The Creative Tension between Statistics and Economics: Recent Development

Cheng Hsiao

University of Southern California and Nanyang Technological University

Aug 2009

Three Main Functions of Econometrics

1. Adaptation of economic theory for empirical testing
2. Forecast the future
3. Constructing counterfactuals to simulate different policy scenarios and evaluate the effectiveness of different programs

- There is a trend on "Measurement without theory" and "statistical analysis without a priori constraints" to provide impartial assessment.
- Economic and social science data are not generated from laboratory experiments. The causal relations among the observed data are hypothesized. Nor can observations be checked for internal consistency through repetitions. It is very hard to extract information from the observed data without imposing some sort of a priori information.

Arguments Favoring Combining Forecasts

- The true data generating process is unknown. Even the most complicated model is likely to be misspecified and can, at best, provide a reasonable "local" approximation. It is highly unlikely that a single model will dominate uniformly over time.
- The best model may change over time in way that can be difficult to track on the basis of past forecasting performance. Combining forecasts across different models may be viewd as a way to make the forecast more robust against misspecification biases and measurement errors in the data set.
- It is possible that diversification gains from combining across a set of forecasting models will dominate the strategy of only using a single forecasting model.

Notation

- Let $\mathbf{X}_t$ be an $m \times 1$ vector of forecasts with forecasting error covariance matrix Σ.
- Let $\mathbf{w}$ be an $m \times 1$ vector of weights to generate the pooled forecasts is $y_t^* = \mathbf{w}'\mathbf{X}_t$.
- Optimal weight $\mathbf{w}$ is a function of the loss function. We consider a symmetric loss function of minimizing the mean squred prediction error

$$\min \; E\left(y_t - y_t^*\right)^2$$

Eigenvector Approach

$$\min \mathbf{w}'\Sigma\mathbf{w}$$
$$\text{s.t.} \quad \mathbf{w}'\mathbf{w} = 1$$

w is the eigenvector corresponding to the smallest eigenvalue of Σ. When Σ is unknown, weighting based on S typically yields a sub-optimal outcome compared to simple averaging.

$$\begin{aligned} \mathbf{w}'\Sigma\mathbf{w} &\leq \mathbf{w}'S\mathbf{w} \\ \mathbf{w}'(I - S)\mathbf{w} &\leq 0 \end{aligned}$$

True and Estimated Covariance Matrix

T=50,N=10

31.0009	13.3650	19.0268	14.5543	14.8379	17.2885	26.6542	26.6001	26.4729	25.6513
13.3650	9.4834	10.8010	5.6672	9.2294	8.5069	13.7443	9.4908	12.8453	13.2887
19.0268	10.8010	15.2037	8.4799	11.8177	11.7045	18.6232	14.7016	17.7562	17.9790
14.5543	5.6672	8.4799	8.3243	6.3527	8.1272	12.3270	13.6661	12.5031	11.8433
14.8379	9.2294	11.8177	6.3527	11.0509	9.3843	15.1193	10.7130	14.1824	14.6142
17.2885	8.5069	11.7045	8.1272	9.3843	11.2190	15.9551	14.5777	15.5902	15.3744
26.6542	13.7443	18.6232	12.3270	15.1193	15.9551	26.0636	21.8885	24.2976	24.1661
26.6001	9.4908	14.7016	13.6661	10.7130	14.5777	21.8885	26.7852	22.4908	21.0075
26.4729	12.8453	17.7562	12.5031	14.1824	15.5902	24.2976	22.4908	24.7973	23.4090
25.6513	13.2887	17.9790	11.8433	14.6142	15.3744	24.1661	21.0075	23.4090	24.3021

25.3738	12.8129	17.4304	10.0315	13.1678	15.3367	23.2366	19.6440	22.3654	21.4021
12.8129	9.7462	10.7963	4.4147	9.3901	8.5308	14.0076	7.6246	12.1308	12.6284
17.4304	10.7963	14.4966	6.8446	11.0438	11.2377	17.6708	12.6821	16.3753	16.3451
10.0315	4.4147	6.8446	5.6755	4.7698	6.1259	9.0596	9.3614	9.1088	8.5166
13.1678	9.3901	11.0438	4.7698	10.9021	9.1184	14.4882	8.4486	12.9646	13.0563
15.3367	8.5308	11.2377	6.1259	9.1184	10.1683	15.0140	11.6982	14.2529	13.6315
23.2366	14.0076	17.6708	9.0596	14.4882	15.0140	24.2321	17.2047	22.1578	21.5438
19.6440	7.6246	12.6821	9.3614	8.4486	11.6982	17.2047	19.0519	17.6318	15.5937
22.3654	12.1308	16.3753	9.1088	12.9646	14.2529	22.1578	17.6318	21.9052	20.1600
21.4021	12.6284	16.3451	8.5166	13.0563	13.6315	21.5438	15.5937	20.1600	20.5624

