Penultimate version of a paper accepted in the Journal of Mathematical Psychology

On the use of Nonparametric Regression in Assessing Parametric Regression Models

Scott Brown and Andrew Heathcote

Department of Psychology,

The University of Newcastle, Australia

14/12/1: Please do not cite without permission

Penultimate version of a paper accepted in the Journal of Mathematical Psychology

Address for correspondence:

Scott Brown

School of Behavioural Science, Building W

The University of Newcastle, Callaghan, 2308, NSW, Australia

Email: scott@baal.newcastle.edu.au

Abstract

We develop a new method for assessing the adequacy of a smooth regression function, based on nonparametric regression and the bootstrap. Our methodology allows users to detect systematic misfit and to test hypotheses of the form "the proposed smooth regression model is not significantly different from the smooth regression model that generated these data". We also provide confidence bands on the location of nonparametric regression estimates assuming that the proposed regression function is true, allowing users to pinpoint regions of misfit. We illustrate the application of the new method, using local linear nonparametric regression, both where an error model is assumed, and where the error model is an unknown non-stationary function of the predictor.

Much research in quantitative psychology involves the assessment of the adequacy of a proposed regression model – where a response variable is expressed as the sum of an expectation function (we will call this the “regression function”) and an error distribution. We will focus our attention on an important subset of regression functions: smooth functions, defined as those having a certain number of continuous derivatives (i.e., the set $\underline{C}^k$, usually with $\underline{k}$ between one and three). Not all proposed regression functions are smooth, but the set is broad enough to encompass the majority of the regression functions used in quantitative psychology.

Nonparametric regression provides a method of estimating the best possible smooth regression function for a given data set (Wand & Jones, 1995). Using nonparametric regression avoids problematic assumptions on the nature of the true regression function, assuming only that it is smooth. In keeping with the nonparametric theme, we employ the bootstrap to model variability in the function estimates. Bootstrapping has been used to build both pointwise (Knafl, Sacks, & Ylvisaker, 1985; Azzalini, Bowman & Hardle, 1989) and simultaneous (Markus & Visser, 1990) confidence bands for non-parametric regression with the aim of testing proposed regression functions by comparison against nonparametric confidence bands generated from the data. Unfortunately, confidence bands for nonparametric regression estimates are plagued by the problem of bias: the nonparametric regression function is in general a biased estimate of the true regression function (see Fan & Gijbels, 1996) and so a bootstrap that adds error samples to the (biased) nonparametric regression estimate will, in general, be unsuitable (Davison & Hinkley, 1997). Several techniques have been proposed to overcome this problem. Some techniques attempt to quantify the bias in the nonparametric regression estimate and compensate accordingly. However, bias quantification can create yet another layer of estimation problems. Perhaps the most promising work so far has employed under-smoothed nonparametric regression to estimate the regression function (thus reducing the magnitude of the bias) followed by the use of over-smoothed nonparametric regression to estimate the error distribution (Davison & Hinkley, 1997; Hardle & Marron, 1991; Hardle & Nussbaum, 1990).

We do not attempt to solve the problem of creating confidence bands on a nonparametric regression estimate in this paper; instead we propose a method of "sidestepping" this difficult problem. The problems that have beset model testing using nonparametric regression stem from an aim common to all the previously proposed techniques, the construction of a confidence region for the "best smooth regression function", using nonparametric regression to estimate this function. The use of a nonparametric regression estimate as the basis for subsequent bootstrap calculations ensures that bias will always be a problem for these methods. In contrast, the method of regression model testing that we develop here avoids this problem by instead creating confidence bands for the location of the nonparametric regression estimate, under the null hypothesis that the proposed regression model generated the observed data. The advantage of our method is tractability – it can produce genuine confidence regions, rather than the "variability bands" often seen in this area. The disadvantage is that each confidence region is model-specific. Testing different parametric regression models requires re-calculation of the confidence region for each one. This is not too burdensome, since the calculations required are both relatively simple and computationally inexpensive.

The method of testing we develop below assumes the existence of an estimate of the proposed regression function to be checked against a set of observed data. The estimated function will usually be a member of a parametric family of smooth regression functions, with parameters fixed by fitting the data under some optimality criterion. This does not have to be the case, however: the parameters need not be optimally estimated in any usual sense, and the parametric family need not be well defined. The estimated function need only specify a value for each design point.

Our technique also requires that the estimated regression function have a corresponding estimated error distribution, which again will usually take the form of a parametric distribution with estimated parameters. However we recognise that researchers are often interested in testing the form of the regression function alone, keeping assumptions about the nature of the error distribution to a minimum. Hence, we also provide an alternative method based on the Wild bootstrap (Rosenblueth, 1975; Wu, 1986) that matches the moments of the observed error distribution around the estimated regression function at each design point. This method requires fewer assumptions about the parametric form of the error distribution and can accommodate nonstationarity in the moments of the error distribution.

In summary, our methodology provides two ways of assessing the adequacy of a smooth regression function. It allows the user to test null hypotheses of very general form, such as “the proposed regression model is not significantly different from the best possible smooth regression model”. It also provides confidence bands on the location of a nonparametric regression estimate under the null hypothesis that the proposed regression model generated the data, thus allowing researchers to assess precisely how the proposed regression function misfits the data. We recommend that researchers employ the confidence bands as the primary product of this method, and use the hypothesis tests only as a guide, since formal hypothesis tests are not always well suited to the noisy data psychologists often use.

In the following sections, we outline the nature and some properties of our methodology, and then provide a simulated example of its use when a parametric error model is known. Following this, we show how the Wild bootstrap may be used when a parametric error model is not assumed. Finally, we present an example analysis of an empirical psychological data set, in which our new methodology coupled with the Wild bootstrap provides a conclusion that is opposite to that of a more standard model selection technique.

