Estimating Heterogeneous Treatment Effects in Randomized Control Trials

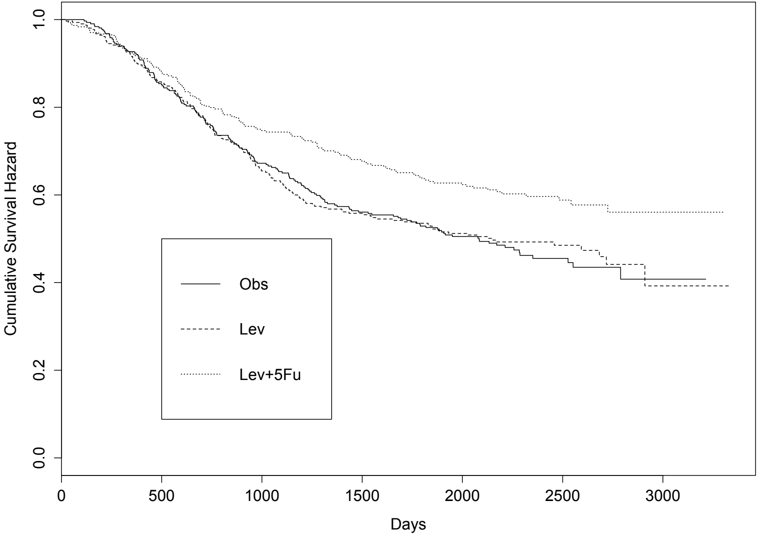
Christopher Adams

Federal Trade Commission

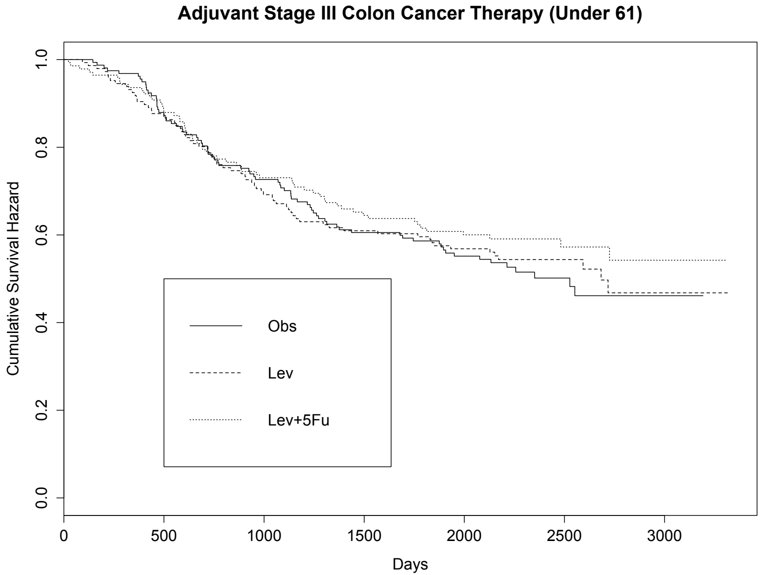
September 26 2014

The views expressed here are those of the author and do not necessarily reflect those of the Federal Trade Commission.

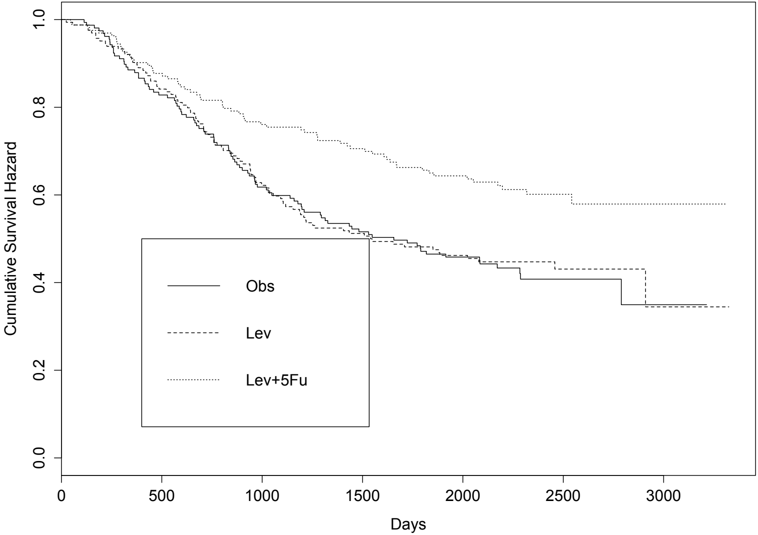
Adjuvant Stage III Colon Cancer Therapy (All patients)



