Ames – February 18, 2004

Nonparametric model calibration estimation in survey sampling

M. Giovanna Ranalli

Department of Statistics, Colorado State University

(Joint work with G.E. Montanari, Dipartimento di Scienze Statistiche, Università di Perugia, Italy)

Outline

Outline

- Introduction to the framework and notation

Outline

- Introduction to the framework and notation
- Auxiliary information: Calibration

Outline

- Introduction to the framework and notation
- Auxiliary information: Calibration
- Allowing for more flexible models: Model Calibration

Outline

- Introduction to the framework and notation
- Auxiliary information: Calibration
- Allowing for more flexible models: Model Calibration
- Nonparametric methods for Model Calibration: Neural Networks and Local Polynomials

Outline

- Introduction to the framework and notation
- Auxiliary information: Calibration
- Allowing for more flexible models: Model Calibration
- Nonparametric methods for Model Calibration: Neural Networks and Local Polynomials
- A small simulation study

Outline

- Introduction to the framework and notation
- Auxiliary information: Calibration
- Allowing for more flexible models: Model Calibration
- Nonparametric methods for Model Calibration: Neural Networks and Local Polynomials
- A small simulation study
- Conclusions and perspectives

The Problem – aim and tools

We want to know more about a Finite Population ...

- $\mathcal{U} = \{u_1, u_2, ..., u_N\}$ $\rightarrow$ finite population of N distinct units labelled by the integers $i = 1, ..., N$.
- y_i $\rightarrow$ value taken by the survey variable y in unit i.
- $\bar{Y} = N^{-1} \sum_{i \in \mathcal{U}} y_i$ $\rightarrow$ population mean of y, parameter of interest.

The Problem – aim and tools

We want to know more about a Finite Population ...

- $\mathcal{U} = \{u_1, u_2, ..., u_N\}$ $\rightarrow$ finite population of N distinct units labelled by the integers $i = 1, ..., N$.
- y_i $\rightarrow$ value taken by the survey variable y in unit i.
- $\bar{Y} = N^{-1} \sum_{i \in \mathcal{U}} y_i$ $\rightarrow$ population mean of y, parameter of interest.

... by means of a Probabilistic sample and Auxiliary Information.

- s $\rightarrow$ sample of size n selected from $\mathcal{U}$ according to a sampling design $p(s)$, which induces first and second order inclusion probabilities π_i and π_{ij}.
- y_i, for $i \in s$ $\rightarrow$ values of y observed on sample units.
- $\boldsymbol{x}_i = (x_{1i}, x_{2i}, \ldots, x_{Qi})'$ $\rightarrow$ value taken by Q auxiliary variables $\boldsymbol{x}$ on unit i and *known* for $i \in \mathcal{U}$.

Calibration Estimators – Deville & Särndal (1992)

Calibration Estimators – Deville & Särndal (1992)

$$\mathsf{CAL} = N^{-1} \sum_{i \in s} w_i y_i$$

Calibration Estimators – Deville & Särndal (1992)

$$\mathsf{CAL} = N^{-1} \sum_{i \in s} w_i y_i$$

The weights w_i are found to be as close as possible to $d_i = \pi_i^{-1}$, while meeting benchmark constraints:

Calibration Estimators – Deville & Särndal (1992)

$$\text{CAL} = N^{-1} \sum_{i \in s} w_i y_i$$

The weights w_i are found to be as close as possible to $d_i = \pi_i^{-1}$, while meeting benchmark constraints:

$$\min_{w_i} \sum_{i \in s} \Phi_i(w_i, d_i) \qquad \text{s.t.} \qquad \boxed{N^{-1} \textstyle\sum_{i \in s} w_i \boldsymbol{x}_i = N^{-1} \sum_{i \in \mathcal{U}} \boldsymbol{x}_i}$$

Calibration Estimators – Deville & Särndal (1992)

$$\mathsf{CAL} = N^{-1} \sum_{i \in s} w_i y_i$$

The weights w_i are found to be as close as possible to $d_i = \pi_i^{-1}$, while meeting benchmark constraints:

$$\min_{w_i} \sum_{i \in s} \Phi_i(w_i, d_i) \qquad \text{s.t.} \qquad \boxed{N^{-1} \sum_{i \in s} w_i \boldsymbol{x}_i = N^{-1} \sum_{i \in \mathcal{U}} \boldsymbol{x}_i}$$

Different distance measures can be employed in the minimization procedure as long as they meet some basic requirements.

How to get a solution

How to get a solution

✓ The minimization problem is solved by $\phi_i(w_i, d_i) - \boldsymbol{x}_i\boldsymbol{\ell} = 0$, where $\phi_i(w, d) = \partial\Phi(w, d)/\partial w$ and $\boldsymbol{\ell}$ is a vector of Lagrangian multipliers.

How to get a solution

- ✓ The minimization problem is solved by $\phi_i(w_i, d_i) - \boldsymbol{x}_i\boldsymbol{\ell} = 0$, where $\phi_i(w, d) = \partial\Phi(w, d)/\partial w$ and $\boldsymbol{\ell}$ is a vector of Lagrangian multipliers.
- ✓ If a solution exists, basic requirements guarantee it is unique and can be written as $w_i = d_i F_i(\boldsymbol{x}_i\boldsymbol{\ell})$, where $d_i F_i(\cdot)$ is the reciprocal mapping of $\phi(\cdot, d_i)$.

How to get a solution

✓ The minimization problem is solved by $\phi_i(w_i, d_i) - \boldsymbol{x}_i\boldsymbol{\ell} = 0$, where $\phi_i(w, d) = \partial\Phi(w, d)/\partial w$ and $\boldsymbol{\ell}$ is a vector of Lagrangian multipliers.

✓ If a solution exists, basic requirements guarantee it is unique and can be written as $w_i = d_i F_i(\boldsymbol{x}_i\boldsymbol{\ell})$, where $d_i F_i(\cdot)$ is the reciprocal mapping of $\phi(\cdot, d_i)$.