The Smooth Function Test

In the following, we refer to “differences” between models, as measured by a discrepancy metric. However, we do not wish to make a commitment to any particular discrepancy metric – the user may select whichever is most suitable. The only condition imposed is that the discrepancy metric must be suitable for calculating the difference between parametric and nonparametric regression estimates. Likelihood ratios will be used in our first example and root-mean-squared (RMS) residual size in the second example. If the user wishes to account for model complexity, for example, the Aikike Information Criterion may be useful, although this requires the estimation of approximate degrees of freedom for the nonparametric estimate (see Bowman & Azzalini, 1997).

Figure 1 schematically represents two sets of data (filled shapes) and four regression functions[1] (unfilled shapes). The two circles at the bottom represent the regression function for the data-generating model (DGM) and the regression function for the proposed regression model. The observed data, which are a sample from the DGM, are represented by the filled square. The unfilled square represents a nonparametric estimate of the regression function of the DGM, calculated from the observed data. The difference between this estimate and the actual regression function (dotted line) is a smoothing bias: a difference introduced by the imperfect nature of the sample-then-smooth method of estimating the regression function of the DGM. The validity of using nonparametric regression to estimate the regression function of the DGM from the observed data depends on our assumption of smoothness.

The researcher is interested in testing the hypothesis that the proposed regression model is not significantly different from the DGM; an assumption that implies that the distance marked A is not significantly greater than the null distance (not always zero, depending on the discrepancy metric employed). Direct assessment of the distance A is not possible because the regression function for the DGM is not observed. Instead, we use our nonparametric estimate of this function (the unfilled square in Figure 1) and estimate the distance A by the distance from this estimate to the proposed parametric regression function (the distance A' in Figure 1).

Even if the discrepancy metric has analytic expressions available for its distribution under certain conditions, these will almost certainly not apply to the distance A' because of the bias induced by the nonparametric regression. For example, the usual chi-squared distribution for likelihood ratios will not apply because the nonparametric estimate of the best smooth function is not a maximum likelihood estimate (and even the estimated parametric function need not be a maximum likelihood estimate). To overcome these difficulties we empirically estimate the distribution of the distance A' under the null hypothesis that the distance A is the null distance, using the bootstrap. Thus, we repeatedly resample a set of data from the proposed regression model (the filled triangle in Figure 1), and apply our nonparametric regression technique to the resampled data, taking care to use the same smoothing parameters as used previously. This operation results in a nonparametric estimate of the proposed regression function (the unfilled triangle in Figure 1). As before, this estimate is, in general, imperfect due to smoothing bias (this bias is marked B' in Figure 1). Under the null hypothesis that the difference A is null, the distances A' and B' will be drawn from the same distribution. Hence, a significance test of the proposed regression model can be constructed by noting that the probability of observing a distance the size of A' will simply be given by its rank within the set of distances B'.

Confidence bands for the expected location of a nonparametric regression estimate, under the null hypothesis that the proposed regression model is correct, can be calculated using the same set of bootstrap samples employed by the significance test. Pointwise (100-2α)% confidence bands may be calculated in the naïve manner: by storing each of the nonparametric regression estimates generated from the resampled data sets and calculating the empirical α-th and (100-α)-th quantiles at each point. The estimation of simultaneous confidence bands using the bootstrap is more complex; in the interests of brevity, we simply refer interested reader to Markus and Visser (1990) and to Hardle and Marron (1991).

Asymptotic Unbiasedness

The asymptotic bias of the method outlined above is determined by the asymptotic bias of the discrepancy metric the user has chosen, as long as the distances marked with dotted lines in Figure 1 are asymptotically null. That is, the significance tests and confidence bands generated using the methodology above will be asymptotically unbiased whenever the model discrepancy metric employed is asymptotically unbiased, as long as the smoothing biases (dotted lines in Figure 1) are asymptotically[2] zero.

The important caveat that the smoothing biases are asymptotically zero does not hold for all nonparametric estimators, depending strongly on how the smoothing parameter is chosen. However, asymptotic unbiasedness does hold for many readily available nonparametric estimators. For instance, if local polynomial regression is employed, the biases will be asymptotically zero whenever the bandwidth selection procedure ensures that bandwidths are asymptotically zero. Many of the commonly used bandwidth selection algorithms (e.g. cross-validation, biased cross-validation, direct plug-in methods, rule-of-thumb calculator) will satisfy this condition (Fan & Gijbels, 1996). Note that “eyeball” smoothing parameter selection, as is commonly employed for smaller data sets and one-off calculations, cannot guarantee asymptotic unbiasedness. This is because it is not guaranteed that such methods result in asymptotically zero smoothing bandwidths, even though this may be probable.

Non-asymptotic properties, such as convergence rates, are not bound by the above argument. For instance, the rate of convergence of a discrepancy metric to its asymptotic value will depend upon both the usual rate of convergence for that statistic and the rate of convergence of the nonparametric regression estimate to its asymptotic value. In particular, the actual rate of convergence can be no better than the worst of these two contributing rates.

Example 1: Known Parametric Error Model

Suppose we have a DGM consisting of a regression function $y = e^{-5x^2} \sin(2px)$, where $\underline{x} \in [-1,1]$, and additive normal noise with a standard deviation of 0.25; we will call this model $\underline{T}$ (for true) and its regression function $\underline{T}_R$. A set of data generated by this model is shown in Figure 2 (unfilled triangles) along with the regression function of the DGM ($\underline{T}_R$, dotted line). Also shown in Figure 2 is a nonparametric estimate of the regression function (solid line) that was calculated using local linear regression with an Epanechnikov kernel and a bandwidth of 0.21 (calculated using the “direct plug-in” method). Note that, as expected, the nonparametric estimate exhibits large bias where the data exhibit large curvature, around x=-0.2 and x=0.2. See Wand and Jones (1995) for details of local linear regression (p.114) and the Epanechnikov kernel (p.30).

