

Matching Methods for Causal Inference

Gary King
Institute for Quantitative Social Science
Harvard University

Talk at University of Georgia, 3/3/2011

- Problem: Model dependence (review)
- Solution: Matching to preprocess data (review)
- Problem: Many matching methods & specifications
- Solution: The Space Graph helps us choose
- Problem: The most commonly used method can increase imbalance!
- Solution: Other methods do not share this problem
- (Coarsened Exact Matching is usually best)
- \rightsquigarrow Lots of insights revealed in the process

Model Dependence Example

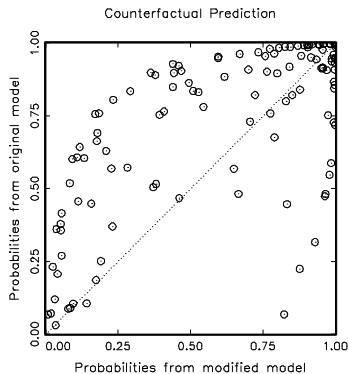
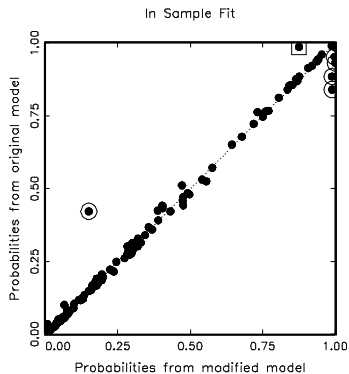
Replication: Doyle and Sambanis, APSR 2000

- **Data:** 124 Post-World War II civil wars
- **Dependent variable:** peacebuilding success
- **Treatment variable:** multilateral UN peacekeeping intervention (0/1)
- **Control vars:** war type, severity, duration; development status; etc.
- **Counterfactual question:** UN intervention switched for each war
- **Data analysis:** Logit model
- **The question:** How *model dependent* are the results?

Two Logit Models, Apparently Similar Results

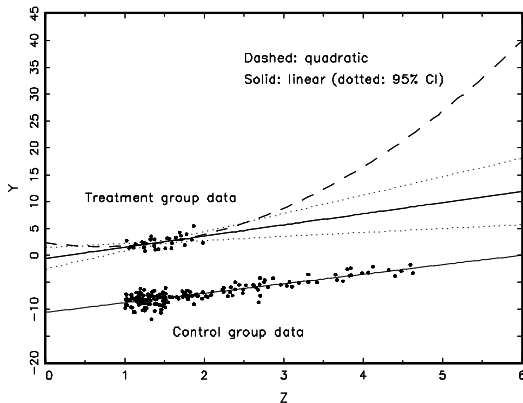
Variables	Original “Interactive” Model			Modified Model		
	Coeff	SE	P-val	Coeff	SE	P-val
Wartype	-1.742	.609	.004	-1.666	.606	.006
Logdead	-.445	.126	.000	-.437	.125	.000
Wardur	.006	.006	.258	.006	.006	.342
Factnum	-1.259	.703	.073	-1.045	.899	.245
Factnum2	.062	.065	.346	.032	.104	.756
Trnsfcap	.004	.002	.010	.004	.002	.017
Develop	.001	.000	.065	.001	.000	.068
Exp	-6.016	3.071	.050	-6.215	3.065	.043
Decade	-.299	.169	.077	-0.284	.169	.093
Treaty	2.124	.821	.010	2.126	.802	.008
UNOP4	3.135	1.091	.004	.262	1.392	.851
Wardur*UNOP4	—	—	—	.037	.011	.001
Constant	8.609	2.157	0.000	7.978	2.350	.000
N		122			122	
Log-likelihood		-45.649			-44.902	
Pseudo R^2		.423			.433	

Doyle and Sambanis: Model Dependence



Model Dependence: A Simpler Example

(King and Zeng, 2006: fig.4 *Political Analysis*)

