A New Difference-based Weighted Mixed Liu Estimator in Partially Linear Models

Fikri Akdeniz
fikriakdeniz@cag.edu.tr

Çag University, Mersin-TURKEY
Department of Mathematics and Computer Science

11-06-2014

Shrinkage Estimators: Ridge and Liu

- Classical linear regression model is defined as:

$$y = X\beta + \varepsilon, \quad E(\varepsilon) = 0, Cov(\varepsilon) = \sigma^2 I \tag{1}$$

- OLS estimator of coefficients are $\widehat{\beta} = (X^\top X)^{-1} X^\top y$.
- OLS estimator for nearly collinear data ($X^T X$ is ill-conditioned) is very unreliable with very high variance.
- Small perturbations in data change everything!
- Multicollinearity between the predictors: shrinkage estimators
- The most popular shrinkage estimator is the ridge regression estimator which is defined by (Hoerl and Kennard, 1970)

$$\widehat{\beta}(k) = (X^\top X + kI)^{-1} X^\top y$$

where $k > 0$ is the tuning (biasing) parameter.

Ridge Regression

Ridge regression is like least squares but shrinks the estimated coefficients towards zero. Given a response vector $y \in \mathbb{R}$ and a predictor matrix $X \in \mathbb{R}^{n \times p}$, the ridge regression coefficients are defined as

$$\begin{aligned}\widehat{\beta}(k) &= \underset{\beta \in \mathbb{R}^p}{\arg\min} \sum (y_i - x_i^T \beta)^2 + k \sum_{j=1}^{p} \beta_j^2 \\ &= \underset{\beta \in \mathbb{R}^p}{\arg\min} \underbrace{\|y - X\beta\|_2^2}_{\text{Loss}} + k \underbrace{\|\beta\|_2^2}_{\text{Penalty}}\end{aligned}$$

Here $k \geq 0$ is a tuning parameter which controls the strength of the penalty term. Note that

- For $k = 0$, we get the OLS estimate
- For $k = \infty$, we get $\widehat{\beta}(k) = 0$
- For k in between 0 and ∞, model fit and amount of shrinkage are balanced

More of Ridge

- The canonical form of model (1) is:
  Let Λ and T be the matrices of eigenvalues and eigenvectors of $X^\top X$

$$y = XTT^\top \beta + \varepsilon = Z\gamma + \varepsilon$$

  where $Z = XT, \gamma = T^\top \beta$ and $Z^\top Z = \Lambda$,
  $T^\top X^\top XT = \Lambda = diag(\lambda_1, \lambda_2, \ldots, \lambda_p)$, λ_i is the ith eigenvalue of $X^\top X$

- The ridge resression estimator for the canonical model is:

$$\widehat{\beta}(k) = (\Lambda + kI)^{-1}\Lambda\widehat{\beta}$$

$$\widehat{\beta}(k)_i = \frac{\lambda_i}{\lambda_i + k}\widehat{\beta}, \quad i = 1, \ldots, p$$

- This illustrates the essential feature of ridge regression: shrinkage

Bias and Variance of Ridge Regression

Theorem

For any design matrix X, the quantity $\Lambda + \lambda I$ is always invertible, thus there is always a unique solution $\widehat{\beta}(k)$.

Theorem

The variance of the ridge regression estimate is: $Var(\widehat{\beta}(k)_i) = \sigma^2 \frac{\lambda_i}{(\lambda_i + k_i)^2}$

Theorem

The bias of the ridge regression estimate is: $Bias(\widehat{\beta}(k)_i) = -\frac{k}{\lambda_i + k}$

- The total squared bias $\sum_i Bias^2(\widehat{\beta}(k)_i)$ is a monotone increasing sequence with respect to k(amount of shrinkage) while the total variance $\sum_i \mathrm{Var}(\widehat{\beta}(k)_i)$ is a monotone decreasing sequence with respect to k.

Theorem

(Existence theorem)There always exists a $k > 0$ such that the MSE of $\widehat{\beta}(k)$ is always less than the MSE of $\widehat{\beta}$

- Ridge regression is a linear estimator ($\widehat{y} = H_{ridge} y$), with

$$H_{ridge} = X(X^T X + kI)^{-1} X^T$$

- One may define its degrees of freedom to be $tr(H_{ridge})$
- Furthermore one can show that $df_{ridge} = \sum \frac{\lambda_i}{\lambda_i + k}$ where λ_i are the eigenvalues of $X^T X$.

Effect of Multicollinearity

Let us consider two parameter simple linear regression such that $y = \beta_1 x_1 + \beta_2 x_2$

Correlation transformation:

$$x_{tj}^* = \frac{x_{tj} - \overline{x}_j}{s_j\sqrt{n-1}}, \overline{x}_j = \frac{\sum x_{ij}}{n},$$

$$s_j^2 = \frac{\sum_{t=1}^{n}(x_{tj} - \overline{x}_j)^2}{n-1}, y_t^* = \frac{y_t - \overline{y}}{s_y\sqrt{n-1}},$$

$$\mathbf{X}^{*T}\mathbf{X}^* = \begin{bmatrix} {x^*}_1^T {x^*}_1 & {x^*}_1^T {x^*}_2 \\ {x^*}_2^T {x^*}_1 & {x^*}_2^T {x^*}_2 \end{bmatrix} = \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix}$$

$$Var(\widehat{\beta}) = \sigma^2(X^TX)^{-1} = \frac{1}{1-r^2}\begin{bmatrix} 1 & -r \\ -r & 1 \end{bmatrix}, Var(\widehat{\beta}_j) = \frac{1}{1-r^2}$$

r	0	0.3	0.5	0.95	0.99
$Var(\widehat{\beta}_i)$	1	1.1	1.33	10.26	50.25

Another Shrinkage Estimator: Liu

Given a response vector $y \in \mathbb{R}$ and a predictor matrix $X \in \mathbb{R}^{n \times p}$, the Liu regression coefficients are defined as

$$\widehat{\beta}(\eta) = \underset{\beta \in \mathbb{R}^p}{\arg\min} \sum (y_i - x_i^T \beta)^2 + \sum_{j=1}^{p} (\eta \widehat{\beta}_j - \beta_j)^2$$

$$= \underset{\beta \in \mathbb{R}^p}{\arg\min} \underbrace{\|y - X\beta\|_2^2}_{\text{Loss}} + \underbrace{\left\|\eta \widehat{\beta} - \beta\right\|_2^2}_{\text{Penalty}}$$

where $\widehat{\beta}$ is OLS estimator. Here $0 \leq \eta \leq 1$ is a tuning parameter which controls the strength of the penalty term. Note that

