Pergamon

Electoral Studies 19 (2000) 577–613

Electoral Studies

www.elsevier.com/locate/electstud

Electoral inquiry section

Loess: a nonparametric, graphical tool for depicting relationships between variables☆

William G. Jacoby *

Department of Government & International Studies, University of South Carolina, Columbia, SC 29208, USA

Abstract

Loess is a powerful but simple strategy for fitting smooth curves to empirical data. The term "loess" is an acronym for "local regression" and the entire procedure is a fairly direct generalization of traditional least-squares methods for data analysis. Loess is nonparametric in the sense that the fitting technique does not require an a priori specification of the relationship between the dependent and independent variables. Although it is used most frequently as a scatterplot smoother, loess can be generalized very easily to multivariate data; there are also inferential procedures for confidence intervals and other statistical tests. For all of these reasons, loess is a useful tool for data exploration and analysis in the social sciences. And, loess should be particularly helpful in the field of elections and voting behavior because theories often lead to expectations of nonlinear empirical relationships even though prior substantive considerations provide very little guidance about precise functional forms.

Keywords: Generalized additive models; Loess; Lowess; Nonparametric regression; Scatterplot smoothing; Statistical graphics

1. Introduction

The purpose of this paper is to discuss the loess procedure for fitting smooth curves to scatterplots. Loess provides a graphical summary of the relationship between a

☆ The data used in the examples presented in this paper, along with the S-Plus routines used to produce the graphs, can be found on the Worldwide Web, at http://www.cla.sc.edu/gint/faculty/jacoby

* Tel.: +803-777-3109; fax: +803-777-8255.
E-mail address: william-jacoby@sc.edu (W.G. Jacoby).

PII: S0261-3794(99)00028-1

dependent variable and one or more independent variables. The distinctive feature of this procedure is that it “allows the data to speak for themselves”. Loess is nonparametric, so the fitted curve is obtained empirically rather than through stringent prior specifications about the nature of any structure that may exist within the data. Therefore, loess-enhanced scatterplots often reveal relatively complex relationships that could easily be overlooked with traditional statistical modeling procedures.

Loess and other nonparametric estimation strategies are useful in social scientific research because current substantive theories usually provide little detail about the kinds of structural patterns that should exist within empirical data. In other words, hypotheses suggest *which* variables should be related to each other, and often, the *direction* of any such relationships: For example, “education levels should be positively related to voting turnout”. Beyond statements like this, however, there are generally no predictions about functional forms. Researchers therefore fall back on simple specifications, for want of theory-based directions to the contrary — a situation that Beck and Jackman (1998) have recently called “linearity by default”. This creates a potentially serious problem because those detailed theories which *do* exist suggest that nonlinear relationships are pervasive throughout the field of elections, voting, and mass political behavior (e.g. Przeworski and Soares, 1971; Zaller, 1992; Brown, 1995). Thus, a nonparametric technique like loess should be very useful for discerning such nonlinearities and explicating their forms.

The rest of this paper provides a detailed presentation of the loess method, along with the major practical considerations involved in its use. Most of the discussion will focus on the simplest case — using loess as a descriptive, exploratory tool for fitting smooth curves to scatterplots. This is undoubtedly the kind of situation where loess is employed most frequently. However, the technique is much more general than this. So, some attention will also be given to statistical inference and multivariate loess. Overall, loess is a very useful tool for discerning systematic structure within empirical data. As such, this technique should help researchers develop theories that provide accurate, powerful representations of real-world phenomena.

2. Scatterplot smoothing

The two-dimensional scatterplot is the basic graphical display method for bivariate data. At the same time, the scatterplot is the “building block” for more complex graphical depictions of multivariate data (Jacoby, 1998). One of the great strengths of the scatterplot is that it enables visual assessments of relationships or functional dependencies between the variables included in the display.[1] In operational terms, functional dependence exists when points that have different coordinates on one scale

[1] The visual nature of this assessment is very important because it avoids the a priori assumptions that provide the basis for more traditional, numerical summaries of statistical relationships, such as the linearity assumption underlying the use of Pearson product–moment correlations. Instead, direct visualization of bivariate data facilitates the identification of relationships and underlying patterns that may not conform to any simple structure (Chambers et al., 1983).

axis of the scatterplot also tend to exhibit systematically different coordinates on the other scale axis. If two variables are related to each other (that is, they are functionally dependent), then the plotted points will not be distributed uniformly throughout the entire plotting area. Instead, the point cloud will form some discernible pattern or shape.

Evaluating functional dependence in a scatterplot should be a straightforward task. However, it is often quite difficult in practice. The problem is that noisy data values, sparse data points, and weak interrelationships can inhibit visual identification of any such patterns. Furthermore, even if a general pattern can be discerned in the graph, it is almost impossible to characterize its precise nature through visual inspection of the scatterplot, alone.

One useful strategy for dealing with the preceding problems involves fitting a smooth curve to the points in the scatterplot. The purpose of the curve is to summarize the central tendency of the *Y* variable's distribution at different locations within the *X* variable's distribution. If the two variables are unrelated to each other, then the smooth curve will be a flat line (the center of the *Y* distribution does not change, regardless of the *X* value). If the two variables are related, then the smooth curve should exhibit some other, non-horizontal shape.

There are two general strategies for fitting a smooth curve: parametric and nonparametric fitting (Cleveland, 1993). The former, parametric fitting, requires the analyst to specify the functional form of the relationship in advance. The fitting algorithm then optimizes the correspondence between the specified form and the empirical data, usually by estimating the set of equation coefficients that produce the best fit between the two. Regression analysis is, by far, the best-known parametric smoothing procedure. It uses the least-squares criterion to fit a straight line to a set of data points.

Parametric fitting is a very effective way to summarize a relationship when the structure in the data conforms to the type of function that is fitted by the smoothing algorithm. But, this is exactly the problem — the "correct" functional form is almost always unknown, at least at the outset of the analysis. As a result, the researcher runs a serious risk of fitting a smooth curve that misrepresents the structure within the data.

Nonparametric smoothers directly address the preceding problem. They can be used to locate a smooth curve among the data points without requiring any advance specification of the functional relationship between the variables. Instead, the fitting algorithm simply tries to follow the empirical concentration of the plotted points. The resultant fitted "line" should pass through the most dense areas of the data region in the scatterplot, regardless of the shape of the curve that is required in order to do so.

Currently, the most popular nonparametric smoother is *loess* (Cleveland and Devlin, 1988).[2] As William S. Cleveland notes, "... loess has some highly desirable statistical properties, (it) is easy to compute, and... (it) is easy to use" (Cleveland, 1993, p. 94). The term 'loess' is an acronym for *lo*cally weighted regr*ess*ion. The

[2] Goodall (1990) provides a useful and succinct overview of many other nonparametric smoothers.

method to which it refers is a generalization of the technique known as 'lowess', for *lo*cally *we*ighted *s*catterplot *s*moother (Cleveland, 1979).

3. An example of loess smoothing

In order to demonstrate the utility of the loess procedure, we will examine a substantive example, using state-level data on education and voter turnout in the 1992 American presidential election. This is an ideal topic for our present purposes, because it epitomizes the ambiguities that often exist in our theoretical propositions. The relationship between education and mass political participation is widely acknowledged by social scientists. However, Nie et al. (1996) point out that even though we know these variables *are* related, we still do not know *why* or *how* education affects participatory activity like voting. Accordingly, it is not clear exactly what the proper functional specification should be. Some aggregate-level analyses have used linear models (e.g. Kim et al., 1975; Patterson and Caldeira, 1983) while others have allowed for nonlinearities (at least implicitly, through the use of dummy independent variables) in the relationship between education and voter turnout (e.g. Powell, 1986).

Fig. 1 shows 1992 state voter turnout rates plotted (on the vertical axis) against the percentage of high school graduates in the respective state populations (on the horizontal axis). The general diagonal orientation of the point cloud suggests that education levels and voter turnout are positively related to each other. But, with no further information, it is impossible to provide any more detail about the exact nature of this relationship from the visual information alone.

Fig. 2 shows the same scatterplot, but it also contains a loess curve superimposed among the data points. The procedure used to fit this curve will be explained below. For now, it is merely necessary to emphasize two things. First, the curve does follow the central tendency of the *Y* variable's values across the range of the *X* variable. In doing so, the curvilinear nature of the relationship between education level and voter turnout is revealed immediately. Second, this curve was obtained without any prior specification about the functional form of the relationship. Instead, the sigmoid (i.e. 'S-like') shape of the smooth curve was produced by the loess procedure, with very little in the way of 'instructions' from the analyst.[3]

The curvilinear fit shown in Fig. 2 is important in substantive terms. As one would expect, states with more highly educated populations tend to exhibit higher voter turnout levels. But, this relationship is not constant. The impact of education on electoral participation is most pronounced among those states with moderate graduation rates; the slope of the fitted curve is steepest in the interval between about

[3] A note of reassurance for skeptics: the nonlinearity is *not* due to unusual observations such as the four states with low graduation percentages and high turnout rates in the upper-left area of the plotting region. For one thing, the loess curve is fitted with a robust estimation procedure that decreases the influence of such outliers. For another, the sigmoid curve remains even if these four data points are eliminated from the scatterplot and the loess calculations.

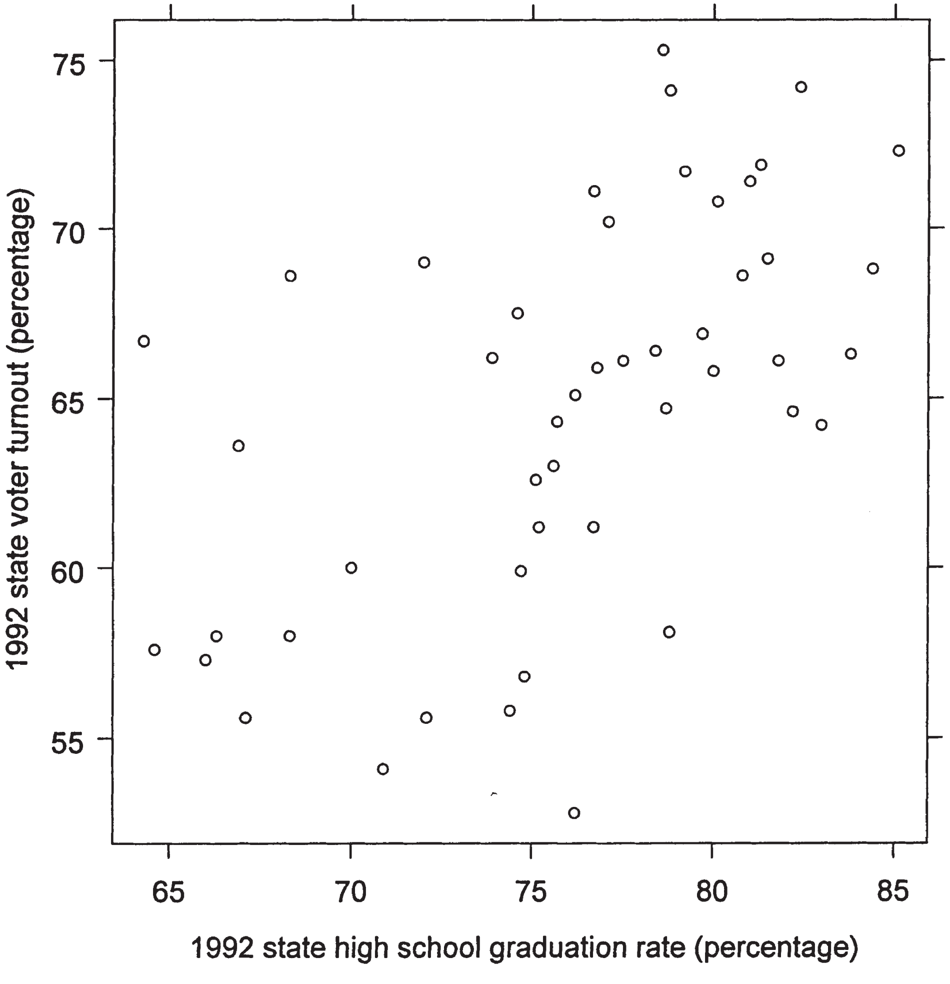

Fig. 1. Basic scatterplot showing the relationship between state education levels and state turnout rates in 1992. Data source: 1993 Statistical Abstract of the United States.

74% and 78% on the horizontal axis. In contrast, the slope becomes very shallow near the right and left sides of the curve. Hence, the precise high school graduation rate makes little difference among states with either poorly- or well-educated populations. Average voter turnout tends to be quite low among the former (the mean turnout for these states hovers between about 59% and 62%) and high among the latter (with mean turnout rates of about 68%).

