Nonparametric Methods: Part 2

Abi Adams

Hilary 2019

Overview

- Recap
- Multivariate Kernel Regression
- Partially Linear Model
- Single Index Model

Recap

- You normally think of estimating regression functions of the form:
$$Y_i = X_i\beta + e_i$$
- We saw that nonparametric methods do not specify a functional form a priori
$$Y_i = m(X_i) + e_i$$
- This allows the data to 'speak for itself' but does not lend itself to giving one answer ('How does X affect y??)

Recap

- Kernel methods are based on smoothing the data, weighting observations closer to the point of interest more closely than those far away
- The key decision you have to make with these methods is what constitutes ‘close’ and ‘far away’ — this is controlled by the bandwidth, h
- The bandwidth should be selected optimally by cross-validation

Kernel Regression

- Using

$$\widehat{m}(x) = \sum_{i=1}^{n} w_i(x)\, Y_i$$

and

$$w_i(x) = \frac{\frac{1}{nh} K\left(\frac{X_i - x}{h}\right)}{\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)}$$

gives the Kernel Regression function estimator (Nadaraya (1964), Watson (1964)):

$$\widehat{m}(x) = \frac{\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right) Y_i}{\frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{X_i - x}{h}\right)}$$

Confidence Intervals

- Standard errors are also defined pointwise

$$se_{\widehat{m}(x)} = \left[\frac{c\widehat{\sigma}^2}{nh\widehat{f}(x)} \right]]$$

where

  - c is a constant that depends on the kernel function used (e.g. Gaussian $c = 0.5\sqrt{\pi}$.
  - Estimator of the variance is:

$$\widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \widehat{m}(x))^2$$

- The 95% pointwise confidence interval is then:

$$\widehat{m}(x) + / - 1.96 se_{\widehat{m}(x)}$$

Cross Validation

- The choice of bandwidth has caused much head scratching within the literature on nonparametric regression
- Very important to the overall quality of our regression output but it is hard to work out exactly how to optimally choose its value.
- One popular and practical approach to bandwidth selection is cross-validation. This method makes the choice of bandwidth directly dependent on the data.

Cross Validation

- Method starts with a direct estimate of the squared error of the regression and then picks the bandwidth to minimize this estimate. The sum of squared errors is given as:

$$\sum (y_i - \hat{y}_i)^2 = \sum \hat{e}_i^2 \tag{1}$$

Cross Validation

- One selects h to minimise the sum of squared "leave one out" residuals:

$$CV = \sum_{i=1}^{N} (y_i - \hat{m}_{\neg i}(X_i))^2 \tag{2}$$

where $\hat{m}_{\neg i}(X_i)$ is the estimator obtained by omitting the i^{th} observation, $\{X_i, y_i\}$.

Multivariate kernels

- There is no fundamental methodological difference between kernel regression in the univariate and multivariate settings.

$$\hat{m}_H(\mathbf{x}) = \frac{\sum_{i=1}^{N} K_H(\mathbf{x} - \mathbf{X}_i) y_i}{\sum_{i=1}^{N} K_h(\mathbf{x} - \mathbf{X}_i)} \tag{3}$$

- The multivariate kernel function K_H is typically chosen as the as the product of univariate kernels:

$$K(\mathbf{u}) = \prod_{j=1}^{d} K(u_j) \tag{4}$$

The Curse of Dimensionality

- The rate of convergence of nonparametric estimates to the truth falls in the number of dimensions of the problem.
- Plotting high dimensional outputs is rarely ever possible making the interpretation of any results very difficult.

The Curse of Dimensionality

- There are three main approaches that researchers use to reduce the complexity of high dimensional problems.
  1. Variable selection in a nonparametric regression.
  2. Using semi- or nonparametric indices.
  3. Using nonparametric link functions.

Chevalier & Ellison: Career Concerns of Mutual Fund Managers

CAREER CONCERNS OF MUTUAL FUND MANAGERS*

JUDITH CHEVALIER AND GLENN ELLISON

We examine the labor market for mutual fund managers. Using data from 1992–1994, we find that "termination" is more performance-sensitive for younger managers. We identify possible implicit incentives created by the termination-performance relationship. The shape of the termination-performance relationship may give younger managers an incentive to avoid unsystematic risk. Direct effects of portfolio composition may also give younger managers an incentive to "herd" into popular sectors. Consistent with these incentives, we find that younger managers hold less unsystematic risk and have more conventional portfolios. Promotion incentives and market responses to managerial turnover are also studied.

Chevalier & Ellison: Career Concerns of Mutual Fund Managers

IV. The Shape of the Termination-Performance Relationship

In this section we examine in more detail how the likelihood of managerial termination varies with the manager's recent performance, estimating the shape of the termination-performance relationship. We do so both to understand better when managers are replaced and because nonlinearities in the termination-performance relationship might alter the manager's incentives to undertake risk.