✓ The vector $\boldsymbol{\ell}$ can be determined through the calibration constraints:

$$\bar{\boldsymbol{x}} = \frac{1}{N}\sum_{i\in s} w_i \boldsymbol{x}_i = \sum_{i\in s} d_i F_i(\boldsymbol{x}_i\boldsymbol{\ell})\boldsymbol{x}_i \quad \Rightarrow \quad g_s(\boldsymbol{\ell}) = \sum_{i\in s} d_i(F_i(\boldsymbol{x}_i\boldsymbol{\ell}) - 1)\boldsymbol{x}_i = \bar{\boldsymbol{x}} - \hat{\bar{\boldsymbol{x}}}.$$

How to get a solution

- ✓ The minimization problem is solved by $\phi_i(w_i, d_i) - \boldsymbol{x}_i\boldsymbol{\ell} = 0$, where $\phi_i(w, d) = \partial\Phi(w, d)/\partial w$ and $\boldsymbol{\ell}$ is a vector of Lagrangian multipliers.
- ✓ If a solution exists, basic requirements guarantee it is unique and can be written as $w_i = d_i F_i(\boldsymbol{x}_i\boldsymbol{\ell})$, where $d_i F_i(\cdot)$ is the reciprocal mapping of $\phi(\cdot, d_i)$.
- ✓ The vector $\boldsymbol{\ell}$ can be determined through the calibration constraints:

$$\bar{\boldsymbol{x}} = \frac{1}{N}\sum_{i\in s} w_i\boldsymbol{x}_i = \sum_{i\in s} d_i F_i(\boldsymbol{x}_i\boldsymbol{\ell})\boldsymbol{x}_i \quad \Rightarrow \quad g_s(\boldsymbol{\ell}) = \sum_{i\in s} d_i(F_i(\boldsymbol{x}_i\boldsymbol{\ell})-1)\boldsymbol{x}_i = \bar{\boldsymbol{x}}-\hat{\bar{\boldsymbol{x}}}.$$

- ✓ Therefore, given a sample s and chosen a distance function $\Phi_i(\cdot, d)$

How to get a solution

- ✓ The minimization problem is solved by $\phi_i(w_i, d_i) - \boldsymbol{x}_i \boldsymbol{\ell} = 0$, where $\phi_i(w, d) = \partial\Phi(w, d)/\partial w$ and $\boldsymbol{\ell}$ is a vector of Lagrangian multipliers.
- ✓ If a solution exists, basic requirements guarantee it is unique and can be written as $w_i = d_i F_i(\boldsymbol{x}_i \boldsymbol{\ell})$, where $d_i F_i(\cdot)$ is the reciprocal mapping of $\phi(\cdot, d_i)$.
- ✓ The vector $\boldsymbol{\ell}$ can be determined through the calibration constraints:

$$\bar{\boldsymbol{x}} = \frac{1}{N}\sum_{i\in s} w_i \boldsymbol{x}_i = \sum_{i\in s} d_i F_i(\boldsymbol{x}_i \boldsymbol{\ell})\boldsymbol{x}_i \quad \Rightarrow \quad g_s(\boldsymbol{\ell}) = \sum_{i\in s} d_i (F_i(\boldsymbol{x}_i \boldsymbol{\ell}) - 1)\boldsymbol{x}_i = \bar{\boldsymbol{x}} - \hat{\bar{\boldsymbol{x}}}.$$

- ✓ Therefore, given a sample s and chosen a distance function $\Phi_i(\cdot, d)$
  1. determine the uniquely corresponding function $F_i(\cdot)$;
  2. solve $g_s(\boldsymbol{\ell}) = \bar{\boldsymbol{x}} - \hat{\bar{\boldsymbol{x}}}$ for $\boldsymbol{\ell}$;
  3. obtain the calibration estimator as CAL $= N^{-1}\sum_{i\in s} w_i y_i = N^{-1}\sum_{i\in s} d_i F_i(\boldsymbol{x}_i \boldsymbol{\ell}) y_i$

An example - not chosen by chance ...

An example - not chosen by chance ...

1. Assume you choose a quadratic distance function

$$\Phi_i(w_i, d_i) = \frac{(w_i - d_i)^2}{2 d_i q_i},$$

   where the q_i's are known constants unrelated to d_i.

An example - not chosen by chance ...

1. Assume you choose a quadratic distance function

$$\Phi_i(w_i, d_i) = \frac{(w_i - d_i)^2}{2d_i q_i},$$

where the q_i's are known constants unrelated to d_i.

The partial derivative is given by $\phi_i(w_i, d_i) = (w_i/d_i - 1)/q_i$ and its reciprocal mapping by $d_i F_i(u) = d_i(1 + q_i u)$;

An example - not chosen by chance ...

1. Assume you choose a quadratic distance function

$$\Phi_i(w_i, d_i) = \frac{(w_i - d_i)^2}{2d_i q_i},$$

   where the q_i's are known constants unrelated to d_i.

   The partial derivative is given by $\phi_i(w_i, d_i) = (w_i/d_i - 1)/q_i$ and its reciprocal mapping by $d_i F_i(u) = d_i(1 + q_i u)$;

   therefore $w_i = d_i(1 + q_i \boldsymbol{x}_i \boldsymbol{\ell})$.

An example - not chosen by chance ...

1. Assume you choose a quadratic distance function

$$\Phi_i(w_i, d_i) = \frac{(w_i - d_i)^2}{2 d_i q_i},$$

   where the q_i's are known constants unrelated to d_i.

   The partial derivative is given by $\phi_i(w_i, d_i) = (w_i/d_i - 1)/q_i$ and its reciprocal mapping by $d_i F_i(u) = d_i(1 + q_i u)$;

   therefore $w_i = d_i(1 + q_i \boldsymbol{x}_i \boldsymbol{\ell})$.

2. The Lagrangian multipliers are determined by means of $\sum_{i \in s} d_i q_i \boldsymbol{x}_i' \boldsymbol{x}_i \boldsymbol{\ell} = (\bar{\boldsymbol{x}} - \hat{\bar{\boldsymbol{x}}})'$ as $\boldsymbol{\ell} = \left(\sum_{i \in s} d_i q_i \boldsymbol{x}_i' \boldsymbol{x}_i \right)^{-1} (\bar{\boldsymbol{x}} - \hat{\bar{\boldsymbol{x}}})'$.

An example - not chosen by chance ... (follows)

3. And the calibration estimator follows to be

An example - not chosen by chance ... (follows)

3. And the calibration estimator follows to be

$$\text{CAL} \quad = \quad N^{-1} \sum_{i \in s} w_i y_i =$$

An example - not chosen by chance ... (follows)

3. And the calibration estimator follows to be

$$
\begin{aligned}
\text{CAL} &= N^{-1}\sum_{i\in s} w_i y_i = \\
&= N^{-1}\sum_{i\in s} d_i\big(1 + q_i \boldsymbol{x}_i \big(\sum_{i\in s} d_i q_i \boldsymbol{x}_i' \boldsymbol{x}_i\big)^{-1} (\bar{\boldsymbol{x}} - \hat{\bar{\boldsymbol{x}}})'\big) y_i =
\end{aligned}
$$

An example - not chosen by chance ... (follows)