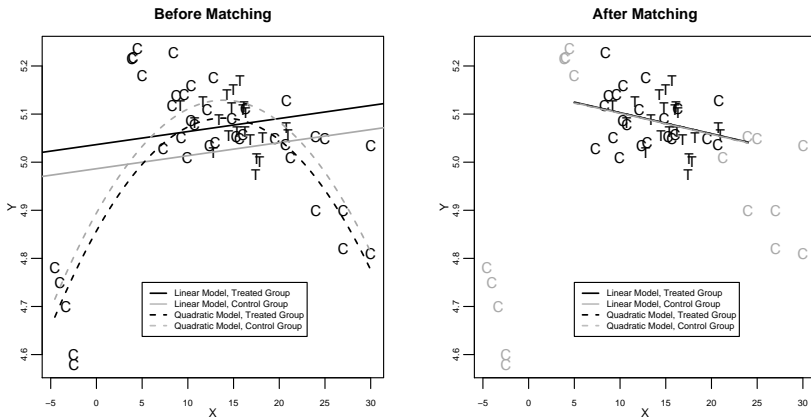


What to do?

- Preprocess I: Eliminate extrapolation region
- Preprocess II: Match (prune bad matches) within interpolation region
- Model remaining imbalance

Matching within the Interpolation Region

(Ho, Imai, King, Stuart, 2007: fig.1, *Political Analysis*)



Matching reduces model dependence, bias, and variance

What Matching Does

- Notation:

Y_i Dependent variable

T_i Treatment variable (0/1)

X_i Pre-treatment covariates

- Treatment Effect for treated ($T_i = 1$) observation i :

$$\begin{aligned} \text{TE}_i &= Y_i(T_i = 1) - Y_i(T_i = 0) \\ &= \text{observed} - \text{unobserved} \end{aligned}$$

- Estimate $Y_i(0)$ with Y_j from matched ($X_i \approx X_j$) controls

$$\hat{Y}_i(0) = Y_j(0) \text{ or a model } \hat{Y}_i(0) = \hat{g}_0(X_j)$$

- Prune unmatched units to improve **balance** (so X is unimportant)

- Qol: Sample Average Treatment effect on the Treated:

$$\text{SATT} = \frac{1}{n_T} \sum_{i \in \{T_i=1\}} \text{TE}_i$$

- or Feasible Average Treatment effect on the Treated: FSATT

Method 1: Mahalanobis Distance Matching

1 Preprocess (Matching)

- $\text{Distance}(X_i, X_j) = \sqrt{(X_i - X_j)' S^{-1} (X_i - X_j)}$
- Match each treated unit to the nearest control unit
- Control units: not reused; pruned if unused
- Prune matches if $\text{Distance} > \text{caliper}$

2 Estimation Difference in means or a model

Method 2: Propensity Score Matching

1 Preprocess (Matching)

- Reduce k elements of X to scalar $\pi_i \equiv \Pr(T_i = 1|X) = \frac{1}{1+e^{-X_i\beta}}$
- $\text{Distance}(X_i, X_j) = |\pi_i - \pi_j|$
- Match each treated unit to the nearest control unit
- Control units: not reused; pruned if unused
- Prune matches if $\text{Distance} > \text{caliper}$

2 Estimation Difference in means or a model

Method 3: Coarsened Exact Matching

1 Preprocess (Matching)

- Temporarily coarsen X as much as you're willing
 - e.g., Education (grade school, high school, college, graduate)
 - Easy to understand, or can be automated as for a histogram
- Apply exact matching to the coarsened X , $C(X)$
 - Sort observations into strata, each with unique values of $C(X)$
 - Prune any stratum with 0 treated or 0 control units
- Pass on original (uncoarsened) units except those pruned

2 Estimation Difference in means or a model

- Need to weight controls in each stratum to equal treateds
- Can apply other matching methods within CEM strata (inherit CEM's properties)

The Bias-Variance Trade Off in Matching

- **Bias** (& model dependence) = $f(\text{imbalance}, \text{importance}, \text{estimator})$
 \rightsquigarrow we measure **imbalance** instead
- **Variance** = $f(\text{matched sample size}, \text{estimator})$
 \rightsquigarrow we measure **matched sample size** instead
- **Bias-Variance trade off** \rightsquigarrow **Imbalance- n Trade Off**
- Measuring Imbalance
 - Classic measure: Difference of means (for each variable)
 - Better measure (difference of multivariate histograms):

$$\mathcal{L}_1(f, g; H) = \frac{1}{2} \sum_{\ell_1 \dots \ell_k \in H(\mathbf{X})} |f_{\ell_1 \dots \ell_k} - g_{\ell_1 \dots \ell_k}|$$

