Achieving Optimal Covariate Balance Under General Treatment Regimes

Marc Ratkovic

Princeton University

May 24, 2012

Motivation

For many questions of interest in the social sciences, experiments are not possible
$\Rightarrow$ Possible bias in effect estimates

Regression adjustment or inverse weighting can be used to adjust for selection bias
$\Rightarrow$ Model dependence

Matching reduces bias and model dependence by identifying a set of untreated observations that are similar to the treated observations

Problems with Existing Matching Methods

Existing matching methods, such as propensity matching, Genetic matching, and Coarsened Exact Matching,

- rely on many user inputs
- are sensitive to these choices
- have no formal statistical properties
- can only handle a binary treatment

Benefits of the Proposed Method

The proposed method

- ~~rely on many user inputs~~
  is fully automated
- are sensitive to these choices
- have no formal statistical properties
- can only handle a binary treatment

Benefits of the Proposed Method

The proposed method

- ~~rely on many user inputs~~
  is fully automated
- ~~are sensitive to these choices~~
  makes no functional form assumptions
- have no formal statistical properties

- can only handle a binary treatment

Benefits of the Proposed Method

The proposed method

- ~~rely on many user inputs~~
  is fully automated
- ~~are sensitive to these choices~~
  makes no functional form assumptions
- ~~have no formal statistical properties~~
  identifies the largest balanced subset
- can only handle a binary treatment

Benefits of the Proposed Method

The proposed method

- ~~rely on many user inputs~~
  is fully automated
- ~~are sensitive to these choices~~
  makes no functional form assumptions
- ~~have no formal statistical properties~~
  identifies the largest balanced subset
- ~~can only handle a binary treatment~~
  can also accommodate continuous treatments

The Setup

The Setup

- Treatment: T_i
  - Binary treatment: $T_i \in \{0, 1\}$
  - Continuous treatment: $T_i \in (a, b)$
- Potential outcome: $Y_i(t)$
- Pre-treatment covariates: X_i
- IID observations $(Y_i(T_i), T_i, X_i)$ observed

Assumptions

- No interference among units
- Treatment occurs with uncertainty
- No omitted variables

Goal of Matching:

Identify a subset of the data such that the covariates are balanced

- $T_i \perp\!\!\!\perp X_i$

Assumptions and Estimands

Goal of Matching:

Identify a subset of the data such that the covariates are balanced

- $T_i \perp\!\!\!\perp X_i$

Common estimands identified on a balanced subset of the data:

- Average Treatment Effect:
  $E(Y_i(1) - Y_i(0))$
- Average Treatment Effect on the Treated:
  $E(Y_i(1) - Y_i(0) | T_i = 1)$

Basic Insight of Proposed Method

The proposed method formulates a Support Vector Machine that identifies a balanced subset of the data.

Basic Insight of Proposed Method

The proposed method formulates a Support Vector Machine that identifies a balanced subset of the data.

The logic of the proposed method proceeds in three steps:

- The optimality condition for an SVM sets an inner product between the treatment level and a covariate to zero
- Centering the treatment and covariate transforms this inner product to balance-in-mean or zero covariance.
- Balancing along a nonparametric basis extends the mean/covariance result to joint independence.

A Simple Example: The Binary Matching SVM

Assume an observed covariate vector, X_i, and target function $X_i^\top \beta$.

A Simple Example: The Binary Matching SVM

Assume an observed covariate vector, X_i, and target function $X_i^\top \beta$.

Center X_i on the treated observations as:

$$X_i^* = X_i - \frac{\sum_i X_i \cdot \mathbf{1}(T_i = 1)}{\sum_i \mathbf{1}(T_i = 1)}$$

A Simple Example: The Binary Matching SVM

Assume an observed covariate vector, X_i, and target function $X_i^\top \beta$.

