Chapter 2
Piecewise Linear L1 Modeling

Kenneth O. Cogger

2.1 Introduction

Lester Taylor (Taylor and Houthakker 2010) instilled a deep respect for estimating parameters in statistical models by minimizing the sum of absolute errors (the L1 criterion) as an important alternative to minimizing the sum of the squared errors (the Ordinary Least Squares or OLS criterion).

He taught many students about the beauty of L1 estimation, including the author. His students were the first to prove asymptotic normality in Bassett and Koenker (1978) and then developed quantile regression (QR), an important extension of L1 estimation, in Koenker and Bassett (1978).

For the case of a single piece, L1 regression is a linear programming (LP) problem, a result first shown by Charnes et al. (1955). Koenker and Bassett (1978) later developed QR and showed that it is a generalization of the LP problem for L1 regression. This LP formulation is reviewed in Appendix 1.

Cogger (2010) discusses various approaches to *piecewise* linear estimation procedures in the OLS context, giving references to their application in Economics, Marketing, Finance, Engineering, and other fields. This chapter demonstrates how *piecewise* linear models may be estimated with L1 or QR using mixed integer linear programming (MILP). If an OLS approach is desired, a mixed integer quadratic programming (MIQP) approach may be taken.

That piecewise OLS regression is historically important is demonstrated in Sect. 2.2, although the estimation difficulties are noted. Section 2.3 develops a novel modification of MILP that easily produces L1 and QR regression estimates in piecewise linear regression with one unknown hinge; an Appendix describes the generalization to any number of unknown hinges. Section 2.4 presents some computational results for the new algorithm. The final section concludes.

K. O. Cogger (✉)
University of Kansas/Peak Consulting Inc., 32154 Christopher Lane,
Conifer, Kansas, CO 80433, USA
e-mail: cogger@peakconsulting.com

J. Alleman et al. (eds.), *Demand for Communications Services - Insights and Perspectives*,
The Economics of Information, Communication, and Entertainment,
DOI: 10.1007/978-1-4614-7993-2_2,

2.2 Piecewise Linear Regression

2.2.1 General

The emphasis in this chapter is on the use of L1 and QR to estimate piecewise linear multiple regressions. When there is only one possible piece, such estimation is an LP problem. The LP formulations are reviewed in Appendix 1.

The term, “regression,” was first coined by Galton (1886). There, Galton studied the height of 928 children and the association with the average height of their parents. Based on the bivariate normal distribution, he developed a confidence ellipse for the association as well as a straight line describing the relationship. Quoting from Galton, “When mid-parents are taller than mediocrity, their children tend to be shorter than they.” In modern language, “mediocrity” becomes “average” and his conclusion rests on the fact that his straight line had a slope with value less than one.

Figure 2.1 illustrates his original data along with the straight line he drew along the major axis of the confidence ellipse. Galton viewed this apparent tendency for children to revert toward the average as a “regression,” hence his title for the paper, “Regression Toward Mediocrity in Hereditary Stature.” One should note that in Fig. 2.1, the vertical axis is Parental height and the horizontal axis is Child height. In modern terms, this would be described as a regression of Parent on Child, even though his conclusive language might best be interpreted in terms of a regression of Child on Parent.