The loess curve in Fig. 2 is also important from a methodological perspective. It clearly shows that a linear model would provide a misleading depiction of the

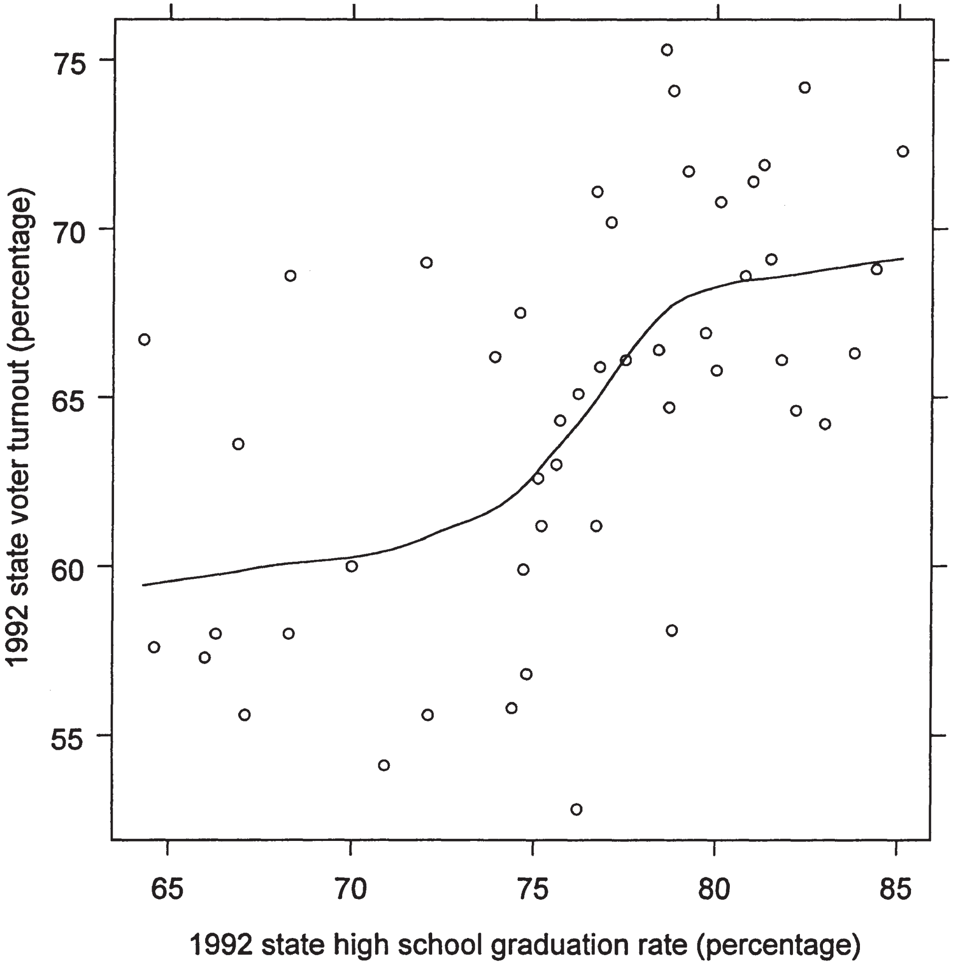

Fig. 2. Loess curve fitted to data on 1992 state education levels and voter turnout rates. Note: the loess curve is fitted with α=0.65, locally linear functional form (λ=1), and robustness iterations.

relationship between education levels and voter turnout rates.[4] But, the relatively simple shape of the empirical curve also suggests that a logistic transformation of the dependent variable might be appropriate. Indeed, if one is willing to make an ecological inference, then these results may also provide some support for the use of logistic or probit specifications in individual-level models of voter turnout. Thus,

[4] When ordinary least squares is employed to fit a linear model to these data, the resultant equation is: $Turnout_i$=19.91+0.59 $Education_i+e_i$, where $Turnout_i$ is the voter turnout rate and $Education_i$ is the high school graduation rate for the *i*th state. The R^2 for this equation is 0.31. The value of the *F* statistic is 20.72 (1 and 47 degrees of freedom), with a *p*-value of 0.00004. While this equation represents the best-fitting *line* for these data, it fails to incorporate any aspects of the curvilinear structure.

the nonparametric smooth curve provides useful information that could be incorporated into a parametric model specification for these (or other) data.

Despite its potential importance, the curvilinearity in the relationship between education and voter turnout would have been completely invisible in a standard, linear regression analysis — at least until an examination of the residuals. Unfortunately, the latter step is often omitted from empirical research efforts. And even when residual plots are produced, it is easy to overlook patterns that may exist within them. The loess procedure helps the researcher avoid the intermediate step of fitting a model that turns out to be an inaccurate representation of the data. Instead, the evidence from the loess curve can be used to formulate a more accurate description of the data in the first place.

4. Fitting a loess smooth curve

The loess procedure is computationally intensive; in other words, there is a large number of distinct steps involved in fitting even a simple loess curve to a small dataset. Nevertheless, the calculations themselves are straightforward. They should be readily understandable to anyone who is familiar with ordinary least squares regression analysis. The discussion in this section will provide a brief overview of the methodology underlying loess. Complete details and a simple, step-by-step example of the fitting procedure can be found in the Appendix.

Assume that the data consist of n observations on two variables, X and Y. These data are displayed in a bivariate scatterplot, with the scale for X on the horizontal axis and the scale for Y on the vertical axis. The plotted points are the ordered pairs (x_i, y_i), where i ranges from 1 to n.

The procedure starts by selecting a series of m locations or evaluation points, v_j, with j running from 1 to m. These evaluation points are equally-spaced across the range of X.[5] Next, loess performs a series of m weighted regression analyses, one at each of the v_j. These regressions are “local” in the sense that each one only uses the subset of observations that fall closest to that evaluation point along the horizontal axis of the scatterplot. The researcher specifies the proportion of the total data that is included within each subset using a loess parameter called α (to be explained below). The local regressions can use either linear or quadratic equations. The researcher specifies the functional form using the loess λ parameter (also explained below). In either case, the observations included in each local regression are inversely weighted according to their distance from the evaluation point along the X axis. The weights insure that observations closer to v_j will have more influence on the placement of the local regression line (or curve, with the quadratic form) than observations that fall farther away within the local region. The local regressions can also incorpor-

[5] The exact number of evaluation points is relatively unimportant, so long as there are enough of them to provide sufficient detail about the variability in the conditional distributions of the Y variable. In practice, the value of m is usually determined by the software employed to fit the loess curve.

ate an optional robust estimation procedure in order to reduce the influence of unusual data points.

The coefficients from each local regression are used to estimate a predicted or fitted value, designated $\hat{g}(v_j)$ for that evaluation point. After all of the local regressions are completed, the m different ordered pairs, $(v_j,\hat{g}(v_j))$ are plotted in the scatterplot, superimposed over the n data points that are already shown in the plot. Finally, adjacent fitted points — that is, the $(v_j,\hat{g}(v_j))$ for successive v_js — are connected by line segments. The evaluation points are located relatively close to each other along the horizontal axis, so the series of connected line segments actually appears to be a smooth curve passing through the data points.

The loess procedure is sometimes conceptualized as a "vertical sliding window" that moves across the horizontal scale axis of the scatterplot. The window stops and estimates a separate regression equation (using weighted least squares) at each of the m different v_js. Since the regressions only involve the data points that fall within the window, the estimated slopes (and hence, the fitted values) can change to follow the contours of the data. This is precisely the feature that gives loess the flexibility to conform to relatively complicated, nonlinear shapes within the point cloud of a scatterplot.

5. Fitting parameters for the loess smooth curve

The loess procedure is nonparametric in the sense that the analyst does not specify the functional form of the final smooth curve. However, there are some parameters that must be supplied prior to the fitting procedure in order to guarantee that the loess curve really does pass through the center of the empirical data points. Selecting the values for these parameters is a subjective process, but the considerations that are involved in the decisions are quite straightforward.

5.1. The smoothing parameter, α

In a loess fit, the α parameter determines the width of the sliding window. More specifically, α gives the proportion of observations that is to be used in each local regression. Accordingly, this parameter is specified as a value between 0 and 1. The α value used for the loess curve in Fig. 2 is 0.65; so, each of the local regressions used to produce that curve incorporates 65% of the total data points. Fig. 3 shows the effect of changing the α parameter. The four panels show loess curves that are fit to exactly the same data (again, the information on state high school graduation and voter turnout rates that was used in Figs. 1 and 2); however, the α values are varied from 0.15 in the first panel, up to 1.00 in the last.

Obviously, the fitted curve becomes smoother with larger values of this parameter. This occurs for two reasons. First, wider fitting windows (i.e. larger α) mean that idiosyncratic observations will tend to cancel each other out, and therefore have proportionately less influence on the local regressions. Second, larger α values mean that fewer observations will change when moving from one fitting window to the

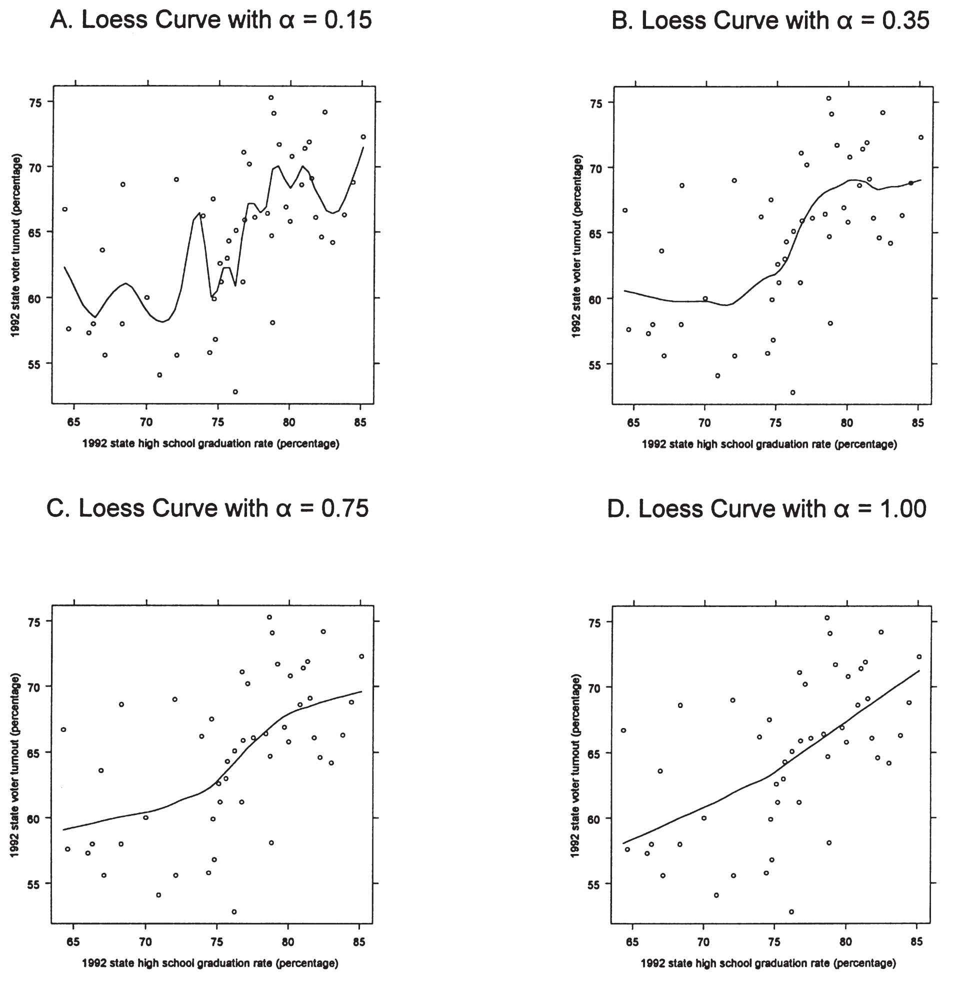

Fig. 3. Effect of the α parameter on the loess smooth curve.

next. Both of these factors should tend to stabilize the local regression lines and fitted values, thereby producing a smoother loess curve.

Informally, one could think of the loess curve as a string that is laid across the range of the *X* values within the data. The α value controls the "slackness" of this string, with larger values pulling it tighter and therefore producing a straighter curve. For this reason, α is sometimes called the "tension" parameter in the loess fit.

With these different possibilities, which loess curve is the "best" or "most appropriate" smoothed version of these data? The curves in Fig. 3A and B are probably not very useful. With α set to low values like 0.15 or 0.35, the window width is extremely narrow, and the local regressions are highly sensitive to "noise" variation within the data values. This produces the undulating curves, both of which obscure the overall structure in the data. At the other extreme, the curves in Figs. 3C and D are very smooth, but they also fail to pass through the center of the entire point cloud; most of the data points in the central region of X values fall below the loess curve. The problem is that the wide windows produced by α values of 0.75 or 1.00 prevent the local regressions from adjusting enough to follow curvilinearity within the data.[6] An intermediate value of α should provide a compromise between the "over-fitting" of the first two panels in Fig. 3, and the "lack of fit" that occurs in the last two panels. This is precisely why the loess curve back in Fig. 2 is based on an α of 0.65.

Decisions about the proper α value must be made on a case-by-case basis. The general objective is to produce a loess curve that is as smooth as possible, but still captures all of the important structure that exists within the data. A strategy for doing so will be presented below, in the section on residual plots.

5.2. The degree of the loess polynomial, λ

The λ parameter specifies the degree of the polynomial that the loess procedure fits to the data. If λ=1, then linear equations are fit within each of the windows. When λ=2, quadratic equations are used. The latter complicate the fitting process somewhat, but they are sometimes necessary in order to produce a smooth curve that follows the data to an acceptable degree.

Fig. 4 illustrates how locally quadratic fitting can produce a more accurate loess

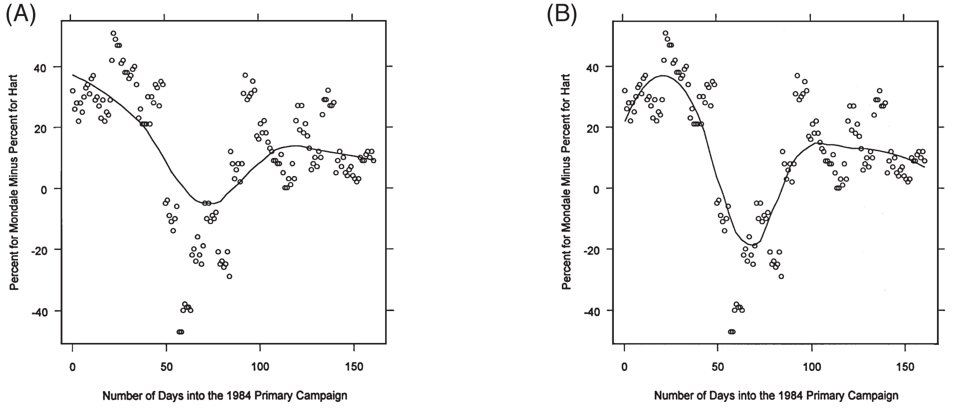

Fig. 4. Effect of the λ parameter on a loess smooth curve: (A) loess curve fitted with λ=1 and α=0.50; (B) loess curve fitted with λ=2 and α=0.50. Note: the data show public preferences between Walter Mondale and Gary Hart (among Democrats only) during the 1984 presidential primary campaign. Data are adapted from the 1984 CPS Continuous Monitoring Survey.