- For $\eta = 1$, we get the OLS estimate
- For $\eta = 0$, we get ridge estimate with $k = 1$
- For η in between 0 and 1, model fit and amount of shrinkage are balanced

Another Shrinkage Estimator: Liu (Linearly Unified)

- Liu(Linearly Unified) estimator is a linear combination of ridge and Stein-Rule estimators:

$$\begin{aligned}\widehat{\beta}(\eta) &= (X^TX+I)^{-1}(X^Ty+\eta\widehat{\beta}), \quad 0<d<1 \\ &= \underbrace{(X^TX+I)^{-1}X^Ty}_{\widehat{\beta}(k)} + \underbrace{(X^TX+I)^{-1}\eta\widehat{\beta}}_{c\widehat{\beta}}\end{aligned}$$

- Liu estimator in canonical form is defined as (Liu,1993):

$$\widehat{\beta}(\eta) = (\Lambda+I)^{-1}(\Lambda+\eta I)$$

Theorem

The variance of the Liu regression estimate is: $Var(\widehat{\beta}(\eta)_i) = \sigma^2\frac{(\eta+\lambda_i)^2}{(1+\lambda_i)^2\lambda_i}$

Theorem

The bias of the Liu regression estimate is: $Bias(\widehat{\beta}(\eta)_i) = -\frac{(1-\eta)}{(\lambda_i+1)}$

Theorem

(Existence theorem)There always exists a $0 < \eta < 1$ *such that the MSE of* $\widehat{\beta}(\eta)$ *is always less than the MSE of* $\widehat{\beta}$*.*

- Liu estimator is a linear estimator ($\widehat{y} = H_{liu}y$), with

$$H_{liu} = X(X^TX + I)^{-1}(I + \eta X^TX)X^T$$

- One may define its degrees of freedom to be $tr(H_{liu})$.
- Furthermore one can show that $df_{liu} = \sum \frac{(1+\eta\lambda_i)\lambda_i}{\lambda_i+1}$ where λ_i are the eigenvalues of X^TX.

Determinants of electricity demand

Monthly aggregated(20 cities) electricity consumption(EC) in Germany: 01.1996-08.2010

Unique physical attributes of electricity are

- non-storability
- uncertain and inelastic demand
- steep supply function

Electricity providers are interested in understanding and hedging demand fluctuations!

Determinants of electricity demand

To determine relationship between electricity consumption (EC) and temperature we have to account for:

- price:electricity price, gas price etc.
- seasonal effects: monthly effects etc.
- economic activity: income etc.

Preassumptions:

- EC and price:$-$
- EC and income:$+$
- EC and temperature:unknown

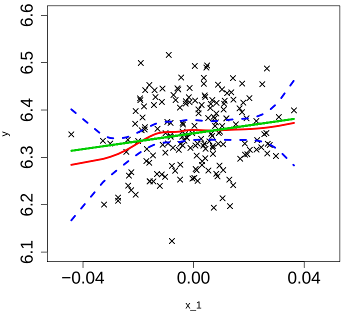

Figure : Plot of income vs. electricity consumption, linear fit(green),local polynomial fit(red), 95% confidence bands(blue).

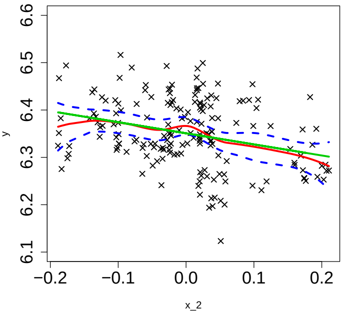

Figure : Plot of relative price vs. electricity consumption, linear fit(green),local polynomial fit(red), 95% confidence bands(blue).

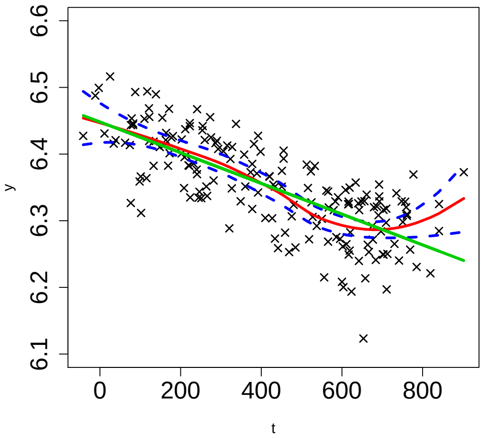

Figure : Plot of temperature vs. electricity consumption, linear fit(green),local polynomial fit(red), 95%confidence bands(black).

Semiparametric Models

- The ordinary regression model can be written as

$$Y = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon. \tag{2}$$

- Assumption: The predictors are linearly related to the response.

Semiparametric Models

- The ordinary regression model can be written as

$$Y = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon. \tag{2}$$

- Assumption: The predictors are linearly related to the response.
- Semiparametric partially linear model is defined as

$$Y = \mathbf{X}^{\mathrm{T}}\boldsymbol{\beta} + f(U) + \varepsilon. \tag{3}$$

- Introduced by Engle, Granger, Rice and Weiss(1986): analyzed the relationship between temperature and electricity usage.
- Became popular in the statistical literature due to the seminal works of Robinson(1988) and Speckman (1988).
- Advantages: Increased flexibility and reduced modeling bias.
- Partially linear models alleviate the curse of dimensionality.

Nonparametric Models versus Semiparametric Models

- Nonparametric models:
  - Relax the restrictive assumptions of parametric models.
  - Are too flexible to permit concise conclusions.
- Semiparametric models:
  - Are natural extensions of ordinary linear regression models with multiple predictors and nonparametric models with several covariates.
  - Retain the explanatory power of parametric models and the flexibility of nonparametric models.