3. And the calibration estimator follows to be

$$\begin{aligned}
\text{CAL} &= N^{-1}\textstyle\sum_{i\in s} w_i y_i = \\
&= N^{-1}\textstyle\sum_{i\in s} d_i\big(1 + q_i \boldsymbol{x}_i \big(\sum_{i\in s} d_i q_i \boldsymbol{x}_i' \boldsymbol{x}_i\big)^{-1} (\bar{\boldsymbol{x}} - \hat{\bar{\boldsymbol{x}}})'\big) y_i = \\
&= \hat{\bar{Y}} + (\bar{\boldsymbol{x}} - \hat{\bar{\boldsymbol{x}}})\big(\textstyle\sum_{i\in s} d_i q_i \boldsymbol{x}_i' \boldsymbol{x}_i\big)^{-1} \textstyle\sum_{i\in s} d_i q_i \boldsymbol{x}_i' y_i =
\end{aligned}$$

An example - not chosen by chance ... (follows)

3. And the calibration estimator follows to be

$$
\begin{aligned}
\text{CAL} &= N^{-1}\textstyle\sum_{i\in s} w_i y_i = \\
&= N^{-1}\textstyle\sum_{i\in s} d_i(1 + q_i \boldsymbol{x}_i (\sum_{i\in s} d_i q_i \boldsymbol{x}_i' \boldsymbol{x}_i)^{-1} (\bar{\boldsymbol{x}} - \hat{\bar{\boldsymbol{x}}})') y_i = \\
&= \hat{\bar{Y}} + (\bar{\boldsymbol{x}} - \hat{\bar{\boldsymbol{x}}})(\textstyle\sum_{i\in s} d_i q_i \boldsymbol{x}_i' \boldsymbol{x}_i)^{-1} \sum_{i\in s} d_i q_i \boldsymbol{x}_i' y_i = \\
&= \hat{\bar{Y}} + (\bar{\boldsymbol{x}} - \hat{\bar{\boldsymbol{x}}})\hat{\boldsymbol{\beta}}_c
\end{aligned}
$$

An example - not chosen by chance ... (follows)

3. And the calibration estimator follows to be

$$
\begin{aligned}
\text{CAL} &= N^{-1}\textstyle\sum_{i\in s} w_i y_i = \\
&= N^{-1}\textstyle\sum_{i\in s} d_i\left(1 + q_i \boldsymbol{x}_i \left(\sum_{i\in s} d_i q_i \boldsymbol{x}_i' \boldsymbol{x}_i\right)^{-1} (\bar{\boldsymbol{x}} - \hat{\bar{\boldsymbol{x}}})'\right) y_i = \\
&= \hat{\bar{Y}} + (\bar{\boldsymbol{x}} - \hat{\bar{\boldsymbol{x}}})\left(\textstyle\sum_{i\in s} d_i q_i \boldsymbol{x}_i' \boldsymbol{x}_i\right)^{-1} \textstyle\sum_{i\in s} d_i q_i \boldsymbol{x}_i' y_i = \\
&= \hat{\bar{Y}} + (\bar{\boldsymbol{x}} - \hat{\bar{\boldsymbol{x}}})\hat{\boldsymbol{\beta}}_c
\end{aligned}
$$

Which is equivalent to the Generalized REGression estimator, GREG, if we choose the q_i's to represent the variance structure of the working model.

Calibration Estimators - properties

Calibration Estimators - properties

✓ In an asymptotic framework as in Isaki and Fuller (1982) and under regularity conditions (Deville & Särndal, 1992, p.379), CAL is shown to be design $\sqrt{n}$-consistent for $\bar{Y}$, in the sense that CAL$-\bar{Y} = O_p(n^{-1/2})$.

Calibration Estimators - properties

✓ In an asymptotic framework as in Isaki and Fuller (1982) and under regularity conditions (Deville & Särndal, 1992, p.379), CAL is shown to be design $\sqrt{n}$-consistent for $\bar{Y}$, in the sense that CAL$-\bar{Y} = O_p(n^{-1/2})$.

✓ Moreover, all the calibration estimators are asymptotically equivalent to the one that employs a quadratic distance measure. In fact, CAL – GREG$= O_p(n^{-1})$.

Calibration Estimators - properties

✓ In an asymptotic framework as in Isaki and Fuller (1982) and under regularity conditions (Deville & Särndal, 1992, p.379), CAL is shown to be design $\sqrt{n}$-consistent for $\bar{Y}$, in the sense that CAL$-\bar{Y} = O_p(n^{-1/2})$.

✓ Moreover, all the calibration estimators are asymptotically equivalent to the one that employs a quadratic distance measure. In fact, CAL – GREG$= O_p(n^{-1})$.

✓ Therefore CAL and GREG share the same limiting distribution and then

$$AV(\mathsf{CAL}) = \sum_{i \in \mathcal{U}} \sum_{j \in \mathcal{U}} (\pi_{ij} - \pi_i \pi_j) \frac{E_i}{\pi_i} \frac{E_j}{\pi_j},$$

with $E_i = y_i - \boldsymbol{x}_i \tilde{\boldsymbol{\beta}}_c$ and $\tilde{\boldsymbol{\beta}}_c = \left(\sum_{i \in \mathcal{U}} q_i \boldsymbol{x}_i' \boldsymbol{x}_i\right)^{-1} \sum_{i \in \mathcal{U}} q_i \boldsymbol{x}_i' y_i$.

Model Calibration - the idea (Wu & Sitter, 2001)

Model Calibration - the idea (Wu & Sitter, 2001)

1. Assume a (linear), nonlinear, generalized linear model well describes the relationship between $\boldsymbol{x}$ and y;

Model Calibration - the idea (Wu & Sitter, 2001)

1. Assume a (linear), nonlinear, generalized linear model well describes the relationship between $\boldsymbol{x}$ and y;
2. obtain obtain fitted values of y for all units in the population according to the model and accounting for the sampling plan;

Model Calibration - the idea (Wu & Sitter, 2001)

1. Assume a (linear), nonlinear, generalized linear model well describes the relationship between $\boldsymbol{x}$ and y;
2. obtain obtain fitted values of y for all units in the population according to the model and accounting for the sampling plan;
3. build a calibration estimator by calibrating on population mean of such fitted values (and not with respect to $\bar{\boldsymbol{x}}$).

Model Calibration - in symbols

Model Calibration - in symbols

1. Working model

$$\begin{cases} E_\xi(y_i) = \mu(\boldsymbol{x}_i, \boldsymbol{\theta}), \quad i \in \mathcal{U},\ \mu(\cdot) \text{ is known} \\ V_\xi(y_i) = \sigma^2 v(\boldsymbol{x}_i), \quad i \in \mathcal{U},\ v(\cdot) \text{ is known} \end{cases}$$

✓ $\mu(\boldsymbol{x}_i, \boldsymbol{\theta}) \longrightarrow$ nonlinear model

✓ $g(\mu_i) = \boldsymbol{x}_i\boldsymbol{\theta} \longrightarrow$ (linear or) generalized linear model

Model Calibration - in symbols

1. Working model

$$\begin{cases} E_\xi(y_i) = \mu(\boldsymbol{x}_i, \boldsymbol{\theta}), & i \in \mathcal{U},\ \mu(\cdot) \text{ is known} \\ V_\xi(y_i) = \sigma^2 v(\boldsymbol{x}_i), & i \in \mathcal{U},\ v(\cdot) \text{ is known} \end{cases}$$

✓ $\mu(\boldsymbol{x}_i, \boldsymbol{\theta}) \longrightarrow$ nonlinear model

✓ $g(\mu_i) = \boldsymbol{x}_i \boldsymbol{\theta} \longrightarrow$ (linear or) generalized linear model

2. Obtain fitted values $\hat{\mu}_i = \mu(\boldsymbol{x}_i, \hat{\boldsymbol{\theta}})$, $\hat{\boldsymbol{\theta}} = f\{y_i, \boldsymbol{x}_i, \pi_i\}$ by e.g. estimating equations modified to account for the design.