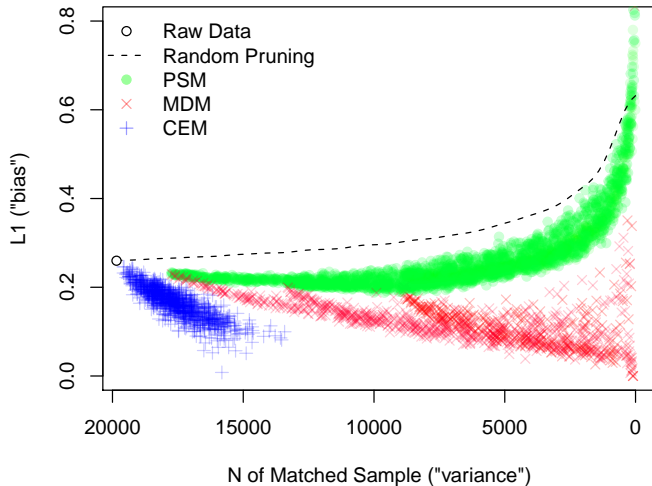
Comparing Matching Methods

- MDM & PSM: Choose matched n , match, check imbalance
- CEM: Choose imbalance, match, check matched n
- Best practice: iterate
- Choose matched solution & matching method becomes irrelevant
- Our idea: Compute lots of matching solutions, identify the frontier of lowest imbalance for each given n , and choose a matching solution

A Space Graph: Real Data

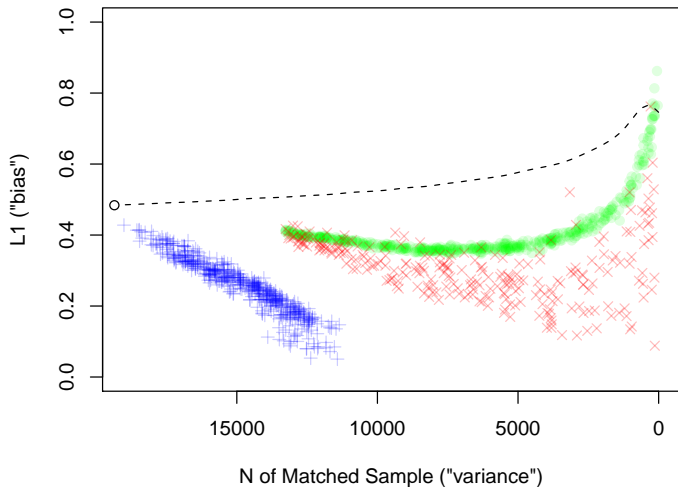
King, Nielsen, Coberley, Pope, and Wells (2011)