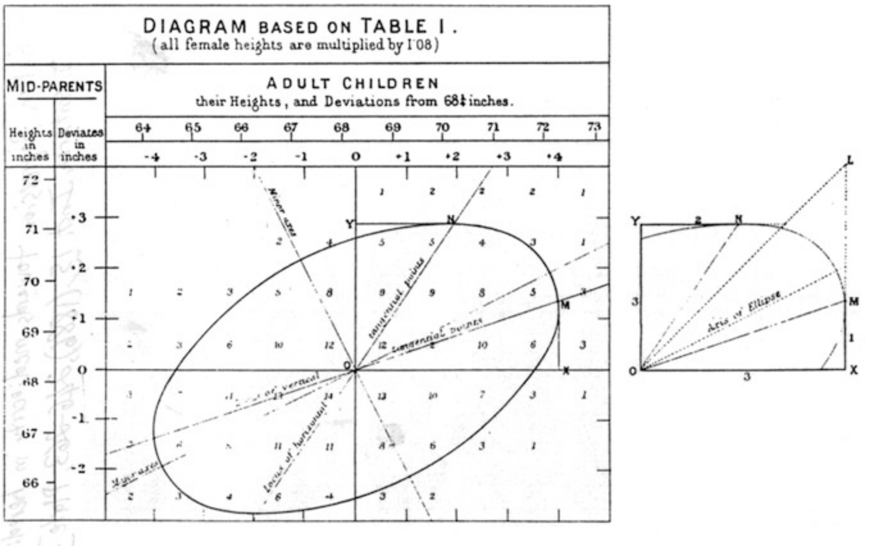

Fig. 2.1 Excerpted diagram from Galton’s seminal regression paper

Galton's paper is of historical primacy. Wachmuth et al. (2003) suggested that Galton's original data are better described by a piecewise linear model. Their resulting OLS piecewise linear fit is:

$$\text{Parent} = 49.81 + 0.270 * \text{Child} + 0.424 * (\text{Child} - 70.803) * \text{I}(\text{Child} > 70.803)$$
$$(\text{P} = 8.28\text{E} - 25)\ (\text{P} = 1.41\text{E} - 04)$$

which can be generalized as a predictor.

$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 (x - H) \cdot I(x > H) \tag{2.1}$$

where $I(x > H) = \left\{ \begin{array}{ll} 1; & x > H \\ 0; & x \leq H \end{array} \right\}$ and H is referred to as a hinge. The first usage of the term, "regression," might therefore better refer to *piecewise linear* regression, which is obviously nonlinear in β_2 and H.

The Wachsmuth et al. (2003) study was based on OLS and is slightly incorrect in terms of its estimates, but probably not in its conclusions. Its estimates were apparently based on the use of Systat, a statistical software program edited by Wilkinson, one of the authors.[1]

2.2.2 Specifics

Infrequently, the hinge location H is known. For example, in economic data, one might know about the occurrence of World War II, the oil embargo of 1973, the recent debt crisis in 2008, and other events with known time values; if economists study time series data, their models can change depending on such events. In climate data, a change was made in the measurement of global CO_2 in 1958 and may influence projected trends. In the study of organizations, known interventions might have occurred at known times, such as Apple Corporation's introduction of (various models) of iPod, iPad, etc.

When the value of H is *known*, L1, QR, OLS and other statistical techniques are easy to use. Simply use a binary variable with known value $\text{B} = 1$ if $\text{x} > H$ (0 otherwise) in Eq. (2.1) and generate another known variable for use. This produces a linear model with two independent variables, x and $\text{z} = (\text{x}–H)\text{B}$. For L1 and QR piecewise multiple regressions with *known* hinges, the LP formulation is described in Appendix 2.

[1] An empirical suggestion of the usefulness of piecewise linear models is that a Google search on "piecewise linear regression" turned up hundreds of thousands of hits. Similarly large numbers of hits occur for synonyms such as "broken stick regression", "two phase regression", "broken line regression", "segmented regression", "switching regression", "linear spline", and the Canadian and Russian preference, "hockey stick regression".

More frequently, the hinge locations H are *unknown* and must be estimated from the data, as with the Galton, Hudson, and other data. Wachsmuth et al. (2003) used a two-stage OLS procedure first developed by Hudson (1966) and later improved upon slightly by Hinkley (1969, 1971). Their sequential procedure requires up to $\sum_{i=0}^{H} 2^i \cdot \binom{H}{i} \cdot \binom{n-H-2}{i}$ separate OLS regressions–where H is the number of hinge points and n is the sample size. For the Galton data ($n = 928$) with one unknown hinge point (two pieces), 1,851 separate OLS regressions may be required with two unknown hinge points (three pieces), 1,709,401 separate OLS regressions may be required. Multicollinearity is known to be present in some of these required OLS regression problems. Multicollinearity can perhaps be overcome in these OLS issues by the use of singular value decomposition (SVD). However, this may be incompatible with the sequential procedure of Hudson (1966) which assumed the existence of various matrix inverses. For moderate to large n, computation time also becomes a concern.