Center X_i on the treated observations as:

$$X_i^* = X_i - \frac{\sum_i X_i \cdot \mathbf{1}(T_i = 1)}{\sum_i \mathbf{1}(T_i = 1)}$$

Transform T_i from $\{0, 1\}$ to $\{-1, 1\}$:

$$T_i^* = 2T_i - 1$$

A Simple Example: The Binary Matching SVM

Define the "hinge loss" $|z|_+ = \max(z, 0)$

Loss function:

$$\mathcal{L}(\beta) = \sum_i |1 - T_i^* X_i^{*\top} \beta|_+ \quad \text{s.t. } X_i^{*\top} \beta \cdot \mathbf{1}(T_i = 1) < 1$$

A Simple Example: The Binary Matching SVM

Define the "hinge loss" $|z|_+ = \max(z, 0)$

Loss function:

$$\mathcal{L}(\beta) = \sum_i |1 - T_i^* X_i^{*\top} \beta|_+ \quad \text{s.t. } X_i^{*\top} \beta \cdot \mathbf{1}(T_i = 1) < 1$$

"Hard to classify" and "Easy to classify" cases:

- $T_i^* = 1;\ X_i^{*\top} \beta = 2 : \quad |1 - 1 \cdot 2|_+ = |-1|_+ = 0$
  Easy to classify

A Simple Example: The Binary Matching SVM

Define the "hinge loss" $|z|_+ = \max(z, 0)$

Loss function:

$$\mathcal{L}(\beta) = \sum_i |1 - T_i^* X_i^{*\top}\beta|_+ \quad \text{s.t. } X_i^{*\top}\beta \cdot \mathbf{1}(T_i = 1) < 1$$

"Hard to classify" and "Easy to classify" cases:

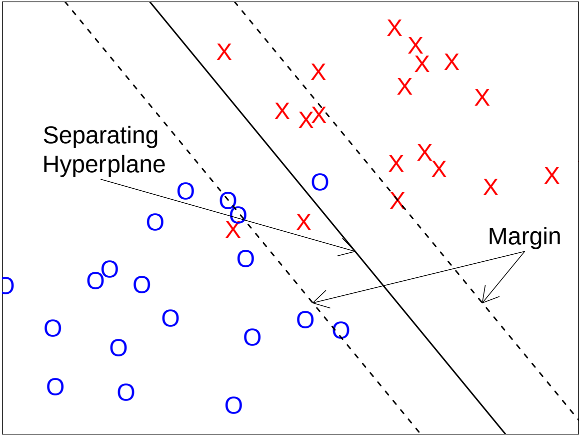
- $T_i^* = 1;\ X_i^{*\top}\beta = 2: \quad |1 - 1 \cdot 2|_+ = |-1|_+ = 0$
  Easy to classify
- $T_i^* = -1;\ X_i^{*\top}\beta = -0.5: \quad |1 - (-1) \cdot (-0.5)|_+ = |0.5|_+ = 0.5$
  Hard to classify

The constraint keeps the loss for all treated observations as non-zero to identify the ATT.

Geometric Intuition of Proposed Method

Properly classified cases outside the margin are "easy-to-classify."

Cases in the margin, or improperly classified, have a treatment assignment estimated with some uncertainty.



A Simple Example: The Binary Matching SVM

Define $\mathcal{M} = \{i : 1 - T_i^* {X_i^*}^\top \beta > 0\}$

A Simple Example: The Binary Matching SVM

Define $\mathcal{M} = \{i : 1 - T_i^* {X_i^*}^\top \beta > 0\}$
Taking and expanding the first order condition:

$$\frac{\partial}{\partial \beta} \sum_i |1 - T_i^* {X_i^*}^\top \beta|_+ = \sum_i T_i^* X_i^* \cdot \mathbf{1}(i \in \mathcal{M}) = 0$$

A Simple Example: The Binary Matching SVM

Define $\mathcal{M} = \{i : 1 - T_i^* X_i^{*\top} \beta > 0\}$
Taking and expanding the first order condition:

$$\frac{\partial}{\partial \beta} \sum_i |1 - T_i^* X_i^{*\top} \beta|_+ = \sum_i T_i^* X_i^* \cdot \mathbf{1}(i \in \mathcal{M}) = 0$$

$$\sum_i X_i^* \cdot \mathbf{1}(T_i = 0, i \in \mathcal{M}) = \sum_i X_i^* \cdot \mathbf{1}(T_i = 1)$$