Model Calibration - the new calibration step

Model Calibration - the new calibration step

3. Build the model calibration estimator

$$\mathrm{MC} = N^{-1} \sum_{i \in s} w_i y_i$$

Model Calibration - the new calibration step

3. Build the model calibration estimator

$$\text{MC} = N^{-1} \sum_{i \in s} w_i y_i$$

with weights w_i found again to be as close as possible to $d_i = \pi_i^{-1}$, but with new benchmark constraints:

Model Calibration - the new calibration step

3. Build the model calibration estimator

$$\mathrm{MC} = N^{-1} \sum_{i \in s} w_i y_i$$

with weights w_i found again to be as close as possible to $d_i = \pi_i^{-1}$, but with new benchmark constraints:

$$\min_{w_i} \sum_{i \in s} \frac{(w_i - d_i)^2}{2 d_i q_i} \qquad \text{s.t.} \qquad \boxed{\begin{array}{c} \sum_{i \in s} w_i = N \\ N^{-1} \sum_{i \in s} w_i \hat{\mu}_i = N^{-1} \sum_{i \in \mathcal{U}} \hat{\mu}_i \end{array}}$$

Model Calibration - the new calibration step

3. Build the model calibration estimator

$$\mathsf{MC} = N^{-1} \sum_{i \in s} w_i y_i$$

with weights w_i found again to be as close as possible to $d_i = \pi_i^{-1}$, but with new benchmark constraints:

$$\min_{w_i} \sum_{i \in s} \frac{(w_i - d_i)^2}{2 d_i q_i} \qquad \text{s.t.} \qquad \boxed{\begin{array}{c} \sum_{i \in s} w_i = N \\ N^{-1} \sum_{i \in s} w_i \hat{\mu}_i = N^{-1} \sum_{i \in \mathcal{U}} \hat{\mu}_i \end{array}}$$

The estimator is obtained as in the previous example as

$$\mathsf{MC} = \hat{\hat{Y}} + \frac{1}{N} \left(\sum_{i \in \mathcal{U}} \hat{\mu}_i - \sum_{i \in s} d_i \hat{\mu}_i \right) \hat{\beta}_{mc}$$

where $\hat{\beta}_{mc}$, is the GREG coefficient of a regression estimator of y_i on $\hat{\mu}_i$.

Nonparametric model calibration - the idea

Nonparametric model calibration - the idea

✓ Allow more flexible modelling than parametric model calibration by assuming more general models and employ nonparametric techniques to obtain fitted values to calibrate on.

Nonparametric model calibration - the idea

✓ Allow more flexible modelling than parametric model calibration by assuming more general models and employ nonparametric techniques to obtain fitted values to calibrate on.

✓ Any nonparametric technique can in principle be employed. We will focus on Neural Networks because this method

1. allows deriving theoretical properties of the resulting estimator,
2. is a popular nonparametric technique for which software is easily available,
3. allows straightforward insertion of multivariate auxiliary information;

Nonparametric model calibration - the idea

✓ Allow more flexible modelling than parametric model calibration by assuming more general models and employ nonparametric techniques to obtain fitted values to calibrate on.

✓ Any nonparametric technique can in principle be employed. We will focus on Neural Networks because this method

1. allows deriving theoretical properties of the resulting estimator,
2. is a popular nonparametric technique for which software is easily available,
3. allows straightforward insertion of multivariate auxiliary information;

✓ and on Local Polynomials, since this has been the first nonparametric method considered for model-assisted survey sampling (Breidt & Opsomer, 2000).

...Quickly, what is a Feedforward Neural Network with skip layer connections??

...Quickly, what is a Feedforward Neural Network with skip layer connections??

It is just a linear combination of sigmoidal functions of linear combinations of the auxiliary variables PLUS a linear model

...Quickly, what is a Feedforward Neural Network with skip layer connections??

It is just a linear combination of sigmoidal functions of linear combinations of the auxiliary variables PLUS a linear model

$$f(\boldsymbol{x}_i) = \qquad \sum_{q=1}^{Q} \gamma_{qm} x_{qi} + \gamma_{0m}$$

...Quickly, what is a Feedforward Neural Network with skip layer connections??

It is just a linear combination of sigmoidal functions of linear combinations of the auxiliary variables PLUS a linear model

$$f(\boldsymbol{x}_i) = \qquad \phi\left(\sum_{q=1}^{Q} \gamma_{qm} x_{qi} + \gamma_{0m}\right)$$

...Quickly, what is a Feedforward Neural Network with skip layer connections??

It is just a linear combination of sigmoidal functions of linear combinations of the auxiliary variables PLUS a linear model

$$f(\boldsymbol{x}_i) = \sum_{m=1}^{M} a_m \, \phi\left(\sum_{q=1}^{Q} \gamma_{qm} x_{qi} + \gamma_{0m}\right) + a_0$$

...Quickly, what is a Feedforward Neural Network with skip layer connections??

It is just a linear combination of sigmoidal functions of linear combinations of the auxiliary variables PLUS a linear model

$$f(\boldsymbol{x}_i) = \sum_{q=1}^{Q} \beta_q x_{qi} + \sum_{m=1}^{M} a_m \, \phi \left(\sum_{q=1}^{Q} \gamma_{qm} x_{qi} + \gamma_{0m} \right) + a_0$$