The main concern in this chapter is L1 and QR estimation of piecewise linear regression with unknown hinge locations. Below, three charts in the case of L1 estimation are provided. The first two charts are based on the contrived data of Hudson and the third is based on the real Galton data. For each chart, the minimum sum of the absolute deviations (regression errors) is shown denoted as SAD or $\sum_{i=1}^{n} |y_i - \hat{y}_i|$ for *fixed* hinges found by exhaustive search; for each H, an LP problem was solved:

Figures 2.2, 2.3, and 2.4 exhibit some features common to L1 and QR estimation of piecewise linear regression functions. First, the SAD functions charted are discontinuous at $H = \min(x)$ and $H = \max(x)$. Second, all derivatives of SAD fail to exist at the distinct values $H = x$ and other values as well. Third, SAD is piecewise linear. Fourth, local maxima and minima can occur. Fifth, for any fixed H, SAD is found in ≤ 1 s with standard LP software.

See Appendix 2 for the standard LP formulation with known hinges H. While not present in Figs. 2.2, 2.3, and 2.4, it is possible for the global minimum of SAD to occur at multiple values of H. Multiple optima are always possible in LP problems. It must be emphasized that Figs. 2.2, 2.3, and 2.4 were obtained by exhaustive search over all fixed values of H in the x range, requiring an LP

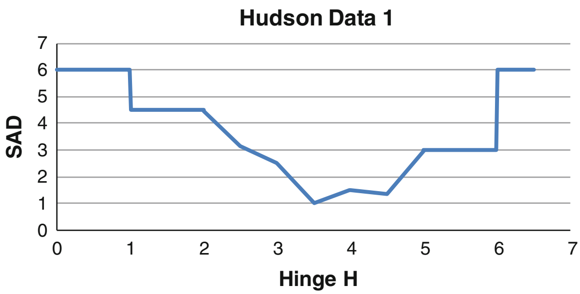

Fig. 2.2 Minimum SAD versus H for Hudson data 1

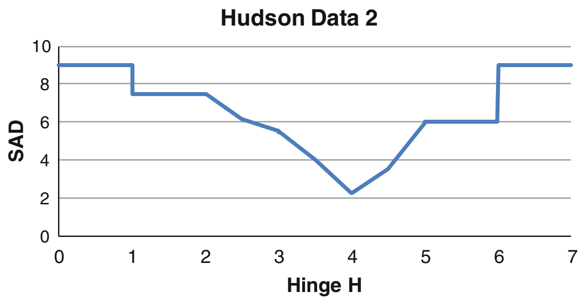

Fig. 2.3 Minimum SAD versus *H* for Hudson data 2

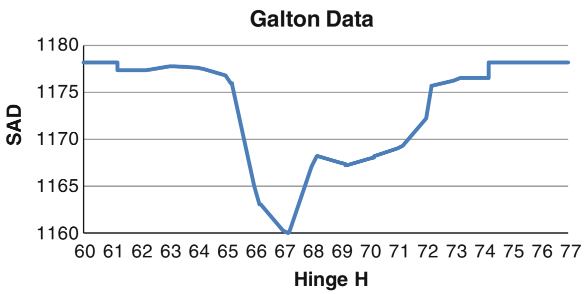

Fig. 2.4 Minimum SAD versus *H* for Galton data

solution for each *H* in the range of the x data values. This illustrates the problems involved in finding the global minimum SAD with various optimum seeking computer algorithms.

Searching for the global optimum *H* in such data is problematic with automated search routines such as Gauss–Newton, Newton–Raphson, Marquardt, and many hybrid techniques available for many years. See Geoffrion (1972) for a good review of the standard techniques of this type. Such techniques fail when the model being fit is piecewise linear due to the nonexistence of derivatives of SAD at many points and/or the existence of multiple local minima.

There are search techniques which do not rely on SAD being well-behaved and do not utilize any gradient information. These would include the "amoeba" search routine of Nelder and Mead (1965), "tabu" search due to Glover (1986), simulated annealing approaches, and genetic algorithm search techniques.[2]

A number of observations can be made. First, the algorithms of Hudson (1966) and Hinkley (1969, 1971) solve only the OLS piecewise linear problem. Second, they provide the maximum likelihood solution if the errors are normally distributed, but they are computationally expensive, are not easily extended to the multiple regression case, are not practically extended to multiple hinges, and software is not available for implementation.

[2] The author has not applied any of these to piecewise linear estimation.

2.3 A New Algorithm for L1 and QR Piecewise Linear Estimation

L1 or QR multiple regression with no hinges (one piece) is an LP problem as reviewed in Appendix 1. The choice of either L1 or QR is dictated by the choice of a fixed parameter $\theta \in [0, 1]$.