We conclude that adjuvant therapy with levamisole and fluorouracil should be standard treatment for Stage C colon carcinoma. Since most patients in our study were treated by community oncologists, this approach should be readily adaptable to conventional medical practice. (N Engl J Med 1990; 322:352-8.)


Adjuvant Stage III Colon Cancer Therapy (Under 61)
Cumulative Survival Hazard
0.0
0.2
0.4
0.6
0.8
1.0
0
500
1000
1500
2000
2500
3000
Days
Obs
Lev
Lev+5Fu

Adjuvant Stage III Colon Cancer Therapy (61 and Over)



In exploratory subset analyses, levamisole–fluorouracil treatment appeared to have the greatest advantage among male patients (in both survival and recurrence), older patients (recurrence), patients with tumors that were well differentiated to moderately well differentiated (survival and recurrence), patients in whom more than four nodes were involved (survival), and patients treated 21 to 35 days after surgery (recurrence). These results show two striking contradictions to those of subset analyses reported in the NCCTG study, in which levamisole plus fluorouracil was found to be most effective in reducing the risk of recurrence among female patients and younger patients. This underscores the importance of the statement by the authors of that study, that "subset analyses must be interpreted with great caution."[3]

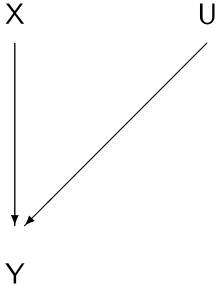
Overview

- Identification of HTEs
  - 1-Signal Model
  - 3-Signal Models
  - 2-Signal Models
- Estimator
  - Additivity Constrained NMF
  - Monte Carlo Simulation
- Adjuvant Therapy for S3 Colon Cancer
- Returns to Schooling
- Instrumental Variables (if time)

Identification

Causal Graph

Mediated Causal Graph



1-Signal Graph

1-Signal Model

- Finite Mixture Model

$$Pr(y < Y) = \sum_{k=1}^{K} \pi(u_k) F(Y|u_k) \tag{1}$$

- If $\mathcal{Y}$ has I elements

$$\mathbf{p}_y = \mathbf{F}\pi \tag{2}$$

where

  - $\mathbf{p}_y$ is a $I \times 1$ vector
  - π is a $K \times 1$ vector
  - $\mathbf{F}$ is a $I \times K$ matrix

1-Signal Model

- Example

$$\begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \\ F_{31} & F_{32} \end{bmatrix} \times \begin{bmatrix} \pi_1 \\ \pi_2 \end{bmatrix} \quad (3)$$

- where
- $\sum_{i=1}^{3} p_i = 1$
- $\sum_{i=1}^{3} F_{ik} = 1$ for all $k \in \{1, 2\}$
- $\sum_{k=1}^{2} \pi_k = 1$

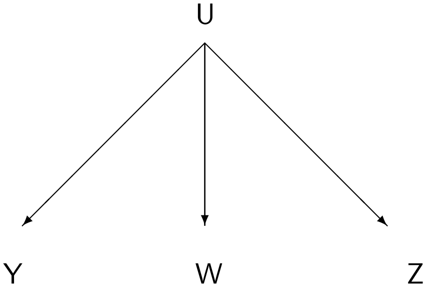
1-Signal Model

- Finite Mixture Model

$$Pr(y < Y) = \sum_{k=1}^{K} \pi(u_k) F(Y|u_k) \tag{4}$$

- $F(.|u_k)$ and $\pi(u_k)$ are not identified (in general)