Semiparametric Partial Linear Model

- To avoid identifiability problems we assume that a variable contained in X is not contained also in U, and in general, that no component of X can be mapped to any component of U.
- Stone (1985) states that the three fundamental aspects of statistical models are flexibility, dimensionality and interpretability.
  - **Flexibility** is the ability of the model to provide accurate fits in a wide variety of stuations, inaccuracy here leading to bias in estimation.
  - **Dimensionality** can be thought of in terms of the variance in estimation, the curse of dimensionality being that the amount of data required to avoid an acceptable large variance increases rapidly with increasing dimensionality.
  - **Interpretability** lies in the potential for shedding light on the underlying structure.

Model and Differencing

Semiparametric partial linear model

$$y_i = x_i^\top \beta + f(u_i) + \varepsilon_i, \qquad i = 1, \ldots, n \tag{4}$$

in vector/matrix notation

$$Y = \mathbf{X}^\mathrm{T} \boldsymbol{\beta} + f(U) + \varepsilon \tag{5}$$

where y_i's are observations at u_i, $0 \le u_1 \le u_2 \le \ldots \le u_n \le 1$,
$x_i^\top = (x_{i1}, x_{i2}, \ldots, x_{ip})$,$y = (y_1, \ldots, y_n)^\top$, $X = (x_1, \ldots, x_n)$,
$f = \{f(u_1), \ldots, f(u_n)\}^\top$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^\top$.

$\mathrm{E}(\varepsilon|x, u) = 0$, $\mathrm{Var}(\varepsilon|x, u) = \sigma^2$, f has bounded first derivative($f' < L$)

Differencing removes the nonparametric component (well.. almost).

How does the approximation work?

mth order differencing equation (Yatchew(1997))

$$\sum_{j=0}^{m} d_j y_{i-j} = \left(\sum_{j=0}^{m} d_j x_{i-j}\right)\beta + \left(\sum_{j=0}^{m} d_j f(u_{i-j})\right) + \left(\sum_{j=0}^{m} d_j \varepsilon_{i-j}\right),$$

where $d_0, d_1, \ldots, d_m$ are the differencing weights.
Suppose u_i are equally spaced on the unit interval and $f' \leq L$. By the mean value theorem for some $u_i^* \in [u_{i-1}, u_i]$

$$f(u_i) - f(u_{i-1}) = f^{\top}(u_i^*)(u_i - u_{i-1}) \leq \frac{L}{n}$$

For $m = 1$ from equation (20)

$$\begin{aligned} y_i - y_{i-1} &= (x_i - x_{i-1})\beta + f(u_i) - f(u_{i-1}) + \varepsilon_i - \varepsilon_{i-1} \\ &= (x_i - x_{i-1})\beta + O\left(\frac{1}{n}\right) + \varepsilon_i - \varepsilon_{i-1} \\ &=\approx (x_i - x_{i-1})\beta + \varepsilon_i - \varepsilon_{i-1}. \end{aligned}$$

$$D = \begin{pmatrix} d_0 & d_1 & d_2 & \cdots & d_m & 0 & \cdots & \cdots & 0 \\ 0 & d_0 & d_1 & d_2 & \cdots & d_m & 0 & \cdots & 0 \\ \vdots & \vdots & & & & & & & \\ 0 & \cdots & \cdots & d_0 & d_1 & d_2 & \cdots & d_m & 0 \\ 0 & 0 & \cdots & \cdots & d_0 & d_1 & d_2 & \cdots & d_m \end{pmatrix}$$

$d_0, d_1, \ldots, d_m$ are differencing weights that minimise

$$\sum_{k=1}^{m} \left(\sum_{j=1}^{m-k} d_j d_{k+j} \right)^2,$$

such that

$$\sum_{j=0}^{m} d_j = 0 \quad \text{and} \quad \sum_{j=0}^{m} d_j^2 = 1 \qquad (6)$$

are satisfied. The constraints (6) ensure that the nonparametric effect is removed as $n \to \infty$ and $\text{Var}(\tilde{\varepsilon}) = \text{Var}(\varepsilon) = \sigma^2$ respectively.

Differencing the model

$$
\begin{aligned}
Dy &= DX\beta + Df + D\varepsilon \approx DX\beta + D\varepsilon \\
\widetilde{y} &\approx \widetilde{X}\beta + \widetilde{\varepsilon} \qquad (7)
\end{aligned}
$$

where $\widetilde{y} = Dy$, $\widetilde{X} = DX$ and $\widetilde{\varepsilon} = D\varepsilon$.

$$
\begin{aligned}
\widehat{\beta}_D &= \left\{(DX)^\top(DX)\right\}^{-1}(DX)^\top Dy \qquad (8) \\
&= \left(\widetilde{X}^\top\widetilde{X}\right)^{-1}\widetilde{X}^\top\widetilde{y}
\end{aligned}
$$

(Yatchew, 2003).

$$
\widehat{\sigma}^2 = \frac{\widetilde{y}^\top(I - P^\perp)\widetilde{y}}{tr\{D^\top(\mathrm{I} - P^\perp)D\}} \qquad (9)
$$

with $P^\perp = \widetilde{X}(\widetilde{X}^\top\widetilde{X})^{-1}\widetilde{X}^\top$, I $(p \times p)$: identity matrix and $tr(\cdot)$: trace of a matrix (Eubank et al., 1998).