Estimation of Eigenvectors

Difference and Squared Difference in True and Estimated Eigenvectors

1	2	3	4	5	6	7	8	9	10
0.1075	0.1264	-0.1311	0.4312	0.5628	-0.6519	0.1843	0.7791	-0.0259	0.0118
0.5286	-0.3493	-0.0276	-0.0265	-0.6436	-0.4236	0.2254	0.3398	-0.0324	-0.0252
-0.3229	-0.0136	0.7905	0.6725	-0.3825	0.0749	0.6371	-0.0098	0.1288	-0.0216
-0.2537	0.3142	0.7916	-0.5915	0.0731	-0.3417	0.6330	-0.4176	0.0196	0.0328
-0.2027	-0.5831	-0.3507	-0.2523	0.4525	-0.6678	0.5578	-0.6395	-0.0127	-0.0153
-0.0738	1.1561	0.1792	-0.0518	-0.2380	0.4252	0.2994	-0.2818	0.0209	-0.0117
0.1199	-0.0587	-0.1077	-0.9263	-0.2853	0.4295	-0.5279	0.2482	-0.0068	-0.0170
0.2477	-0.6940	-0.4036	0.0029	-0.3112	-0.0976	0.1473	-0.3271	0.0203	0.0431
-0.5392	-0.0241	0.4920	0.1324	-0.0529	0.3888	-0.8656	0.1989	0.0676	-0.0040
0.3349	0.1961	-0.7841	0.3487	0.5581	0.4784	-0.3720	-0.4214	0.0232	0.0031

1	2	3	4	5	6	7	8	9	10
0.0116	0.0160	0.0172	0.1860	0.3168	0.4250	0.0340	0.6069	0.0007	0.0001
0.2794	0.1220	0.0008	0.0007	0.4142	0.1795	0.0508	0.1154	0.0011	0.0006
0.1043	0.0002	0.6248	0.4523	0.1463	0.0056	0.4059	0.0001	0.0166	0.0005
0.0643	0.0987	0.6266	0.3499	0.0053	0.1167	0.4007	0.1744	0.0004	0.0011
0.0411	0.3401	0.1230	0.0637	0.2048	0.4460	0.3111	0.4090	0.0002	0.0002
0.0054	1.3365	0.0321	0.0027	0.0567	0.1808	0.0897	0.0794	0.0004	0.0001
0.0144	0.0035	0.0116	0.8581	0.0814	0.1845	0.2786	0.0616	0.0000	0.0003
0.0613	0.4816	0.1629	0.0000	0.0968	0.0095	0.0217	0.1070	0.0004	0.0019
0.2907	0.0006	0.2420	0.0175	0.0028	0.1511	0.7493	0.0396	0.0046	0.0000
0.1121	0.0384	0.6149	0.1216	0.3115	0.2289	0.1384	0.1776	0.0005	0.0000

MSFE with B~N(0,1)

(T0,T1)=	30 60	30 60	60 120	60 120	120 240	120 240
K=	1	5	1	5	1	5
AMSE1	1.0255*	1.0861*	1.0117*	1.1361*	1.0509*	2.0873*
AMSE2	2.1596	2.3439	1.3603	1.5359	1.199	2.4072
FULLSA	1.1162**	1.2422**	1.0536**	1.2587**	1.0791**	2.738
BG	1.5124	1.6654	1.2136	1.4104	1.167	2.3729
GRc	2.2322	2.4314	1.379	1.5601	1.2082	2.4287
GR	2.1118	2.2795	1.3512	1.5398	1.2021	2.4186
HLSA	1.5876	1.7205	1.2214	1.4144	1.1682	2.3923
HLT	2.1609	2.3422	1.3934	1.5554	1.2052	2.4268
HLE	1.579	2.3349	1.1073	2.6744	1.8097	7.4514
MAAIC	1.2007	1.4043***	1.0868	1.315	1.1214***	2.2494**
MAAICC	1.1885	1.4093	1.0835	1.323	1.1224	2.2524***
MABIC	1.171***	1.4237	1.0717***	1.3655	1.136	2.2838
AIC	1.2309	1.4577	1.0977	1.3361	1.1299	2.2636
AICC	1.2157	1.4572	1.0932	1.3459	1.1314	2.2678
BIC	1.1891	1.4733	1.0772	1.3938	1.1441	2.2981
PC1	1.3612	1.461	1.1567	1.294**	1.1238	2.2581
PC2	1.3612	1.461	1.1567	1.294**	1.1238	2.2581
PC3	1.3612	1.461	1.1567	1.294**	1.1238	2.2581
IC1	1.3612	1.461	1.1567	1.294**	1.1238	2.2581
IC2	1.3612	1.461	1.1567	1.294**	1.1238	2.2581
IC3	1.3612	1.461	1.1567	1.294**	1.1238	2.2581

