMP} \geq a$ can be used as a test statistic. The intuition is that for a very small h, there will be large variance and thus $\hat{\delta}_h^{PMP}$ will be close to 0.5. As the bandwidth increases, $\hat{\delta}_h^{PMP} \to b \in \{0, 1\}$ because in the limit, a local polynomial regression yields an estimate that is a straight line. Under the null hypothesis, it will not take a lot of smoothing (i.e. h is "small") before $\hat{\delta}_h^{PMP} \geq a$ because it is just ϵ that needs to be smoothed away. Under the alternative, more smoothing (i.e. a "larger" h) is needed because smoothing ϵ is not enough to reach a; underlying features in the regression function also need to be smoothed away. This test will have better properties the further a is from 0.5, because otherwise the test statistic will be small under the null and alternative. An advantage of this testing strategy over the plug-in estimator is that the critical bandwidth test statistic does not depend on a method for choosing a bandwidth because the test statistic *is* the bandwidth.

A disadvantage of using a critical bandwidth approach is that it is inherently suited for hypothesis testing and not for creating confidence intervals. Because δ^{PMP} is a parameter whose value is interpretable, point estimates and corresponding confidence intervals are perhaps the most interesting targets of inference. We can work around

this disadvantage by constructing a confidence set at level α as

$$CI = \left\{a | \text{fail to reject } H_0 \colon \delta^{PMP} \geq a \text{ at the } \alpha \text{ level}\right\}$$

Although this seems computationally intensive, the process could be easily parallelized.

3.4 Multivariate Hypotheses

Extending the concept of PMP to multivariate hypotheses is straightforward with partial linear models (PLMs) (Robinson, 1988) and generalized additive models (GAMs) (Hastie and Tibshirani, 1986). PLMs are a natural extension, especially when PMP is used as a complement to OLS. A PLM restricts some variables as having a linear relationship with Y, and others to have an unknown additive relationship, such as

$$Y = f(X) + \boldsymbol{Z}\beta + \epsilon,$$

where f is left unspecified. Estimating the PMP of X, conditional on $\boldsymbol{Z}$ is straightforward. Moving on the semi-parametric spectrum further towards non-parametrics, GAMs allow each variable to have an unspecified but still additive functional form, although they enter the relationship additively, such as

$$Y = f(X) + \sum_j g_j(Z_j) + \epsilon.$$

When estimating a PMP in a GAM, one must remember that a difference when comparing with OLS estimates could come from a variety of sources. For example, suppose that $\hat{\beta}_X^{OLS}$ is large and positive and $\hat{\delta}_X^{PMP}$ is less than 0.5, and let us suppose that X is distributed uniformly to abstract away from marginal density issues. One

story consistent with this situation is that f is very large for a small part of the x-axis, and negative but small in absolute value for a large part of the X-axis. However, a different story consistent with the same situation, is that the linear specification used for OLS is restrictive and as a result the unrestricted f in the PLM picks up some of the effect of variables in $\boldsymbol{Z}$ on Y that could not be captured by their own coefficients because they have a relationship that is not captured by the functional form specified in $\boldsymbol{Z}\beta$.

Using GAMs provides a less biased $f(x)$ because the control variables are given as many degrees of freedom as the variable of interest. Thus, there is not much concern of f picking up one of the effects of Z because the g_j will flexibly pick up those effects.

3.5 An Application in Subjective Well-Being

Using data from Chapter 2 on the relationship between subjective well-being (SWB) and age, Table 3.3 shows how PMP can be combined with OLS to provide a detailed picture of the relationship between variables. The first two columns in the table summarize the unconditional relationship between life satisfaction and age. The OLS coefficient, $\hat{\beta}^{OLS}$, suggests a significant but small negative effect of age on life satisfaction. Life satisfaction ranges from 1 to 10, with a standard deviation of 2.4, so an increase of age of 100 years would be needed to decrease life satisfaction by a standard deviation. The $\hat{\delta}^{PMP}$ column shows that age is positively associated with life satisfaction for about 33% of the population. In most situations, when $\hat{\beta}^{OLS}$ is negative, $\hat{\delta}^{PMP}$ will be less than 50%, as is the case here.

The $\hat{\delta}^{PMP}$ magnitude of 33% suggests that there is a non-monotone relationship between life satisfaction and age. Adding polynomials of age to the OLS regression would likely notably increase the R-squared. However, instead of just trying to get a better fit, a natural approach is to investigate the underlying drivers of

	$\hat{\beta}^{OLS}$	$\hat{\delta}^{PMP}$	$\hat{\beta}^{OLS}$	$\hat{\delta}^{PMP}$
age	−0.0022	32.7%	−0.0028	4.2%
	(0.0003)		(0.0002)	
financial sat.			0.52	100%
			(0.00)	

Table 3.3: OLS and PMP Estimation of SWB on Age and Financial Satisfaction
This table shows results from OLS in the $\hat{\beta}^{OLS}$ columns and from PMP in the $\hat{\delta}^{PMP}$ columns. The dependent variable is a measure of life satisfaction. Data are from the World Values Survey and the European Values Survey.

the non-monotonicity. Estimating the PMP is one approach to both finding the underlying drivers and reporting them in an easy way. In the table, the second pair of $\left(\hat{\beta}^{OLS}, \hat{\delta}^{PMP}\right)$ in the "age" row show what happens to the relationship between life satisfaction and age when financial satisfaction is conditioned on. The OLS coefficient on age becomes slightly more negative, and PMP estimate decreases to 4.2% suggesting that the relationship is now mostly monotonically decreasing. It appears that financial satisfaction can thus explain the non-montone relationship between life satisfaction and age.