3-Signal Graph



3-Signal Model

- Finite Mixture Model

$$Pr(y < Y, w < W, z < Z) = \sum_{k=1}^{K} F(Y|u_k)G(W|u_k)H(u_k, Z) \quad (5)$$

- If $\mathcal{Y}$ has I elements, $\mathcal{W}$ has J elements, $\mathcal{Z}$ has L elements.

$$\mathbf{P}_z = \mathbf{F}\mathbf{D}_z\mathbf{G}^T \quad (6)$$

- where
  - $\mathbf{P}_z$ is a $I \times J$ prob. matrix conditional on $Z = z$
  - $\mathbf{D}_z$ is a $K \times K$ diagonal matrix conditional on $Z = z$
  - $\mathbf{G}^T$ is a $K \times J$ matrix

3-Signal Model

- Example

$$\begin{bmatrix} P_{11z} & P_{12z} \\ P_{21z} & P_{22z} \\ P_{31z} & P_{32z} \end{bmatrix} = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \\ F_{31} & F_{33} \end{bmatrix} \begin{bmatrix} \pi_{1z} & 0 \\ 0 & \pi_{2z} \end{bmatrix} \begin{bmatrix} G_{11} & G_{21} & G_{31} \\ G_{12} & G_{22} & G_{32} \end{bmatrix} \tag{7}$$

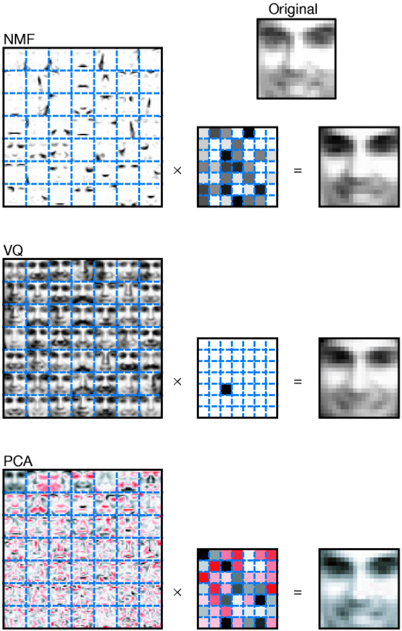
- where
- $\sum_{i=1}^{3} P_{ijz} = 1$ for all $j \in \{1, 2\}$
- $\sum_{i=1}^{3} F_{ik} = 1$ for all $k \in \{1, 2\}$
- $\sum_{k=1}^{2} \pi_{kz} = 1$
- $\sum_{j=1}^{2} G_{kj} = 1$ for $k \in \{1, 2\}$

3-Signal Model

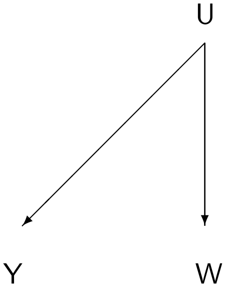
$$\mathbf{P}_z = \mathbf{F}\mathbf{D}_z\mathbf{G}^T \tag{8}$$

- If
  - $L \geq 2$ (num of Z's)
  - certain rank conditions hold
- then $\mathbf{F}$, $\mathbf{D}_z$, $\mathbf{D}_{z'}$, and $\mathbf{G}$ are identified.
- Kruskal (1977), An et al (2010), etc.

NMF


Original
NMF
×
=
VQ
×
=
PCA
×
=

2-Signal Graph



2-Signal Model

- Exclusion restriction

$$F(Y|u_k, W) = F(Y|u_k, W') \tag{9}$$

- for all Y, $W \neq W'$ and u_k
- Henry et al. (2014, 2013)

2-Signal Model

- Finite Mixture Model

$$Pr(y < Y, w < W) = \sum_{k=1}^{K} \pi(u_k) F(Y|u_k) G(W|u_k) \quad (10)$$

- As matrices

$$\mathbf{P} = \mathbf{F}\mathbf{D}_{\pi}\mathbf{G}^{T} \quad (11)$$

- or

$$\mathbf{P} = \mathbf{W}\mathbf{H} \quad (12)$$

2-Signal Model

- In general, bi-llinear matrix decompositions are not unique

$$\mathbf{P} = \mathbf{W}\mathbf{H} = \mathbf{W}\mathbf{A}\mathbf{A}^{-1}\mathbf{H} \tag{13}$$

- or

$$\mathbf{P} = \tilde{\mathbf{W}}\tilde{\mathbf{H}} \tag{14}$$

- where $\tilde{\mathbf{W}} = \mathbf{W}\mathbf{A}$ and $\tilde{\mathbf{H}} = \mathbf{A}^{-1}\mathbf{H}$
- for all $\mathbf{A}$ of full-rank.