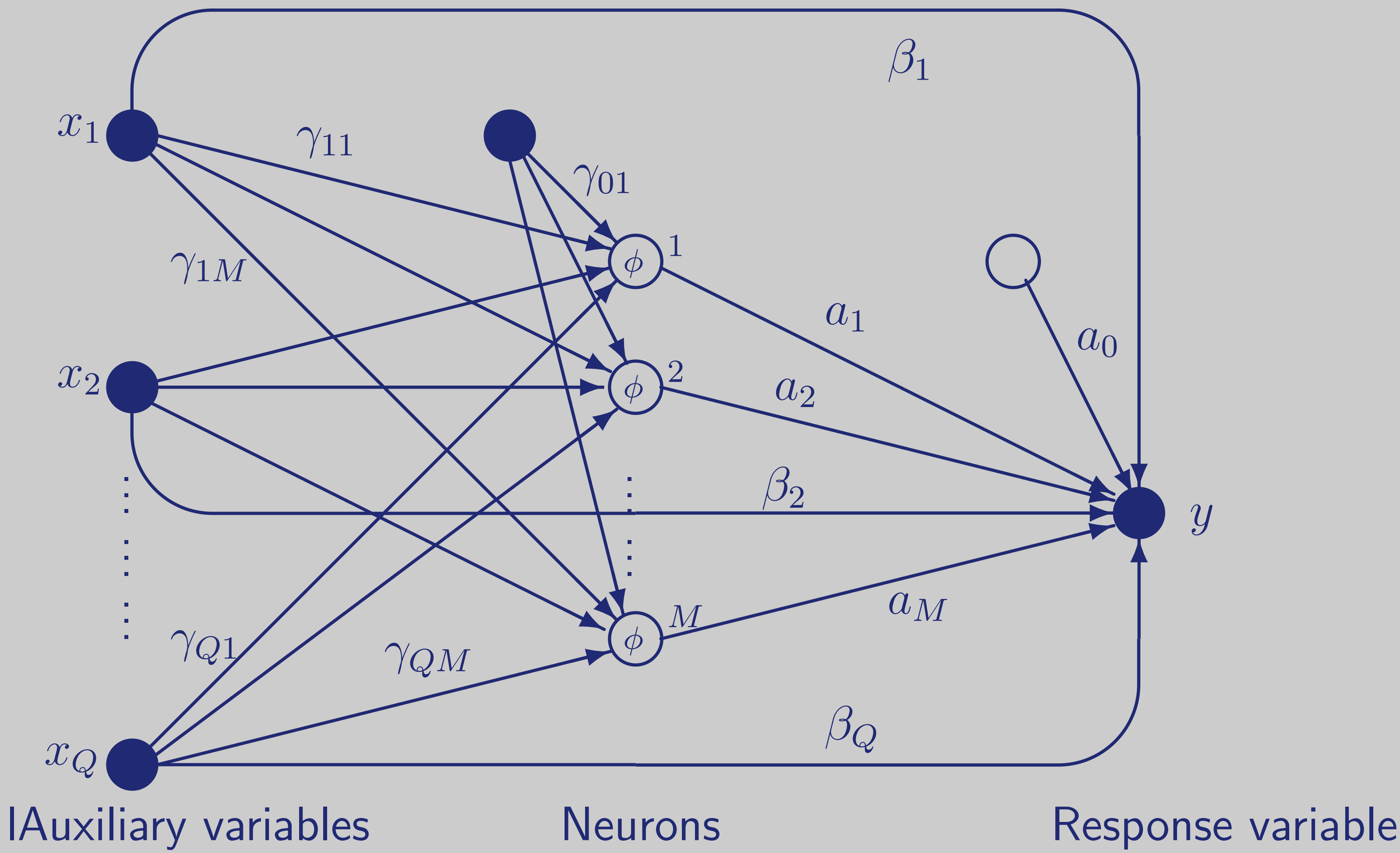
...Quickly, what is a Feedforward Neural Network with skip layer connections??

It is just a linear combination of sigmoidal functions of linear combinations of the auxiliary variables PLUS a linear model

$$f(\boldsymbol{x}_i) = \sum_{q=1}^{Q} \beta_q x_{qi} + \sum_{m=1}^{M} a_m \, \phi\left(\sum_{q=1}^{Q} \gamma_{qm} x_{qi} + \gamma_{0m}\right) + a_0$$

We will denote by $\boldsymbol{\theta} = \{\beta_1, \ldots, \beta_Q, a_0, a_1, \ldots, a_M, \gamma_{01}, \ldots, \gamma_{0M}, \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_M\}$ the set of all parameters of the net.

A diagram that usually represents a NNet



Nonparametric model calibration - the steps

Nonparametric model calibration - the steps

1. Working model - Neural Network model

$$\begin{cases} E_\xi(y_i) = f(\boldsymbol{x}_i), & i \in \mathcal{U} \\ V_\xi(y_i) = v(\boldsymbol{x}_i), & i \in \mathcal{U} \end{cases}$$

Nonparametric model calibration - the steps

1. Working model - Neural Network model

$$\begin{cases} E_\xi(y_i) = f(\boldsymbol{x}_i), & i \in \mathcal{U} \\ V_\xi(y_i) = v(\boldsymbol{x}_i), & i \in \mathcal{U} \end{cases}$$

2. Obtain fitted values. First get the population parameter as the minimizer of

$$\tilde{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N} \frac{1}{v_i} \left(y_i - f(\boldsymbol{x}_i, \boldsymbol{\theta})\right)^2 + \lambda \sum_{l=1}^{r} \theta_l^2 \right\},$$

   where λ is a weight decay parameter,

Nonparametric model calibration - the steps

1. Working model - Neural Network model

$$\begin{cases} E_\xi(y_i) = f(\boldsymbol{x}_i), & i \in \mathcal{U} \\ V_\xi(y_i) = v(\boldsymbol{x}_i), & i \in \mathcal{U} \end{cases}$$

2. Obtain fitted values. First get the population parameter as the minimizer of

$$\tilde{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N} \frac{1}{v_i} \left(y_i - f(\boldsymbol{x}_i, \boldsymbol{\theta})\right)^2 + \lambda \sum_{l=1}^{r} \theta_l^2 \right\},$$

where λ is a weight decay parameter, by means of estimating equations

$$\sum_{i=1}^{N} \left\{ \left(y_i - f(\boldsymbol{x}_i, \boldsymbol{\theta})\right) \frac{\partial f(\boldsymbol{x}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{1}{v_i} - \frac{\lambda}{N} \boldsymbol{\theta} \right\} = \mathbf{0}.$$

Nonparametric model calibration - the steps

1. Working model - Neural Network model

$$\begin{cases} E_\xi(y_i) = f(\boldsymbol{x}_i), & i \in \mathcal{U} \\ V_\xi(y_i) = v(\boldsymbol{x}_i), & i \in \mathcal{U} \end{cases}$$

2. Obtain fitted values. First get the population parameter as the minimizer of

$$\tilde{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N} \frac{1}{v_i} (y_i - f(\boldsymbol{x}_i, \boldsymbol{\theta}))^2 + \lambda \sum_{l=1}^{r} \theta_l^2 \right\},$$

where λ is a weight decay parameter, by means of estimating equations

$$\sum_{i=1}^{N} \left\{ (y_i - f(\boldsymbol{x}_i, \boldsymbol{\theta})) \frac{\partial f(\boldsymbol{x}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{1}{v_i} - \frac{\lambda}{N} \boldsymbol{\theta} \right\} = \mathbf{0}.$$

