Achieving Optimal Covariate Balance Under General Treatment Regimes

Marc Ratkovic

Princeton University

March 14, 2012
Summary

Causal inference is central to empirical research.

Experiments in observational settings are not always possible \(\Rightarrow\) biased inference

Matching pre-processes the data such that treated and untreated observations are balanced along pre-treatment covariates.

I show that a modified Support Vector Machine (SVM) identifies a balanced sample, for both a binary and continuous treatment.
Structure of Presentation

Presentation proceeds in four parts:

1. Causal Inference, Matching, and Balance
2. Support Vector Machines and Balance
3. Empirical Examples
   - Binary treatment
   - Continuous treatment
4. Conclusion
# Table of Contents

1. **A Brief Overview of Matching**
   - The Potential Outcomes Framework
   - Balancing Methods

2. **Support Vector Machines and Balance**
   - Contributions
   - A Simple Example
   - Analytic Results
   - Nonparametric Representation

3. **Empirical Examples**
   - Binary Treatment Regime
   - Continuous Treatment Regime

4. **Conclusion**
The Setup

- **Treatment:** $T_i \sim F_T$
  - Binary treatment: $T_i \in \{-1, 1\}$
  - Continuous treatment: $T_i \in (a, b)$

- **Potential outcome:** $Y_i(T_i)$

- **Pre-treatment covariates:** $x_i \sim F_X$

- **Sign functional:** $\eta(\cdot)$ such that $E(\text{sgn}(T_i)|x_i) = \text{sgn}(\eta(x_i))$

- **IID observations** $(T_i, x_i)$ observed
Assumptions and Estimands

Treatment effect:

- Binary: $TE_i = Y_i(1) - Y_i(-1)$
- Continuous: $TE_i = Y_i(T) - Y_i(E(T))$
Assumptions and Estimands

Treatment effect:
- **Binary:** $TE_i = Y_i(1) - Y_i(-1)$
- **Continuous:** $TE_i = Y_i(T) - Y_i(E(T))$

Unbiased estimation of a treatment effect requires
- $P(T_i = T_i|x_i) > 0 \forall T_i$
- $Y_i(T_i) \perp T_i|x_i$
Assumptions and Estimands

Treatment effect:

- Binary: $TE_i = Y_i(1) - Y_i(-1)$
- Continuous: $TE_i = Y_i(T) - Y_i(E(T))$

Unbiased estimation of a treatment effect requires

- $P(T_i = T_i|x_i) > 0 \forall T_i$
- $Y_i(T_i) \perp T_i|x_i$

A subset of the data is balanced if

- $T_i \perp x_i$
Assumptions and Estimands

Treatment effect:
- Binary: $TE_i = Y_i(1) - Y_i(-1)$
- Continuous: $TE_i = Y_i(T) - Y_i(E(T))$

Unbiased estimation of a treatment effect requires
- $P(T_i = T_i|x_i) > 0 \ \forall \ T_i$
- $Y_i(T_i) \perp \perp T_i|x_i$

A subset of the data is balanced if
- $T_i \perp \perp x_i$

Common estimands
- $ATE = E(TE_i)$
- $ATT = E(TE_i|T_i)$
A Brief Overview of Matching
Support Vector Machines and Balance
Empirical Examples
Conclusion

Achieving Balance Through Matching

Why match?

- Observations may select into a treatment level
- Inference may not be robust to model specification
Achieving Balance Through Matching

Why match?
- Observations may select into a treatment level
- Inference may not be robust to model specification

How matching works
- Identifies a set of untreated units similar to treated units
- Achieves balance
Existing Matching Methods

Propensity-based methods (Rosenbaum and Rubin, 1983)

- Propensity score: \( E(T_i = T_i | x_i = x_i) \)
- Estimate propensity, then
  - Matching
  - Subclassification
Existing Matching Methods

Propensity-based methods (Rosenbaum and Rubin, 1983)
- Propensity score: $E(T_i = T_i|x_i = x_i)$
- Estimate propensity, then
  - Matching
  - Subclassification

Covariate balancing methods
- CEM: Multidimensional binning
- GenMatch: Stochastic optimizer, minimizes marginal covariate discrepancy
Existing Matching Methods