2-Signal Model

- If
  - $\mathbf{P}$, $\mathbf{W}$, $\mathbf{H}$, $\tilde{\mathbf{W}}$ and $\tilde{\mathbf{H}}$ are non-negative
  - $\mathbf{W}$, $\mathbf{H}$ rank conditions hold
  - $\mathbf{A}$ is diagonal
- then $\mathbf{W}$ and $\mathbf{H}$ are "essentially unique" (up to scaling and permutation).
- Huang et al (2013)

2-Signal Model

- When is **A** diagonal?
  - Given certain "sparsity conditions."
  - Also if
    - $\sum_{i=1}^{I} \mathbf{F}_{ik} = 1$,
    - $\sum_{j=1}^{J} \mathbf{G}_{jk} = 1$, and
    - certain (possibly less strong) sparsity conditions.

Set and Point Identification

Sufficient Conditions for Uniqueness

If $\mathbf{A}$, $\mathbf{W}$, $\tilde{\mathbf{W}}$, $\mathbf{H}$ and $\tilde{\mathbf{H}}$ are matrices with the following properties,

1. $\tilde{\mathbf{H}} = \mathbf{A}^{-1}\mathbf{H}$,
2. $\tilde{\mathbf{W}} = \mathbf{W}\mathbf{A}$,
3. $\sum_{j=1}^{J} \mathbf{H}_{kj} = 1$ for all $k \in \{1, ..., K\}$,
4. $\sum_{j=1}^{J} \tilde{\mathbf{H}}_{kj} = 1$ for all $k \in \{1, ..., K\}$,
5. $\mathbf{H} > 0$, $\mathbf{W} > 0$, and
6. $\tilde{\mathbf{H}} \geq 0$, $\tilde{\mathbf{W}} \geq 0$

then as $\mathbf{W}_{i_k k} \rightarrow 0$ and $\mathbf{H}_{k j_k} \rightarrow 0$ for at least one $i_k \in \{1, ..., I\}$ and $j_k \in \{1, ..., J\}$, where $i_k \neq i_{k'}$ and $j_k \neq j_{k'}$ for all $k \neq k' \in \{1, ..., K\}$, $\mathbf{A} \rightarrow \mathbf{Q}$ where $\mathbf{Q} \in \mathcal{Q}$.

The Set $\mathcal{A}$

- Given the conditions we can write

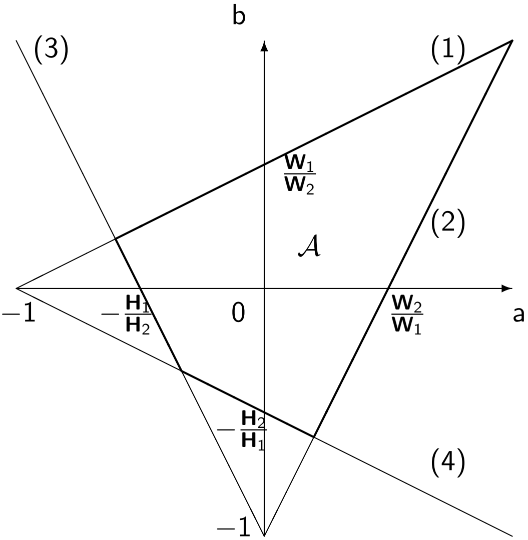
$$\mathbf{A} = \begin{bmatrix} 1+a & -a \\ -b & 1+b \end{bmatrix} \tag{15}$$

- where $a, b \in \Re$.
- Let $\{a, b\} \in \mathcal{A}$ for all vectors associated with **W** and **H** such that the inequalities in the theorem hold.