Then get $\hat{\boldsymbol{\theta}}$ as the solution of the design based sample version of it

$$\sum_{i=1}^{n} \frac{1}{\pi_i} \left\{ (y_i - f(\boldsymbol{x}_i, \boldsymbol{\theta})) \frac{\partial f(\boldsymbol{x}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{1}{v_i} - \frac{\lambda}{N} \boldsymbol{\theta} \right\} = \mathbf{0}.$$

Nonparametric model calibration - the steps

1. Working model - Neural Network model

$$\begin{cases} E_\xi(y_i) = f(\boldsymbol{x}_i), & i \in \mathcal{U} \\ V_\xi(y_i) = v(\boldsymbol{x}_i), & i \in \mathcal{U} \end{cases}$$

2. Obtain fitted values. First get the population parameter as the minimizer of

$$\tilde{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N} \frac{1}{v_i} \left(y_i - f(\boldsymbol{x}_i, \boldsymbol{\theta})\right)^2 + \lambda \sum_{l=1}^{r} \theta_l^2 \right\},$$

where λ is a weight decay parameter, by means of estimating equations

$$\sum_{i=1}^{N} \left\{ (y_i - f(\boldsymbol{x}_i, \boldsymbol{\theta})) \frac{\partial f(\boldsymbol{x}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{1}{v_i} - \frac{\lambda}{N} \boldsymbol{\theta} \right\} = \mathbf{0}.$$

Then get $\hat{\boldsymbol{\theta}}$ as the solution of the design based sample version of it

$$\sum_{i=1}^{n} \frac{1}{\pi_i} \left\{ (y_i - f(\boldsymbol{x}_i, \boldsymbol{\theta})) \frac{\partial f(\boldsymbol{x}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{1}{v_i} - \frac{\lambda}{N} \boldsymbol{\theta} \right\} = \mathbf{0}.$$

Then we can obtain fitted values $\hat{f}_i = f(\boldsymbol{x}_i, \hat{\boldsymbol{\theta}})$ such that $\hat{f}_i = f(\boldsymbol{x}_i, \tilde{\boldsymbol{\theta}}) + O_p(n^{-1/2})$.

Nonparametric model calibration - the calibration step

Nonparametric model calibration - the calibration step

3. Build the calibration estimator as for parametric model calibration, just the fitted values in the calibration constraint are different

$$\text{NNMC} = N^{-1} \sum_{i \in s} w_i y_i$$

Nonparametric model calibration - the calibration step

3. Build the calibration estimator as for parametric model calibration, just the fitted values in the calibration constraint are different

$$\mathrm{NNMC} = N^{-1} \sum_{i \in s} w_i y_i$$

$$\min_{w_i} \sum_{i \in s} \frac{(w_i - d_i)^2}{2 d_i q_i} \qquad \text{s.t.} \qquad \boxed{\begin{array}{c} \sum_{i \in s} w_i = N \\ N^{-1} \sum_{i \in s} w_i \hat{f}_i = N^{-1} \sum_{i \in \mathcal{U}} \hat{f}_i \end{array}}$$

Nonparametric model calibration - the calibration step

3. Build the calibration estimator as for parametric model calibration, just the fitted values in the calibration constraint are different

$$\text{NNMC} = N^{-1} \sum_{i \in s} w_i y_i$$

$$\min_{w_i} \sum_{i \in s} \frac{(w_i - d_i)^2}{2 d_i q_i} \qquad \text{s.t.} \qquad \boxed{\begin{array}{c} \sum_{i \in s} w_i = N \\ N^{-1} \sum_{i \in s} w_i \hat{f}_i = N^{-1} \sum_{i \in \mathcal{U}} \hat{f}_i \end{array}}$$

Then,

$$\text{NNMC} = \hat{\hat{Y}} + \frac{1}{N} \left(\sum_{i \in \mathcal{U}} \hat{f}_i - \sum_{i \in s} d_i \hat{f}_i \right) \hat{\beta}_{nn}$$

where $\hat{\beta}_{nn}$, is the GREG coefficient of a regression estimator of y_i on $\hat{f}_i$. (Cfr with a modified GAM regression estimator in Opsomer *et al.*, 2001)

Local polynomials model calibration - the steps

Local polynomials model calibration - the steps

1. Working model - A very general *univariate* model

$$\begin{cases} E_\xi(y_i) = m(x_i), & i \in \mathcal{U} \\ V_\xi(y_i) = v(x_i), & i \in \mathcal{U} \end{cases}$$

Local polynomials model calibration - the steps

1. Working model - A very general *univariate* model

$$\begin{cases} E_\xi(y_i) = m(x_i), & i \in \mathcal{U} \\ V_\xi(y_i) = v(x_i), & i \in \mathcal{U} \end{cases}$$

2. Get fitted values by means of local polynomials

$$\hat{m}_i = \boldsymbol{e}_1'(\boldsymbol{X}_{si}'\boldsymbol{W}_{si}\boldsymbol{X}_{si})^{-1}\boldsymbol{X}_{si}'\boldsymbol{W}_{si}\boldsymbol{y}_s,$$

where $\boldsymbol{e}_1 = (1, 0, \ldots, 0)'$ is a column vector of length $p+1$, p is the order of the local polynomial fit, $\boldsymbol{y}_s = (y_1, \ldots, y_n)'$, $\boldsymbol{W}_{si} = \operatorname{diag}\{d_j K_h(x_j - x_i)\}_{j \in s}$, K is a kernel function, h is the bandwidth and $\boldsymbol{X}_{si} = [1 \;\; x_j - x_i \;\; \cdots \;\; (x_j - x_i)^p]_{j \in s}$.

Local polynomials model calibration - the steps

Local polynomials model calibration - the steps

3. Build the model calibration estimator as previously

$$\text{LPMC} = N^{-1} \sum_{i \in s} w_i y_i$$

Local polynomials model calibration - the steps

3. Build the model calibration estimator as previously

$$\text{LPMC} = N^{-1} \sum_{i \in s} w_i y_i$$

$$\min_{w_i} \sum_{i \in s} \frac{(w_i - d_i)^2}{2 d_i q_i} \qquad \text{s.t.} \qquad \boxed{\begin{array}{c} \sum_{i \in s} w_i = N \\ N^{-1} \sum_{i \in s} w_i \hat{m}_i = N^{-1} \sum_{i \in \mathcal{U}} \hat{m}_i \end{array}}$$

Local polynomials model calibration - the steps

3. Build the model calibration estimator as previously

$$\text{LPMC} = N^{-1} \sum_{i \in s} w_i y_i$$

$$\min_{w_i} \sum_{i \in s} \frac{(w_i - d_i)^2}{2 d_i q_i} \qquad \text{s.t.} \qquad \boxed{\begin{array}{c} \sum_{i \in s} w_i = N \\ N^{-1} \sum_{i \in s} w_i \hat{m}_i = N^{-1} \sum_{i \in \mathcal{U}} \hat{m}_i \end{array}}$$

Then,

$$\text{LPMC} = \hat{\hat{Y}} + \frac{1}{N} \left(\sum_{i \in \mathcal{U}} \hat{m}_i - \sum_{i \in s} d_i \hat{m}_i \right) \hat{\beta}_{lp}$$

where $\hat{\beta}_{lp}$, is the GREG coefficient of a regression estimator of y_i on $\hat{m}_i$.