For piecewise multiple regression with one hinge, Eq. (2.1) may be converted to

$$\hat{y} = \left\{ \begin{array}{ll} \beta_1' x; & x \leq H \\ \beta_2' x; & x > H \end{array} \right\} \tag{2.2}$$

where the hinge H is defined as

$$\{x : (\beta_2' - \beta_1')x = 0\} \tag{2.3}$$

In simple linear piecewise regression, H will be a scalar value. In the piecewise multiple regression case, H will be a hyperplane. Generally, for a continuous piecewise linear fit,

$$\hat{y} = \max(\beta_1' x,\ \beta_2' x) \text{ OR } \hat{y} = \min(\beta_1' x,\ \beta_2' x) \tag{2.4}$$

If H is *known*, an LP problem may be solved for L1 or QR with an appropriate choice of $\theta \in [0,\ 1]$. Multiple known hinges are also LP problems. See Appendix 2.

For L1 or QR piecewise multiple regression with one *unknown* hinge, there are two possible linear predictors for each observation. This follows from Eq. (2.2). For all observations, a Max or Min operator applies. This follows from Eq. (2.4). Therefore, an overall continuous piecewise linear predictor must satisfy the following:

$$\hat{y} = \left\{ \begin{array}{ll} \hat{y}_1 = \beta_1' x; & x \in S_1 \\ \hat{y}_2 = \beta_2' x; & x \notin S_1 \end{array} \right\} \tag{2.5}$$

$$\hat{y} = \max(\beta_1' x,\ \beta_2' x) \text{ OR } \hat{y} = \min(\beta_1' x,\ \beta_2' x) \tag{2.6}$$

where S_1 is a chosen subset of x. The particular subset chosen in Eq. (2.5) can be described by n binary decision variables $B1_i = \left\{ \begin{array}{ll} 0; & x_i \in S_1 \\ 1; & x_i \notin S_1 \end{array} \right\}$ and can be enforced as linear inequalities in MILP using the Big-M method. The single choice in Eq. (2.6) can be described by a single binary decision variable B1 and can also be enforced as linear inequalities using the Big-M method (Charnes 1952). Combining the following 8n linear constraints enforce Eqs. (2.5) and (2.6) in any feasible solution to the L1 or QR piecewise linear multiple regression problem with one *unknown* hinge:

$$
\begin{aligned}
&\hat{y}_i \le \hat{y}_{1i} + M \cdot B1; \forall i\\
&\hat{y}_i \le \hat{y}_{2i} + M \cdot B1; \forall i\\
&\hat{y}_{1i} \le \hat{y}_i + M \cdot (1 - B1); \forall i\\
&\hat{y}_{2i} \le \hat{y}_i + M \cdot (1 - B1); \forall i\\
&\hat{y}_i \le \hat{y}_{1i} + M \cdot (1 - B1) + M \cdot B1_i; \forall i\\
&\hat{y}_i \le \hat{y}_{2i} + M \cdot (1 - B1) + M \cdot (1 - B1_i); \forall i\\
&\hat{y}_{1i} \le \hat{y}_i + M \cdot B1 + M \cdot B1_i; \forall i\\
&\hat{y}_{2i} \le \hat{y}_i + M \cdot B1 + M \cdot (1 - B1_i); \forall i
\end{aligned}
\tag{2.7}
$$

For large M, it is clear that the constraints in Eq. (2.7) result in the following restrictions on any feasible solution to the L1 or QR piecewise linear multiple regression with one *unknown* hinge (Table 2.1):

Estimating a piecewise multiple linear regression with L1 or QR and one unknown hinge (two linear pieces) may therefore be expressed as the following MILP problem:

$$
\min : \sum_{i=1}^{n} (\theta e_{i+} + (1 - \theta) e_{i-})
$$

Such that:

$$
\hat{y}_{1i} = \beta_1' x_i;\ \forall i
$$

$$
\hat{y}_{2i} = \beta_2' x_i;\ \forall i
$$

$$
y_i - \hat{y}_i = e_{i+} - e_{i-};\ \forall i
$$

$$
e_{i+} \ge 0;\ \forall i
$$

$$
e_{i-} \ge 0;\ \forall i
$$

$$
\begin{aligned}
&\hat{y}_i \le \hat{y}_{1i} + M \cdot B1;\ \forall i\\
&\hat{y}_i \le \hat{y}_{2i} + M \cdot B1;\ \forall i\\
&\hat{y}_{1i} \le \hat{y}_i + M \cdot (1 - B1);\ \forall i\\
&\hat{y}_{2i} \le \hat{y}_i + M \cdot (1 - B1);\ \forall i\\
&\hat{y}_i \le \hat{y}_{1i} + M \cdot (1 - B1) + M \cdot B1_i;\ \forall i\\
&\hat{y}_i \le \hat{y}_{2i} + M \cdot (1 - B1) + M \cdot (1 - B1_i);\ \forall i\\
&\hat{y}_{1i} \le \hat{y}_i + M \cdot B1 + M \cdot B1_i;\ \forall i\\
&\hat{y}_{2i} \le \hat{y}_i + M \cdot B1 + M \cdot (1 - B1_i);\ \forall i
\end{aligned}
$$

$\beta_1,\ \beta_2$	are decision vectors unrestricted in sign
$B1$	is a binary decision variable
$B1_i;\ \forall i$	are binary decision variables.

Table 2.1 Results of enforcing Eq. (2.7) for large M

B1	$B1_i$	Result
0	0	$\hat{y}_i = \min(\hat{y}_{1i},\ \hat{y}_{2i}) = \hat{y}_{1i}$
0	1	$\hat{y}_i = \min(\hat{y}_{1i},\ \hat{y}_{2i}) = \hat{y}_{2i}$
1	0	$\hat{y}_i = \max(\hat{y}_{1i},\ \hat{y}_{2i}) = \hat{y}_{1i}$
1	1	$\hat{y}_i = \max(\hat{y}_{1i},\ \hat{y}_{2i}) = \hat{y}_{2i}$

This MILP problem has $5n + 4$ continuous decision variables, $n + 1$ binary decision variables, and 11n constraints. The choice between L1 and QR piecewise linear regression is made by the choice of $\theta \in [0,\ 1]$. From the notation, the β vectors may be of arbitrary dimension, permitting piecewise multiple linear regressions to be estimated with L1 or QR. If desired, some of the equality constraints may be combined to modestly reduce the number of constraints and continuous variables, but in practice the computation time depends mostly on the number of binary decision variables. M must be a large positive number suitably chosen by the user. Too small a value will produce incorrect results; too large a value will cause numerical instability in MILP software. $M > 2|y_i|;\ \forall i$ provides a reasonable value for M.

The binary search space has 2^{n+1} combinations to search and is a major factor in execution time for MILP if the sample size is large. For the Hudson data, $n = 6$ and the binary search space has only 128 combinations; execution time is <0.1 s for the MILP formulation. For the Galton data, $n = 928$ and the binary search space has 4.538E + 279 combinations to search; execution time takes several days for the MILP formulation. To dramatically reduce execution time for large n, it is wise to recognize any a priori restrictions on the two pieces. Often, such a priori restrictions are quite weak.

The x values for the Galton data, for example, may be arranged in nondescending order. Below the estimated hinge, all cases must have one of the two linear predictors; above the estimated hinge, all cases must have the other predictor. This means that without loss of generality the constraint $B1_i \leq B1_{i-1};\ i = 2 : n$ may imposed which reduces the binary search space to size $n + 1$ rather than 2^{n+1}. With this additional linear constraint added to the MILP, the Galton piecewise linear estimation is solved in about 20 min rather than several days.

The MILP solution does not directly produce an estimate of the hinge. The definition of the hinge in Eq. (2.3) may be used to produce this estimate from the MILP solution. In the case where the two β vectors are of dimension two (piecewise linear regression with a constant term and scalar x values), this solution is a scalar value. Generally, when the two β vectors are of dimension p, the solution of Eq. (2.3) will be a hyperplane of dimension $p{-}1$ as described by Breiman (1993), there is no meaningful scalar hinge value for $p > 2$ and the hinge may be defined as any scalar multiple of the difference between the two β vectors.