The Set $\mathcal{A}$

1. $b \leq (1+a)\frac{\mathbf{W}_1}{\mathbf{W}_2}$
2. $b \geq a\frac{\mathbf{W}_1}{\mathbf{W}_2} - 1$
3. $b \geq -a\frac{\mathbf{H}_2}{\mathbf{H}_1} - 1$
4. $b \geq -(1+a)\frac{\mathbf{H}_2}{\mathbf{H}_1}$

The set $\mathcal{A}$



Necessary and Sufficient Conditions for Point Identification

If $\mathbf{A}$, $\mathbf{F}$, $\tilde{\mathbf{F}}$, $\mathbf{D}_\pi$, $\tilde{\mathbf{D}}_\pi$, $\mathbf{G}$ and $\tilde{\mathbf{G}}$ are matrices with the following properties,

1. $\tilde{\mathbf{G}}^T = \mathbf{A}^{-1}\mathbf{G}^T$,
2. $\tilde{\mathbf{F}}\tilde{\mathbf{D}}_\pi = \mathbf{F}\mathbf{D}_\pi\mathbf{A}$,
3. $\mathbf{G}^T \geq 0$, $\mathbf{F}\tilde{\mathbf{D}}_\pi \geq 0$,
4. $\tilde{\mathbf{G}}^T \geq 0$, $\tilde{\mathbf{F}}\tilde{\mathbf{D}}_\pi \geq 0$
5. $\sum_{i=1}^{I} \mathbf{F}_{i,k} = 1$ and $\sum_{i=1}^{I} \tilde{\mathbf{F}}_{i,k} = 1$ for all $k \in \{1, ..., K\}$,
6. $\sum_{k=1}^{K} \mathbf{D}_{\pi k} = 1$ and $\sum_{k=1}^{K} \tilde{\mathbf{D}}_{\pi k} = 1$ and
7. $\sum_{j=1}^{J} \mathbf{G}^T_{k,j} = 1$ and $\sum_{j=1}^{J} \tilde{\mathbf{G}}^T_{k,j} = 1$ for all $k \in \{1, ..., K\}$,

and define

$$\begin{aligned} \mathcal{I}_k &= \{i \in \{1, ..., I\} | \mathbf{F}_{i,k} \neq 0\} \\ \mathcal{J}_k &= \{j \in \{1, ..., J\} | \mathbf{G}^T_{k,j} \neq 0\} \end{aligned} \tag{16}$$

then $\mathbf{P} = \mathbf{F}\mathbf{D}_\pi\mathbf{G}^T$ is unique up to relabeling if and only if there do not exist $k_1, k_2 \in \{1, ..., K\}$, $k_1 \neq k_2$ such that $\mathcal{I}_{k_1} \subseteq \mathcal{I}_{k_2}$ or $\mathcal{J}_{k_1} \subseteq \mathcal{J}_{k_2}$.

Proof.

- Necessary: Huang et al. (2013) Theorem 3.
- Sufficient: Result above.

Estimator

Constrained NMF

$$
\begin{aligned}
\min_{\mathbf{F},\mathbf{D}_\pi,\mathbf{G}} \quad & ||\mathbf{P} - \mathbf{F}\mathbf{D}_\pi\mathbf{G}^T||_F^2 \\
s.t. \quad & \mathbf{F}_{ik} \geq 0 \text{ for all } i, k \\
& \mathbf{D}_{\pi kk} \geq 0 \text{ for all } k \\
& \mathbf{G}_{jk} \geq 0 \text{ for all } j, k \\
& \textstyle\sum_{i=1}^{I} \mathbf{F}_{ik} = 1 \text{ for all } k \\
& \textstyle\sum_{j=1}^{J} \mathbf{G}_{jk} = 1 \text{ for all } k \\
& \textstyle\sum_{k=1}^{K} \pi_k = 1
\end{aligned}
\tag{17}
$$

- Implemented using continuously updating GMM.