A Simple Example: The Binary Matching SVM

Define $\mathcal{M} = \{i : 1 - T_i^* X_i^{*\top} \beta > 0\}$
Taking and expanding the first order condition:

$$\frac{\partial}{\partial \beta} \sum_i |1 - T_i^* X_i^{*\top} \beta|_+ = \sum_i T_i^* X_i^* \cdot \mathbf{1}(i \in \mathcal{M}) = 0$$

$$\sum_i X_i^* \cdot \mathbf{1}(T_i = 0, i \in \mathcal{M}) = \sum_i X_i^* \cdot \mathbf{1}(T_i = 1)$$

$$\sum_i X_i^* \cdot \mathbf{1}(T_i = 0, i \in \mathcal{M}) = \underbrace{\sum_i X_i^* \cdot \mathbf{1}(T_i = 1) = 0}_{\text{since } X_i^* \text{ is centered on } T_i = 1}$$

A Simple Example: The Binary Matching SVM

Define $\mathcal{M} = \{i : 1 - T_i^* {X_i^*}^\top \beta > 0\}$
Taking and expanding the first order condition:

$$\frac{\partial}{\partial \beta} \sum_i |1 - T_i^* {X_i^*}^\top \beta|_+ = \sum_i T_i^* X_i^* \cdot \mathbf{1}(i \in \mathcal{M}) = 0$$

$$\sum_i X_i^* \cdot \mathbf{1}(T_i = 0, i \in \mathcal{M}) = \sum_i X_i^* \cdot \mathbf{1}(T_i = 1)$$

$$\sum_i X_i^* \cdot \mathbf{1}(T_i = 0, i \in \mathcal{M}) = \underbrace{\sum_i X_i^* \cdot \mathbf{1}(T_i = 1) = 0}_{\text{since } X_i^* \text{ is centered on } T_i = 1}$$

Law of Large Numbers gives

$$E(X_i | T_i = 1) = E(X_i | T_i = 0, i \in \mathcal{M})$$

Balance-in-means is not balance in distribution.

The Binary Treatment SVM

Balance-in-means is <u>not</u> balance in distribution.

To achieve joint independence (**Proposition 1**):

- Change the target functional from $X_i^{*\top}\beta$ to $\eta^*(X_i)$
- Add a regularization term, to balance covariate imbalance and model complexity
- Observations in $\mathcal{M}$ are balanced

The Binary Treatment SVM

Balance-in-means is <u>not</u> balance in distribution.

To achieve joint independence (**Proposition 1**):

- Change the target functional from $X_i^{*\top}\beta$ to $\eta^*(X_i)$
- Add a regularization term, to balance covariate imbalance and model complexity
- Observations in $\mathcal{M}$ are balanced

The proof follows nearly exactly as the linear case, except in a high-dimensional space.

Extension to a Continuous Treatment

Follows nearly exactly from the binary case.

Extension to a Continuous Treatment

Follows nearly exactly from the binary case.

Using identical reasoning, I show that, for observations in $\mathcal{M}$, the treatment and a single covariate is uncorrelated.

Extension to a Continuous Treatment

Follows nearly exactly from the binary case.

Using identical reasoning, I show that, for observations in $\mathcal{M}$, the treatment and a single covariate is uncorrelated.

Extension to a nonparametric function of X_i transforms uncorrelatedness to joint independence (**Proposition 2**).

Properties and Comparisons

- Selects *largest* balanced subset

Properties and Comparisons

- Selects *largest* balanced subset
  - In a simple randomized experiment, the sample size is twice the expected misclassification loss

- Selects *largest* balanced subset
  - In a simple randomized experiment, the sample size is twice the expected misclassification loss
  - Number of balanced observations approaches twice the expected misclassification loss asymptotically

Properties and Comparisons

- Selects *largest* balanced subset
  - In a simple randomized experiment, the sample size is twice the expected misclassification loss
  - Number of balanced observations approaches twice the expected misclassification loss asymptotically
- Answers question of "how many matches"

Properties and Comparisons

- Selects *largest* balanced subset
  - In a simple randomized experiment, the sample size is twice the expected misclassification loss
  - Number of balanced observations approaches twice the expected misclassification loss asymptotically
- Answers question of "how many matches"
- Tuning parameters selected through GACV criterion

Properties and Comparisons

- Selects *largest* balanced subset
  - In a simple randomized experiment, the sample size is twice the expected misclassification loss
  - Number of balanced observations approaches twice the expected misclassification loss asymptotically
- Answers question of “how many matches”
- Tuning parameters selected through GACV criterion
- Identifies observations that appear to follow a simple randomization
  - Most useful when researcher does not know which variables to match finely, exactly, in mean, etc.