Differenced based Liu-type estimator

Akdeniz-Duran et.al [1] proposed the difference-based Liu-type estimator minimizing (w.r.t. β)

$$L = (\tilde{y} - \widetilde{X}\beta)^\top(\tilde{y} - \widetilde{X}\beta) + (\eta\widehat{\beta}_D - \beta)^\top(\eta\widehat{\beta}_D - \beta)$$

as

$$\widehat{\beta}_D(\eta) = (\widetilde{X}^\top\widetilde{X} + I)^{-1}(\widetilde{X}^\top\tilde{y} + \eta\widehat{\beta}_D) \tag{10}$$

where η, $0 \leq \eta \leq 1$ is a biasing parameter and when $\eta = 1$, $\widehat{\beta}_D(\eta) = \widehat{\beta}_D$.

$$\text{Bias}\{\widehat{\beta}_D(\eta)\} = -(1-\eta)(\widetilde{X}^\top\widetilde{X} + \mathrm{I})^{-1}\beta. \tag{11}$$

[1] Akdeniz-Duran, E., Haerdle, W., Osipenko, M. (2012). Difference based ridge and Liu type estimators in semiparametric regression models. *Journal of Multivariate Statistics*

Generalization of the Model

- Semiparametric partial linear model for observations $\{y_i, x_i, u_i\}_{i=1}^n$

$$y_i = x_i^\top \beta + f(u_i) + \varepsilon_i, \;\; i = 1, \ldots, n \tag{12}$$

- Previously we assumed $\mathrm{E}(\varepsilon\varepsilon^\top) = \sigma^2 \mathrm{I}$.
- Real data often reveals heteroscedasticity and in time series context error term exhibits autocorrelation.
- Now we assume $\mathrm{E}(\varepsilon\varepsilon^\top) = \sigma^2 V$, V not necessarily diagonal. Thus $\mathrm{E}(\widetilde{\varepsilon}\widetilde{\varepsilon}^\top) = \sigma^2 V_D$ where $V_D = DVD^\top$.
- Generalized difference-based estimator is

$$\widehat{\beta}_{GD} = (\widetilde{X}^\top V_D^{-1} \widetilde{X})^{-1} \widetilde{X}^\top V_D^{-1} \widetilde{y}.$$

EGeneralization of the Model cont'd

- Properties of the generalized difference-based estimator of β depends on the characteristics of the information matrix $\widetilde{X}^\top V_D^{-1}\widetilde{X} = G$.
- The estimate of the σ^2 is

$$\widehat{\sigma}^2 = \frac{1}{n}(\widetilde{y} - \widetilde{X}\widehat{\beta}_{GD})^\top V_D^{-1}(\widetilde{y} - \widetilde{X}\widehat{\beta}_{GD}) \tag{13}$$

It is easy to show that

$$s^2 = \frac{1}{n-p}(\widetilde{y} - \widetilde{X}\widehat{\beta}_{GD})^\top V_D^{-1}(\widetilde{y} - \widetilde{X}\widehat{\beta}_{GD}) \tag{14}$$

is an unbiased estimator of σ^2 .

- If $G(p \times p), p << n - m$ matrix is ill-conditioned with a large condition number, then the $\widehat{\beta}_{GD}$ produces large sampling variances. Use restricted estimation or shrinkage estimation!

Generalized Difference-based Restricted Estimator

- The available prior information sometimes can be expressed in the form of exact, stochastic or inequality restrictions.
- We assumed [2] an exact linear restriction on the parameters

$$R\beta = r$$

$R(q \times p)$ matrix, $r(q \times 1)$ vector.

- Thus the generalized difference-based restricted parameter estimate is given by

$$\widehat{\beta}_{GRD} = \widehat{\beta}_{GD} + G^{-1}R^\top(RG^{-1}R^\top)^{-1}(r - R\widehat{\beta}_{GD})$$

where $G = \widetilde{X}^\top V_D^{-1}\widetilde{X}$.

[2] Akdeniz, F., Akdeniz-Duran, E., Roozbeh,M., Arashi, M. (2013). Efficiency of the generalized difference-based Liu estimators in semiparametric regression models with correlated errors. *Journal of Statistical Computation and Simulation*

Generalized Difference-based Stochastic Restricted Estimator

- We do not always have *exact* prior information such as $R\beta = r$ due to economic relations, industrial structures, production planning, etc.
- Thus we assume a stochastic linear constraint $r = R\beta + \epsilon$, $\epsilon \sim (0, \sigma^2 W)$ where W is assumed to be known and positive definite.
- Prior information are to be assigned not necessarily equal weights(ω).
- In order to incorporate the restrictions in the estimation of parameters, we minimize (w.r.t. β)

$$(\tilde{y} - \tilde{X}\beta)^\top V_D^{-1}(\tilde{y} - \tilde{X}\beta) + \omega(r - R\beta)^\top W^{-1}(r - R\beta)$$

- $0 < \omega < 1$: prior information receives less weight in comparison to the sample information
- $\omega > 1$: higher weight to the prior information(of little practical interest).

This leads to the following solution for β:

$$\widehat{\beta}_{GDWM}(\omega) = (G + \omega R^\top W^{-1} R)^{-1} (\widetilde{X} V_D^{-1} \widetilde{y} + \omega R^\top W^{-1} r),$$

Since

$$(G + \omega R^\top W^{-1} R)^{-1} = G^{-1} - \omega G^{-1} R^\top (W + \omega R G^{-1} R^\top)^{-1} R G^{-1}$$

we have *generalized difference-based weighted mixed estimator* as

$$\widehat{\beta}_{GDWM}(\omega) = \widehat{\beta}_{GD} + \omega G^{-1} R^\top (W + \omega R G^{-1} R^\top)^{-1} (r - R\widehat{\beta}_{GD}). \quad (15)$$

For $\omega = 1$ we obtain the *generalized difference-based mixed estimator*:

$$\widehat{\beta}_{GDM} = (G + R^\top W^{-1} R)^{-1} (\widetilde{X}^\top V_D^{-1} \widetilde{y} + R^\top W^{-1} r), \quad (16)$$

This estimator in (16) gives equal weight to sample and prior information.