Suppose that we have somehow hit upon the DGM ($\underline{T}$) and that we wish to test its adequacy using the methodology developed above. We begin by calculating the discrepancy between $\underline{T}_R$ and the nonparametric estimate obtained from the observed data; using log-likelihood-ratio for the discrepancy metric, this turns out to be A’ = -5.18. An empirical approximation to the sampling distribution for this statistic is generated by the bootstrap, under the null hypothesis that $\underline{T}$ is the DGM: that is, we repeatedly sample data from $\underline{T}$, smooth those data, then find the discrepancy between that smooth and $\underline{T}_R$. The observed value turns out to be well within the bootstrapped distribution ($\underline{p}$>.1) and so we cannot reject the hypothesis that $\underline{T}$ is the best smooth model. In this example, and those below, we used 5000 bootstrap samples to ensure very small sample variability in distribution function and confidence band estimates. For the task of estimating tail probabilities and pointwise confidence bands, as few as 500 samples typically suffice.

To further confirm the adequacy of the proposed function, we can calculate confidence bands on the location of the nonparametric regression estimate under the assumption that the data were generated by model $\underline{T}$. Pointwise 95% confidence bands for $\underline{T}_R$ are shown in Figure 3 (dotted lines). Note that the nonparametric estimate of the best smooth model calculated from the observed data lies everywhere within the confidence bands, confirming the proposed regression function’s adequacy.

Some readers may be concerned that the regression function of the DGM ($\underline{T}_R$) does not always lie within the confidence bands. This is to be expected because of smoothing bias. The confidence bands fix the location of the *smoothed estimate* of the regression function, which is different from the "true" function because of smoothing bias.

Now suppose that we wish to test a competing regression model, consisting of the regression function $y = \frac{160x}{\Gamma(4.5(|x|+1))}$, and the same additive $\underline{N}(0,0.25)$ error structure as before; call this model $\underline{F}$ (for false) and its regression function $\underline{F}_R$. Figure 4 shows the observed data along with the regression function of the DGM ($\underline{T}_R$, dotted line) and the new proposed regression function ($\underline{F}_R$, solid line). Note that $\underline{F}_R$ provides a good approximation to the data – without prior knowledge it would be difficult to tell whether $\underline{F}_R$ provided an adequate description of these data.

Using the same settings as before (i.e., the same nonparametric regression algorithm, bandwidth and bootstrap sample size) we test the adequacy of $\underline{F}$ by testing the null hypothesis that it is not significantly worse than DGM (assuming that is smooth), and also by constructing pointwise confidence bands on the expected location of the nonparametric estimate under $\underline{F}$. The log-likelihood-ratio of $\underline{F}_R$ is $\underline{A}'$ = -15.3. Bootstrap calculations as before show that the probability of observing a likelihood ratio as small as or smaller than this value under the null hypothesis is only $\underline{p}$=.044: it seems that perhaps $\underline{F}$ is not as good as the best smooth model.

Confidence bands can provide us with more information on the adequacy of regression model $\underline{F}$. The confidence bands on the position of the nonparametric estimate under the (false) assumption that $\underline{F}$ is the DGM appear in Figure 5 (dotted lines) along with the nonparametric estimate of the best smooth regression function generated from the observed data (solid line). Notice that the best smooth model estimate strays outside the confidence bands at two points (around $\underline{x}$ = -0.5 and $\underline{x}$ = 0.7). These deviations show that $\underline{F}_R$ underestimates the best smooth model around $\underline{x}$ = -0.5 and overestimates it around $\underline{x}$ = 0.7. This is precisely what we could deduce if we knew the difference between $\underline{T}_R$ and $\underline{F}_R$ (see Figure 4).

Estimating an Error Model with the Wild Bootstrap

It is sometimes necessary to test the adequacy of a regression function without prior knowledge of an associated parametric error distribution. For this purpose we propose the use of a semi-parametric resampling scheme, the Wild bootstrap. This method is similar to that used by Hardle and Marron (1991) for the construction of multivariate simultaneous confidence bands for non-parametric regression models, but avoids the problem of nonparametric regression bias, as in the above method.

The Wild bootstrap aims to independently estimate the distributions from which each of the observed residuals were drawn, thus respecting distributional differences across the design points, such as heteroscedacity. The Wild bootstrap assumes some general form for the distribution of each residual, such as an n-point Dirac function (other distributions may be more useful, particularly those incorporating autocorrelation, but this one is both simple and tractable; see Wu, 1986).

For example, suppose we have a set of residuals $\{\varepsilon_i: i=1\ldots\underline{N}\}$, and assume that these are realisations from two-point Dirac distributions: $\varepsilon_i \sim \rho_i\delta_{ai} + (1-\rho_i)\delta_{bi}$. Here, δ_{ai} and δ_{bi} represent Dirac functions with mass at $\underline{a}_i$ and $\underline{b}_i$ respectively, and ρ_i is a mixing probability. Next, assume that the expected moments at design point $\underline{i}$, given the observed residual ε_i, are $E(\eta_i) = 0$, $E(\eta_i^2) = \varepsilon_i^2$, $E(\eta_i^3) = \varepsilon_i^3$, and so on. Other assumptions about the expected moments are of course possible, but this one has some desirable properties, such as ensuring a zero mean, and (almost) matching the variance of most parent distributions (see below). By matching the first three moments of the two-point distribution to these estimates we can fix the three parameters of the resampling distributions independently for each design point. For the i-th residual, this means the parameters of the two-point Dirac distribution given above are:

$$r_i = \frac{5+\sqrt{5}}{10} \qquad a_i = \frac{1-\sqrt{5}}{2} e_i \qquad b_i = \frac{1+\sqrt{5}}{2} e_i$$

Several improvements may be made to the basic Wild resampling scheme. Most importantly, Wild resampling typically underestimates variance of the parent distribution. Davison and Hinkley (1997) show that this defect can be remedied by replacing the observed residuals with "leveraged" residuals: $\varepsilon_i/(1-h_i)$, where h_i is the leverage for the i-th residual, estimated from the parametric model along with the raw residual. We use these leveraged residuals in the example below.