Returning the Experimental Result from Experimental Data

The 1975-1978 National Supported Work Study (Lalonde 1986)

- Treatment: job training, close management, peer support
- Recipients: welfare recipients, ex-addicts, young school dropouts, and ex-offenders
- n=445: 260 treated; 185 control
- PSID data used for matching, n=2490
- X: age, years of education, race, marriage status, high school degree, 1974 earnings, 1975 earnings, zero earnings in 1974, zero earnings in 1975

Analyses

Competitors

- Logistic propensity matching (Ho, et al. 2011)
- Genetic Matching (Sekhon 2011)
- Coarsened Exact Matching (Iacus, et al. 2011)
  - CEM estimates ATT through extrapolation
- BART+ Optimal Matching (Hill, et al. 2011; Hansen 2004)

Analyses

Competitors

- Logistic propensity matching (Ho, et al. 2011)
- Genetic Matching (Sekhon 2011)
- Coarsened Exact Matching (Iacus, et al. 2011)
  - CEM estimates ATT through extrapolation
- BART+ Optimal Matching (Hill, et al. 2011; Hansen 2004)

Outcomes

- 1978 earnings
- 1978 earnings - 1975 earnings

Analyses

Competitors

- Logistic propensity matching (Ho, et al. 2011)
- Genetic Matching (Sekhon 2011)
- Coarsened Exact Matching (Iacus, et al. 2011)
  - CEM estimates ATT through extrapolation
- BART+ Optimal Matching (Hill, et al. 2011; Hansen 2004)

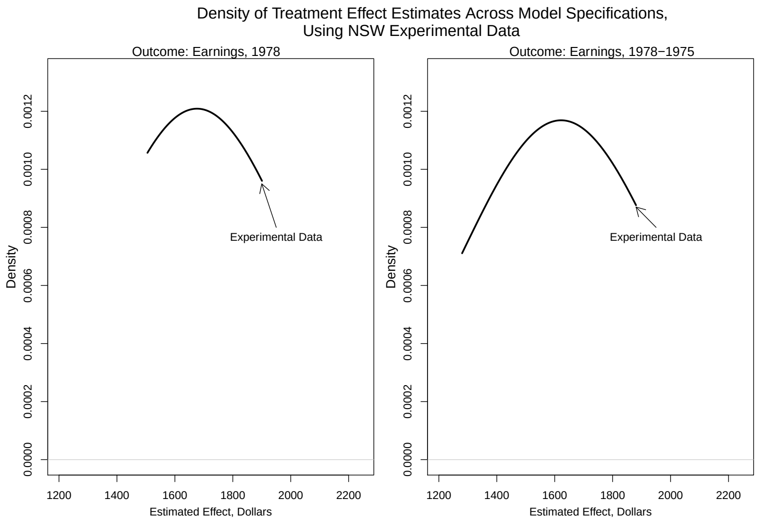
Outcomes

- 1978 earnings
- 1978 earnings - 1975 earnings

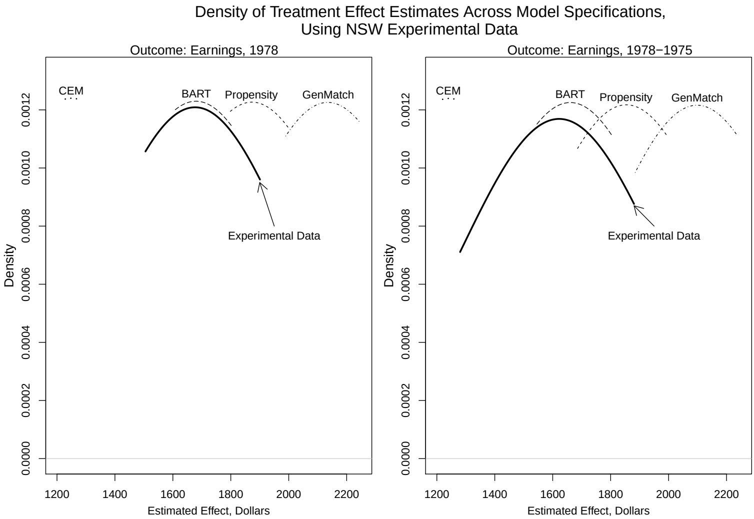
Datasets

- Experimental treated and untreated observations
- Experimental treated observations; observational untreated observations