Generalized Difference-based Liu Estimator

(Generalized) Liu estimator proposed by Liu(1993) and Akdeniz and Kaciranlar(1995) is defined as

$$\begin{aligned}\widehat{\beta}_{GDL}(\eta) &= (G+I)^{-1}(\widetilde{X}^\top V_D^{-1}\widetilde{y} + \eta\widehat{\beta}_{GD}), 0 \le \eta \le 1, \\ &= (G+I)^{-1}(G+\eta I)\widehat{\beta}_{GD} = F_\eta \widehat{\beta}_{GD} = F_\eta G^{-1}\widetilde{X}^\top V_D^{-1}\widetilde{y} \qquad (17)\end{aligned}$$

where $\widehat{\beta}_{GD} = (\widetilde{X}^\top V_D^{-1}\widetilde{X})^{-1}\widetilde{X}^\top V_D^{-1}\widetilde{y}$ is the generalized difference-based estimator of β and $F_\eta = (G+I)^{-1}(G+\eta I)$.
Observing that F_η and G^{-1} are commutative, we have

$$\widehat{\beta}_{GDL}(\eta) = G^{-1}F_\eta \widetilde{X}^\top V_D^{-1}\widetilde{y} \qquad (18)$$

Generalized Difference-based Weighted Mixed Liu Estimator

We have obtained

$$\widehat{\beta}_{GDWM}(\omega) = \widehat{\beta}_{GD} + \omega G^{-1}R^\top (W + \omega R G^{-1} R^\top)^{-1}(r - R\widehat{\beta}_{GD}).$$

substituting $\widehat{\beta}_{GD}(\eta)$ with $\widehat{\beta}_{GDL}(\eta)$ in $\widehat{\beta}_{GDWM}(\omega)$, we describe a *generalized difference-based weighted mixed Liu estimator (GDWMLE)*, as follows:

$$\begin{aligned}\widehat{\beta}_{GDWML}(\omega,\eta) &= \widehat{\beta}(\omega,\eta) \\ &= \widehat{\beta}_{GDL}(\eta) + \omega G^{-1}R^\top (W + \omega R G^{-1} R^\top)^{-1}(r - R\widehat{\beta}_{GDL}(\eta))\end{aligned}$$

arranging terms we finally have

$$\widehat{\beta}_{GDWML}(\omega,\eta) = (G + \omega R^\top W^{-1} R)^{-1}(F_d \widetilde{X}^\top V_D^{-1}\widetilde{y} + \omega R^\top W^{-1} r) \quad (19)$$

(Hubert and Wijekoon(2006) and Yang et.al(2009)).

In fact, from the definition of $\widehat{\beta}(\omega, \eta)$, we can see that it is a general estimator which includes the $\widehat{\beta}_{DWM}$, $\widehat{\beta}_{DL}$ and the mixed regression estimator $\widehat{\beta}_{GDM}$ as special cases. Namely,
if $\eta = 1$ then $\widehat{\beta}(\omega, \eta = 1) = \widehat{\beta}_{GDWM}(\omega)$,
if $\omega = 0$ then $\widehat{\beta}(\omega = 0, \eta) = \widehat{\beta}_{GDL}(\eta)$,
if $\eta = 1$ and $\omega = 1$ then $\widehat{\beta}(\omega = 1, \eta = 1) = \widehat{\beta}_{GDM}$.

Lemmas

Lemma

(Farebrother, 1976) Let A be a positive definite matrix, namely $A > 0$, and let α be some vector, then $A - \alpha\alpha^\top \geq 0$ if and only if $\alpha^\top A^{-1}\alpha \leq 1$.

Lemma

Let $n \times n$ matrices $M > 0$, $N \geq 0$, then $M > N$ if and only if $\lambda_{max}(NM^{-1}) < 1$ (Rao,2008).

Lemma

(Trenkler and Toutenburg,1990) Let $\widehat{\beta}_j = A_j y$, $j = 1, 2$ be two competing estimators of β. Suppose that $\Delta = Cov(\widehat{\beta}_1) - Cov(\widehat{\beta}_2) > 0$. Then $MSEM(\widehat{\beta}_1) - MSEM(\widehat{\beta}_2) \geq 0$ if and only if $b_2^\top(\Delta + b_1 b_1^\top)^{-1} b_2 \leq 1$, where b_j denotes bias vector of $\widehat{\beta}_j$.

MSEM Superiority of $\widehat{\beta}_{GDWM}(\omega)$ and $\widehat{\beta}(\omega, \eta)$ over $\widehat{\beta}_{GD}$

- Expectation and dispersion matrices of the $\widehat{\beta}(\omega, \eta)$ are

$$E(\widehat{\beta}(\omega, \eta)) = BA\beta \ \ and \ \ Var(\widehat{\beta}(\omega, \eta)) = \sigma^2 B\widetilde{A}B^\top \tag{20}$$

- The bias of $\widehat{\beta}(\omega, \eta)$ is

$$Bias(\widehat{\beta}(\omega, \eta)) = E(\widehat{\beta}(\omega, \eta)) - \beta = B(F_\eta - I)G\beta \tag{21}$$

- The mean squared error matrix of $\widehat{\beta}(\omega, d)$ is

$$\begin{aligned} MSEM(\widehat{\beta}(\omega, \eta)) &= Var(\widehat{\beta}(\omega, \eta)) + Bias(\widehat{\beta}(\omega, \eta))Bias(\widehat{\beta}(\omega, d))^\top \\ &= \sigma^2 B\widetilde{A}B^\top + b_1 b_1^\top \end{aligned} \tag{22}$$

$B =: (G + \omega R^\top W^{-1} R)^{-1}, A =: (F_\eta G + \omega R^\top W^{-1} R)$ *and* $\widetilde{A} =:$ $(F_\eta D {F_\eta}^\top + \omega^2 R^\top W^{-1} R)$, $b_1 = B(F_\eta - I)G\beta$.