Which is the difference between LPMC and LPRE?

The Local polynomial regression estimator (Breidt & Opsomer, 2000) is given by

$$\mathrm{LPRE} = \hat{\hat{Y}} + \frac{1}{N}\left(\sum_{i\in\mathcal{U}} \hat{m}_i - \sum_{i\in s} d_i \hat{m}_i\right).$$

Which is the difference between LPMC and LPRE?

The Local polynomial regression estimator (Breidt & Opsomer, 2000) is given by

$$\text{LPRE} = \hat{\hat{Y}} + \frac{1}{N}\left(\sum_{i\in\mathcal{U}} \hat{m}_i - \sum_{i\in s} d_i \hat{m}_i\right).$$

The only difference lays in the supplementary regression step provided by $\hat{\beta}_{lp}$. LPMC can be also thought of a GREG estimator with a working model of the type

$$E_\xi(y_i) = a + b\, m(x_i).$$

Which is the difference between LPMC and LPRE?

The Local polynomial regression estimator (Breidt & Opsomer, 2000) is given by

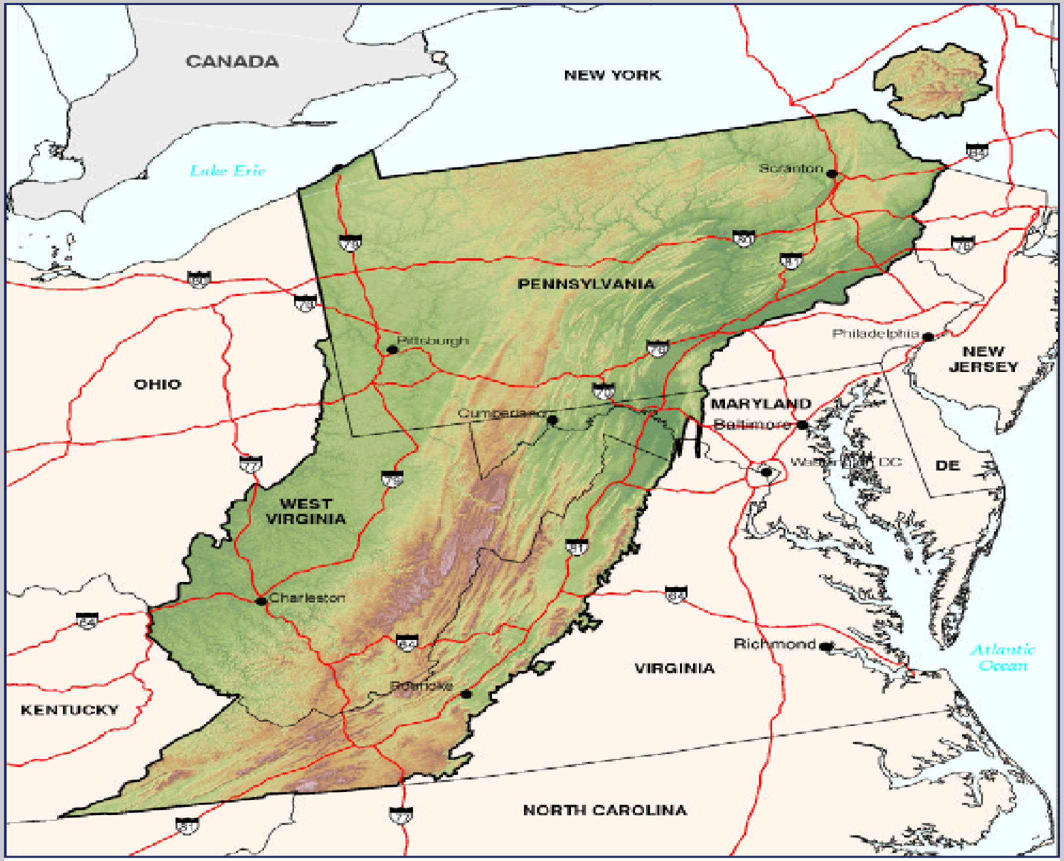
$$\text{LPRE} = \hat{\hat{Y}} + \frac{1}{N}\left(\sum_{i\in\mathcal{U}} \hat{m}_i - \sum_{i\in s} d_i \hat{m}_i\right).$$

The only difference lays in the supplementary regression step provided by $\hat{\beta}_{lp}$. LPMC can be also thought of a GREG estimator with a working model of the type

$$E_\xi(y_i) = a + b\,m(x_i).$$

If the nonparametric technique provides biased estimates of the mean function or the working model is not valid, then this step will lead to more efficient estimates by correcting this bias.