Propensity-based methods (Rosenbaum and Rubin, 1983)
- Propensity score: \( E(T_i = T_i | x_i = x_i) \)
- Estimate propensity, then
  - Matching
  - Subclassification

Covariate balancing methods
- CEM: Multidimensional binning
- GenMatch: Stochastic optimizer, minimizes marginal covariate discrepancy

Binary and continuous treatment regimes
- Binary treatment: well-studied
- Continuous treatment: lack of a reference group
Table of Contents

1 A Brief Overview of Matching
   - The Potential Outcomes Framework
   - Balancing Methods

2 Support Vector Machines and Balance
   - Contributions
   - A Simple Example
   - Analytic Results
   - Nonparametric Representation

3 Empirical Examples
   - Binary Treatment Regime
   - Continuous Treatment Regime

4 Conclusion
The proposed method

- Maximizes balance across all covariates simultaneously
- Accommodates continuous treatments
Contributions

The proposed method
- Maximizes balance across all covariates simultaneously
- Accommodates continuous treatments

Support Vector Machines
- Popular classifier that outperforms logistic regression
- Kernel trick allows for nonparametric specification
- Identifies marginal cases
Geometric Intuition

Separating Hyperplane

Margin

Marc Ratkovic

Optimal Matching
A Simple Example: The Standard SVM

Assume $T_i \in \{\pm 1\}$, a single covariate $x_i$, target functional $x_i \beta$
A Simple Example: The Standard SVM

Assume $T_i \in \{\pm 1\}$, a single covariate $x_i$, target functional $x_i \beta$

SVM objective, with $|z|_+ = \max(z, 0)$:

$$\mathcal{L}^{SVM}(\beta) = \sum_{i=1}^{n} |1 - T_i x_i \beta|_+$$
A Simple Example: The Standard SVM

Assume $T_i \in \{\pm 1\}$, a single covariate $x_i$, target functional $x_i \beta$

SVM objective, with $|z|_+ = \max(z, 0)$:

$$\mathcal{L}_{\text{SVM}}(\beta) = \sum_{i=1}^{n} |1 - T_i x_i \beta|_+$$

$\mathcal{M}_{\text{SVM}}$: marginal SVM cases, $\{i : 1 - T_i x_i \beta > 0\}$

$$\frac{\partial \mathcal{L}_{\text{SVM}}(\beta)}{\partial \beta} = \sum_{i \in \mathcal{M}_{\text{SVM}}} T_i x_i = 0 \Rightarrow \sum_{T_i = 1, i \in \mathcal{M}_{\text{SVM}}} x_i = \sum_{T_i = -1, i \in \mathcal{M}_{\text{SVM}}} x_i$$
A Simple Example: The Standard SVM

Assume $T_i \in \{\pm 1\}$, a single covariate $x_i$, target functional $x_i \beta$
SVM objective, with $|z|_+ = \max(z, 0)$:

$$
\mathcal{L}^{\text{SVM}}(\beta) = \sum_{i=1}^{n} |1 - T_i x_i \beta|_+
$$

$\mathcal{M}^{\text{SVM}}$: marginal SVM cases, $\{i : 1 - T_i x_i \beta > 0\}$

$$
\frac{\partial \mathcal{L}^{\text{SVM}}(\beta)}{\partial \beta} = \sum_{i \in \mathcal{M}^{\text{SVM}}} T_i x_i = 0 \Rightarrow \sum_{T_i = 1, i \in \mathcal{M}^{\text{SVM}}} x_i = \sum_{T_i = -1, i \in \mathcal{M}^{\text{SVM}}} x_i
$$

Optimality condition for an SVM implies $E(T_i x_i | x_i \in \mathcal{M}^{\text{SVM}}) = 0$
A Simple Example: The Binary Matching SVM

Centered covariate, \( x_i^* = x_i - \sum_{i:T_i=1} x_i / \sum_i 1(T_i = 1) \)
A Simple Example: The Binary Matching SVM

Centered covariate, $x_i^* = x_i - \sum_{i:T_i=1} x_i / \sum_i 1(T_i = 1)$

$$\mathcal{L}^{Bin}(\beta) = \sum_{i=1}^{n} |1 - T_i x_i^* \beta|_+ \text{ s.t. } x_i^* \beta < 1 \forall \{i : T_i = 1\}$$
A Simple Example: The Binary Matching SVM