[6] Beginners often think that a loess curve fitted with an α value of 1.0 should always produce a straight line. However, this is *not* true, in general. Even though the 'local' regressions each incorporate 100% of the data, the neighborhood weights guarantee that observations close to each evaluation point will have greater influence on the fitted value at that point than do observations farther away. This, in turn, allows enough flexibility to produce a nonlinear loess curve, given the appropriate data.

curve in certain situations. The figure shows data on public preferences (among self-identified Democrats only) between Walter Mondale and Gary Hart, over the course of the 1984 presidential primary campaign. The first panel contains a loess curve with λ=1. Even though the smoothing parameter is set to a reasonable value (α=0.50), the curve fails to track the data points accurately in the 'peaks' and 'valleys' that occur in the scatterplot. Specifically, note that the fitted line has a negative slope near the left side of the plot even though most of the data points at the extreme left fall below the line. Another serious problem exists in the interval ranging from about 50 to 100 on the horizontal axis. Here, the curve does not dip low enough to follow closely the sizable set of data points that occur in that region of the scatterplot.

The second panel in Fig. 4 shows the same scatterplot. However, the loess curve has now been fitted with λ=2 (the α remains at 0.5). With this modification, the curve tends to pass through the center of the point cloud at all locations throughout the set of plotted points. And, exactly as expected, the most pronounced changes occur in the regions at the extreme left, and in the center of the plot. The smooth curve in this plot suggests that Mondale's support peaked and then fell off during the early days of the 1984 campaign. This differs markedly from the locally linear fit, which seemed to indicate that Mondale simply began the campaign with steadily declining support among Democrats. The locally-quadratic fit also shows that the decline in Mondale's support was very sharp during the middle of the campaign period; this feature is much more pronounced than it was in the loess fit based upon locally linear equations.

In practice, the specification of the λ parameter is usually fairly easy; the decision can often be made upon visual inspection of the scatterplot, alone. If the point cloud conforms to a generally monotonic pattern (either increasing or decreasing), then λ should be set to 1 for locally linear fitting. If the data exhibit some nonmonotone pattern, with local minima and/or maxima, then λ should be set to a value of 2 for locally quadratic equations.

The reasoning behind these recommendations is as follows: if X and Y exhibit a monotonic relationship, then the point clouds within the local fitting windows should always exhibit the same general orientation. When this occurs, varying the intercepts and slopes of the locally linear regressions should be sufficient to produce a smooth curve that follows the data accurately. On the other hand, a nonmonotonic relationship implies that the general orientation of the bivariate point cloud changes direction somewhere within the data region of the scatterplot. And, such reversals cannot be handled very effectively with linear equations. The quadratic specification allows for sharper inflections within the locally-fitted curve. This, in turn, produces the flexibility that is required to insure that the final loess curve passes through the center of a nonmonotone point cloud.

5.3. The robustness step in the loess smoother

The robustness step in the loess fitting procedure is, strictly speaking, optional. Nevertheless, it is often included in the calculation of loess smooth curves. Like other least-squares methods, loess can be adversely and strongly influenced by unusual observations (e.g. Belsley et al., 1980). This problem is exacerbated by the fact that the local regressions typically involve a subset of the overall data. Therefore, any discrepant data points will comprise a sizable proportion of the observations used in the local estimation and their degree of influence will also increase accordingly. The robustness step of the loess procedure downweights the observations that are most likely to have an adverse effect on the local regressions — those with large residuals. After doing so, the smooth curve is more likely to track the more concentrated areas of data points, rather than 'chasing the outliers' in the scatterplot.

Fig. 5 provides an example that illustrates how the robust fitting option can affect

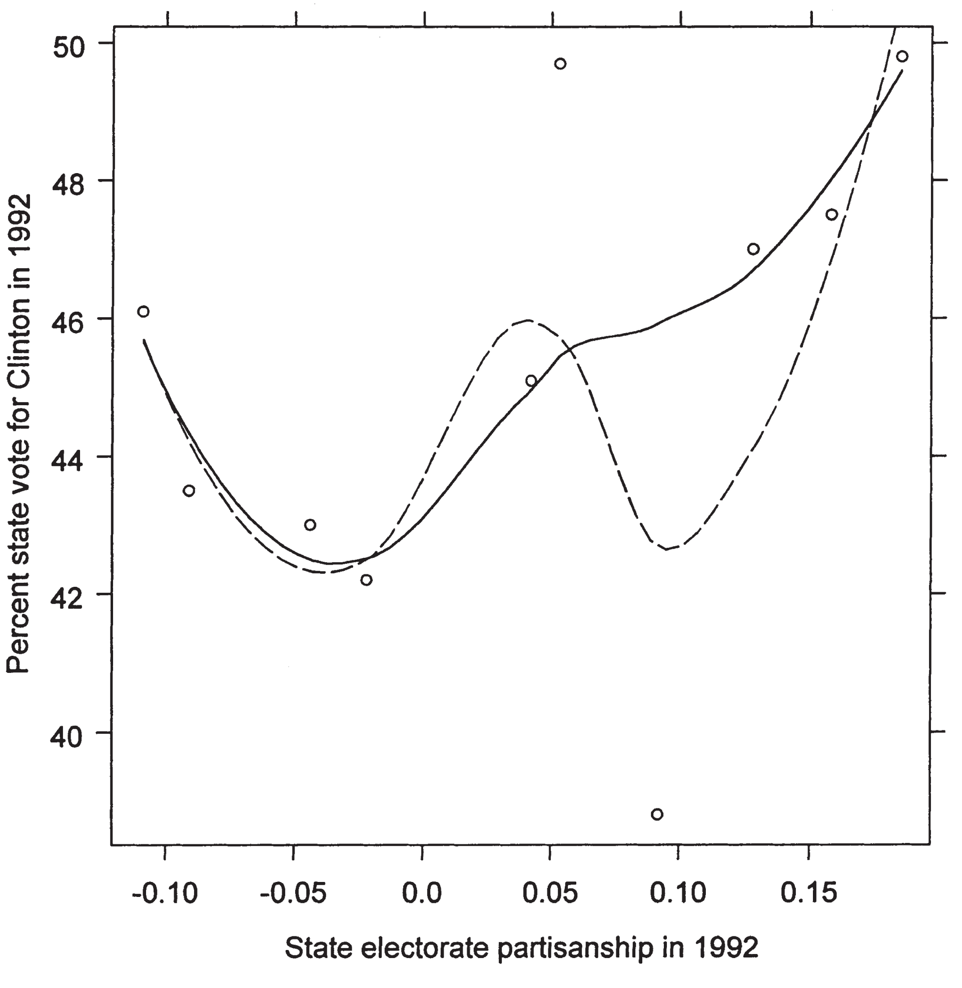

Fig. 5. The effect of robustness weights on a loess curve. Notes: the solid line shows the robust loess curve. The dotted line shows the loess curve obtained when the robustness weights are omitted from the fitting procedure. Data source: Clinton vote percentages are obtained from the 1993 Statistical Abstract of the United States. Data on state electoral partisanship are provided by Gerald C. Wright.

a loess curve. The display shows information about partisanship and 1992 presidential voting in ten northeastern states. Specifically, Clinton's vote percentages in each state are plotted against state electorate partisanship scores. The latter variable is coded so that larger values indicate states whose citizens are more likely to identify themselves as Democrats rather than Republicans, and vice versa for smaller values.[7] The solid line in the figure represents a loess curve with the optional robustness weights incorporated into the fitting process. The dotted line shows the loess curve obtained without the robustness step. Both of these curves are fit with α set at 0.75 and a λ of 2.

First, let us consider the robust loess curve. It shows a relatively simple (although unexpectedly nonmonotonic) functional relationship between electorate partisanship and Clinton voting. Specifically, the pattern in the data could be approximated using a second-degree polynomial with an inflection point at approximately -0.04 on the horizontal axis. In substantive terms, Clinton's support was strongest within the most Democratic states; this is shown by the positive slope of the solid curve in the right side of Fig. 1. But, Clinton's vote percentages also increased among the states with the most Republican electorates; this is signalled by the negative slope of the solid curve near the left side of the display. This latter result is certainly a little surprising, and it signals the need to incorporate other factors besides partisanship into any explanations of 1992 presidential voting patterns.[8]

Next, consider the loess curve fitted without the robustness weights (the dotted line in Fig. 5), which 'undulates' across the range of partisanship scores. A fourth-degree polynomial would be required to fit a smooth function to these data. This would be excessively complex, since there are only ten data points in the first place. Therefore, an unwary observer might conclude that there is no coherent relationship between state partisanship and Clinton voting. But, closer inspection shows that the local minima and maxima in the interior of the curve are entirely due to two data points. These correspond to states with moderately Democratic electorates, but extremely high and low Clinton vote percentages, respectively. The nonrobust loess curve exhibits two sharp reversals which are caused by these points. In contrast, the robust loess fitting procedure effectively 'ignored' the outliers, as a result of the downweighting procedure for the observations with large residuals.

When the objective is to produce a parsimonious graphical summary of the bivariate data, then it is probably best to routinely include the robustness weights in the loess fitting procedure. The only potential disadvantage of robust loess estimation is that the residuals may not be 'well-behaved'; that is, they may not be normally distributed with a mean of zero and constant variance. But, this is really only a serious concern in situations where the analyst is trying to accomplish more than a simple graphical summary of bivariate data (e.g. statistical inference, as discussed

[7] This variable is created using state-level public opinion data. Specifically, the proportion of survey respondents within each state who called themselves 'Republican' is subtracted from the proportion calling themselves 'Democratic'. These data were provided by Gerald C. Wright and they are discussed in greater detail in Erikson et al. (1993).

[8] In fact, electorate ideology accounts for the inflection among Republican states.

below). Omitting the robustness weights definitely entails certain risks. At the very least, a nonrobust loess fit may generate a summary of the data that is more complicated than it really needs to be. At worst, a nonrobust loess fit can produce a misleading representation of the predominant structure within the data.

6. Plotting loess residuals

The residuals from a loess fit can be employed as a useful diagnostic tool in order to determine whether the smooth curve adequately incorporates all of the interesting structure in the data. The strategy for doing so is identical to that used in traditional, linear regression analysis. The residuals are scrutinized for systematic patterns that may remain after an hypothesized structural representation has been fitted to the empirical data.

The loess residuals are defined as the difference between the observed values of the *Y* variable, and the corresponding fitted values for the respective occurrences of the *X* variable values:

$$e_i = y_i - \hat{g}(x_i) \tag{1}$$

Eq. (1) is very similar to the familiar formula for calculating residual values in regression analysis. However, there is one important difference. The *m* evaluation points used to find the loess curve (the v_js) are imaginary values which are usually different from the *n* observed values of the independent variable, *X*. Therefore, the fitted values for the empirical observations, $\hat{g}(x_i)$ are typically obtained by interpolating between the two closest occurrences of the equally-spaced evaluation points.

Once the loess residuals are calculated, they are plotted against either the corresponding fitted values or (more commonly) the values of the original *X* variable. Then, a loess curve is fitted to the points within the residual plot. This new application of the loess smoother should produce a flat line located at the zero value on the vertical axis in the residual plot. The reasoning is as follows. The loess residuals measure the variability in *Y* that remains after the dispersion of the fitted values (and hence, the smooth curve) is taken into account. Any systematic functional dependencies between *X* and *Y* should be picked up by the original smooth curve fitted to the bivariate data. To the extent that the loess fitting process does so successfully, there should be no discernible patterns of any kind among the residuals; this, in turn, would produce a horizontal line when a smooth curve is fitted to the residual plot (Cleveland, 1993).