Simulation Study - MAHA region

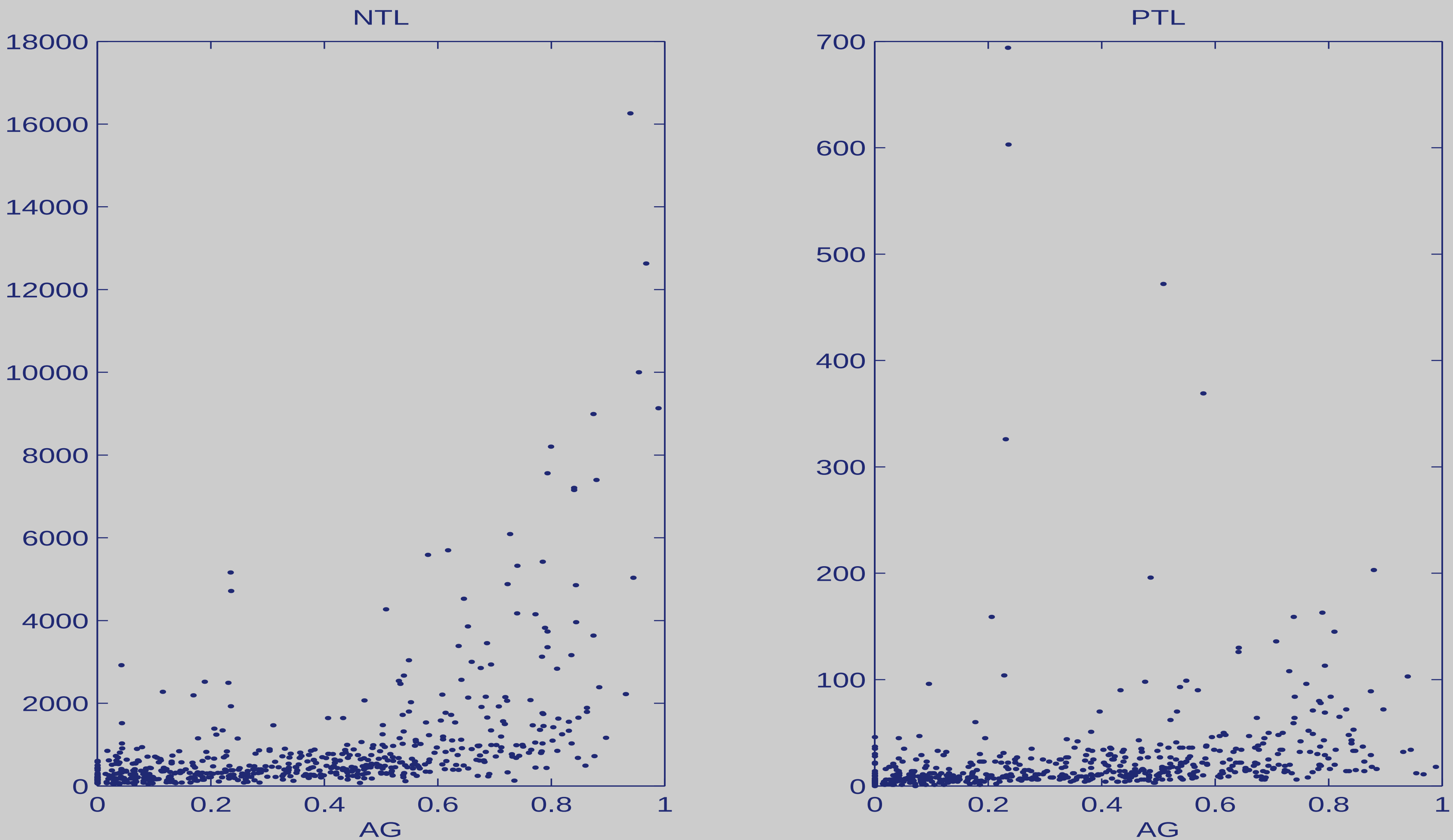


The Mid-Atlantic Highlands study region includes the area from the Blue Ridge Mountains in the east to the Ohio River in the west and from the Catskill Mountains in the north to Virginia in the south.

Simulation Study - data

- Mid-Atlantic Highlands Area (MAHA) EPA Region 3: PA, VA, WV, DE, MD.
- Data collected by EPAs EMAP.
- Streams sampled from 1993 through 1996.
- Total Nitrogen (NTL) and Total Phosphorus (PTL) sampled at 574 sites.
- Proportion of watershed devoted to Agriculture (AG) available from remote sensing.

Scatterplots of the two survey variables wrt AG



Simulation Study - setup

- $N = 574$, auxiliary variable: AG, survey variables: NTL and PTL.
- A thousand simple random samples without replacement of dimension $n = 100$ have been drawn from $\mathcal{U}$.

Simulation Study - setup

- $N = 574$, auxiliary variable: AG, survey variables: NTL and PTL.
- A thousand simple random samples without replacement of dimension $n = 100$ have been drawn from $\mathcal{U}$.
- For each sample we computed the following estimators

HT the sample mean;

CAL the regression estimator;

NNMC with four combinations of complexity parameters M and λ

LPMC, LPRE with a local constant and a local linear fit with bandwidths $h = 0.10$ and $h = 0.25$

Simulation Study - setup

- $N = 574$, auxiliary variable: AG, survey variables: NTL and PTL.
- A thousand simple random samples without replacement of dimension $n = 100$ have been drawn from $\mathcal{U}$.
- For each sample we computed the following estimators

  HT the sample mean;

  CAL the regression estimator;

  NNMC with four combinations of complexity parameters M and λ

  LPMC, LPRE with a local constant and a local linear fit with bandwidths $h = 0.10$ and $h = 0.25$
- Performance evaluated by means of Smse$(\cdot)$ $=\widehat{\text{Mse}}(\cdot)/\widehat{\text{Mse}}(\text{CAL})$

Simulation Study - results

		NTL	PTL
HT		1.344	1.037
CAL		1.000	1.000
NNMC	M=3 λ=5e-4	0.810	1.003
	M=4 λ=5e-4	0.812	1.010
	M=6 λ=1e-3	0.809	1.010
	M=8 λ=1e-3	0.808	1.003
LPMC	p=0 h=0.1	0.807	1.051
	p=0 h=0.25	0.809	1.002
	p=1 h=0.1	20.241	2.443
	p=1 h=0.25	0.814	1.008
LPRE	p=0 h=0.1	0.812	1.132
	p=0 h=0.25	0.845	1.002
	p=1 h=0.1	21.531	2.465
	p=1 h=0.25	0.815	1.008

Conclusions

- The use of Nonparametric methods to obtain fitted values for non-sampled units provides more efficient estimates by allowing for more complex modelling than with parametric calibration.
- Neural Networks have shown efficient and robust behaviors in both univariate and (multivariate - not shown here) settings. Such behavior seems to be less influenced by the choice of complexity parameters w.r.t. other nonparametric methods.
- With respect to nonparametric regression estimation, the supplementary calibration step performed by model calibration provides more efficient estimators when the fitted values underfit the data (low degree polynomial and/or large bandwidth)

Ongoing research and perspectives

- Performance of NNMC is being explored with multiple auxiliary information and compared to other nonparametric techniques (GAM, MARS, DART)
- Model calibration weights depend on the survey variable, therefore a single set of weights is not available for more than one. Also studying the effects of *multiple simultaneous calibration* (idea in Opsomer *et al.*, 2001).
- Variance estimation is an issue when nonparametric methods are employed: usually largely underestimated.

Essential bibliography

Breidt F.J. and Opsomer J.D. (2000), "Local polynomial regression estimators in survey sampling", *The Annals of Statistics*, **28**, 1026–1053.

Deville J.C. and Särndal C.E. (1992), "Calibration estimators in survey sampling", *JASA*, **87**, 376–382.

Opsomer J.D., Moisen G.G. and Kim J.Y. (2001), "Model-assisted estimation of forest resources with generalized additive models", *ASA Proceedings of the Section on Survey Research Methods*.

Wu C. and Sitter R.R. (2001), "A model-calibration to using complete auxiliary information from survey data", *JASA*, **96**, 185–193.

The work reported here was in part developed under the STAR Research Assistance Agreement CR-829095 awarded by the U.S. Environmental Protection Agency (EPA) to Colorado State University. This presentation has not been formally reviewed by EPA. The views expressed here are solely those of the presenter and STARMAP. EPA does not endorse any products or commercial services mentioned in this presentation.