MSEMs of the Estimators

$$MSEM(\widehat{\beta}_{GD}) = Var(\widehat{\beta}_{GD}) = \sigma^2 G^{-1} \quad (23)$$

$$MSEM(\widehat{\beta}_{GDM}) = Var(\widehat{\beta}_{GDM}) = \sigma^2 (G + R^\top W^{-1} R)^{-1} \quad (24)$$

$$MSEM(\widehat{\beta}_{GDL}(\eta)) = \sigma^2 F_\eta G^{-1} F_\eta^\top + b_2 b_2^\top, \quad (25)$$

with $b_2 = Bias(\widehat{\beta}_{GDL}(\eta)) = (F_\eta - I)\beta$.

$$\begin{aligned} MSEM(\widehat{\beta}_{GDWM}(\omega)) &= MSEM(\widehat{\beta}(\omega)) \\ &= Var(\widehat{\beta}(\omega)) = \sigma^2 B(G + \omega^2 R^\top W^{-1} R) B^\top \end{aligned} \quad (26)$$

MSEM Comparison between $\widehat{\beta}(\omega)$ and $\widehat{\beta}(\omega, \eta)$

Theorem

The generalized difference-based weighted mixed Liu estimator $\widehat{\beta}(\omega, \eta)$ is superior to the generalized difference-based weighted mixed estimator $\widehat{\beta}(\omega)$ in the MSEM sense, namely $MSEM(\widehat{\beta}(\omega)) - MSEM(\widehat{\beta}(\omega, \eta)) \geq 0$ if and only if $\sigma^{-2} b_1^\top B^{\star -1} b_1 \leq 1$.

Proof.

$$MSEM(\widehat{\beta}(\omega)) - MSEM(\widehat{\beta}(\omega, \eta)) = \sigma^2 B \Delta_1 B^\top - b_1 b_1^\top$$

where $\delta_i = \lambda_i(G) - \frac{\lambda_i(G)(\lambda_i(G)+\eta)^2}{(\lambda_i(G)+1)^2} = \frac{(1-\eta)(1+2\lambda_i(G)+\eta)\lambda_i(G)}{(\lambda_i(G)+1)^2}$. Since $0 < \eta < 1$ and $\lambda_i(G) > 0$, then $\delta_i > 0$. We note that $\delta_i > 0$ is monotonic decreasing in η. Observing that $B =: (G + \omega R^T W^{-1} R)^{-1} > 0$ we get $\Delta_1 > 0$ and $B^\star =: B\Delta_i B^\top > 0$.By lemma 1, we get $MSEM(\widehat{\beta}(\omega)) - MSEM(\widehat{\beta}(\omega, \eta)) \geq 0$ if and only if $\sigma^{-2} b_1^\top B^{\star -1} b_1 \leq 1$. □

MSEM Comparison between $\widehat{\beta}_{GDL}(\eta)$ and $\widehat{\beta}(\omega,\eta)$

Theorem

When $\lambda_{max}(NM^{-1}) < 1$, the generalized difference-based weighted mixed Liu estimator $\widehat{\beta}(w,\eta)$ is superior to the generalized difference-based Liu estimator $\widehat{\beta}_{GDL}(\eta)$ in the MSEM sense, namely $MSEM(\widehat{\beta}_{GDL}(\eta)) - MSEM(\widehat{\beta}(\omega,\eta)) \geq 0$ if and only if $b_1^\top(\sigma^2\Delta_2 + b_2 b_2^\top)^{-1} \leq 1$.

Proof.

$$MSEM(\widehat{\beta}_{GDL}(\eta)) - MSEM(\widehat{\beta}(\omega,\eta)) = \sigma^2\Delta_2 + b_2 b_2^\top - b_1 b_1^\top.$$

where $M = F_\eta G^{-1} F_\eta^\top, N = B(F_\eta G F_\eta{}^\top + \omega R^\top W^{-1} R)B^\top$ and $\Delta_2 = M - N$. It is obvious that, $M = F_\eta G^{-1} F_\eta^\top > 0$, $N = B(F_\eta G F_\eta^\top + \omega^2 R^\top W^{-1} R)B^\top > 0$. Therefore, when $\lambda_{max}(NM^{-1}) < 1$, we get $\Delta_2 > 0$ by applying Lemma 2. By Lemma 3, we have $MSEM(\widehat{\beta}_{GDL}(\eta)) - MSEM(\widehat{\beta}(\omega,\eta)) \geq 0$ if and only if $b_1^\top(\sigma^2\Delta_2 + b_2 b_2^\top)^{-1} b_1 \leq 1$. □

Variance Comparison between $\widehat{\beta}_{GD}$ and $\widehat{\beta}_{GDWM}$

Theorem

The generalized difference-based weighted mixed estimator $\widehat{\beta}_{GDWM}$ is superior to the generalized difference-based estimator in the sense that $Var(\widehat{\beta}_{GD}) - Var(\widehat{\beta}_{GDWM}) \geq 0$ if and only if $\left(\frac{2}{\omega} - 1\right) W + RG^{-1}R^{\top} > 0$.

Proof.

$$\Delta_3 = Var(\widehat{\beta}_{GD}) - Var(\widehat{\beta}_{GDWM})$$
$$= \sigma^2\omega^2 BR^{\top}W^{-1}\left[\left(\frac{2}{\omega} - 1\right) W + RG^{-1}R^{\top}\right] W^{-1}RB^{\top}.$$

The difference is positive definite when B is positive definite and $\left[\left(\frac{2}{\omega} - 1\right) W + RG^{-1}R^{\top}\right] > 0$ which is positive definite as long as $\omega < 2$. When $q < p$, R has full row rank, therefore $R^{\top}$ has full column rank and it follows that in this case we can only conclude that $\Delta_3 \geq 0$. □

Application 1

The data set is from a household demand for gasoline in Canada (Yatchew(2003)). We allow price and age to appear nonparametrically: The basic specification is given by Model:

$$dist = f(price, age) + \beta_1 income + \beta_2 drivers + \beta_3 hhsize \\ + \beta_4 youngsingle + \beta_5 urban + \text{monthly dummies} + \varepsilon$$

dist: log of distance traveled per month by household, **price**: log of price of 1 lt gasoline, age: log of age, **hhsize**: size of the household etc.
specification test result of joint significance of the nonparametric variables: 5.96
order of differencing: m=10 with d=(0.9494,-0.1437,-0.1314,-0.1197,-0.1085,-0.0978,-0.0877,-0.0782,-0.0691,-0.0606,-0.527)