An n-point Dirac distribution is limited in several ways: most importantly, it has support on only a finite set of points. This means that likelihood cannot, in general, be calculated for nonparametric regression estimates. Researchers who wish to employ likelihood-based discrepancy metrics can do so by employing a different distribution function. For example, a mixture of two normal distributions will provide support on the entire real line, but the parameters of these distributions will need to be fixed by matching more than three expected moments, so may be subject to greater sampling variability.

Example 2: Estimated Error Model

Recently there has been some debate about the mathematical form of the law of practice: that is, the regression function that relates response latency to practice level in speeded responding tasks (Heathcote, Brown & Mewhort, 2000). In this application, researchers are primarily interested in the form of the regression function without reference to the error distribution, so we use the Wild bootstrap. By illustrating our methodology in application to two competing smooth regression functions – the exponential and power functions – we also demonstrate that our methodology can clarify model selection results.

Figure 6 presents a single subject's data[3] for a practice experiment (triangles), along with the best-fitting (least squares) exponential regression function (dotted line) and a nonparametric estimate of the best smooth regression function (solid line). The ordinate represents response time (RT, in milliseconds) and the abscissa represents number of practice trials (N). We will use RMS as the model discrepancy metric, to avoid problems with calculating likelihood under the two-point Wild bootstrap. The discrepancy between the proposed exponential function and the non-parametric estimate is A' = 485ms.

Again, resampling was used to build an empirical estimate of the sampling distribution of A', under the assumption that the estimated exponential function and Wild error distribution generated the data. The observed value was small compared with those expected under the null hypothesis: the probability of observing a value larger than that observed was 0.955. Thus, we have no reason to doubt the null hypothesis that the proposed exponential function is not significantly different from the best smooth model of the regression function.

An alternative smooth regression function for the practice data is the best-fitting three-parameter power function. On visual inspection, this function (dotted line in Figure 7) seems to provide a good fit to the data. Indeed, the discrepancy for the power function suggests it provides a better description of the data: its RMS is approximately 2% better than the RMS for the exponential function. Note however, that the asymptotic expected response time for the power function is zero, an implausible value[4].

Figure 8 shows pointwise 95% confidence bands on the location of a nonparametric regression estimate, under the assumption that the estimated power function and Wild error distribution generated the data. The confidence bands show that the power function is not a good smooth regression function for these data: it underestimates the data at around $N=15$ and overestimates the data from about $N=40$ onwards. Hypothesis testing provides further support for this conclusion: the probability of observing a discrepancy as large as that between the power function and the nonparametric regression function is only $p=.047$ (recall that the corresponding statistic for the exponential function was $p=.955$).

In this example, our methodology has detected a mis-fit due to a systematic pattern in the residuals as a function of the predictor, rather than simply in terms of a single goodness of fit measure. The reason for the contrasting results is the use of nonparametric regression, rather than simply raw data, to estimate the best smooth model. The power function lies closer to the raw data than does the exponential function: RMS for the difference between the power function and the raw data is 3304ms, compared to 3340ms for the exponential. However the exponential function lies very much closer to the nonparametric regression estimate of the best smooth model than does the power function: RMS for the difference between the power function and the nonparametric estimate of the best smooth model is 1839ms, compared to only 485ms for the exponential function. These findings, and the implausible asymptote estimate for the best-fitting power function, suggest that it does not give a satisfactory account of the data.

Discussion

Psychological data often exhibit a complex pattern of dependence on a predictor (e.g., Figure 6). Regression models usually partition this complex dependence into a smooth function of the predictor – the regression function – and an error model that accounts for high frequency variation around the regression function. Other decompositions are possible, such as accounting for at least some of the high frequency variation with a chaotic deterministic function (e.g., Kelly, Heathcote, Heath & Longstaff, 2001, see Heath, 2000, for methods of estimating these functions). However, the majority of modelling in quantitative psychology follows the regression model approach and uses smooth parametric regression functions (e.g., Cutting, 2000).

The methodology developed above represents a new tool for assessing the adequacy of smooth regression functions. By assuming only that the regression function of the DGM is smooth (i.e. has a certain number of everywhere-continuous derivatives) we can test hypotheses on the adequacy of proposed regression models, and also construct confidence bands that can help pinpoint the nature of a proposed function's inadequacy. As stated above, the correct operation of this technique depends on the chosen nonparametric regression estimator being asymptotically unbiased; many common algorithms fulfil this condition.

The methods we provide can, as illustrated, provide insights beyond those provided by conventional techniques. Further, they require relatively little programming or statistical expertise on the part of the user. All that is required is an ability to calculate nonparametric regression estimates, and to employ a basic bootstrap. Nonparametric regression routines are part of many widely available statistical packages. Custom written software is also freely available on the internet to implement the more esoteric nonparametric regression algorithms. The local-linear regression algorithm used above was written by the authors as a Matlab script, and is available on the internet (follow the links to the software download page from http://ntserver.newcastle.edu.au/ncl/).

A point that may trouble some readers pertains to the differences between a DGM and the best fitting model from a parametric family. In the methodology we have developed, we assume only that the true regression model is smooth, and make no assumptions about the parametric family to which the posited regression model belongs. However, in the model selection literature the discrepancy between the DGM and the estimated parametric model (distance A in Figure 1) is commonly decomposed into the sum of two distances (e.g. Forster, 2000; Zucchini, 2000). This decomposition could be illustrated in Figure 1 by plotting one extra model: the best-fitting model from the hypothesised parametric family. Then the distance A would be expressed as the sum of the distance from the best smooth model to the best-fitting model of the given family plus the distance from the best-fitting model to the estimated model. These two distances are sometimes called the error due to model approximation and the error due to model estimation, respectively.