Centered covariate, \( x_i^* = x_i - \sum_{i:T_i=1} x_i / \sum_i 1(T_i = 1) \)

\[
\mathcal{L}_{Bin}(\beta) = \sum_{i=1}^{n} |1 - T_i x_i^* \beta| + \text{s.t. } x_i^* \beta < 1 \forall \{i : T_i = 1\}
\]

\( \mathcal{M}^{Bin} \): marginal cases, \( \{i : 1 - T_i x_i^* \beta > 0\} \)

\[
\frac{\partial \mathcal{L}_{Bin}(\beta)}{\partial \beta} = \sum_{i \in \mathcal{M}^{Bin}} T_i x_i^* = 0 \Rightarrow \sum_{T_i=1} x_i^* = \sum_{T_i=-1, i \in \mathcal{M}^{Bin}} x_i^* = 0
\]
Centered covariate, $x_i^* = x_i - \sum_{i:T_i=1} x_i / \sum_i 1(T_i = 1)$

$$\mathcal{L}^{Bin}(\beta) = \sum_{i=1}^{n} |1 - T_i x_i^* \beta| + \text{s.t. } x_i^* \beta < 1 \forall \{i : T_i = 1\}$$

$\mathcal{M}^{Bin}$: marginal cases, $\{i : 1 - T_i x_i^* \beta > 0\}$

$$\frac{\partial \mathcal{L}^{Bin}(\beta)}{\partial \beta} = \sum_{i \in \mathcal{M}^{Bin}} T_i x_i^* = 0 \implies \sum_{T_i=1} x_i^* = \sum_{T_i=-1, i \in \mathcal{M}^{Bin}} x_i^* = 0$$

Centering $x_i$ gives balance-in-mean for marginal observations: $E(x_i|T_i = 1) = E(x_i|T_i = -1, x_i \in \mathcal{M}^{Bin})$
A Simple Example: The Continuous Treatment SVM

Centered covariate: $$x_i^* = x_i - \sum_{i \in M_{\text{Cont}}} x_i / |M_{\text{Cont}}|$$

Centered treatment: $$T_i^* = T_i - \sum_{i \in M_{\text{Cont}}} T_i / |M_{\text{Cont}}|$$
A Simple Example: The Continuous Treatment SVM

Centered covariate: \( x_i^* = x_i - \sum_{i \in M^{\text{Cont}}} x_i / |M^{\text{Cont}}| \)

Centered treatment: \( T_i^* = T_i - \sum_{i \in M^{\text{Cont}}} T_i / |M^{\text{Cont}}| \)

\[
\mathcal{L}^{\text{Cont}}(\beta) = \sum_{i=1}^{n} |(T_i^*)^2 - T_i^* x_i^* \beta|_+ 
\]
A Simple Example: The Continuous Treatment SVM

Centered covariate: \( x_i^* = x_i - \sum_{i \in M^{Cont}} x_i / |M^{Cont}| \)
Centered treatment: \( T_i^* = T_i - \sum_{i \in M^{Cont}} T_i / |M^{Cont}| \)

\[
L^{Cont}(\beta) = \sum_{i=1}^{n} |(T_i^*)^2 - T_i^* x_i^* \beta| +
\]

\( M^{Cont}: \) marginal cases, \( \{i : (T_i^*)^2 - T_i^* x_i^* \beta > 0\} \)

\[
\frac{\partial L^{Cont}(\beta)}{\partial \beta} = \sum_{i \in M^{Cont}} T_i^* x_i^* = 0 \Rightarrow \text{cov}(T_i, x_i | x_i \in M^{Cont}) = 0
\]
Appropriate centering of the covariate and treatment variable lead to balance-in-means or uncorrelatedness.

The result clearly holds for multiple, uncorrelated covariates.
Appropriate centering of the covariate and treatment variable lead to balance-in-means or uncorrelatedness.

The result clearly holds for multiple, uncorrelated covariates.