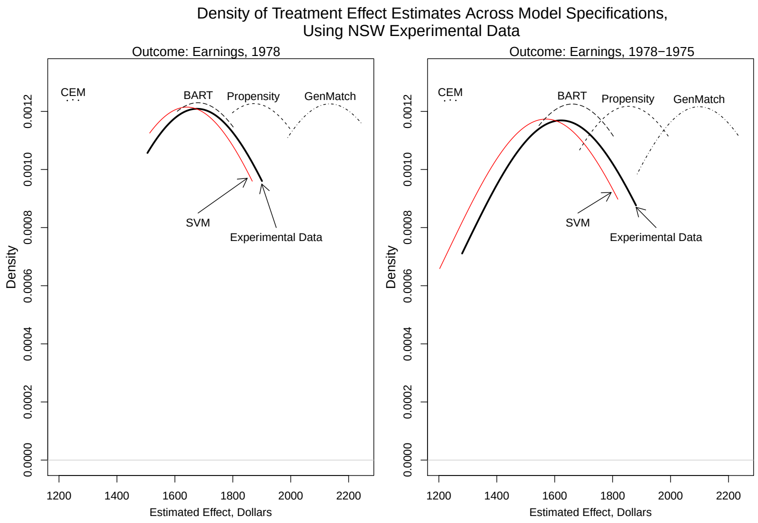
Experimental Results



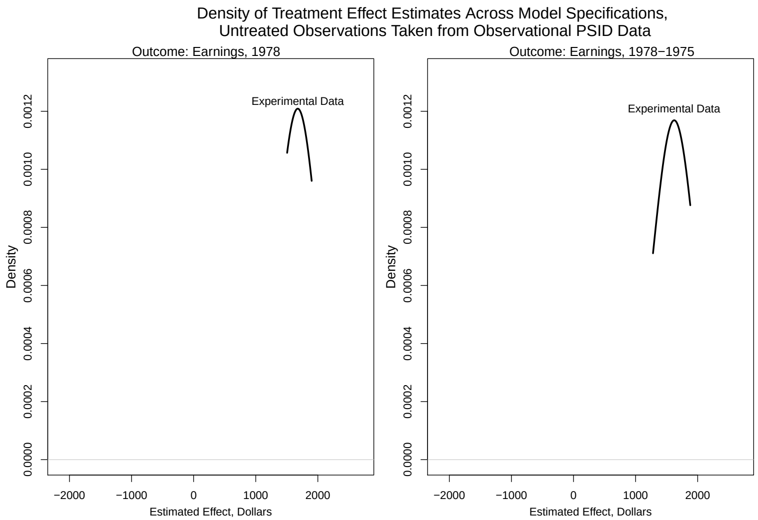
Experimental Results



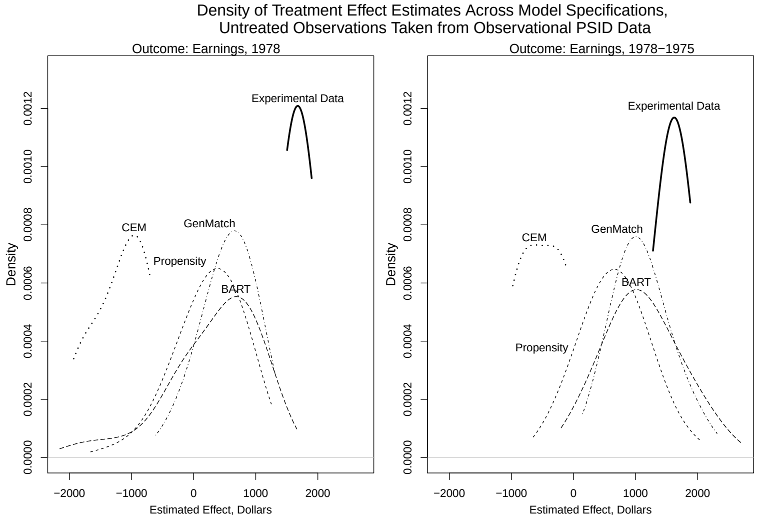
Experimental Results



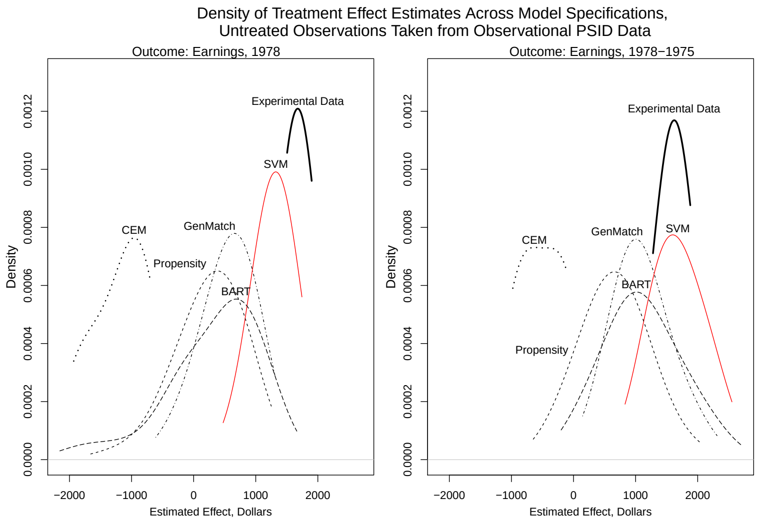
Observational Results