While this decomposition is very often important, we argue that there exist situations in which it is also useful to consider the sum of these two distances – the non-decomposed distance, A – as we have done. A test of the null hypothesis that this summed distance is zero will be sensitive to both model mis-specification and to poor parameter estimation. Thus, a model will be rejected when it is the wrong model, but it will also be rejected when its parameters are poorly estimated. This could be a useful test in practice, allowing the identification of model estimates for which the parameter estimation algorithm (whatever that may be) has not worked effectively.

Of course, the disadvantage of testing the summed distance is that it is possible to confound poor parameter estimation with poor model fit.

The smooth function test developed here can also facilitate model discrimination in cases where comparative model testing, based on a single discrepancy measure, is indeterminate, as in Example 2 above. The smooth function test is sensitive to systematic dependence between a predictor and deviations from the best smooth function, and so can sometimes adjudicate between functions that have similar overall discrepancy from the data. The nature of systematic distortions in the shape of a regression function can be determined using the confidence bands developed here. Systematic shape distortions, along with other factors like the plausibility of parameter estimates and the pattern of variation of parameter estimates as a function of experimental conditions, provide important constraints for model assessment over and above the overall goodness-of-fit of the a regression function.

The methodology developed here cannot identify a single "true" regression function. For example, where a simple parametric function proves adequate, so will all more complex parametric functions that nest it. Our methodology can be used to determine the set of functions that provide an adequate description of the data, but where the "simplest" function is preferred a discrepancy function that includes a complexity penalty must be employed (e.g. Bozdogan, 2000; Forster, 2000). A simple nonlinear function might provide an adequate fit because the range of the predictor is restricted, while a more complex function may be required in the general case. In the latter case, only further data that samples a greater range of design points, and so provides better constraint for estimating the more complex function, can adjudicate adequately.

References

Azzalini, A., Bowman, A.W., & Hardle, W (1989) On the use of non-parametric regression for model checking. Biometrika, 76, 1-11

Bozdogan, H. (2000) Aikike's information criterion and recent developments in information complexity. Journal of Mathematical Psychology, 44, 62-92

Bowman, A.W. & Azzalini, A. (1997) Applied smoothing techniques for data analysis: The kernel approach with S-Plus illustrations. Oxford: Clarendon Press.

Cutting, J. E. (2000). Accuracy, scope, and flexibility of models. Journal of Mathematical Psychology, 44, 3-19.

Davison, A.C. & Hinkley, D.V. (1997) Bootstrap methods and their application. Cambridge: Cambridge University Press.

Fan, J. & Gijbels, I. (1996) Local polynomial modelling and its applications. London: Chapman & Hall.

Forster, M.R. (2000) Key concepts in model selection: Performance and generalizability. Journal of Mathematical Psychology, 44 205-232

Hardle, W. & Marron, J.S. (1991) Bootstrap simultaneous error bars for non-parametric regression. Annals of Statistics, 19, 778-796

Hardle, W. & Nussbaum, M. (1990) Bootstrap confidence bands. In Jockel, K.-H., Rothe, G. & Sendler, W. (Eds.) Bootstrapping and Related Techniques. New York: Springer Verlag

Heath, R.A. (2000) Nonlinear dynamics: Techniques and applications in psychology. New York: Erlbaum.

Heathcote, A., Brown, S. & Mewhort, D.J.K. (2000) Repealing the power law: The case for an exponential law of practice. Psychonomic Bulletin and Review, 7, 185-207.

Kelly, A., Heathcote, A., Heath, R. A. & Longstaff, M. (2001). Response time dynamics: Evidence for linear and low-dimensional nonlinear structure in human choice sequences. Quarterly Journal of Experimental Psychology, 54, 805-840

Knafl, G., Sacks, J. & Ylvisaker, D. (1985) Confidence bands for regression functions. Journal of the American Statistical Association, 80, 683-691

Markus, M.T. & Visser, R.A. (1990) Applying the bootstrap to generate confidence regions in multiple correspondence analysis. In Jockel, K.-H., Rothe, G. & Sendler, W. (Eds.) Bootstrapping and Related Techniques. New York: Springer Verlag

Rosenblueth, E. (1975) Point estimates for probability moments. Proceedings of the National Academy of Sciences of the United States of America, 72, 3812-3814

Rickard, T. C. (1997). Bending the Power law: A CMPL theory of strategy shifts and the automatization of cognitive skills. Journal of Experimental Psychology: General, 126, 288-311.

Wand, M. P., & Jones, M. C. (1995). Kernel Smoothing. London: Chapman & Hall.

Wu, C.F.J. (1986) Jackknife, bootstrap and other resampling methods in regression analysis. Annals of Statistics, 14, 1261-1343

Zucchini, W. (2000) An introduction to model selection. Journal of Mathematical Psychology, 44, 41-62

Figure Captions

Figure 1. Schematic depiction of the models and their smoothed estimates. Dotted lines indicate biases due to the smoothing operation. Filled shapes represent data sets, unfilled shapes represent regression functions.

Figure 2. Simulated data for Example 1 (triangles). The dotted line represents the regression function of the DGM: $y = e^{-5x^2} \sin(2px)$. The solid line is a nonparametric estimate of the regression function.

Figure 3. Observed data for Example 1 (triangles), nonparametric regression estimate (solid line) and 95% confidence bands (dotted lines) for the nonparametric estimate under the assumption that model T generated the data.

Figure 4. Observed data for Example 1 (triangles), DGM regression function (dotted line), and new regression function to be tested (solid line).

Figure 5. Simulated data (triangles), the nonparametric regression estimate from observed data (solid line) and 95% confidence bands (dotted lines) for the nonparametric estimate under the assumption that model F is the DGM.

Figure 6. Practice data (triangles) along with best-fitting exponential function (dotted line) and non-parametric estimate (solid line). The ordinate is

response time (RT, in ms) and the abscissa is number of practice trials (N). The exponential function is $E(\underline{RT})=728.1+2544*\exp(-0.05253*\underline{N})$. The non-parametric estimate was calculated using a bandwidth of 9 practice trials.