Next, I embed the target functional in a Hilbert space, which

- Is a high-dimensional space where any well-behaved curve is linear
- Admits a linear expansion of the target functional in terms of uncorrelated eigenfunctionals
- Can be represented on the data through the kernel trick
The Binary Treatment SVM

**Lemma 1**: Joint Independence between Treatment Assignment and Covariates with a Binary Treatment

Assume $\eta(x_i)$ lives in a Hilbert space and is smooth, bounded, and twice differentiable
The Binary Treatment SVM

**Lemma 1:** Joint Independence between Treatment Assignment and Covariates with a Binary Treatment

Assume $\eta(x_i)$ lives in a Hilbert space and is smooth, bounded, and twice differentiable

The proposed method minimizes $E(|1 - T_i \eta(x_i)|_+)$ where

- $\eta(x_i)$ admits representation $\sum_j \alpha_j \psi_j(x_i)$
- $E(\psi_j|T_i = 1) = 0$
- $\eta(x_i|T_i = 1) < 1$
- $M^{Bin} = \{x_i : \{T_i = 1\} \cup \{\eta(x_i) > -1; T_i = -1\}\}$
Lemma 1: Joint Independence between Treatment Assignment and Covariates with a Binary Treatment

Assume $\eta(x_i)$ lives in a Hilbert space and is smooth, bounded, and twice differentiable

The proposed method minimizes $E(|1 - T_i \eta(x_i)|_\infty)$ where

- $\eta(x_i)$ admits representation $\sum_j \alpha_j \psi_j(x_i)$
- $E(\psi_j | T_i = 1) = 0$
- $\eta(x_i | T_i = 1) < 1$
- $M^{Bin} = \{x_i : \{T_i = 1\} \cup \{\eta(x_i) > -1; T_i = -1\}\}$

Then, $T_i \perp x_i$ if $x_i \in M^{Bin}$
Lemma 2: Joint Independence between Treatment Assignment and Covariates with a Continuous Treatment

Now, assume $T_i$ has support $(a, b)$, and $E(T_i|x_i \in M^{Cont}) = 0$
The Continuous Treatment SVM

**Lemma 2**: Joint Independence between Treatment Assignment and Covariates with a Continuous Treatment

Now, assume $T_i$ has support $(a, b)$, and $E(T_i | x_i \in M^{Cont}) = 0$

The proposed method minimizes $E(|T_i^2 - T_i \eta(x_i)|_+)$ where

- $\eta(x_i)$ admits representation $\sum_j \alpha_j \psi_j(x_i)$
- $E(\psi_j | x_i \in M^{Cont}) = E(T_i | x_i \in M^{Cont}) = 0$
- $M^{Cont} = \{x_i : 1 > \eta(x_i)/T_i\}$
**Lemma 2**: Joint Independence between Treatment Assignment and Covariates with a Continuous Treatment

Now, assume \( T_i \) has support \((a, b)\), and \( E(T_i|x_i \in \mathcal{M}^{Cont}) = 0 \)

The proposed method minimizes \( E(|T_i^2 - T_i\eta(x_i)|_+) \) where

- \( \eta(x_i) \) admits representation \( \sum_j \alpha_j \psi_j(x_i) \)
- \( E(\psi_j|x_i \in \mathcal{M}^{Cont}) = E(T_i|x_i \in \mathcal{M}^{Cont}) = 0 \)
- \( \mathcal{M}^{Cont} = \{x_i : 1 > \eta(x_i)/T_i\} \)

Then, \( T_i \perp \perp x_i \) if \( x_i \in \mathcal{M}^{Cont} \)
Representor Theorem

Given:

- a bandwidth parameter $\theta$
Representor Theorem

Given:

- a bandwidth parameter $\theta$
- points of evaluation $k \in K$, and $R_{kern}$ the $|K| \times |K|$ reproducing kernel.

- $K = \{k : T_k = 1\}$, for binary treatment
- $K = \{k : k \in M_{Cont}\}$, for continuous treatment
Representor Theorem

Given:

- a bandwidth parameter $\theta$
- points of evaluation $k \in K$, and $R_{kern}$ the $|K| \times |K|$ reproducing kernel.
  - $K = \{k : T_k = 1\}$, for binary treatment
  - $K = \{k : k \in M^{Cont}\}$, for continuous treatment
- $V = (X_K^T X_K)^{-1}$
- $R_{kern} = [r_{k_1,k_2}] = \exp\left(-\theta \|x_{k_1} - x_{k_2}\|_V^2\right)$
- nonparametric bases: $R_{coef}$, the $n \times k$ matrix, columns centered on $K$
Objective Function