Observational Results



Observational Results

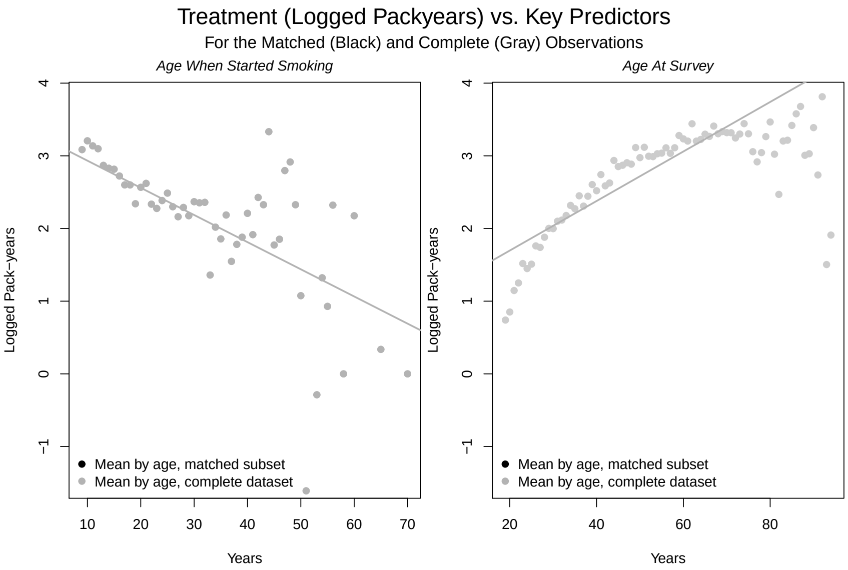


Smoking and Medical Expenditures

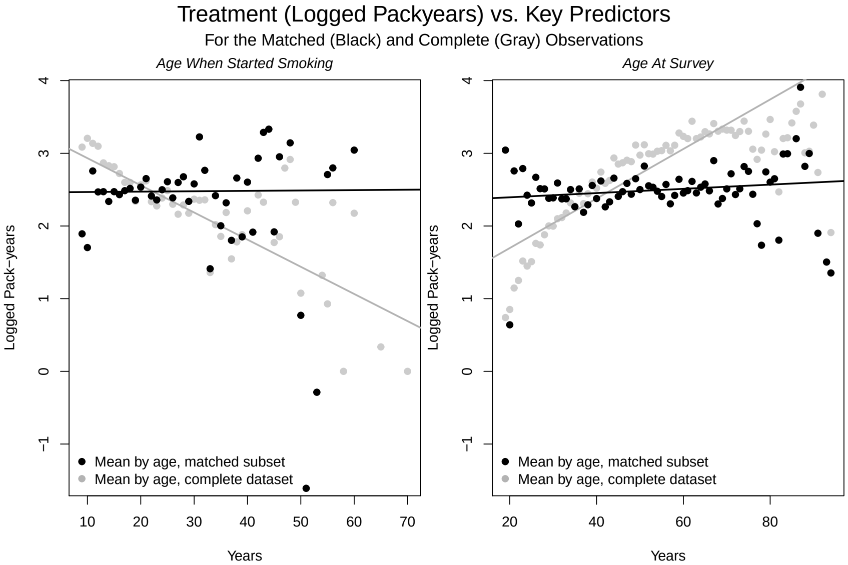
The 1987 National Medical Expenditure Survey (Johnson, et al. 2003; Imai and van Dyk 2004)