Healthways Data

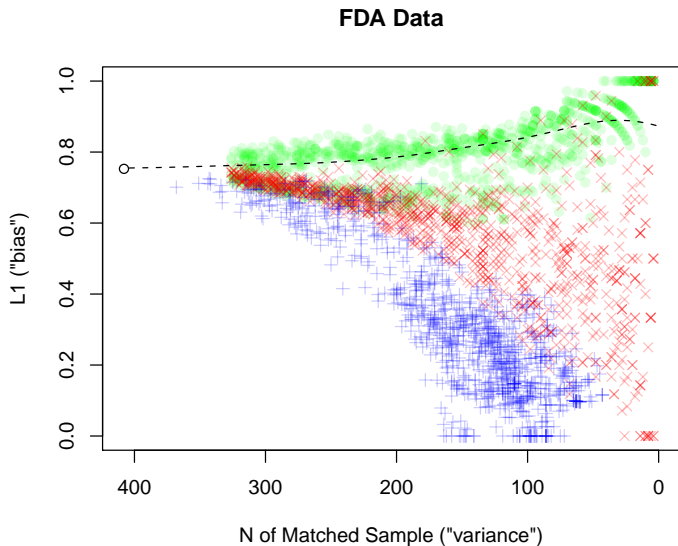


A Space Graph: Real Data

Called/Not Called Data

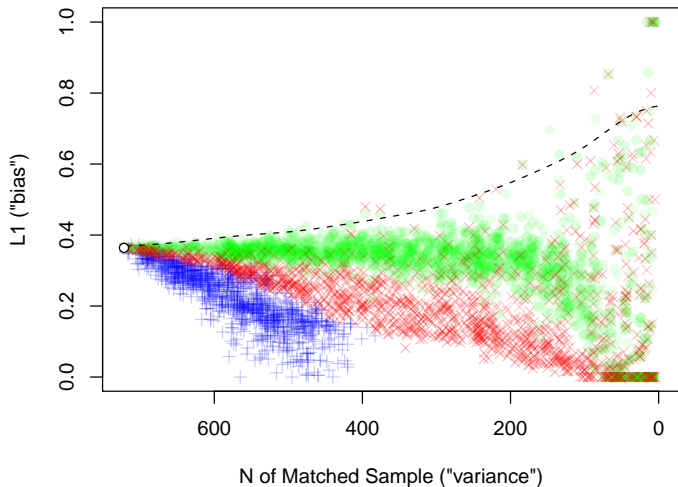


A Space Graph: Real Data

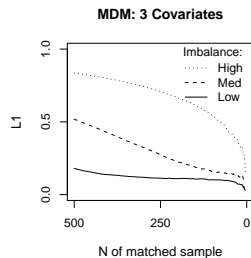
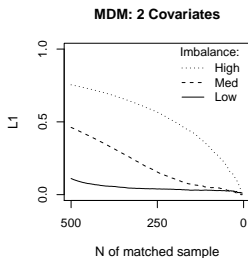
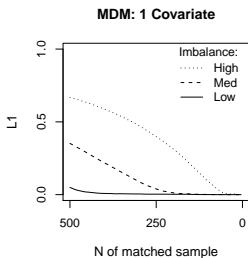


A Space Graph: Real Data

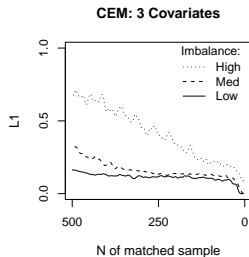
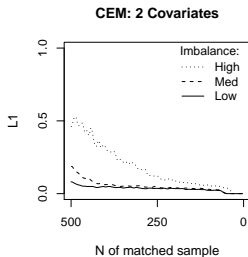
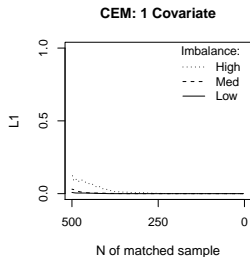
Lalonde Data Subset



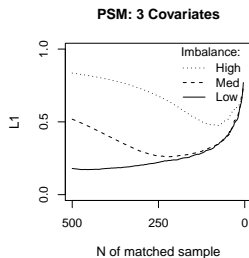
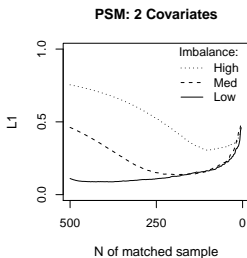
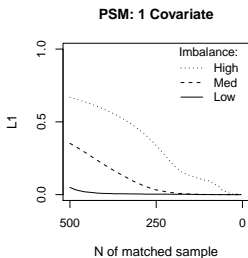
A Space Graph: Simulated Data — Mahalanobis



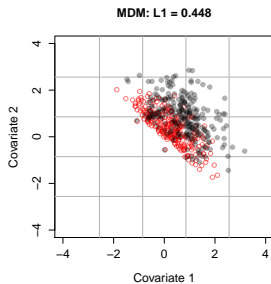
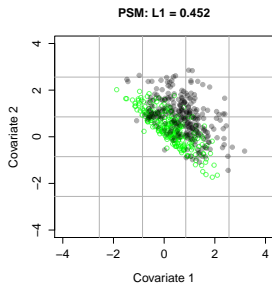
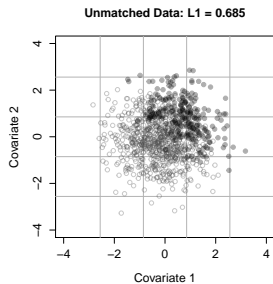
A Space Graph: Simulated Data — CEM