Regularization: $\textit{Expected Loss} = \textit{Sample Loss} + \textit{Complexity Term}$
Objective Function

Regularization: *Expected Loss* = *Sample Loss* + *Complexity Term*

Binary Treatment:

\[ E(|1 - T_i \eta(x_i)|_+) = \frac{1}{n} \sum_{i=1}^{n} |1 - T_i \eta(x_i)|_+ + \lambda c^\top R_{kern} c \]

s.t. \( \eta(x_i) < 1 \ \forall \ T_i = 1 \)
Objective Function

Regularization: \( \text{Expected Loss} = \text{Sample Loss} + \text{Complexity Term} \)

Binary Treatment:

\[
E( |1 - T_i \eta(x_i)|_+) = \frac{1}{n} \sum_{i=1}^{n} |1 - T_i \eta(x_i)|_+ + \lambda c^\top R_{kern} c
\]

\[ \text{s.t. } \eta(x_i) < 1 \ \forall \ T_i = 1 \]

Continuous Treatment:

\[
E( |T_i^2 - T_i \eta(x_i)|_+) = \frac{1}{n} \sum_{i=1}^{n} |T_i^2 - T_i \eta(x_i)|_+ + \lambda c^\top R_{kern} c
\]

\[ \text{s.t. } \sum_{i \in M_{\text{Cont}}} T_i = \sum_{i \in M_{\text{Cont}}} R_{\text{coef},i} = 0 \]

where \( \eta(x_i) = R_{i,\text{coef}} c \)
For a fixed \( \{\lambda, \theta\} \):

- Coordinate descent algorithm
- Solve \( j^{th} \) coordinate subproblem; update sequentially
- Iterate to convergence
For a fixed \( \{\lambda, \theta\} \):

- Coordinate descent algorithm
- Solve \( j^{th} \) coordinate subproblem; update sequentially
- Iterate to convergence

Selecting among \( \{\lambda, \theta\} \):

- GACV statistic
- Alternating line search
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)
- Coarsened Exact Matching (Iacus, et al. 2011)
- Genetic Matching (Sekhon 2011)
Analyses

Competitors
- Logistic propensity matching (Ho, et al. 2011)
- Coarsened Exact Matching (Iacus, et al. 2011)
- Genetic Matching (Sekhon 2011)

Outcomes
- 1978 earnings
- 1978 earnings - 1975 earnings
Analyses

Competitors

- Logistic propensity matching (Ho, et al. 2011)
- Coarsened Exact Matching (Iacus, et al. 2011)
- Genetic Matching (Sekhon 2011)

Outcomes

- 1978 earnings
- 1978 earnings - 1975 earnings

Datasets

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

Density of Treatment Effect Estimates Across Model Specifications, Using NSW Experimental Data

Outcome: Earnings, 1978

Outcome: Earnings, 1978–1975

Experimental Data

Marc Ratkovic

Optimal Matching
Experimental Results

Density of Treatment Effect Estimates Across Model Specifications, Using NSW Experimental Data

Outcome: Earnings, 1978

Outcome: Earnings, 1978−1975

Marc Ratkovic

Optimal Matching
Experimental Results

Density of Treatment Effect Estimates Across Model Specifications, Using NSW Experimental Data

Outcome: Earnings, 1978

Outcome: Earnings, 1978–1975

Marc Ratkovic
Optimal Matching
Observational Results

Density of Treatment Effect Estimates Across Model Specifications, Untreated Observations Taken from Observational PSID Data

Outcome: Earnings, 1978

Outcome: Earnings, 1978–1975

Marc Ratkovic
Optimal Matching
Observational Results

Density of Treatment Effect Estimates Across Model Specifications, Untreated Observations Taken from Observational PSID Data

Outcome: Earnings, 1978

- BART
- CEM
- GenMatch
- Propensity

Outcome: Earnings, 1978–1975

- BART
- CEM
- GenMatch
- Propensity
Observational Results

Density of Treatment Effect Estimates Across Model Specifications, Untreated Observations Taken from Observational PSID Data