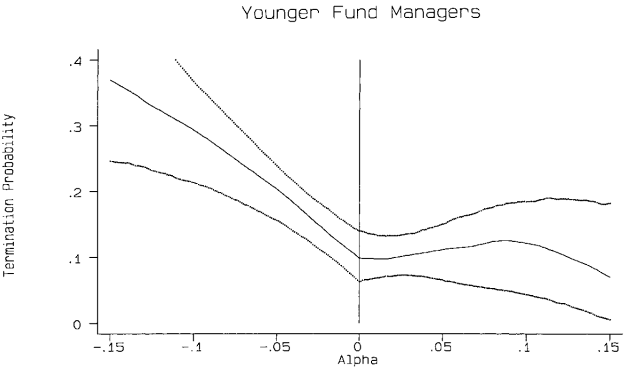
The idea that the shape of the performance contract facing a mutual fund manager may have incentive effects is not new. For example, Starks [1987] and Grinblatt and Titman [1989] show that mutual fund fee schedules that are nonlinear in fund performance may distort the fund's risk incentives. Chevalier and Ellison [1997] suggest that nonlinearities in the relationship between the flow of new funds into mutual funds and fund performance may also lead to distortions in the fund's risk incentives. However, this literature does not consider incentives of the fund managers; these could well differ from those of the fund company.

Chevalier & Ellison: Career Concerns of Mutual Fund Managers

$$\begin{aligned} Termination_{it} = {} & f(Alpha_{it}) + \beta_1 Alpha_{it-1} + \beta_2 Alpha_{it-2} \\ & + \beta_3 ManagerAge_{it} + \beta_4 GrowIncDummy_{it} \\ & + \beta_5 Age60+_{it} + \beta_6 Year92_t + \beta_7 Year93_t + \epsilon_{it}, \end{aligned}$$

with ϵ_{it} assumed to have expectation zero conditional on the right-hand-side variables. We apply the procedure of Robinson [1988] to obtain estimates of the coefficients β on the control variables and an estimate of the function f.[10] To allow for differences depending on the manager's age, we estimate the equation separately on two subsamples: the 651 fund-years for which the manager is less than 45 years of age and the 669 fund-years for which the manager is at least 45.

Chevalier & Ellison: Career Concerns of Mutual Fund Managers



Chevalier & Ellison: Career Concerns of Mutual Fund Managers

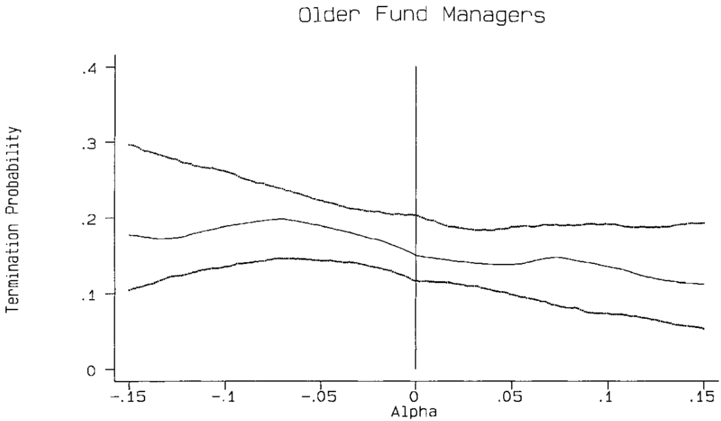


FIGURE I

Marginal Treatment Effects

AMERICAN ECONOMIC ASSOCIATION

Membership About AEA Log In

Journals Annual Meeting Careers Resources EconLit

ne › Journals › American Economic Review › AER October 2011 › Estimating Marginal Returns to Education

ournals

merican Economic Review

About the *AER*

Forthcoming Articles

Issues

Archived Internet Comments

Estimating Marginal Returns to Education

Pedro Carneiro

James J. Heckman

Edward J. Vytlacil

AMERICAN ECONOMIC REVIEW
VOL. 101, NO. 6, OCTOBER 2011
(pp. 2754-81)

Partially linear model

- The partially linear model has two parts: a parametric component $X\beta$ and a nonparametric component $m_z(Z)$.

$$y_i = \mathbf{X}_i\beta + m_z(Z_i) + e_i \tag{5}$$

with $E(e_i) = 0$

- We will focus on the seminal approach suggested by Robinson (1988), which involves "concentrating out" the unknown function m_z using a double residual regression.
- This estimator for β is consistent, asympotically normal and converges at the parametric rate.