Table 2.2 MILP on the Hudson (1966) data

	x = (1,2,3,4,5,6) y = (1,2,4,4,3,1) (Hudson 1)	x = (1,2,3,4,5,6) y = (1,2,4,7,3,1) (Hudson 2)
Estimates		
β_{01}	10.000	14.500
β_{11}	−1.500	−2.250
β_{02}	−0.500	−0.500
β_{12}	1.500	1.500
SAD	1.000	2.250
Estimated hinge	3.500	4.000
Execution time (s)	0.043	0.037

At the time of the Hinkley (1969, 1971) and Hudson (1966) papers, linear programming was in its infancy. Hillier and Lieberman (1970) noted at that time, "Some progress has been made in recent years, largely by Ralph Gomory, in developing [algorithms] for [integer linear programming]." Even 6 years later, Geoffrion (1972) observed, "A number of existing codes are quite reliable in obtaining optimal solutions within a short time for general all-integer linear programs with a moderate number of variables—say on the order of 75 …… and mixed integer linear programs of practical origin with [up to 50] integer variables and [up to 2,000] continuous variables and constraints are [now] tractable." It is not surprising that Hudson and others in the mid-1960s were not looking at alternatives to OLS.

At present, large MILP problems may be handled. Excellent software is available that can handle MILP for problem sizes limited only by computer memory limits. The next section reports some computational results. Appendix 3 shows how the MILP formulation is easily extended to more than one unknown hinge.

2.4 Computational Results

Table 2.2 shows that solutions from the MILP formulation are correct for both Hudson data sets. The recommended values of M from the second section were used. There is complete agreement with Figs. 2.2 and 2.3 which were produced with exhaustive manual search. The suggestion in the second section for reducing the size of the binary search space to $n + 1 = 7$ was also incorporated.

Table 2.3 shows that the MILP solution is correct for the Galton data. Again, the recommended values of M from the second section were used and reduced the binary search space to size $n + 1 = 929$. There is complete agreement with Fig. 2.4 which required exhaustive manual search.

Table 2.3 MILP on the Galton data

	Galton
Estimates	
β_{01}	67.5
β_{11}	0
β_{02}	33.9
β_{12}	0.5
SAD	1160
Estimated hinge	67.2
Execution time (s)	1339.28

2.5 Computer Software

Implementing the new algorithm depends on computer software. There is now much available. Probably the most widespread is *Solver* in Excel on Windows and Macs, but it is limited in the formulation to $n = 48$ cases unless upgrading to more expensive versions.

LP_Solve is free and capable up to the memory limits of a computer; this package may be imported into R, Java, AMPL, MATLAB, D-Matrix, Sysquake, SciLab, Octave, FreeMat, Euler, Python, Sage, PHP, and Excel. The R environment is particularly notable for its large number of statistical routines. *GAMS* and other commercial packages are also available for the OLS formulation. The *LP_Solve* program has an easy IDE environment for Windows, not requiring any programming skills in R, SAS, etc. All computation in this section was done using *LP_Solve*.

2.6 Conclusions

Piecewise linear estimation is important in many studies. This chapter develops a new practical MILP algorithm for such estimation which is appropriate for piecewise linear L1 and QR estimation; it may be extended to OLS estimation by MIQP by changing the objective function from linear to quadratic.

Software is widely available to implement the algorithm. Some is freeware, some is commercial, and some is usually present on most users' Excel platform (Solver), but the latter is quite limited in sample sizes.

Statistical testing of piecewise linear estimators with L1 and QR is not discussed in this chapter but is an important topic for future research. It is plausible that in large samples, the asymptotic theory of L1 and QR will apply.

Since the piecewise linear L1 and QR estimates will always produce better (or no worse) fits than standard linear models, it is suggested that all previous studies using standard linear models could be usefully revisited using the approach.

Appendix 1. Standard L1 or QR Multiple Regression

The estimation of a single multiple regression with L1 or QR is the following LP problem:

$$\min : \sum_{i=1}^{n} (\theta e_{i+} + (1-\theta)e_{i-}).$$

Such that:

$$y_i - \beta' x_i = e_{i+} - e_{i-};\ \forall i$$

$$e_{i+} \geq 0;\ \forall i$$

$$e_{i-} \geq 0;\ \forall i,$$

β unrestricted.

In this primal LP problem, the x_i are known p-vectors and the y_i are known scalar values. β is a p-vector of decision variables. For L1, choose $\theta = 0.5$; for QR choose any $\theta \in [0, 1]$. This well-known LP formulation has $2n + p$ decision variables and n linear equality constraints. For this primal LP formulation, duality theory applies and the dual LP problem is:

$$\max : \sum_{i=1}^{n} \lambda_i y_i.$$

Such that:

$$X'\lambda = 0$$

$$\theta - 1 \leq \lambda_i \leq \theta :\ \forall i.$$

This LP problem has n decision variables, p linear equality constraints, and n bounded variables, so it is usually a bit faster to solve for large n. Importantly, the optimal values in λ may be associated with important test statistics developed by Koenker and Bassett.