A graphical analysis of the above statements confirms the analysis based on $\left(\hat{\beta}^{OLS}, \hat{\delta}^{PMP}\right)$. Figure 3.2 plots the estimated unconditional relationship between life satisfaction and age (the red line) as well as the estimated relationship conditional on financial satisfaction (the blue line). The red line shows a U-shape, but after conditioning on financial satisfaction, the previous increase in later years bends down and the relationship is mostly monotonically decreasing.

A difference to keep in mind when comparing Figure 3.2 to Table 3.3 is that the PMP estimates are with respect to the marginal distribution of age. Although similar to the PMP, the figure gives a better idea of $Pr_U\left\{f'(X) > 0\right\}$, which is discussed in Section 3.2.6. The skewness of the distribution of age towards younger ages (the median is 39) thus explains why in the table $\hat{\delta}^{PMP}$ is 33% but in the figure it seems that the red line is increasing more than just 33% of the x-axis.

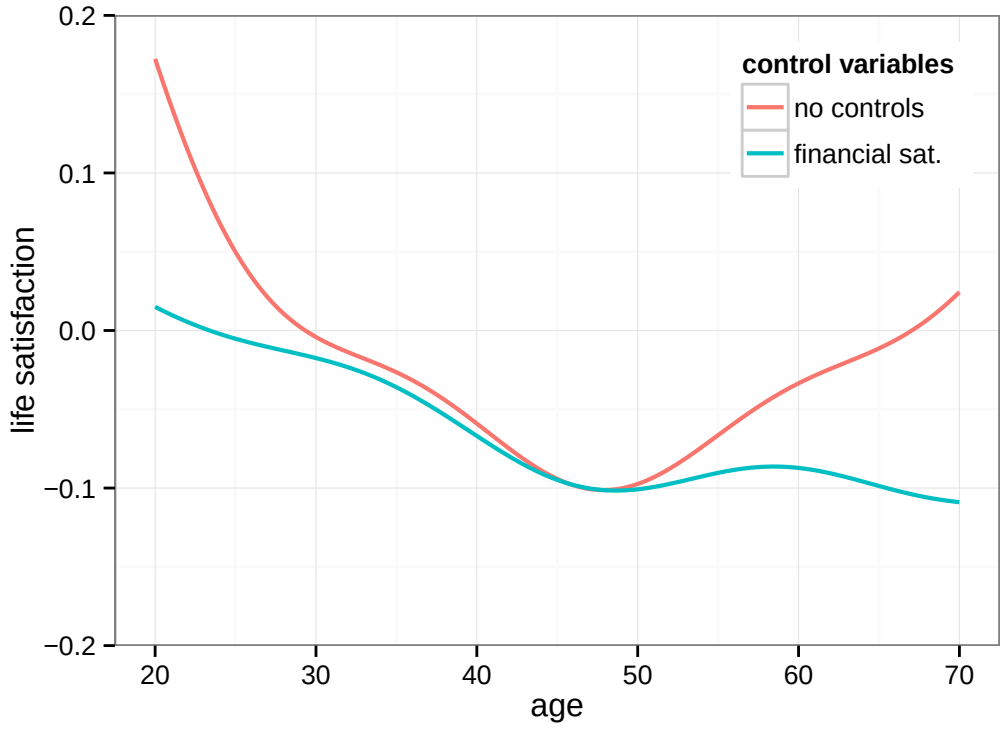


Figure 3.2: Graphical Depiction of Non-Monotone Relationship

3.6 Conclusion

By adding PMP estimation to the toolbox, researchers have a tool to check for non-monotone effects. Inference on PMP can be useful as a descriptive summary of non-monotonicity, as a test for monotonicity without an all-in approach, and can in some sense be used to check the model and specification (e.g. if polynomials should be added). The intuitive interpretation of PMP makes it a useful statistic to present to policy makers and the general audience. Estimating the PMP with non-parametric and semi-parametric methods is straightforward with plug-in estimator. Extending the PMP to a multivariate context is simple, allowing for practical applications.

References

Fan, J.: 1992, Design-adaptive nonparametric regression, *Journal of the American statistical Association* **87**(420), 998–1004.

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

Hardle, W. and Marron, J.: 1991, Bootstrap simultaneous error bars for nonparametric regression, *The Annals of Statistics* pp. 778–796.

Härdle, W. and Stoker, T. M.: 1989, Investigating smooth multiple regression by the method of average derivatives, *Journal of the American Statistical Association* **84**(408), 986–995.

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

Powell, J. L., Stock, J. H. and Stoker, T. M.: 1989, Semiparametric estimation of index coefficients, *Econometrica: Journal of the Econometric Society* pp. 1403–1430.

Ripley, B. D.: 1987, Stochastic simulation.

Robinson, P. M.: 1988, Root-n-consistent semiparametric regression, *Econometrica: Journal of the Econometric Society* pp. 931–954.

Silverman, B. W.: 1981, Using kernel density estimates to investigate multimodality, *Journal of the Royal Statistical Society* **43**, 97–99.

Soyer, E. and Hogarth, R. M.: 2012, The illusion of predictability: How regression statistics mislead experts, *International Journal of Forecasting* **28**(3), 695–711.

Venables, W. N. and Ripley, B. D.: 2002, *Modern Applied Statistics with S*, fourth edn, Springer, New York. ISBN 0-387-95457-0.

URL: *http://www.stats.ox.ac.uk/pub/MASS4*