Outcome: Earnings, 1978

Outcome: Earnings, 1978–1975
Summary

The proposed method has been shown to:
- Return experimental results from a field experiment
- Outperform competitors with untreated observations drawn from an observational dataset
Smoking and Medical Expenditures

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

- **Treatment**: $\log(\text{pack} - \text{years})$: 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

Quantile Plot of Coefficient p–values from Regressing the Treatment On Pretreatment Covariates, Versus a Uniform Distribution

Weighted Data

- Matched subset,
  \( R^2 = 0.0039; \ p = 0.7899 \)
- Full dataset,
  \( R^2 = 0.3413; \ p < 2.2 \times 10^{-16} \)

Unweighted Data

- Matched subset,
  \( R^2 = 0.0047; \ p = 0.5899 \)
- Full dataset,
  \( R^2 = 0.3401; \ p < 2.2 \times 10^{-16} \)
Assessing Balance

Treatment (Logged Pack-years) vs. Key Predictors
For the Matched (Black) and Complete (Gray) Observations

- Mean by age, matched subset
- Mean by age, complete dataset

Marc Ratkovic
Optimal Matching
Assessing Balance

Treatment (Logged Packyears) vs. Key Predictors
For the Matched (Black) and Complete (Gray) Observations

Age When Started Smoking

Age At Survey

Marc Ratkovic
Optimal Matching
Estimated Effect

Medical Expenditures Relative to Non-Smokers
Versus Pack-years

![Graph showing estimated effect of medical expenditures relative to non-smokers versus pack-years. The graph includes a solid line representing the conditional mean and a dashed line representing the 95% Bayesian confidence interval. The x-axis represents pack-years ranging from 0.05 to 200, and the y-axis represents treatment effect in dollars ranging from -1000 to 1100.](image-url)
Summary

The proposed method has been shown to

- Identify a subset of observations for which the treatment level is uncorrelated with pre-treatment covariates
- Maintain a sufficiently large subset of observations to allow for modeling and inference
# Table of Contents

1. A Brief Overview of Matching  
   - The Potential Outcomes Framework  
   - Balancing Methods
2. Support Vector Machines and Balance  
   - Contributions  
   - A Simple Example  
   - Analytic Results  
   - Nonparametric Representation
3. Empirical Examples  
   - Binary Treatment Regime  
   - Continuous Treatment Regime
4. Conclusion
The proposed method adapts the SVM technology to the matching problem.

The method:

- Examines SVM marginal observations
- Accommodates both binary and continuous treatments
- Identifies a subset of observations for which pre-treatment covariates and treatment level are jointly independent
- Is fully automated
- Has desirable theoretical properties
- Performs well on benchmark datasets
Thank you!
Thank you!
Thank you!
GACV

\[ GACV_{SVM}(\lambda, \theta) = \]
\[ \frac{1}{n} \sum_{i=1}^{n} w_i | 1 - T_i \hat{T}_i^{SVM} |_+ \]
\[ + \frac{1}{n} \sum_{j \in J_{samp}} w_j \hat{c}_i T_i \left( 1 + 1(T_j \hat{T}_j^{SVM} < -1) \right) \]

Continuous case: \( w_i = T_i^*^2 \)
Binary case: \( w_i = n_1/n; T_i = 1; w_i = n_{-1,M^{Bin}}/n, T_i = -1 \)
GACV

\[ GACV_{bin}(\lambda, \theta) = \]
\[ \frac{1}{n} \left\{ \frac{n_1^2}{n} + \sum_{T_i^{bin} = -1} \left( \frac{n_{-1,j}}{n} \right) \left| 1 - T_i \hat{T}_i \right|_+ + \right. \]
\[ \frac{1}{n} \sum_{i \in I_{samp}} \left( \frac{n_1 |\hat{c}_i|}{n} \right) \left( 1 + 1 (\hat{T}_i < -1) \right) \right\} \]

\[ GACV_{cont}(\lambda, \theta) = \]
\[ \frac{1}{n} \sum_{i=1}^{n} (T_i^*)^2 \left| 1 - \frac{\hat{T}_i}{T_i^*} \right|_+ \]
\[ + \frac{1}{n} \sum_{i \in I_{samp}} |\hat{c}_iT_i^*| \left( 1 + 1 (|\hat{T}_i| > |T_i^*|) \right) \]