Fig. 6 shows the residual plot from the original loess curve that was fitted to the data on state education level and voter turnout (shown in Fig. 2). The points in this figure are obtained by plotting the loess residual values (on the vertical axis) against state high school graduation percentages (on the horizontal axis) for each state. The dotted horizontal line is a visual baseline, corresponding to a residual value of zero. The loess fit to these residuals is shown as a solid curve, which is fairly straight and horizontal. This result provides strong evidence that the simple curvilinear relationship depicted in Fig. 2 does provide an adequate representation of the struc-

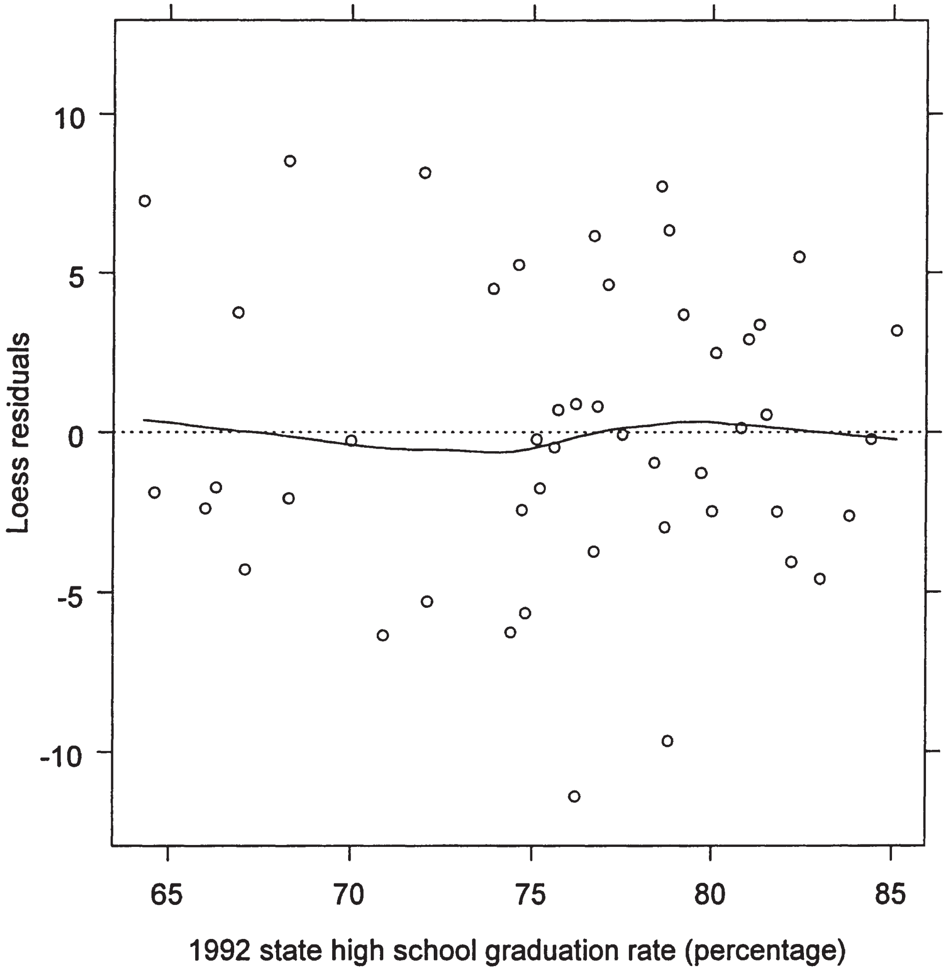

Fig. 6. Residual plot from loess curve fitted to state education and vote turnout data. Note: the values on the vertical axis are the residuals from the loess curve in Fig. 2 (that is, fitted to the data with α=0.65, λ=1, and robustness weights). The dotted horizontal line is a baseline, located at the origin on the vertical axis. The solid line is a loess curve, fitted to the residuals with α=0.75, λ=1, and robustness weights.

ture in the bivariate data. The original loess curve picked up all significant shifts in the central tendencies of the conditional *Y* distributions across the entire range of *X* variable values. There is no 'left-over' structure to be found within the residuals, so the loess curve that fitted to them in Fig. 6 looks nearly flat and featureless.

Residual plots can provide the analyst with useful guidance for controlling the loess fitting process (Cleveland 1993, 1994). This is particularly important for selecting the proper value of the smoothing parameter, α, because the appropriate λ value and the need for the robustness weights can often be determined through visual inspection of the original scatterplot. An effective general strategy for finding the best smoothing parameter value is to start with a curve based upon a relatively small α (say, 0.25). The residual plot for this smooth curve will almost always indicate an adequate fit.[9] The problem is that the fitted curve will not be very smooth, since some of this fit will involve the uninteresting noise component of the data. Therefore, gradually increase the α value (by about 0.10), repeat the loess estimation procedure, and check the new residual plot (including its own loess curve) for patterns. Continue this process until the residual plots begin to indicate that the loess curve is missing important features in the data points. The appropriate α value should be the one just before a non-horizontal loess curve begins to appear in the residual plot. The loess curve (in the scatterplot of the X and Y variables) associated with this α value should be the smoothest fit that still tracks all of the important structural features within the data.

Fig. 7 shows two examples of residual plots obtained from loess curves fitted with the wrong α value. Once again, the figure uses the data on state education and voter turnout. The first panel (Fig. 7A) shows the residuals obtained when α is set to 0.15. The loess curve fitted to these residuals is flat and located at the origin of the vertical axis, indicating that all of the structure has, indeed, been removed from the bivariate data. However, the 'over-fitted' loess curve that produced these residuals (shown in Fig. 3A) is unsatisfactory because it is very complicated and difficult to interpret. The second panel (Fig. 7B) shows the residuals obtained when a loess curve is fitted with α=1.00. Here, the data have been 'over-smoothed' and there is lack of fit, signalled by the clear, curvilinear pattern in the loess curve that is superimposed

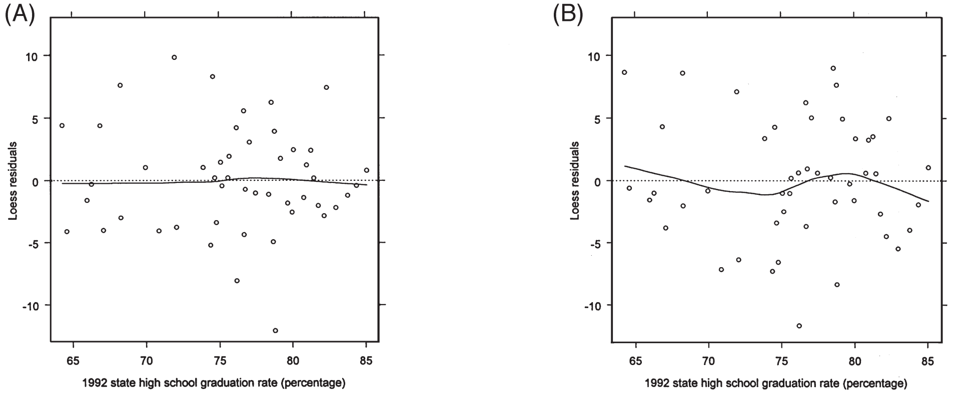

Fig. 7. Residual plots for loess curves fitted with the wrong α value: (A) α=0.15; (B) α=1.00. Note: the panels in this figure show residual plots obtained for the loess curves from Figs. 3A and D, respectively. In each panel, the dotted line is a baseline and the solid line is a loess curve fitted to the residuals, with α=0.75, λ=1, and robustness iterations.

[9] If the loess α parameter is set to a low value and the residual plot still indicates an inadequate fit to the data, then it is probably necessary to employ locally-quadratic fitting, robustness weights, or both in the estimation process. Again, however, this is usually apparent before the researcher gets to this stage of the analysis.

over the residuals. This reveals structure that was not incorporated into the smooth curve fitted to the original data. In general, the objective is to find a nonparametric fit that falls in between these two extremes, with a relatively simple curve that still represents the important features of the data. The loess curve fitted to the state voter turnout data with α=0.65 seems to achieve this goal.

It is difficult to provide useful guidelines for a priori selection of the α value. Typically, α values will fall somewhere between about 0.40 and 0.80, depending upon the nature of the bivariate relationship and the amount of noise that exists within the data. But, there are many exceptions to this generalization. In actual practice, the α value is almost always determined using the iterative process described here.

7. Goodness of fit for a loess smooth curve

When a loess smooth curve is fitted to data, attention is usually focused on the shape of the resultant curve because that feature is most revealing of the structure within the data. However, it is also useful to consider how well the smooth curve characterizes the empirical data values. This latter phenomenon is usually called ‘goodness of fit’, although that term is only partially appropriate in the case of non-parametric smoothers like loess.

A summary fit statistic similar to an R^2 value can be obtained by taking the ratio of the sum of squares in the loess fitted values to the total sum of squares in the dependent variable:

$$R^2_{\text{loess}} = \frac{\sum_{i=1}^{n}(\hat{g}(x_i) - \overline{\hat{g}(x_i)})^2}{\sum_{i=1}^{n}(y_i - \bar{Y})^2} \tag{2}$$

On the right-hand side of Eq. (2), $\hat{g}(x_i)$ is the loess fitted value for observation i, $\overline{\hat{g}(x_i)}$ is the mean of the loess fitted values, y_i is the dependent variable value for observation i, and $\bar{Y}$ is the sample mean for the dependent variable.

The R^2_{loess} for the smoothed relationship between high school graduate rate and voter turnout (as shown in Fig. 2) is 0.361. This indicates that the variance in the loess fitted values is slightly more than one-third the size of the total variance in state voter turnout figures. Such a relatively low R^2_{loess} value would lead to the conclusion that the smooth curve summarizes part, but not all, of the total dispersion in the dependent variable.

There are three important caveats that must be kept in mind whenever one tries to interpret an R^2_{loess} value. First, the R^2_{loess} cannot, strictly speaking, be interpreted as *variance explained* because the loess fitting procedure does not partition the total sum of squares in Y neatly into additive components representing the sums of squares in the fitted values and the residuals, respectively. Therefore, it is inappropriate to say that the R^2_{loess} value gives the proportion of variance that is, in any way, 'accounted for' by the *X* variable. Still, this does not imply that the statistic is useless or meaningless. Users should simply give the R^2_{loess} value a more limited interpretation. It conveys the size of the fitted value variance, expressed as a ratio of the total variance in *Y*. While the latter is not really variance explained in the traditional sense, it does provide an effective summary of the degree to which the loess fitted values track the empirical data points in the scatterplot.

Second, the R^2_{loess} statistic can produce misleading or even meaningless results when robustness weights are used in the fitting process. The difficulties arise when there are outliers in the data, but the exact nature of the problem can vary markedly. For example, univariate outliers in *Y*'s distribution can inflate the total sum of squares in *Y* relative to the sum of squares for the fitted values. In this sense, reliance on the traditional conception of *variance* (which is, of course, based upon summing the squared deviations from the center of a distribution) produces a misleading representation of the actual amount of interesting, meaningful *dispersion* within the data values. Alternatively, bivariate outliers can lead to a sum of squares for the fitted values that is larger than the total sum of squares for *Y*.[10] This would, in turn, produce an R^2_{loess} value larger than 1.00 which is uninterpretable. In both of these situations, the R^2_{loess} value would fail to produce an accurate measure of the degree to which the loess curve summarizes the empirical data.

Third, there is usually no reason to expect that the R^2_{loess} value will be particularly large. The very purpose of nonparametric smoothing is to uncover structural patterns within relatively noisy data. This implies that the residuals from any smooth curve fitted to the data points will tend to be fairly large and that the value of any goodness-of-fit statistic will be small. Stated differently, if the nature of the functional dependency between the *X* and *Y* variables was clear, then there would be no need for the loess fitting procedure in the first place.

The general problem is that the loess smoother differs from parametric fitting procedures (like OLS regression) in a fundamental way: it does not fit a particular, narrowly-defined model to the data (Weisberg, 1996). Therefore, the very concept of 'goodness of fit' is problematic. With a nonparametric smoother, the variance explained in *Y* is less important than the degree to which the resultant smooth curve follows the prominent features of the bivariate data. As a result, the R^2_{loess} is seldom reported in the results of analyses that employ the loess fitting procedure.

8. Loess and statistical inference

The discussion so far has assumed that loess is being used as a strictly descriptive tool. However, the statistical theory for local regression models has been worked out, so it is possible to incorporate an inferential component into a loess analysis.

[10] Bivariate outliers can be defined as observations that fall within the extreme tails of the *conditional* distribution of *Y*, given a particular x_i value. These kinds of outliers are often not apparent in the univariate distributions of *X* or *Y*.

Doing so facilitates generalizations about the structure of the population from which the observed data were drawn. Inferential tools also enable the researcher to assess the degree of uncertainty about the precise form of the smooth curve fitted to the bivariate data.

Statistical inference with a loess smooth curve is usually grounded in least-squares theory (Cleveland and Devlin, 1988), and it requires several assumptions. Specifically, the observed fitted values, $\hat{g}(x_i)$, are now viewed as estimates that should approximate, as closely as possible, the true but unobserved fitted values, $g(x_i)$. Furthermore, the residuals about these fitted values should be gaussian. That is, the $y_i - g(x_i)$ should be independently and identically distributed according to a normal distribution, with a mean of zero and a constant variance. When these assumptions are met, direct generalizations of traditional least-squares methods can be employed to perform statistical tests.

For example, analysis of variance can be used to compare two nested loess fits to a common dataset. Assume that 'loess,1' refers to a smooth curve fitted using parameters that are nested within those of another smooth curve, designated 'loess,2'. Then, just as in traditional linear regression, a test statistic can be constructed to measure the improvement in fit across the two loess curves (adjusted by degrees of freedom) relative to the lack of fit remaining in the more complicated of the two loess curves:

$$F = \frac{(\mathrm{RSS}_{\mathrm{loess},1} - \mathrm{RSS}_{\mathrm{loess},2})/(\mathrm{df}_{\mathrm{loess},2} - \mathrm{df}_{\mathrm{loess},1})}{(\mathrm{RSS}_{\mathrm{loess},2})/(n - \mathrm{df}_{\mathrm{loess},2})} \tag{3}$$

In Eq. (3), $\mathrm{RSS}_{\mathrm{loess},1}$ and $\mathrm{RSS}_{\mathrm{loess},2}$ are the sums of squared loess residuals for the two curves while $\mathrm{df}_{\mathrm{loess},1}$ and $\mathrm{df}_{\mathrm{loess},2}$ are the degrees of freedom associated with the respective curves. As usual, n is the total number of observations. Under the null hypothesis of no improvement in fit across the two models, the preceding test statistic follows an F distribution with $(\mathrm{df}_{\mathrm{loess},2} - \mathrm{df}_{\mathrm{loess},1})$ and $(n - \mathrm{df}_{\mathrm{loess},2})$ degrees of freedom. A special case of Eq. (3) arises when a single loess curve is tested against the null hypothesis of no functional dependence between Y and X:

$$F = \frac{(\mathrm{TSS}_{\mathrm{Y}} - \mathrm{RSS}_{\mathrm{loess}})/(\mathrm{df}_{\mathrm{loess}} - 1)}{(\mathrm{RSS}_{\mathrm{loess}})/(n - \mathrm{df}_{\mathrm{loess}})} \tag{4}$$

In Eq. (4), TSS_Y is the total sum of squares in Y and the remaining terms are defined as in Eq. (3). The various residual sums of squares are calculated just as they are in linear regression. But, the degrees of freedom associated with the loess curves are a bit trickier. They do not correspond to readily observable quantities (like the number of independent variables in a linear regression equation) and the computational details are fairly complex. But, the calculations are direct mathematical generalizations of those employed in OLS regression (Cleveland et al., 1988, 1993). And, the degrees of freedom have a straightforward interpretation: They correspond to the *equivalent number of parameters* for the loess curve. This is the number of terms that would be required to produce a parametric function similar to the fitted loess curve. A curve with $\mathrm{df}_{\mathrm{loess}}=2$ looks like a quadratic function (a curve with one reversal in direction), a curve with $\mathrm{df}_{\mathrm{loess}}=3$ approximates a third-degree polynomial (a curve with two inflection points), and so on. Note, however, that the df_{loess} values are usually noninteger quantities, so the interpretation should only be regarded as a heuristic convenience.