Partially linear model

- To convey the essential theory behind the Robinson estimator, take expectations of the partially linear model with respect to Z_i to yield:

$$\begin{aligned} E(y_i|Z_i) &= E(\mathbf{X}_i\beta|Z_i) + E(m_z(Z_i)|Z_i) + E(e_i|Z_i) \quad (6) \\ &= E(\mathbf{X}_i|Z_i)\beta + m_z(Z_i) \quad (7) \end{aligned}$$

- Subtracting this from the full model removes the unknown function m_z to leave:

$$\begin{aligned} y_i - E(y_i|Z_i) &= (\mathbf{X}_i - E(X_i|Z_i))\beta + e_i \quad (8) \\ e_{yi} &= e_{xi}\beta + e_i \quad (9) \end{aligned}$$

Partially linear model

- Robinson (1988) suggested first estimating $E(y_i|Z_i)$ and $E(\mathbf{X}_i|Z_i)$ by Nadaraya-Watson regressions and then using the associated residuals $\hat{e}_{yi}$ and $\hat{\mathbf{e}}_{xi}$ to estimate $\hat{\beta}$.

$$\begin{aligned} y_i &= m_y(Z_i) + e_{yi} &\quad (10)\\ \mathbf{X}_i &= \mathbf{m}_x(Z_i) + e_{xi} &\quad (11)\end{aligned}$$

with the estimators

$$\begin{aligned} \hat{m}_y(z) &= \frac{\sum_{i=1}^{N} K_h(z - Z_i) y_i}{\sum_{i=1}^{N} K_h(z - Z_i)} &\quad (12)\\ \hat{m}_x^d(z) &= \frac{\sum_{i=1}^{N} K_h(z - Z_i) X_i^d}{\sum_{i=1}^{N} K_h(z - Z_i)} &\quad (13)\end{aligned}$$

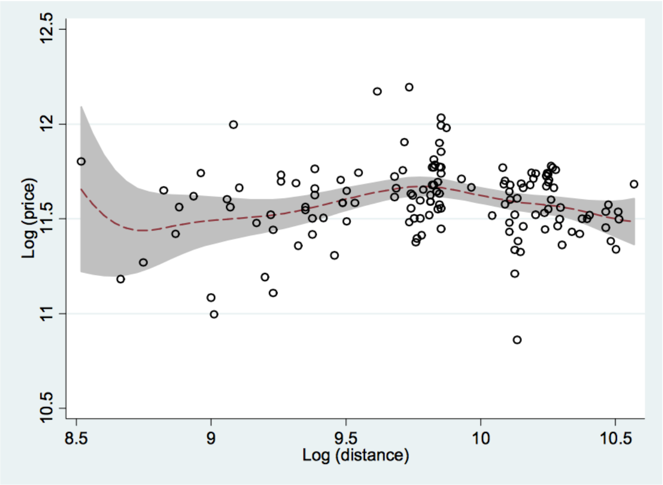
Partially Linear in Stata

```
. semipar lprice larea lland rooms bath age if y81==1, nonpar(ldist)

                                                Number of obs =       142
                                                R-squared     =    0.6863
                                                Adj R-squared =    0.6748
                                                Root MSE      =    0.1859

------------------------------------------------------------------------------
      lprice |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       larea |   .3266051    .070965     4.60   0.000     .1862768    .4669334
       lland |   .0790684   .0318007     2.49   0.014     .0161847    .1419521
       rooms |    .026588   .0266849     1.00   0.321    -.0261795    .0793554
       baths |   .1611464   .0400458     4.02   0.000     .0819585    .2403342
         age |  -.0029953   .0009564    -3.13   0.002    -.0048865   -.0011041
------------------------------------------------------------------------------
```

Partially Linear in Stata



Testing a Parametric Approximation

- You might want to see if you can approximate a nonparametric functions by some parametric polynomial alternative
- Hardle and Mammen (1993) develop a test statistic for comparing the nonparametric and parametric regression fits using the squared deviations between them

$$T_N = N\sqrt{h}\sum_{i=1}^{N}(\hat{m}_z(Z_i) - \hat{m}_z(Z_i,\theta))^2 \tag{14}$$

- To obtain critical values for the test, use a wild bootstrap
- Implemented in Stata

Single Index Models

- Single index models (SIMs) are another important and popular class of semiparametric methods.
- In these models, $\mathbb{E}(y|X) = g(\mathbf{X}\beta)$, where $g(\cdot)$ is often referred to as the link function.
- These models are only nonparametric in one dimension — which is very useful for alleviating the Curse!

Single Index Models

- You will have come across SIMs many times before
- Many parametric models are single index- Logit, Probit, Tobit. . .
- These models based on strong functional form and distributional assumptions, which you might want to relax — or at least examine the robustness of your results to relaxing these restrictions.