Appendix 2. L1 or QR Piecewise Multiple Regression with Known Hinges

With one known hinge, Eq. (2.2) describes the predictor and Eq. (2.3) defines the hinge. Let x be a p-vector of known values of independent variables. Typically, the first element of x is unity for a constant term in the multiple regression. The hinge given by Eq. (2.3) will then be a p-vector, H, which is here assumed known. Define the p-vector $z = \begin{Bmatrix} x; & x \leq H \\ x - H; & x > H \end{Bmatrix}$ with individual calculations for each element of x and H. Since x and H are known, z has known values. This results in the LP problem:

$$\min : \sum_{i=1}^{n} (\theta e_{i+} + (1 - \theta) e_{i-})$$

Such that:

$$y_i - \beta_1' x_i - \beta_2' z_i = e_{i+} - e_{i-};\ \forall i$$

$$e_{i+} \geq 0;\ \forall i$$

$$e_{i-} \geq 0;\ \forall i$$

β_1, β_2 unrestricted.

For more than one known hinge, this LP can be easily extended; simply add additional β vectors and additional z vectors for each additional hinge to the formulation.

Appendix 3. L1 or QR Piecewise Multiple Regression with Unknown Hinges

The solution for $H = 1$ hinge and two pieces is clearly found with the MILP formulation in the second section. Let this solution be denoted by $\hat{y}_i = \hat{y}(1)_i;\ \forall i$ [with notation changes to Eq. (2.7)] which chooses one of the linear pieces $(\hat{y}_{1i},\ \hat{y}_{2i})$ as the regression for each i.

For $H = 2$ hinges, there are three possible pieces $(\hat{y}_{1i},\ \hat{y}_{2i},\ \hat{y}_{3i})$. This reduces to a choice between one of two linear pieces $\left(\hat{y}_{3i},\ \hat{y}(1)_i\right)$ and a second set of binary variables and constraints such as Eq. (2.7) (with notation changes) enforces this choice to solve the problem for $H = 2$. This solution can be denoted by $\hat{y}(2)_i;\ \forall i$.

This inductive argument can be continued for $H = 3, 4$, etc. For any number of hinges, an MILP formulation can be created with $H(n + 1)$ binary variables, the main determinant of computing time.

References

Bassett G, Koenker R (1978) Asymptotic theory of least absolute error regression. J Am Stat Assoc 73:618–622

Breiman L (1993) Hinging hyperplanes for regression, classification, and function approximation. IEEE Trans Inf Theory 39(3):999–1013

Charnes A (1952) Optimality and degeneration in linear programing. Econometrica 20(2):160–170

Charnes A, Cooper WW, Ferguson RO (1955) Optimal estimation of executive compensation by linear programming. Manage Sci 1:138–151

Cogger K (2010) Nonlinear multiple regression methods: a survey and extensions. Intell Syst Account Finance Manage 17:19–39

Galton F (1886) Regression towards mediocrity in hereditary stature. J Anthropol Inst Great Br Irel 15:246–263

Geoffrion A (1972) Perspectives on optimization. Addison-Wesley, Reading

Glover F (1986) Future paths for integer programming and links to artificial intelligence. Comput Oper Res 13:533–549

Hillier F, Lieberman G (1970) Introduction to operations research. Holden-Day, San Francisco

Hinkley D (1969) Inference about the intersection in two-phase regression. Biometrika 56:495–504

Hinkley D (1971) Inference in two-phase regression. J Am Stat Assoc 66:736–743

Hudson D (1966) Fitting segmented curves whose join points have to be estimated. J Am Stat Assoc 61:1097–1129

Koenker R, Bassett G (1978) Regression quantiles. Econometrics 16:33–50

Nelder J, Mead R (1965) A simplex method for function minimization. Comput J 7:308–313

Taylor L, Houthakker H (2010) Consumer demand in the United States: prices, income, and consumer behavior. Kluwer Academic Publishers, Netherlands

Wachsmuth A, Wilkinson L, Dallal G (2003) Galton's Bend: a previously undiscovered nonlinearity in Galton's family stature regression data. Am Stat 57(3):190–192