Let us go back to the loess curve from the voter turnout data (in Fig. 1). Inspection of the residuals indicates that they do, in fact, conform to the necessary assumptions even though the curve was fitted with robustness iterations.[11] So, the first inferential question is whether this curve really measures anything more than 'noise' due to sampling variability. The TSS_Y=1612.32, RSS_{loess}=985.17, df_{loess}=3.5, and n=48. Substituting these values into Eq. (4):

$$F=\frac{(1612.32-985.17)/(3.5-1)}{(985.17)/(48-3.5)}=11.33$$

The critical value of the F statistic at the 0.05 level, with degrees of freedom rounded to 4 and 44, is only 2.58. Thus, the null hypothesis of no relationship can be rejected.

The next question is whether the curvilinear loess fit provides any improvement over the simpler, linear fit obtained through OLS. Fox (1999) points out that a linear model is nested within a nonlinear model for a given dataset. Therefore, Eq. (3) can be used to compare the loess curve and the OLS regression fitted to the state voter turnout data. Fig. 8A shows the OLS line for the dataset. The residual sum of squares for this OLS fit is 1106.34, and there are 2 degrees of freedom for a bivariate regression equation. These values are substituted for $RSS_{loess,1}$ and $df_{loess,1}$ in Eq. (3). The values for $RSS_{loess,2}$ and $df_{loess,2}$ are taken from the bivariate loess fit, so they

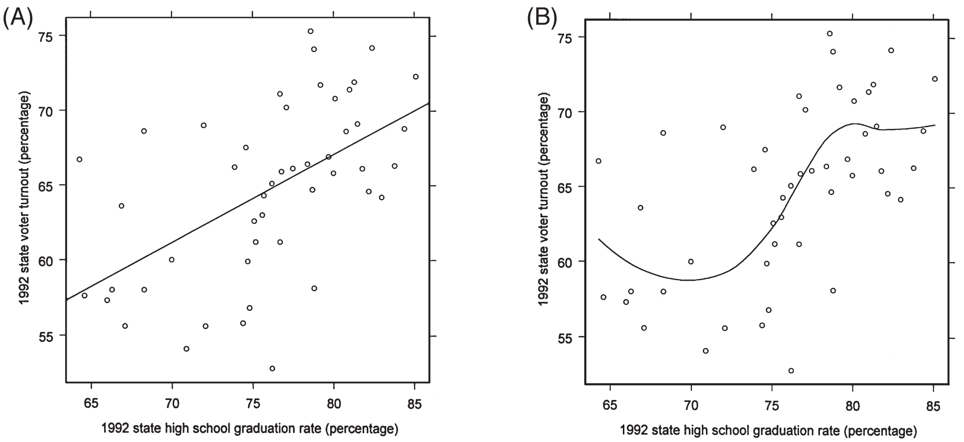

Fig. 8. Alternative smooth curves fitted to 1992 state education levels and voter turnout rates: (A) OLS regression line; (B) loess curve fitted with α=0.65, λ=2 and robustness iterations.

[11] Specifically, they appear to approximate a normal distribution, and the spread does not change very much across the range of the independent variable. Details on the loess residual diagnostics for these data are available from the author.

are 985.17 and 3.5, just as before. Substituting these values into Eq. (3) produces the following F test statistic:

$$F=\frac{(1106.34-985.17)/(3.5-2)}{(985.17)/(48-3.5)}=3.65$$

The critical value of F for the 0.05 level, with 2 and 44 degrees of freedom (once again, rounding to integers), is 3.21. Thus, we can safely reject the null hypothesis of no improvement in fit in the bivariate loess curve, relative to the linear regression model. Stated somewhat differently, the nonlinearity in the functional dependence of voter turnout on education levels is probably not due to noise variability emanating from sampling error.

One could next ask whether quadratic fitting improves the representation of the data. In order to address this question, we can fit a loess curve to the turnout data with α=0.65 and λ=2; this is shown in Fig. 8B. The RSS_{loess} for the local-quadratic curve is 980.14 and df_{loess}=5.6. The inflections are a bit more pronounced in this curve, but the basic shape remains quite similar to that shown in Fig. 2. The loess curve with λ=1 is nested within the curve fitted with λ=2, so the various summary statistics and degrees of freedom for the two curves can be used to calculate the F statistic as follows:

$$F=\frac{(985.17-980.14)/(5.6-3.5)}{(980.18)/(48-5.6)}=0.10$$

The critical value of F at the 0.05 level, with 2 and 43 degrees of freedom (once again, rounding to integers), is 1.65. So, the null hypothesis cannot be rejected. The more complicated loess curve obtained through quadratic fitting does not seem to provide a significant improvement over the simpler curve, obtained with λ=1.

Once the appropriate fitting parameters have been determined, it is often useful to construct a confidence interval around a loess smooth curve. In principle, this is accomplished by viewing each loess fitted value as a predicted value from a regression equation. For each of the m predicted values (recall that one predicted value is obtained for each of the v_j), the standard error is calculated, and it is multiplied by the appropriate t value in order to find the confidence limits for that $\hat{g}(v_j)$. This procedure is carried out for the entire set of m fitted values. Then, all adjacent upper and lower confidence limits are simply connected with line segments in order to produce the final confidence band.[12]

Fig. 9 shows the 95% confidence band for the loess smooth curve fitted to the data on state welfare spending and electorate ideologies. Notice that the band is narrowest in the center of the point cloud, and becomes wider near the edges of the plotting region. This is basically a curved version of the 'hourglass' shape that would

[12] Strictly speaking, the 'pointwise' confidence limits described here do not define the 'global' confidence band for the full loess curve. However, the former are much easier to calculate than the latter. Furthermore, they seem to work very well, in practice, for summarizing the uncertainty involved in the location of the loess curve (Hastie and Tibshirani, 1990; Beck and Jackman, 1997).

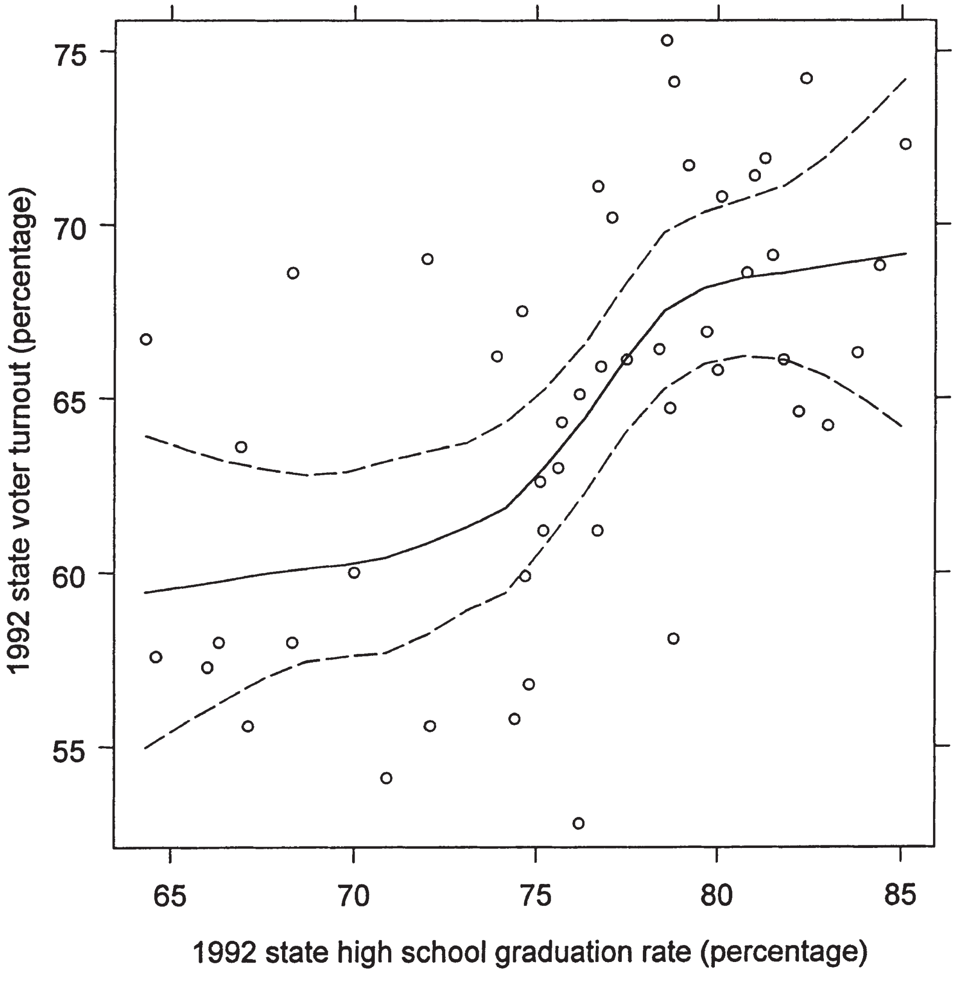

Fig. 9. 95% confidence band around loess smooth curve fitted to data on state education and voter turnout rates. Notes: the solid line is the loess curve, fitted with α=0.65, λ=1, and robustness weights. The dotted lines show the limits of the 95% confidence band. The band is constructed from the pointwise least-squares confidence limits for the loess fitted values.

be obtained with a traditional regression analysis of these data, and it reflects the greater uncertainty about observations that fall near the edges of the independent variable's distribution. The band generally follows the contours of the loess curve, providing visual confirmation of the results from the first *F* test reported above. From this evidence it appears unlikely that the nonlinear features in the curve (e.g. the relatively flat segments near the left and right sides of the plot, along with the relatively steep portion of the curve in the middle) are due to sampling fluctuations alone.

There are at least two potential problems with the preceding approach to statistical inference with loess. First, the assumption of gaussian residuals is relatively stringent;

it seems to contradict the general ‘spirit’ of nonparametric smoothing, which is based upon minimal a priori specifications about the structure within the data (Tukey, 1977). Second, the loess residuals will often be non-gaussian when robustness weights are included in the estimation procedure (as they were in the original smooth curve fitted to the state voter turnout data), particularly when the data contain serious outliers. Therefore, it is important to investigate the loess residuals very closely, in order to determine whether the gaussian assumptions are tenable. If they are not, then alternative strategies for statistical inference should be employed. For example, Efron and Tibshirani (1993) demonstrate the use of a bootstrap confidence band around a loess smooth curve, and this approach shows great promise for more complex local regression models as well. The general point is that statistical inference is possible with a loess smooth curve. The widely-held belief that loess is strictly an exploratory tool is, to put it bluntly, wrong. Indeed, the statistical tractability of loess is precisely one of the features that provides it with powerful advantages over other nonparametric smoothing procedures.

9. Loess and multivariate data

Although the discussion so far has focused on bivariate scatterplot smoothing, loess can also be a useful tool for situations where a dependent variable is hypothesized to be a function of several independent variables. In fact, there are at least two different approaches that can be used: *multivariate loess* (or, more precisely, ‘local multiple regression’) and *generalized additive models*. Let us briefly consider each of these strategies.

9.1. Multivariate loess

The principles of multivariate loess smoothing are identical to the bivariate case. The only changes involve relatively minor details. Specifically, the fitting window, the distances from the evaluation points (the v_js) to the observations (the x_is), and the local weights (the w_{ij}s) are now calculated within a k-dimensional subspace spanned by the k independent variables, X_1, X_2,..., X_k, rather than along the single horizontal dimension of the bivariate scatterplot.

The general objective of the multivariate loess fitting procedure is the same as most other multivariate modeling strategies: to assess the functional dependence of a dependent variable on each of a set of independent variables, while simultaneously taking the effects of all other independent variables into account. The end result of a multivariate loess analysis is a smooth surface that tends to coincide with the center of the data point cloud, throughout the entire k+1 dimensional space formed by the independent variables X_1 through X_k and the dependent variable, Y. The considerations about loess fitting parameters, robustness weights, residual plots, goodness of fit, and statistical inference all remain identical to those discussed earlier, for the bivariate case.

The most difficult aspects of multivariate loess do not stem from the estimation procedure; rather, they involve the presentation of the final results. The problem is that the central output from a loess fit is graphical, rather than numerical in nature. And, it can be very tricky to produce an accurate and easily-interpretable visual representation of multivariate information in a static, two-dimensional display medium like a printed page or a computer screen.