Single Index Models

- Semiparametric SIMs keep a linear form for the index ($\mathbf{X}\beta$) but allow the link function ($g(\cdot)$) to be any smooth function.
- Two different approaches are typically taken to estimate β and $g(\cdot)$:
  1. Iterative approximation of β using semiparametric least squares (SLS) or pseudo maximum likelihood estimation (PMLE).
  2. Direct estimation of β using the average derivative of the regression function.

Single Index Models

- Although the nitty-gritty varies across the different methods, the basic approach to estimation can be summarised as follows:
    1. Estimate β by $\hat{\beta}$.
    2. Compute index values $\hat{v} = \mathbf{X}\hat{\beta}$.
    3. Estimate the link function $g(\cdot)$ using nonparametric regression of y on $\hat{v}$.

Identification issues

- Note that **X** cannot include an intercept — the function $g(\cdot)$ will include any location and level shift.
- Further, the level of β is not identified and thus a normalisation criterion is required.
  - This is typically achieved by setting one element of β equal to 1.
- Require **X** to include a continuously distributed variable that is informative for y — it is impossible to identify the continuous function m on a discrete support.

Semiparametric Least Squares

- Semiparametric least squares was introduced by Ichimura (1993).
- The approach is motivated by the insight that, if $g(\cdot)$ were known, we could estimate β using nonlinear least squares.
- This would involve picking β to minimise the criterion function:

$$S(\beta, g) = \sum_{i=1}^{N} (y_i - g(\mathbf{X}_i \beta))^2 \tag{15}$$

- Sadly for us, this estimator is infeasible- we do not know the structure of the nonparametric link function (this is what we are trying to find out!).

Semiparametric Least Squares

- Ichimura proposed replacing the unknown function $g(\cdot)$ with the 'leave-one-out estimator', resulting in a technique with a similar flavour to the cross validation method
- In fact, Hardle, Hall and Ichimura (1993) suggest picking β and h to jointly minimise $S(\beta)$.
- Under the approach, the parameter vector β is chosen to minimise:

$$S(\beta) = \sum_{i=1}^{N} (y_i - \hat{g}_{\neg i}(\mathbf{X}_i\beta))^2 \tag{16}$$

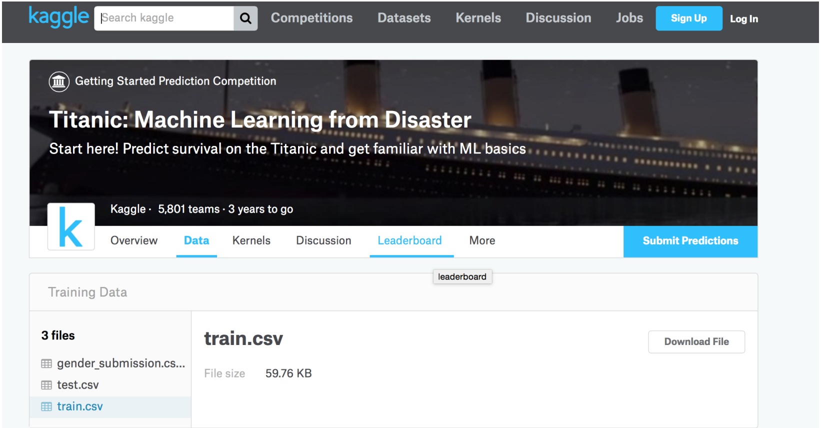
where

$$\hat{g}_{\neg i}(\mathbf{X}_i\beta) = \frac{\sum_{i \neq j} K_h\left((\mathbf{X}_i - \mathbf{X}_j)\beta\right) y_i}{\sum_{i \neq j} K_h\left((\mathbf{X}_i - \mathbf{X}_j)\beta\right)} \tag{17}$$

Semiparametric Least Squares

- There are two elements to the estimation process for β:
  1. For a given β and h, evaluate $S(\beta)$;
  2. Jointly select the β and h that minimize $S(\beta)$.

Single Index Model in Stata



Single Index Model in Stata

```
. xi: sml survived female age i.pclass
i.pclass          _Ipclass_1-3        (naturally coded; _Ipclass_1 omitted)

Iteration 0:   log likelihood = -485.15013

…

Iteration 6:   log likelihood = -471.17626

SML Estimator - Klein & Spady (1993)               Number of obs   =       1046
                                                   Wald chi2(4)    =      27.30
Log likelihood = -471.17626                        Prob > chi2     =     0.0000

------------------------------------------------------------------------------
    survived |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   3.220109   .6381056     5.05   0.000     1.969445    4.470772
         age |  -.0334709   .0076904    -4.35   0.000    -.0485438   -.0183981
  _Ipclass_2 |  -1.360299    .370819    -3.67   0.000    -2.087091   -.6335076
  _Ipclass_3 |  -3.605414   .8002326    -4.51   0.000    -5.173842   -2.036987
------------------------------------------------------------------------------
```

Single Index Model in Stata