Monte Carlo Results ($N = 30,000$)

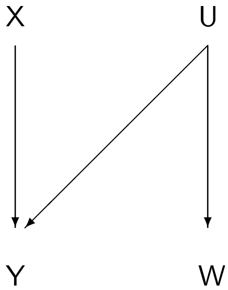
Monte Carlo Results

$$\mathbf{F} = \begin{bmatrix} 0.29 & 0.95 \\ 0 & 0.05 \\ 0.71 & 0 \end{bmatrix} \tag{18}$$

$$\mathbf{G} = \begin{bmatrix} 0 & 0.2 \\ 0.19 & 0.8 \\ 0.81 & 0 \end{bmatrix} \tag{19}$$

	Actual	Mean	*SD*	Mean	*SD*
π_1	0.7500	0.7337	(0.0862)	0.7493	(0.0032)
π_2	0.2500	0.2663	(0.0862)	0.2507	(0.0032)
$\mathbf{F}_{11}$	0.2900	0.3522	(0.1648)	0.2891	(0.0034)
$\mathbf{F}_{21}$	0.0001	0.0607	(0.1487)	0.0001	(0.0001)
$\mathbf{F}_{31}$	0.7099	0.5871	(0.2316)	0.7109	(0.0034)
$\mathbf{F}_{12}$	0.9499	0.8184	(0.2581)	0.9488	(0.0026)
$\mathbf{F}_{22}$	0.0500	0.1229	(0.1847)	0.0506	(0.0025)
$\mathbf{F}_{32}$	0.0001	0.0558	(0.1043)	0.0006	(0.0007)
$\mathbf{G}_{11}$	0.0001	0.0689	(0.1752)	0.0001	(0.0001)
$\mathbf{G}_{21}$	0.1899	0.2588	(0.1806)	0.1903	(0.0030)
$\mathbf{G}_{31}$	0.8100	0.6723	(0.2566)	0.8096	(0.0030)
$\mathbf{G}_{12}$	0.1999	0.2108	(0.1094)	0.1998	(0.0044)
$\mathbf{G}_{22}$	0.8000	0.6841	(0.2201)	0.7980	(0.0051)
$\mathbf{G}_{32}$	0.0001	0.0892	(0.1722)	0.0023	(0.0027)
SOS	0.0000	0.2507	(0.4348)	0.00003	(0.00003)
T		100		56	

RCT-2-Signal Graph



Adjuvant Chemotherapy for S3 Colon Cancer

RCT Data

- Moertel (1990)
- NEJM study on adjuvant therapy for stage III colon cancer patients
- R dataset called "colon" in the "survival" package

Heterogeneous Treatment Effect of Colon Cancer Drugs

	Type 1			Type 2		
	$\pi_1 = 0.77$			$\pi_2 = 0.23$		
	Obs	Lev	Lev+5FU	Obs	Lev	Lev+5FU
$<$ 1 Yr	0.04	0.04	0.05	0.20	0.28	0.17
1-3 Yrs	0.14	0.20	0.13	0.71	0.47	0.32
3-4 Yrs	0.09	0.05	0.07	0.09	0.11	0.00
$>$ 4 Yrs	0.73	0.71	0.75	0.00	0.15	0.51
$<$ 4, MW	0.65			0.00		
$<$ 4, W or P	0.16			0.54		
$>$ 4, MW	0.18			0.11		
$>$ 4, W or P	0.02			0.35		
SOS	0.0418					

Treatment Effect for Type 1 [5th percentile, 95th percentile]