If there are two independent variables, then the fitted loess surface can be illustrated directly, using a three-dimensional wireframe plot. Specifically, the fitted values, $\hat{g}(v_j)$, are plotted in three-dimensional space, for each evaluation point, v_j, in the plane formed by the two independent variables; if there are m evaluation points for each independent variable, then there will be a total of m^2 plotted points. Adjacent points parallel to each of the independent variable axes are connected with line segments, thereby forming the wireframe surface.[13]

Fig. 10 shows an example of a wireframe plot, using 1992 state voter turnout as a function of high school graduation rates and the proportion of African-Americans in the state population. Note that the latter variable is logged in order to correct for positive skewness in its distribution. This graph shows that the state-level relationship between education and voter turnout remains nonlinear, even after state racial composition is taken into account. However, the nature of the curve varies somewhat: it 'flexes' in opposite directions while moving from the front-right facet of the plotting cube to the left-rear facet. In states with small African-American populations, voter turnout increases with high school graduation rates, until the latter reach a value of about 75%. After that, further increases in education have no apparent effect on voter turnout. Among states with relatively large black populations, low education levels are unrelated to voter turnout. But, when high school graduation rates reach 70% or so, the slope of the surface turns sharply upward. Thus, in states with high proportions of African-Americans, high levels of education are strongly related to voter turnout rates.

[13] The great attraction of three-dimensional wireframe plots is their ease of interpretation, even by relatively unsophisticated observers. Nevertheless, these graphs do have some drawbacks: First, the perspective view used to create the illusion of three dimensions may impair accurate visual perception of the quantitative information in the plotting region (Cleveland and McGill, 1984). A second, related concern is that details of the functional dependence among the variables may be hidden by 'folds' in the loess surface. Third, it is almost impossible to include the data points in the three-dimensional plot along with the wireframe surface; doing so produces a very cluttered graph that is quite difficult to interpret. And fourth, the wireframe plot is obviously limited to analyses with only two independent variables. Fortunately, all of the preceding problems can be handled very easily. For example, a *dynamic* display, i.e. a three-dimensional graph in which the plotting region seems to spin or rotate, enables the analyst to change the viewing perspective in real time. This overcomes most of the limitations involving visual perception of the information in the graph (Becker et al., 1988). At the same time, *conditioning plots* can be used to handle multiple independent variables. These are multipanel graphical displays that show the relationship between an independent variable and a dependent variable at specific values of one or more additional independent or control variables. Conditioning plots literally 'hold other variables constant' while illustrating the functional dependence of one variable on another. Thus, there are several strategies that can be used to 'look into' the multidimensional space that is formed by a given set of multivariate data (Cleveland, 1993; Jacoby, 1998); all of these are potentially useful for showing the results of a multivariate loess analysis.

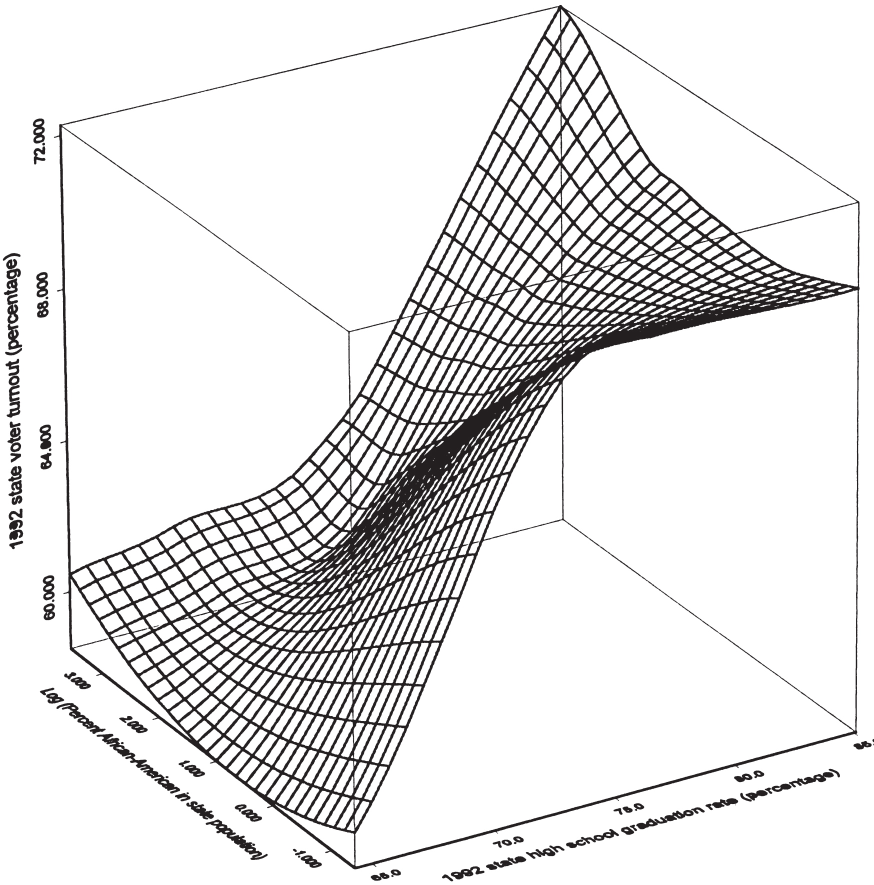

Fig. 10. Multivariate loess surface summarizing the functional dependence of 1992 state voter turnout on high school graduation rate and logged percentage African-American in state population. Notes: the data are obtained from the 1993 Statistical Abstract of the United States. The loess surface is fitted with α=0.65, λ=1, and robustness weights.

The relationship between logged percentages of African-Americans and voter turnout, controlling for education, is relatively weak. At the lowest and highest levels of education (i.e. near the front-left and back-right facets of the plotting cube), turnout seems to be highest among states with large *or* small black populations, while it is somewhat lower among states with intermediate numbers of African-Americans. Among states with intermediate high school graduation rates (i.e. around 75%), the relationship between race and turnout seems to be negative: within this subset of states, voter turnout declines with increasing log proportions of African-Americans. Again, however, there is a slight upturn among those states with the largest black populations. Overall, the predominant slope of the loess curve in Fig. 10 shows that state education level has a much more pronounced effect on state voter turnout than does the racial composition of the state population.

The statistical inference strategies discussed in the previous section can be applied to multivariate loess surfaces. The loess curve fitting turnout to high school graduation rates is nested within the surface fitted with graduation rates and logged proportion African-Americans. So, we can examine whether the addition of this second independent variable really provides a significant improvement over the original, bivariate loess curve. The multivariate loess surface shown in Fig. 10 produces an RSS_{loess} of 923.11. The df_{loess} is 5.6. Substituting these values, along with the corresponding quantities from the bivariate loess curve, into Eq. (3) produces:

$$F=\frac{(985.17–923.11)/(5.6–3.5)}{(923.11)/(48–5.6)}=1.36$$

The critical value of F at the 0.05 level, with 6 and 42 degrees of freedom (once again, rounding to integers) is 2.32. Since the empirical test statistic fails to exceed this value, we cannot reject the null hypothesis. Thus, the racial composition of a state's population does not seem to make a significant contribution to the explanation of voting turnout rates, beyond that based upon education levels alone.

From a conceptual perspective, multivariate loess is a straightforward generalization of bivariate smoothing and multiple regression analysis. In practice, however, there are several caveats and potential problems that must be recognized. First, the multivariate loess surface is usually quite complex; that is, the smooth curve with respect to one variable changes across all combinations of values on the other variables. Second, the n data points are generally spread out very sparsely throughout the k+1 dimensional space that contains them. This means that the local fitting window must cover a large area, even though doing so compromises the 'local' nature of the smooth curve. Third, the jury still seems to be out on the utility of multivariate graphics. Some authors are enthusiastic proponents of visualization strategies for multivariate data (e.g. Becker et al., 1996; Jacoby, 1998); others are more circumspect about the complexity of such graphical displays and the difficulties involved in their interpretation (e.g. Wainer, 1988; Fox, 1999). Because of limitations like these, multivariate loess fitting will probably never become a widely-used modelling strategy in the social sciences.

9.2. Generalized additive models

In contrast to multivariate loess, generalized additive models take a somewhat more complicated computational approach in order to generate graphical depictions of functional dependence that are easier to view and interpret. The basic form of a generalized additive model depicting the relationship between variables Y and X_1, X_2,...., X_k is as follows (for observation i):

$$\eta(Y_i)=\alpha+\sum_{j=1}^{k} f_j(X_{ij})+\varepsilon_i \qquad (5)$$

The η term on the left-hand side of Eq. (5) represents a transformation of Y called a *link function*. It specifies how the conditional mean of the dependent variable is related to specific values of the independent variables. Examples include the logit and probit functions for categorical variables. For present purposes, we will focus on the identity link, which simply specifies that $\eta(Y)=Y$. This is a reasonable choice when the dependent variable is relatively continuous and measured at the interval level. In that case, Eq. (5) is sometimes called an *additive regression model*.

On the right-hand side of Eq. (5), α is an intercept, and ε_i is a disturbance (which is subject to the usual assumptions). The most interesting terms are those contained within the summation, which represent the effects of the k independent variables. Specifically, each f_j is a smooth function of its corresponding X_j. The specific nature of these functions is left entirely to the analyst and loess is often used when nonparametric fitting is desired. In any event, these functions are estimated in a way that optimizes a fitting criterion, such as minimizing the sum of squares in the estimated disturbances.

The distinctive feature of an additive model is that it expresses the *multivariate* structure in the data as a sum of k different *bivariate* relationships, each of which expresses the impact of an X_j on Y with the effects of the other $k-1$ independent variables removed. Stated somewhat differently, for any observation, i, the value of the dependent variable (Y_i) is a sum of k loess fitted values, along with the intercept and the disturbance term for that observation:

$$Y_i=\alpha+\hat{g}_1(x_{i1})+\hat{g}_2(x_{i2})+\ldots+\hat{g}_j(x_{ij})+\ldots+\hat{g}_k(x_{ik})+\hat{\varepsilon}_i \qquad (6)$$

In Eq. (6), $\hat{g}_1(x_{i1})$ is the fitted value from the bivariate loess of Y on X_1, with the effects of X_2 through X_k removed; $\hat{g}_2(x_{i2})$ is the fitted value from the bivariate loess of Y on X_2, with the effects of X_1 and X_3 through X_k removed, and so on.

The major output from an additive model is a set of k scatterplots. Each one of these contains a smooth curve (e.g. a loess curve) depicting the net impact of one independent variable on the dependent variable (again, controlling for the remaining independent variables). In other words, for independent variable X_j, the graph may contain the data points, the (x_{ij}, y_i*); the asterisk indicates that the effects of the other independent variables have been removed from Y via the additive modelling estimation procedure (which is usually called ‘backfitting’). Note that these points are often omitted, since the values plotted on the vertical axis are not directly interpretable (see below). The graph will definitely plot the points that define the smooth curve, that is the $(x_{ij},\hat{g}_j(x_{ij}))$, and connect the adjacent points with line segments to form a ‘smooth’ curve. The curve is then interpreted just as in any other bivariate relationship.

The additive modelling strategy can be used to examine the impact of high school graduation rates and African-American proportion on 1992 state voter turnout levels. Specifically, turnout rates are fitted to the education variable using a loess curve with α=0.65 and λ=1, and to the racial composition variable using a loess curve with α=0.5 and λ=2 (these values of the fitting parameters were obtained through trial and error). Fig. 11 shows the two scatterplots that are produced by this additive model. In each panel, the horizontal axis shows the scale of values for the corre-

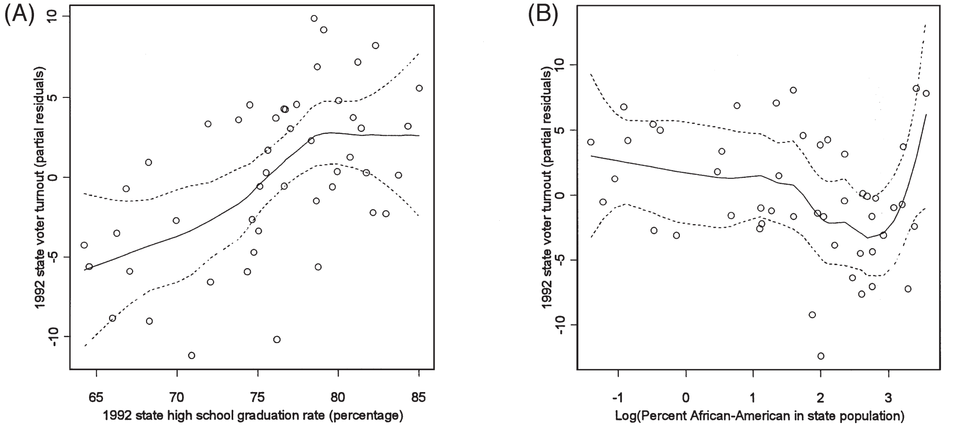

Fig. 11. Loess curves fitted to 1992 state voter turnout data by additive regression model: (A) the relationship between state education levels and state voter turnout; (B) the relationship between proportion of African-Americans in the state population (logged) and state voter turnout.

sponding independent variable. The vertical axes in both panels are labelled as voter turnout rates; but, it is important to emphasize that these *do not* correspond to the raw *Y* values. Instead, they show the 'partial residuals'; that is, transformations of the *Y* values that remove the effect of the other independent variable, and center the values to a mean of zero. The fitted loess curves (shown by the solid line in each panel) are drawn into each plot, along with the approximate 95% confidence bands (dotted lines).