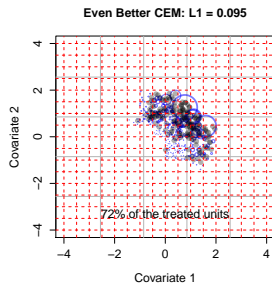
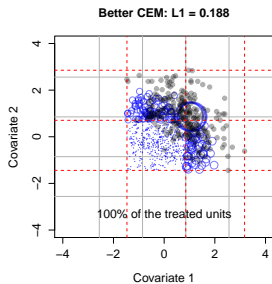
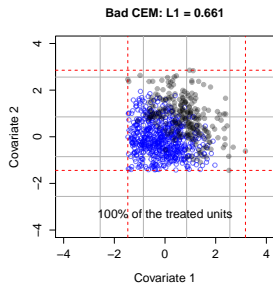
A Space Graph: Simulated Data — Propensity Score



Data where PSM Works Reasonably Well — PSM & MDM



Data where PSM Works Reasonably Well — CEM



CEM Weights and Nonparametric Propensity Score

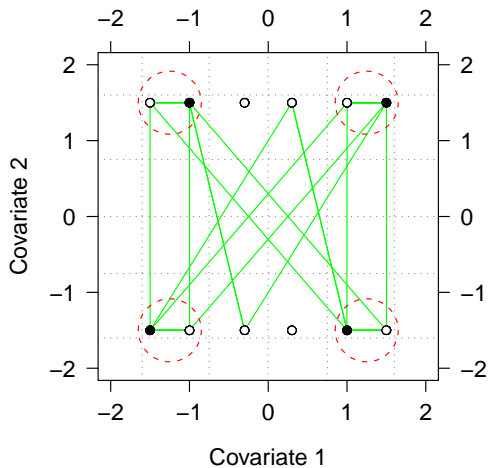
CEM Weight: $w_i = \frac{m_i^T}{m_i^C}$ (Unnormalized)

CEM Pscore: $\hat{\Pr}(T_i = 1|X_i) = \frac{m_i^T}{m_i^T + m_i^C}$

⇒ CEM:

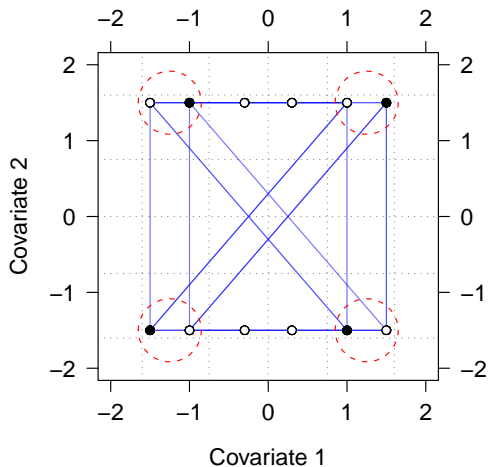
- Gives a better pscore than PSM
- Doesn't match based on crippled information

PSM Approximates Random Matching in Balanced Data



- PSM Matches
- - - CEM and MDM Matches

Destroying CEM with PSM's Two Step Approach



- CEM Matches
- CEM-generated PSM Matches

Conclusions

- Propensity score matching:
 - The problem:
 - Imbalance can be worse than original data
 - Can increase imbalance when removing the worst matches
 - Approximates random matching in well-balanced data (Random matching increases imbalance)
 - The Cause: unnecessary 1st stage dimension reduction
 - Implications:
 - Balance checking required
 - Adjusting for potentially irrelevant covariates *with PSM* is a mistake
 - Adjusting experimental data *with PSM* is a mistake
 - Reestimating the propensity score after eliminating noncommon support may be a mistake
- In four data sets and many simulations:
CEM > Mahalanobis > Propensity Score
- (Your performance may vary)
- CEM and Mahalanobis do not have PSM's problems
- You can easily check with the Space Graph

For papers, software (for R and Stata), tutorials, etc.

<http://GKing.Harvard.edu/cem>