Results 1

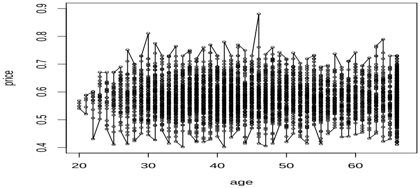

Figure : Ordering of data with respect to price and age

	Diff. Liu		Difference	
	Est.	Std. Error	Est.	Std.Error
income	0.287	0.021	0.291	0.021
drivers	0.532	0.035	0.571	0.033
hhsize	0.122	0.029	0.093	0.027
youngsingle	0.198	0.063	0.191	0.061
urban	-0.333	0.020	-0.332	0.020
R^2	.270		.263	
s^2_{diff}	.496		.501	

Results 2

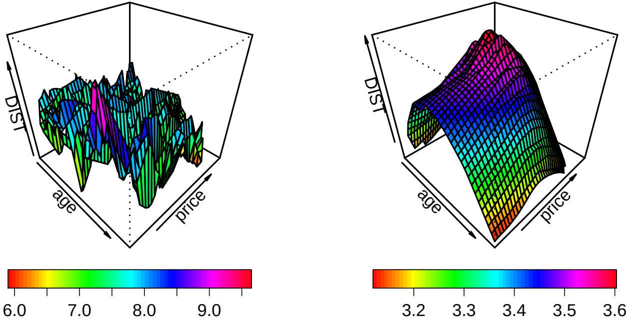

Figure : raw data graph(left), smooth graph after estimating parametric terms with Liu estimator

Having estimated the parametric effects using GDWML estimator, the constructed data $(y_i - z_i\beta_{diff}, age_i, price_i)$ are then smoothed to estimate

Application 2: Hedonic Pricing of Housing Attributes

- Housing prices are very much affected by location (Yatchew(2003))
- The price surface may be unimodal, multimodal or have ridges (prices along the subway are often higher)
- Thus, we include a two-dimensional nonparametric effect.
- Dataset: 92 detached homes in Ottawa.
- y(dependent variable): saleprice
- lotarea, square footage of housing, average neighbourhood income etc. are independent variables

Results 1

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	73.9775	17.9961	4.11	0.0001
fireplac	11.7950	6.1523	1.92	0.0587
garage	11.8383	5.0793	2.33	0.0223
luxbath	60.7362	10.5148	5.78	0.0000
avginc	0.4776	0.2244	2.13	0.0364
disthwy	-15.2774	6.6974	-2.28	0.0252
lotarea	3.2432	2.2766	1.42	0.1581
nrbed	6.5860	4.8962	1.35	0.1823
usespace	21.1285	10.9864	1.92	0.0580
south	7.5268	2.1505	3.50	0.0008
west	-3.2062	2.5296	-1.27	0.2086
R^2	.62			
s^2_{res}	424.3			

Results 2

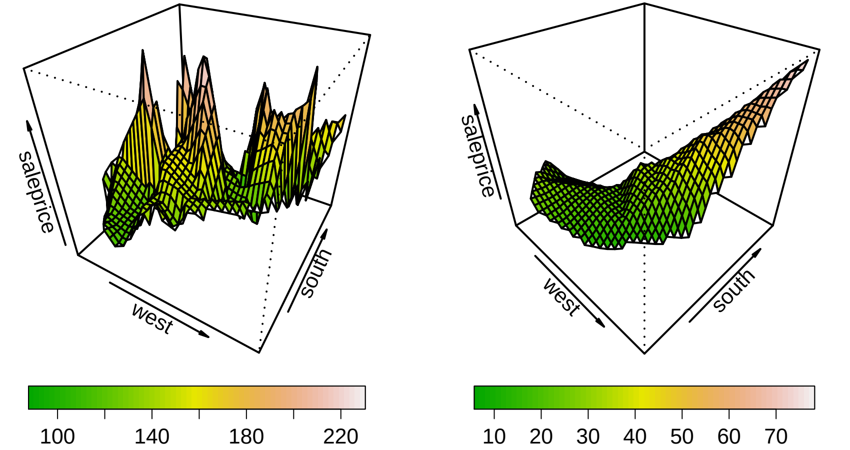

Figure : (a) Graph of location versus saleprice , (b) Local Polynomial regression

Results 3

	Estimate	Std. Error	t value	Pr(>|t|)
fireplac	12.6489	5.7875	2.19	0.0316
garage	12.9252	4.9146	2.63	0.0102
luxbath	57.6187	10.5741	5.45	0.0000
avginc	0.5968	0.2342	2.55	0.0126
disthwy	1.5538	21.3877	0.07	0.9423
lotarea	3.0986	2.2387	1.38	0.1700
nrbed	6.4323	4.7490	1.35	0.1792
usespace	24.7213	10.5893	2.33	0.0220
R^2	.66			
s^2_{diff}	375.5			

Table : Parametric part is estimated by Liu-type regression

Simulation Design

- Various biasing parameters are considered: $\eta = (0, 0.1, 0.2, ..., 1)$
- Various weights $w = (0.1, 0.3, 0.5, 0.7, 0.9)$
- To achieve different degrees of collinearity predictors are generated by:

$$x_{ij} = (1 - \gamma^2)^{1/2} z_{ij} + \gamma z_{z_{ip}}, i = 1, \ldots, n, j = 1, \ldots, p.$$

  where z_{ij} are i.i.d. normal numbers.
- $\gamma = (0.80, 0.90, 0.99)$ where correlation between two explanatory variables are γ^2
- Dependent observations are generated by

$$y_i = \sum_{j=1}^{5} x_{ji}\beta_j + f(t_i) + \varepsilon_i$$

  where $\beta = (1.5, 2, 3, -5, 4)$, $\varepsilon \sim N(0, \sigma^2 V)$ the elements of V are $v_{ij} = \left(\frac{1}{n}\right)^{|i-j|}, \sigma^2 = 4.$

Simulation Design

Nonparametric function is generated by

$$f(t_i) = \sqrt{t_i(1 - t_i)}sin(2.1\pi/(t_i + 0.05))$$

where $t_i = (i - 0.5)/n$ which is called the Doppler function.