Very briefly, the graph in Panel A of the figure suggests that turnout rates increase with education levels in states with relatively low to moderate high school graduate rates. However, in states with high graduation rates (i.e. those above 78% or so), this effect levels off. Notice that the 'floor effect' which was evident in Fig. 2 (i.e. the relatively flat portion of the loess curve in the lower-left region of the graph) largely disappears when the racial composition of state populations is taken into account. Turning to Panel B, race shows a very slight negative effect on voter turnout in most states. But, there is a distinct drop, followed by an even sharper increase in turnout among states with the highest proportions of African-Americans. Closer inspection of the plot reveals that these latter features are largely due to the influence of two unusual observations (the upper-rightmost points in the display). Overall, the effect of state racial composition is not statistically significant. Thus, the substantive conclusions are quite similar to those reported from the multivariate loess surface. Here, though, it is much easier to convey the results, it being necessary only to inspect two simple scatterplots, rather than to describe an undulating surface within a three-dimensional space.

A broader introduction to generalized additive models is contained in Beck and Jackman (1997), while Hastie and Tibshirani (1990) provide a more comprehensive and technical treatment of the topic. For now, an informal statement of the connection between generalized additive models and loess might be the following. The backfitting estimation procedure underlying the additive modelling approach isolates the effects on Y that are associated with each of the k independent variables. Once this is accomplished, loess can be used to summarize and illustrate the functional dependence of Y on each of the X_js. The important point is that bivariate loess fitting can be used, through the mechanism of additive modelling, to summarize multivariate structure within data.

10. Software for loess

Because of its computationally intensive nature, loess smoothing is effectively impossible to carry out by hand. Therefore, most potential users (at least non-programmers) are constrained by the options that are provided by the available software. Fortunately, loess fitting is now widely incorporated into statistical software packages. However, the exact nature, capabilities and flexibility of the routines vary markedly from one program to the next.

Some packages only provide basic scatterplot smoothing, usually as a graphical enhancement to their bivariate scatterplot routine. Thus, SPSS for Windows (Norušis, 1993) and SYSTAT (Wilkinson, 1998) both have LOWESS options that provide robust, locally-linear fitting.[14] The user can specify the value for α, but none of the other fitting parameters can be changed.

Other software packages provide a bit more flexibility in the loess smoothing process. STATA (StataCorp, 1997), SAS/INSIGHT — a dynamic, interactive graphics module in the SAS system (SAS Institute Inc., 1995), and the R-CODE — a small but powerful program for regression graphics (Cook and Weisberg, 1994) all contain bivariate loess fitting.[15] SAS/INSIGHT and the R-CODE allow interactive specification of α, using a slider bar displayed along with the scatterplot. SAS/INSIGHT also contains several options for automatically selecting the 'best' α value for the data. All three of these programs output loess fitted values. Hence, users can calculate loess residuals and construct the various diagnostic plots that are based upon them.

But once again, each of these packages has limitations. For example, STATA and the R-CODE are both restricted to locally-linear fitting (SAS/INSIGHT allows the user to set λ to either 1 or 2). Similarly, none of these programs allows the user any choice with respect to robust estimation. The R-CODE always includes the robustness weights, while STATA and SAS/INSIGHT do not. Thus, STATA,

[14] SYSTAT and SPSS are both available from SPSS, Inc., 444 North Michigan Avenue, Chicago, IL 60611-3962, USA. Worldwide web: http://www.spss.com.

[15] STATA is available from the Stata Corporation, 702 University Drive East, College Station, TX 77840, USA. Worldwide web: www.stata.com. SAS/INSIGHT is part of the SAS system, which is available from SAS Institute Inc., SAS Campus Drive, Cary, NC 27513, USA. Worldwide web: www.sas.com. The R-CODE is a copyrighted software package that is available to anyone who purchases the book 'An introduction to regression graphics' by R. Dennis Cook and Sanford Weisberg. Full citation information is given in the references and the web site is http://www.stat.unm.edu/-rcode/.

SAS/INSIGHT, and the R-CODE all go beyond basic scatterplot smoothing, but none of them really provide the user with complete control over the loess fitting process.

EViews (Quantitative Micro Software, 1997)[16] — a package designed primarily for econometric analysis — contains a very flexible loess routine for enhancing scatterplots. Users can specify virtually all of the loess fitting parameters, including the α and λ values, the use of robustness weights, and the number of evaluation points. The routine can also output loess fitted values. Thus, EViews overcomes most of the limitations in the packages mentioned above. Like the others, however, it is limited to descriptive applications and bivariate data.

Currently, and probably for the foreseeable future, the most powerful loess software is contained in S-PLUS (Statistical Sciences Inc., 1995).[17] This package provides the analyst with control over virtually all aspects of the loess fitting process. S-PLUS also contains two other features that set it apart from all other statistical software packages. First, it contains an extensive set of tools for statistical inference with loess smooth curves and surfaces. Second, S-PLUS is the only package that currently estimates generalized linear models and performs loess smoothing for multivariate data. All of these capabilities, along with a set of very powerful and flexible graphics routines, make S-PLUS truly the 'software of choice' for loess analyses.

There is one other software possibility that should be mentioned. For anyone with even a minimal amount of programming experience, it is a fairly straightforward task to write special-purpose loess routines in a programming language or in the programming environments that are now included within several statistical software packages (Stine and Fox, 1996). Examples of the former include LISP-STAT (Tierney, 1990) and APL2STAT (e.g. Fox and Friendly, 1996). The primary examples of the latter include PROC IML (for *I*nteractive *M*atrix *L*anguage) in SAS (SAS Institute Inc., 1990), ADO files in STATA, and the MATRIX routine in SPSS for Windows. Various analysts have already made a great deal of progress along these lines, and their programs are readily available to interested users. For example, AXIS, a graphical user interface for LISP-STAT, contains commands for scatterplot smoothing with bootstrap resampling (Stine, 1996). And, a SAS/IML macro for bivariate loess has been published by Friendly (1991).[18] Finally, I have written a number of SAS/IML routines for loess fitting (e.g. bivariate smoothing with user control of the fitting parameters, multivariate loess smoothing, bootstrap confidence intervals for a loess curve, etc.), and they are available to readers upon request.

It is important to emphasize that software discussions on any subject tend to become outdated very quickly. This is particularly true with a topic like loess, since graphical presentations of quantitative data currently represent a major area of devel-

[16] EViews is available from Quantitative Micro Software, 4521 Campus Drive, Suite 336, Irvine, CA 92612, USA. Worldwide web: http://www.eviews.com.

[17] S-PLUS is available from Statistical Sciences Inc., a Division of MathSoft, 1700 Westlake Ave., N. Seattle, WA 98109, USA. Worldwide web: http://www.mathsoft.com/splus.html.

[18] This macro is also available through his Worldwide Web home page, at: http://www.math.yorku.ca/SCS/friendly.html.

opment in the statistical sciences. Nevertheless, the brief discussion presented in this section clearly demonstrates that software for loess smoothing and fitting is already widely available. One can only expect that the routines will continue to evolve, becoming more flexible and powerful in the future.

11. Conclusions

Loess has recently received a great deal of attention in statistical circles, where it is recognized as one member from a broader family of procedures called *nonparametric regression models* (Green and Silverman, 1994; Fan and Gijbels, 1996; Fox, 1999). However, loess is far less well known among political scientists. This is unfortunate, because it provides a very flexible approach to the problem of representing structure within a dataset. Accordingly, loess fitting is a useful addition to the social scientist's repertoire of techniques for investigating empirical data.

When loess is employed merely as a scatterplot smoother, it can be very helpful for a number of important research tasks, including the exploration of bivariate and multivariate data, assessment of functional forms for relationships among variables, examination of model assumptions in regression analysis, and representation of complex structures within empirical data. But, extensions of basic loess fitting have utility beyond these relatively simple applications. Indeed, I would argue that loess can be regarded as a fairly comprehensive strategy for modelling functional dependence in many kinds of data analysis contexts.

At the same time, it is important to emphasize that loess is not, in any way, a panacea. Nonparametric smoothers like loess should be viewed as a complement to, rather than a replacement for, traditional parametric smoothing methods (e.g. OLS). These two general approaches — parametric and nonparametric smoothing — should be used together, since they each have their own distinctive strengths and weaknesses.

The strengths of parametric smoothing include the following. Procedures based upon this approach usually result in an equation that provides a concise, easily-understandable summary of the structure underlying the data (assuming the specified model fits well enough). The fitting methods are well-known and understood. And there is usually statistical theory available to incorporate sampling variability into the precision of the smoothed fit.

These strengths of parametric smoothing strategies are offset by some weaknesses. Parametric fitting procedures are not very flexible. That is, they do not handle misspecifications very well. At the same time, they rely on global, rather than local, fitting routines. This makes the final smooth curve relatively sensitive to outliers, discrepant observations, and patterns that may only exist within a limited subset of the data.

The strength of nonparametric smoothers is that they are very flexible about the exact nature of the relationship between the variables. They are also relatively local, in that the position and orientation of the fitted curve in any vicinity of X values is primarily dependent upon the data points in that vicinity. Observations with X values that are relatively distant have little or no effect on that segment of the curve (although they do, of course, affect the placement of the smooth curve in their own vicinity). This property tends to make nonparametric smoothers less sensitive to discrepant points within the scatterplot.

The major weakness of nonparametric smoothers is that they cannot be used to characterize the data in terms of a simple equation. Also, the statistical theory is less well-developed for nonparametric smoothing algorithms than for more traditional fitting methods. And finally, nonparametric fitting methods require the analyst to make several partially arbitrary decisions about the fitting parameters.

Nevertheless, nonparametric smoothers still let the data 'speak for themselves' to a greater extent than traditional, parametric methods. This is particularly important in the field of elections and voting behavior. Theories often lead to expectations of nonlinear empirical relationships, but prior substantive considerations usually provide very little guidance about precise functional forms. Loess smoothing can be used to resolve precisely this kind of dilemma.

Acknowledgements

I would like to thank Harold D. Clarke and Harvey Starr for their comments and suggestions on an earlier version of this paper. Special thanks go to Saundra K. Schneider; this project could not have been completed without her help.

Appendix. The details of fitting a loess smooth curve

Fig. 12 shows an example of loess smoothing, using hypothetical bivariate data on 20 observations. The first panel (Fig. 12A) shows the basic scatterplot. In order to keep the example simple, the data form a clear, curved pattern within the plot. The steps of the loess fitting procedure are as follows.

A.1. Preliminaries

First, define m equally-spaced locations across the range of X values. Call these v_j, where the subscript j ranges from 1 to m. The loess curve will be evaluated at each of the v_js. The loess fitted value (i.e. the vertical coordinate for the curve) at v_j is designated as $\hat{g}(v_j)$.

Second, supply values for two parameters, α and λ. α is a value between 0 and 1, which gives the proportion of observations that are used in each of the local regressions. This is used to find the number of observations used in each local regression (called 'the window'). The latter is defined as $q=\alpha n$, where n is the number of empirical data points, and q is truncated to an integer value, if necessary. The value of λ is either 1 or 2; it gives the degree of the polynomial that is actually fitted to the data.

For the example data, there are 21 v_j values (that is, m=21), uniformly spaced in

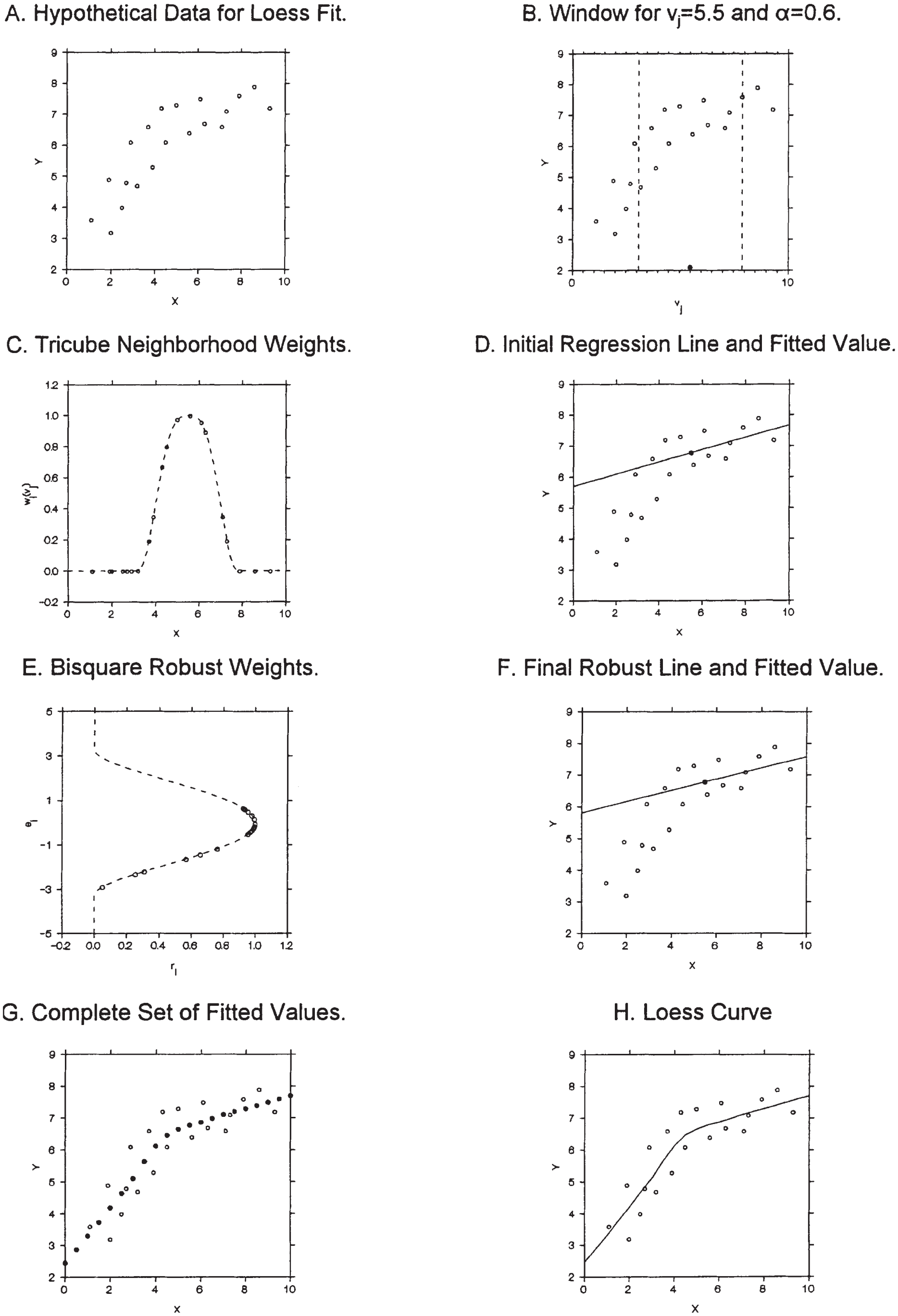

Fig. 12. Illustration of loess fitting procedure using hypothetical data.

the closed interval from 0 to 10.0; hence, v_1=0.0, v_2=0.5, and so on, up to v_{21}=10.0. The locations of the various v_js are shown as tick marks on the horizontal scale of Figure 12B. Note that the v_js would usually range only across the observed values of the X variable. For this example, the evaluation points are evenly spaced across the interval from 0 to 10, in order to keep the v_js at simple, round values.