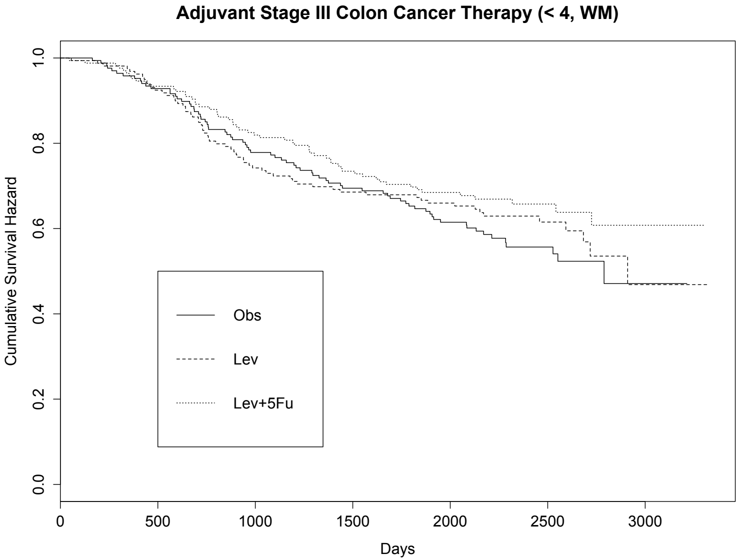
	Obs	Lev	Lev & 5-Fu
Type 1	0.77 [0.66,0.89]		
Less 1 Year	0.04 [0.00,0.08]	0.04 [0.00,0.07]	0.05 [0.00,0.36]
1-3 Years	0.14 [0.00,0.22]	0.20 [0.07,0.27]	0.13 [0.00,0.69]
3-4 Years	0.09 [0.00,0.14]	0.05 [0.00,0.09]	0.07 [0.00,1.00]
More 4 Years	0.73 [0.58,0.89]	0.71 [0.62,0.89]	0.75 [0.00,0.89]

Treatment Effect for Type 2 [5th percentile, 95th percentile]

	Obs	Lev	Lev & 5-Fu
Type 2	0.23		
	[0.11,0.34]		
Less 1 Year	0.20	0.28	0.17
	[0.00,0.29]	[0.11,0.51]	[0.00,0.98]
1-3 Years	0.71	0.47	0.32
	[0.52,1.00]	[0.29,0.79]	[0.00,1.00]
3-4 Years	0.09	0.11	0.00
	[0.00,0.18]	[0.00,0.22]	[0.00,0.67]
More 4 Years	0.00	0.15	0.51
	[0.00,0.12]	[0.00,0.14]	[0.00,1.00]

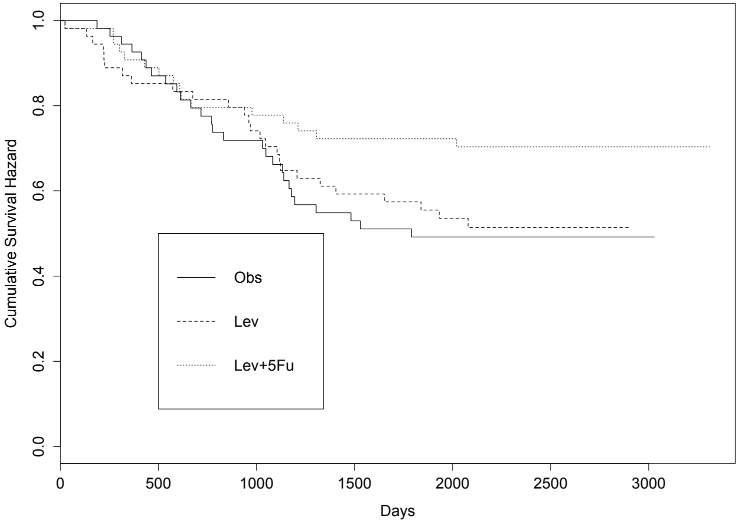
Observed Patient Characteristics By Type for Nodal Involvement and Histological Differentiation [5th percentile, 95th percentile]

	Type 1	Type 2
$<$ 4, Mod. Well	0.65	0.00
	[0.56,0.67]	[0.00,0.30]
$<$ 4, Well & Poor	0.16	0.54
	[0.13,0.23]	[0.38,0.67]
$>$ 4, Mod. Well	0.18	0.11
	[0.14,0.21]	[0.00,0.19]
$>$ 4, Well & Poor	0.02	0.35
	[0.00,0.06]	[0.11,0.45]
Sum of Squares	0.04512	
	[0.05,0.38]	