- Treatment: $log(pack - years)$: packs a day times number of years smoking, logged
- Respondents: Representative sample of US population
- $n = 9,708$ smokers; to be balanced
- $n = 9,804$ non-smokers; reference group
- Outcome: Medical expenditure, dollars
- X: age at survey, age when started smoking, gender, race, education, marital status, census region, poverty status, seat-belt use

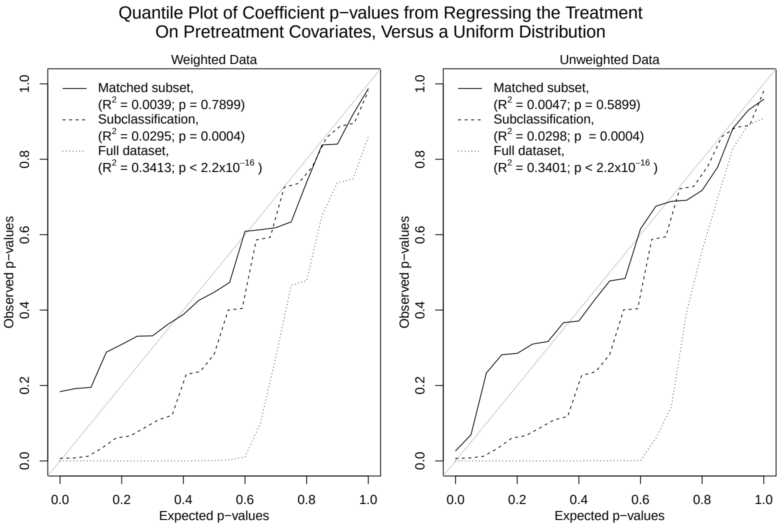
Assessing Balance



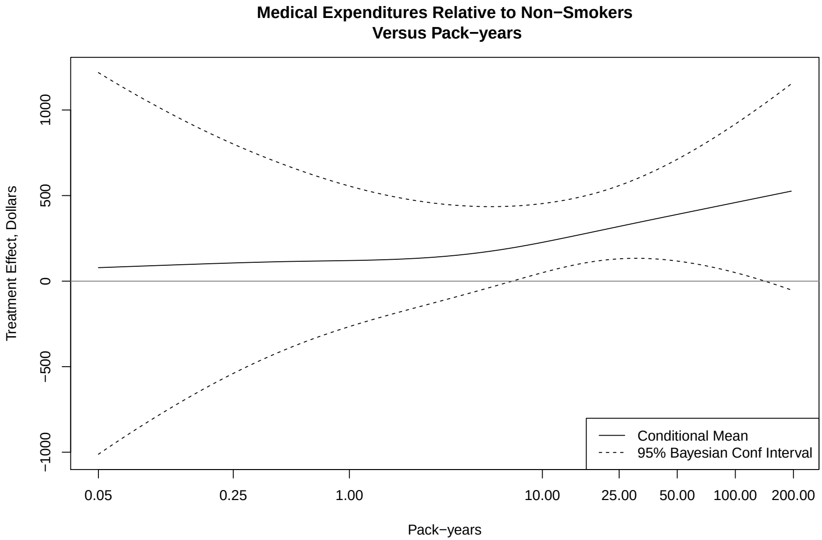
Assessing Balance



Assessing Balance



Estimated Effect



Conclusion

The proposed method adapts the SVM technology to the matching problem.

The method:

- is fully automated
- makes no functional form assumptions
- identifies the largest balanced subset
- can also accommodate continuous treatments