In this case, α is set to 0.6, so αn=12. Thus, the window will always enclose the 12 empirical data points that fall closest to the current v_j on the horizontal axis. Figure 12B shows the window for v_j=5.5 (an arbitrarily-selected value, located by the asterisk along the horizontal axis in the figure). Note that the physical width of this window will change at different values of v_j, as the distance to the twelfth closest data point changes. Also, evaluation points close to either the maximum or minimum X values will be asymmetric; to pick some obvious examples, v_1 will have all 12 points in its window arrayed to the right, while v_{21} will have all 12 points arrayed to its left. In any event, the window will always contain the 12 closest empirical data points, regardless of their direction and/or distance from v_j.

A.2. For each v_j, calculate neighborhood weights

Let $\Delta_{[i]}(v_j)=|x_i-v_j|$, the distance from the point of evaluation (v_j) to the ith observation, x_i. The brackets around the subscript indicate that these distances are sorted from smallest to largest.

Then, $\Delta_{[i]}(v_j)^*=\Delta_{[i]}(v_j)/\Delta_{[q]}(v_j)$ is the same distance, expressed as a proportion of the distance from v_j to the farthest data point within the window. Thus, $\Delta_{[i]}(v_j)^*\leq$ 1.0 if x_i falls within the current window, and $\Delta_{[i]}(v_j)^*>1.0$ if x_i falls outside the window.

The neighborhood weight for observation i is defined using the tricube weight function, as follows:

$$w_i(v_j)=\begin{cases}(1-|\Delta_{[i]}(v_j)^{*3}|)^3 & \text{for } \Delta_{[i]}(v_j)^*<1\\ 0 & \text{otherwise}\end{cases} \tag{A1}$$

Neighborhood weights are calculated for all observations, from i=1 to n. The shape of the tricube weight function, as well as the specific weights assigned to the 20 observations, are shown in Figure 12C. Note how the observations with X values close to 5.5 have large weights, near 1.0. The weights fall off fairly quickly for observations with X values substantially different (in either direction) from the current evaluation point of v_j=5.5.

A.3. For each v_j, estimate the fitted value, $\hat{g}\,(v_j)$

First, find the coefficients, b_k, that minimize the following:

$$\sum_{i=1}^{n} w_i(v_j)\left(y_i-\left[\sum_{k=0}^{\lambda} b_{kj}x_i^k\right]\right)^2 \tag{A2}$$

Since the value of λ was set to 1, a linear equation is fit to the weighted data, and Eq. (A2) can be expressed as:

$$\sum_{i=1}^{n} w_i(v_j)(y_i - b_{0j} - b_{1j}x_i)^2 \tag{A3}$$

Once the coefficients of the preceding equation are found, the fitted value is obtained by taking the predicted value of Y at v_j:

$$\hat{g}(v_j) = \sum_{k=0}^{\lambda} b_{kj} v_j^k \tag{A4}$$

Fig. 12D shows the line fitted at v_j=5.5, along with the point $(v_j, \hat{g}(v_j))$, which is plotted as the solid circle superimposed over the line. Note how the slope of the fitted line is positive, but relatively shallow. This reflects the orientation and shape of the point cloud in that region of the scatterplot. Evaluation points at different locations within the data rectangle would produce different results. For example, if v_j were closer to the left side of the plot (at, say, 2.0), the slope of the fitted line would be positive, but much steeper.

A.4. Optional robustness step for each v_j

In most loess smooths, the initial local regression line and fitted value are replaced by another line and fitted value that are obtained using a robust estimation procedure. In order to do this, first obtain the residuals from the preceding local regression, for all n observations. For observation i:

$$e_i = y_i - \sum_{k=0}^{\lambda} b_{kj} x_i^k \tag{A5}$$

Then, define e_i^* as follows:

$$e_i^* = \frac{e_i}{6\ \text{Median}\ |e_i|} \tag{A6}$$

Use the bisquare weight function to define the robustness weight for each observation, as follows:

$$r_i = \begin{cases} (1 - |(e_i^*)^2|)^2 & \text{for } |e_i^*| < 1 \\ 0 & \text{otherwise} \end{cases} \tag{A7}$$

The shape of the bisquare weight function and the robustness weights assigned to the 20 observations are both shown in Fig. 12E. Use the robustness weights to estimate a new set of coefficients, b_k^*, which minimize the following expression:

$$\sum_{i=1}^{n} r_i w_i(v_j) \left(y_i - \left[\sum_{k=0}^{\lambda} b_{kj}^* x_i^k \right] \right)^2 \tag{A8}$$

Use the newly-estimated b_k^* values to obtain a new fitted value, $\hat{g}(v_j)$:

$$\hat{g}(v_j) = \sum_{k=0}^{\lambda} b^*_{kj} v_j^k \qquad \text{(A9)}$$

Repeat the robustness steps until the values of the estimated coefficients converge. This usually occurs very quickly, after one or two iterations. Fig. 12F shows the robust line and final fitted value for v_j=5.5. In this case, the robust fitted line is not very different from the original line. This occurs because there are no serious outliers or unusual data points within the scatterplot. If there were, then the initial and final fitted lines could differ quite substantially.

A.5. Repeat steps 2, 3, and (optionally) 4 for all m values of v_j

Fig. 12G shows the points, $(v_j,\hat{g}(v_j))$ obtained by carrying out the preceding steps for all of the m=21 evaluation points. The plot shows the actual data points as open circles and the loess fitted points as solid circles. Finally, use line segments to connect the adjacent $(v_j,\hat{g}(v_j))$ points (as shown in Fig. 12H for the example data) and eliminate the points themselves. The latter are superfluous, since they are just a function of the arbitrarily-chosen evaluation points, rather than the actual data values.

References

Anon., 1997. EViews 3.0 user's guide. Quantitative Micro Software, Irvine, CA, USA.

Anon., 1990. SAS/IML software: usage and reference, version 6, 1st ed. SAS Institute Inc., Cary, NC, USA.

Anon., 1995. SAS/INSIGHT user's guide, version 6, 3rd ed. SAS Institute Inc., Cary, NC, USA.

Anon., 1997. STATA reference manual, release 5. Stata Press, College Station, TX, USA.

Anon., 1995. S-PLUS guide to statistical and mathematical analysis. StatSci (a division of MathSoft, Inc.), Seattle, WA, USA.

Beck N, Jackman S, 1997. Getting the mean right is a good thing: generalized additive models. Working Paper at the Political Methodology WWW Site: wizard.ucr.edu/polmeth/working_papers97

Beck, N., Jackman, S., 1998. Beyond linearity by default: generalized additive models. Am. J. Polit. Sci. 42, 596–627.

Becker, R., Cleveland, W.S., Schyu, M., 1996. The visual design and control of trellis displays. J. Comput. Graph. Stat. 5, 123–155.

Becker, R., Cleveland, W.S., Weil, G., 1988. The use of brushing and rotation for data analysis. In: Cleveland, W.S., McGill, M.E. (Eds.), Dynamic graphics for statistics. Wadsworth and Brooks/Cole, Pacific Grove, CA, USA.

Belsley, D.A., Kuh, E., Welsch, R.E., 1980. Regression diagnostics. John Wiley, New York.

Brown, C., 1995. Serpents in the sand: essays on the nonlinear nature of politics and human destiny. University of Michigan Press, Ann Arbor, MI, USA.

Chambers, J.M., Cleveland, W.S., Kleiner, B., Tukey, P.A., 1983. Graphical methods for data analysis. Wadsworth and Brooks/Cole Advanced Books and Software, Pacific Grove, CA, USA.

Cleveland, W.S., 1979. Robust locally weighted regression and smoothing scatterplots. J. Am. Stat. Assoc. 74, 829–836.

Cleveland, W.S., 1993. Visualizing data. Hobart Press, Summit, NJ, USA.

Cleveland, W.S., 1994. The elements of graphing data (revised edition). Hobart Press, Summit, NJ, USA.

Cleveland, W.S., Devlin, S.J., 1988. Locally weighted regression: an approach to regression analysis by local fitting. J. Am. Stat. Assoc. 83, 596–610.

Cleveland, W.S., McGill, R., 1984. Graphical perception: theory, experimentation, and application to the development of graphical methods. J. Am. Stat. Assoc. 79, 531–553.
Cleveland, W.S., Devlin, S.J., Grosse, E., 1988. Regression by local fitting: methods, properties, and computational algorithms. J. Econometrics 37, 87–114.
Cleveland, W.S., Grosse, E., Shyu, W.M., 1993. Local regression models. In: Chambers, J.M., Hastie, T.J. (Eds.), Statistical models in S. Chapman and Hall, London.
Cook, R.D., Weisberg, S., 1994. An introduction to regression graphics. John Wiley, New York.
Efron, B., Tibshirani, R.J., 1993. An introduction to the bootstrap. Chapman and Hall, New York.
Erikson, R.S., Wright, G.C., McIver, J.P., 1993. Statehouse democracy: public opinion and public policy in the American States. Cambridge University Press, Cambridge.
Fan, J., Gijbels, I., 1996. Local polynomial modelling and its applications. Chapman and Hall, London.
Fox J. Nonparametric regression analysis. Typescript, McMaster University, 1999.
Fox, J., Friendly, M., 1996. Data analysis using APL2 and APL2STAT. In: Stine, R., Fox, J. (Eds.), Statistical computing environments for social research. Sage Publications, Thousand Oaks, CA, USA.
Friendly, M., 1991. SAS system for statistical graphics, 1st ed. SAS Institute, Inc., Cary, NC, USA.
Goodall, C., 1990. A survey of smoothing techniques. In: Fox, J., Long, J.S. (Eds.), Modern methods of data analysis. Sage Publications, Newbury Park, CA, USA.
Green, P.J., Silverman, B.W., 1994. Nonparametric regression and generalized linear models: a roughness penalty approach. Chapman and Hall, London.
Hastie, T.J., Tibshirani, R.J., 1990. Generalized additive models. Chapman and Hall, New York.
Jacoby, W.G., 1998. Statistical graphics for visualizing multivariate data. Sage, Thousand Oaks, CA, USA.
Kim, J., Petrocik, J.R., Enokson, S.N., 1975. Voter turnout among the American states: systemic and individual components. Am. Polit. Sci. Rev. 69, 107–123.
Nie, N.H., Junn, J., Stehlik-Barry, K., 1996. Education and democratic citizenship in America. University of Chicago Press, Chicago, IL, USA.
Norušis, M.J., 1993. SPSS for Windows base system user's guide, release 6.0. SPSS, Inc., Chicago, IL, USA.
Patterson, S.C., Caldeira, G.A., 1983. Getting out the vote: participation in gubernatorial elections. Am. Polit. Sci. Rev. 77, 675–689.
Powell, G.B. Jr., 1986. American voter turnout in comparative perspective. Am. Polit. Sci. Rev. 80, 17–43.
Przeworski, A., Soares, G., 1971. Theories in search of a curve. Am. Polit. Sci. Rev. 65, 51–68.
Stine, R., 1996. AXIS: an extensible graphical user interface for statistics. In: Stine, R., Fox, J. (Eds.), Statistical computing environments for social research. Sage Publications, Thousand Oaks, CA, USA.
Stine, R., Fox, J. (Eds.), 1996. Statistical computing environments for social research. Sage Publications, Thousand Oaks, CA, USA.
Tierney, L., 1990. LISP-STAT: an object-oriented environment for statistical computing and dynamic graphics. John Wiley, New York.
Tukey, J.W., 1977. Exploratory data analysis. Addison-Wesley, Reading, MA, USA.
Wainer, H., 1988. Comment on dynamic graphics for data analysis: deja view. In: Cleveland, W.S., McGill, M.E. (Eds.), Dynamic graphics for statistics. Wadsworth and Brooks/Cole, Pacific Grove, CA, USA.
Weisberg, S., 1996. The R-code: a graphical paradigm for regression analysis. In: Stine, R., Fox, J. (Eds.), Statistical computing environments for social research. Sage Publications, Thousand Oaks, CA, USA.
Wilkinson, L., 1998. SYSTAT 8.0: graphics. SPSS Inc., Chicago, IL, USA.
Zaller, J., 1992. The nature and origins of mass opinion. Cambridge University Press, New York.