Figure 7. Practice data (triangles), with the best-fitting three-parameter power function: $E(\underline{RT})=0.0+4226*\underline{N}^{-0.3613}$.

Figure 8. Pointwise 95% confidence bands (dotted lines) for location of the nonparametric estimate the under assumption of a power-function regression model. Also shown is the nonparametric regression estimate calculated from the data (solid line).

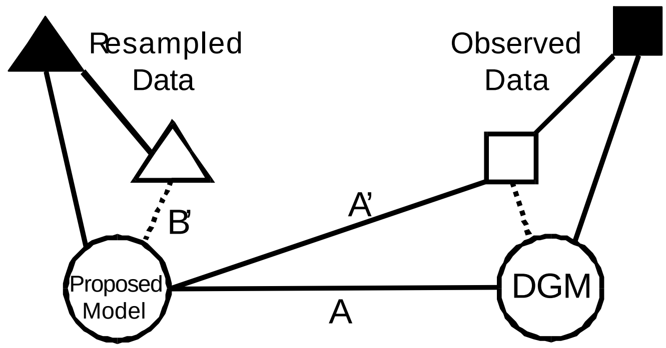


Figure 1

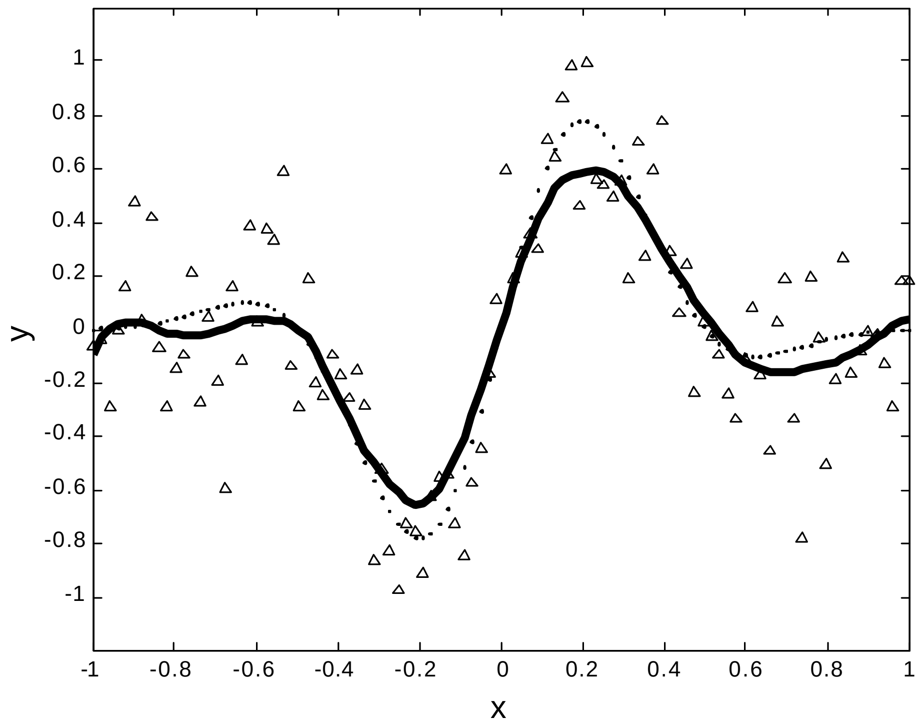


Figure 2

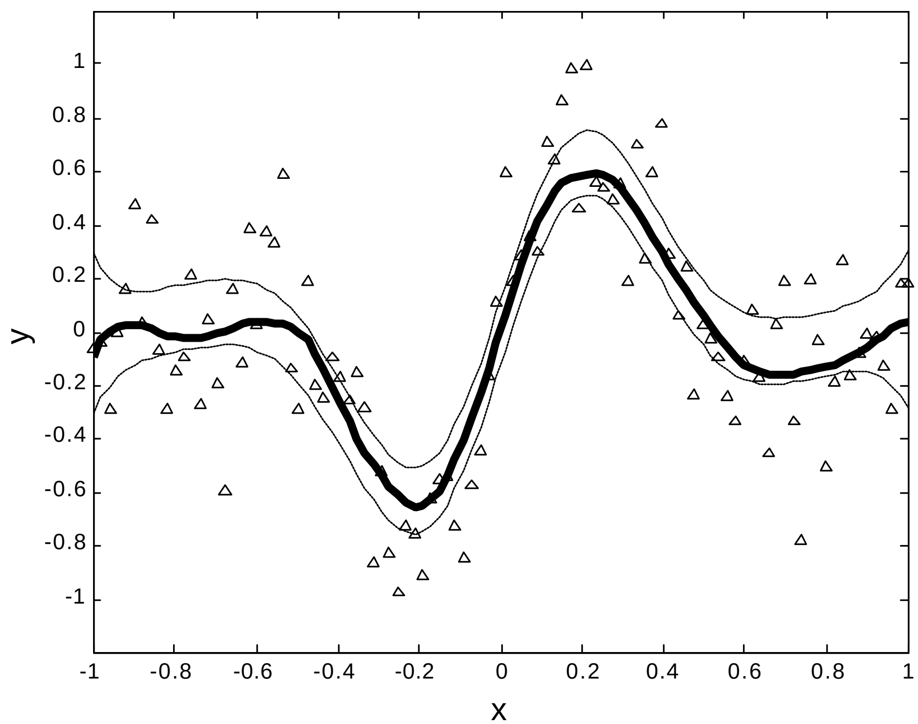


Figure 3

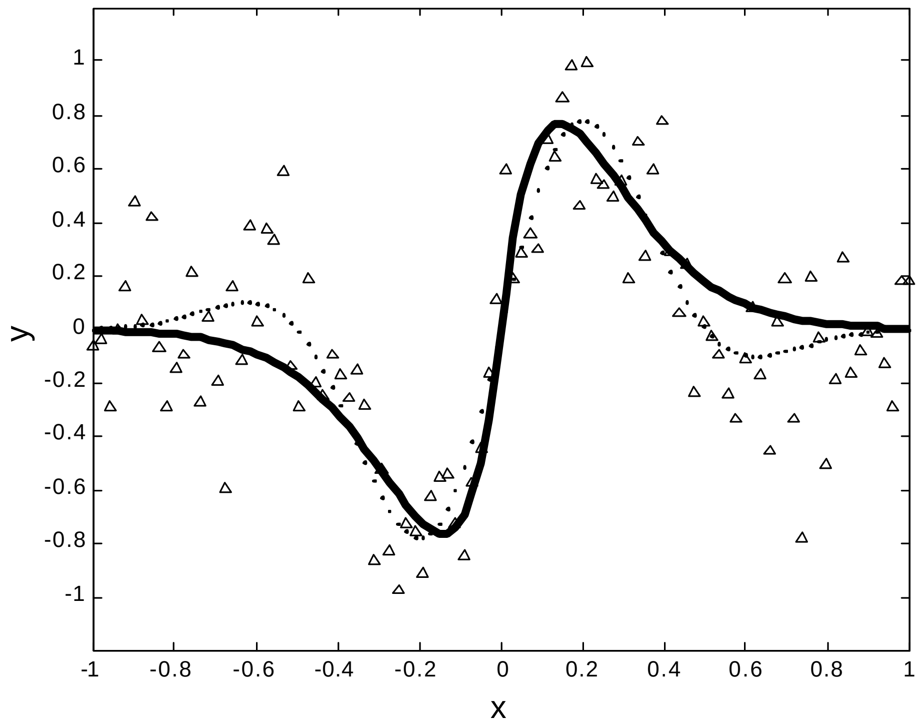


Figure 4

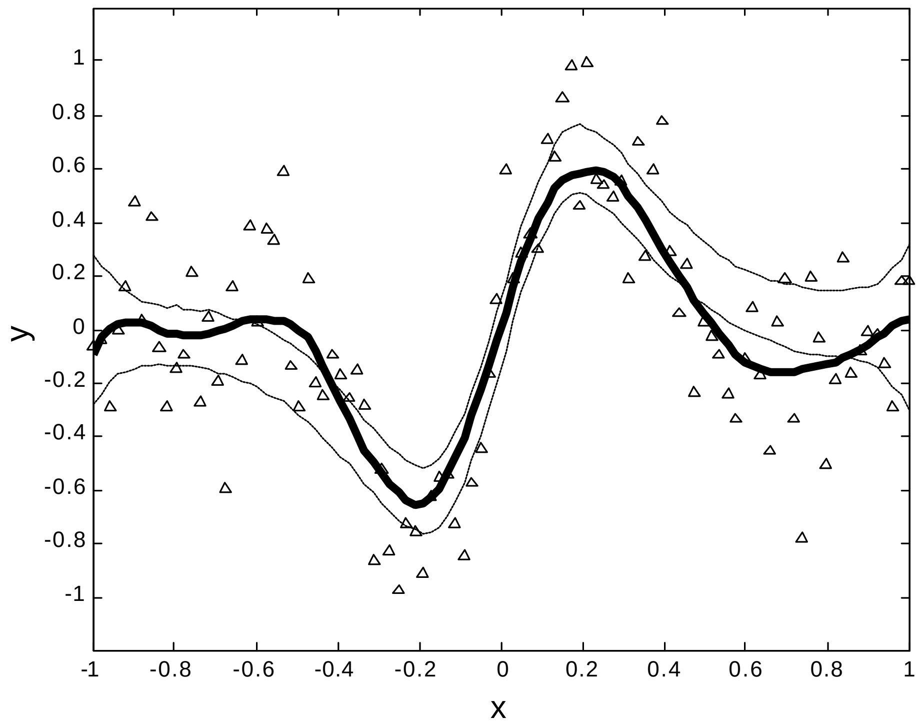
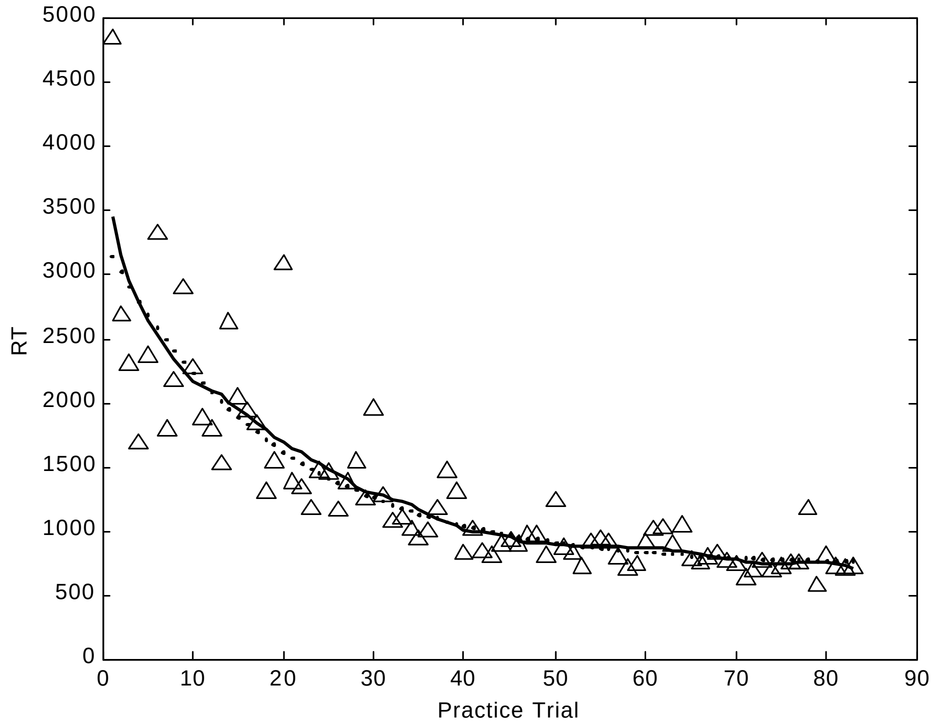
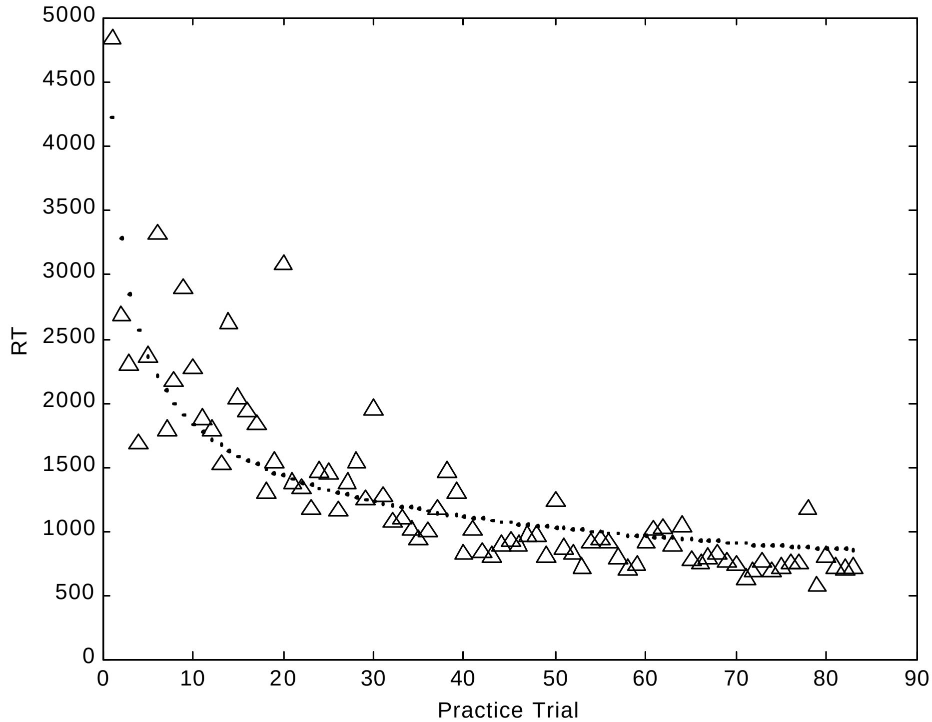