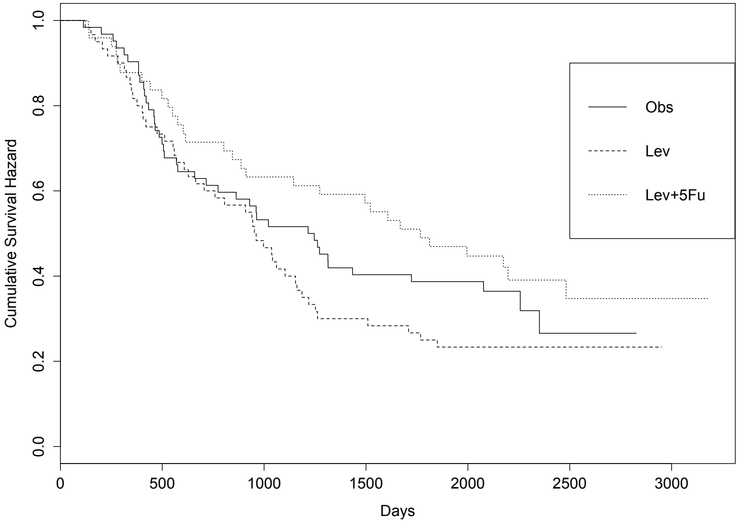


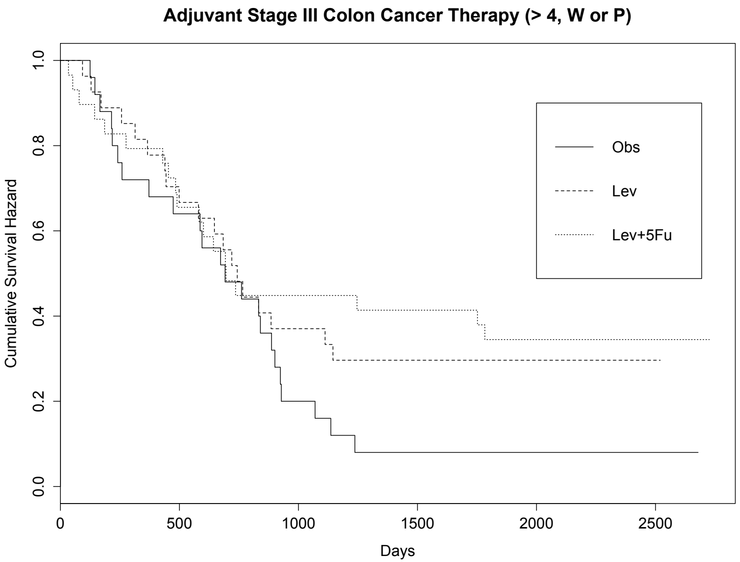
Adjuvant Stage III Colon Cancer Therapy (< 4, WM)
Cumulative Survival Hazard
1.0
0.8
0.6
0.4
0.2
0.0
0
500
1000
1500
2000
2500
3000
Days
Obs
Lev
Lev+5Fu

Adjuvant Stage III Colon Cancer Therapy (< 4, W or P)



Adjuvant Stage III Colon Cancer Therapy (> 4, WM)




Adjuvant Stage III Colon Cancer Therapy (> 4, W or P)
Cumulative Survival Hazard
0.0
0.2
0.4
0.6
0.8
1.0
0
500
1000
1500
2000
2500
Days
Obs
Lev
Lev+5Fu

Conclusion

- Estimate HTEs with mixture models
- Matrix factorization used to identify types
- Treatment effect varies across types
- Similar method for unconfounded and confounded data.

Marc Henry, Koen Jochmans, and Bernard Salanié. Inference on mixtures under tail restrictions. Penn State, December 2013.

Marc Henry, Yuichi Kitamura, and Benard Salanié. Partial identification of finite mixtures in econometric models. Quantitative Economics, 5:123–144, 2014.

Kejun Huang, Nicholas D Sidiropoulos, and Ananthram Swami. Non-negative matrix factorization revisited: Uniqueness and algorithm for symmetric decomposition. IEEE Transactions on Signal Processing, 2013. Forthcoming.