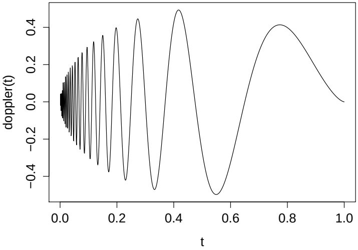

Figure : Nonparametric part of the model

- Difficult to estimate, spatially inhomogenous function its smoothness varies over t.

Simulation Design

- The restriction on the parameters are taken to be

$$R\beta + \epsilon = r, \quad \epsilon \sim N(0,1)$$

  with $R = (0, 1, -1, 1, 0)$ and $r = 0$.
- The fourth-order optimal differencing weights for $m = 4$ are $d_0 = 0.8873, d_1 = -0.3099, d_2 = -0.2464, d_3 = -0.1901, d_4 = -0.1409$ (Yatchew(2003),p.61)
- The difference $(100 - 4) \times 100$ differencing matrix is defined as follows:

$$D = \begin{pmatrix} d_0 & d_1 & d_2 & d_3 & d_4 & 0 & \cdots & \cdots & 0 \\ 0 & d_0 & d_1 & d_2 & d_3 & d_4 & 0 & \cdots & 0 \\ \vdots & \vdots & & & & & & & \\ 0 & \cdots & \cdots & d_0 & d_1 & d_2 & \cdots & d_4 & 0 \\ 0 & 0 & \cdots & \cdots & d_0 & d_1 & d_2 & d_3 & d_4 \end{pmatrix}$$

Simulation Design

- X and y values are ordered with respect to the nonparametric variable.
- The matrix $G = \widetilde{X}^\top V_D^{-1} \widetilde{X}$ has condition numbers $20.92, 33.23, 352.63$ for $\gamma = 0.8$, $\gamma = 0.9$ and $\gamma = 0.99$, respectively, which implies the existence of multicollinearity in the data set.
- Differencing we can perform inference on β as if there were no nonparametric component f in the model to begin with.

Performance criteria

The performances of the estimators are measured by

$$\widehat{\text{SMSE}}(\widehat{\beta}) = \frac{1}{1000}\sum_{l=1}^{1000}\left\|\widehat{\beta}_{(l)} - \beta\right\|^2$$

$$\widehat{\text{ABIAS}}(\widehat{\beta}) = \frac{1}{1000}\sum_{l=1}^{1000}\sum_{i=1}^{p}\left|\widehat{\beta}_{i(l)} - \beta_i\right|$$

where $\widehat{\beta}_{(l)}$ denotes the estimated parameter in the ith iteration

$$\widehat{\text{MSE}}(\widehat{f}(u), f(u)) = \frac{1}{1000}\sum_{i=1}^{1000}\left[\widehat{f}(u_i) - f(u_i)\right]^2$$

Simulation Results

Table : Evaluation of the risk functions for $\gamma = 0.99$ different values of ω

	GDE	GDWME	GDME	GDLE	GDWMLE
mse($\omega = 0.1$)	157.66	157.66	157.17	155.80	155.80
bias($\omega = 0.1$)	21.62	21.62	21.54	21.46	21.46
mse($\omega = 0.3$)	156.81	156.60	156.32	154.96	154.74
bias($\omega = 0.3$)	21.65	21.62	21.56	21.48	21.45
mse($\omega = 0.5$)	156.35	156.00	155.81	154.51	154.16
bias($\omega = 0.5$)	21.53	21.48	21.44	21.36	21.31

Optimum value of η for Liu estimator are calculated using the generalized cross validation.

Final Considerations

- We have proposed estimation methods for partially linear regression framework based on difference-based method.
- Especially, the estimation of the parametric part is considered.
- Stochastic restrictions were put on the parameter space with specified weights.
- The results are generalized for heteroscedastic cases.
- The proposed estimators GDLE and GDWMLE had smaller smses for all cases.

References I

Engle,R.F., Granger,C.W.J., Rice, J. and Weiss,A.
Semiparametric estimates of the relation between weather and electricity sales
Journal of American Statistical Association,81:310-320,1986.

Härdle,W.K., Müller,M., Sperlich,M.S. and Werwatz,A.
Nonparametric and Semiparametric Models
Springer Verlag, Heidelberg, 2004.

Hoerl, A.E. and Kennard, R.W.
Ridge regression: Biased estimation for nonorthogonal problems.
Technometrics,12: 55-67,1970.

Liu, K.
A new class of biased estimate in linear regression.
Communication in Statistics: Theory and Methods,22:393-402,1993.

References II

Green,P.J. and Silverman, B.W.
Nonparametric Regression and Generalized Linear Models.
Chapman and Hall, 1994.

Akdeniz-Duran,E., Haerdle, W. and Osipenko, M.
Difference based ridge and Liu type estimators in semiparametric regression models.
Journal of Multivariate Analysis, 105, 164-175, 2012.

Akdeniz,F.,Akdeniz-Duran,E., Roozbeh, M. and Arashi, M.
Efficiency of the generalized difference-based Liu estimators in semiparametric regression models with correlated errors.
Journal of Statistical Computation and Simulation, 2013, DOI: 10.1080/00949655.2013.806924.

Akdeniz-Duran,E., Akdeniz, F. and Hu, H.
Efficiency of a Liu-type estimator in semiparametric regression models.
Journal of Computational and Applied Mathematics, 235, 1418-1428, 2011.

References III

Akdeniz-Duran,E., Akdeniz, F. and Hu, H.

Efficiency of the generalized difference-based Liu estimators in semiparametric regression models with correlated errors.

Journal of Statistical Computation and Simulation,DOI:10.1080/00949655.2013.806924.

Akdeniz,F., Akdeniz-Duran, E., Roozbeh, M. and Arashi, M.

New difference-based estimator of parameters in semiparametric regression models.

Journal of Statistical Computation and Simulation, Volume 83, Issue 5, 2013.

www.r-project.rog

Thank you.

Questions?

A New Difference-based Weighted Mixed Liu Estimator in Partially Linear Models

Fikri Akdeniz
fikriakdeniz@cag.edu.tr

Çag University, Mersin-TURKEY
Department of Mathematics and Computer Science

11-06-2014