Figure 5



RT
Practice Trial
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
0
10
20
30
40
50
60
70
80
90

Figure 6



5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
RT
0
10
20
30
40
50
60
70
80
90
Practice Trial

Figure 7

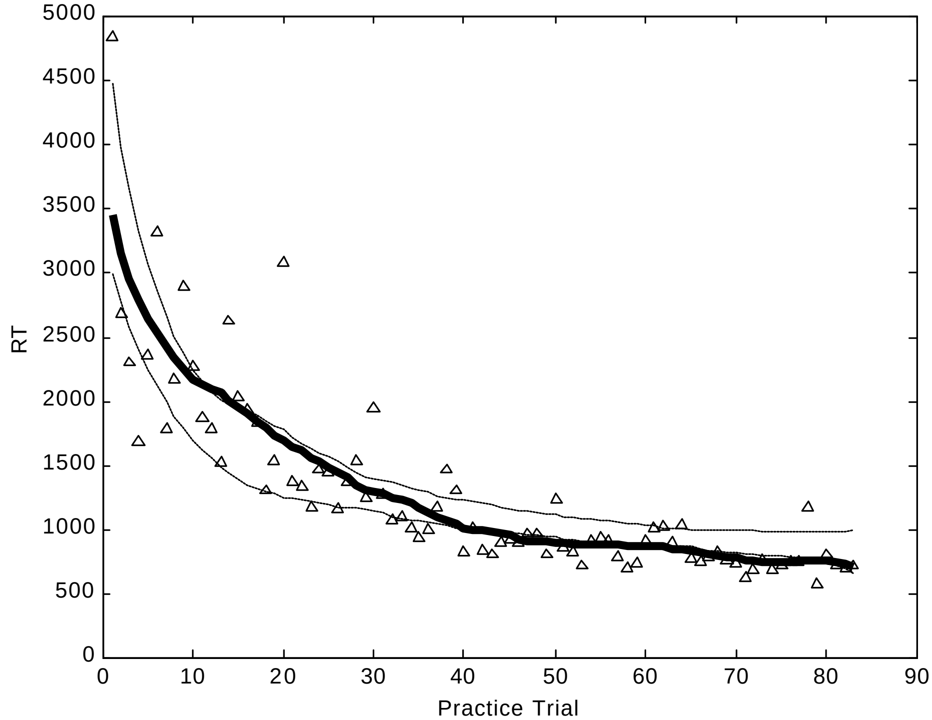

5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
RT
0
10
20
30
40
50
60
70
80
90
Practice Trial

Figure 8

Footnotes

[1] Although data sets and regression functions may seem incongruous on the same diagram, due to incommensurability, recall that under our definitions both can be specified simply as an ordinate value at each design point.

[2] In the sense of sample density on the covariate axis.

[3] The data come from Rickard (1997), subject 9, condition 10. The participant was learning to perform "alphabet arithmetic" under speed stress. That is, they were given a great deal of practice (x-axis) on responding to problems similar to "A+2=C, True/False". These data are available on the web – follow the links to the data repository from http://ntserver.newcastle.edu.au/ncl/

[4] The estimate of the asymptote parameter was bounded below by zero. When this constraint was not imposed the estimated asymptote parameter for the power function was negative.