MSFE with B~N(1,1)

(T0,T1)=	30 60	30 60	60 120	60 120	120 240	120 240
K=	1	5	1	5	1	5
AMSE1	1.1116*	1.2699*	1.0727*	1.5498*	1*	1.1686*
AMSE2	2.3732	2.7309	1.4579	2.0944	1.147	1.3342
FULLSA	1.2158**	1.5246**	1.1289**	1.8573	1.0223	1.3131
BG	1.6678	2.0123	1.3319	1.9597	1.0968	1.2946
GRc	2.4585	2.8241	1.4821	2.125	1.1552	1.3447
GR	2.3183	2.6939	1.4675	2.1194	1.1449	1.3431
HLSA	1.7336	2.0497	1.3385	1.9632	1.0984	1.2995
HLT	2.3085	2.6686	1.4688	2.1031	1.1567	1.3457
HLE	5.4575	7.0562	2.7416	6.4745	1.0224	6.0931
MAAIC	1.4109***	1.7101	1.2346***	1.8471***	1.0316	1.2446**
MAAICC	1.4117	1.7463	1.2392	1.8641	1.0307	1.2453***
MABIC	1.4237	1.7989	1.2644	1.9181	1.0182**	1.2585
AIC	1.4672	1.767	1.2563	1.8711	1.0353	1.2544
AICC	1.4587	1.8048	1.2598	1.8896	1.0342	1.2532
BIC	1.464	1.8581	1.2868	1.9547	1.019***	1.2673
PC1	1.4744	1.7035***	1.2379	1.793**	1.0696	1.2454
PC2	1.4744	1.7035***	1.2379	1.793**	1.0696	1.2454
PC3	1.4744	1.7035***	1.2379	1.793**	1.0696	1.2454
IC1	1.4744	1.7035***	1.2379	1.793**	1.0696	1.2454
IC2	1.4744	1.7035***	1.2379	1.793**	1.0696	1.2454
IC3	1.4744	1.7035***	1.2379	1.793**	1.0696	1.2454

Let (y_{0i}^*, y_{1i}^*) denote the potential outcomes of the i^{th} individual in the untreated and treated state. Then, the treatment effects on the i^{th} individual are just

$$\Delta_i = y_{1i}^* - y_{0i}^*$$

The Average Treatment Effects (ATE)

$$ATE = E\left(y_{1i}^{*} - y_{0i}^{*}\right)$$

The ATE is of interest if one is interested in the effect of treatment for a randomly assigned individual or population mean response to treatment.

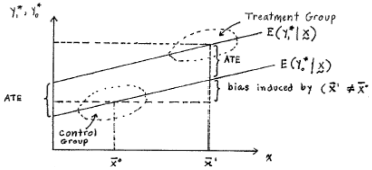
Data

The observed data is in the form of (y_i, d_i), $i = 1, ..., N$, where $d_i = 1$ if the i^{th} individual receives treatment and $d_i = 0$ if not, and

$$y_i = d_i y_{1i}^* + (1 - d_i)\, y_{0i}^*, \text{where } i = 1, ..., N$$

In other words, we do not simultaneously observe y_{1i}^* and y_{0i}^* $\longrightarrow$ there is a missing data problem.

- Selection on observables



- Selection on unobservables

$$E(\varepsilon_{1i}) \neq E(\varepsilon_{1i}/d_i = 1)$$

$$E(\varepsilon_{0i}) \neq E(\varepsilon_{0i}/d_i = 1)$$

Propensity Score Matching Method
(Rosenbaum and Rubin (1983))

Assume no selection on unobservables
(i.e. $E\left(\varepsilon_{ji}\right) = E\left(\varepsilon_{ji} / d_i = j\right) = 0$ where $j = 0, 1$)
Let $0 < p\left(d_i = 1/\mathbf{w}_i\right) = p\left(\mathbf{w}_i\right) < 1$, then
(i) $\left\{\left(y_{1i}^*, y_{0i}^*\right) \perp d_i / \mathbf{w}_i\right\} \Longrightarrow \left\{\left(y_{1i}^*, y_{0i}^*\right) \perp d_i / p\left(\mathbf{w}_i\right)\right\}$
(ii) $\mathbf{w}_i \perp d_i / p\left(\mathbf{w}_i\right)$

Nonparametric Model

Conditional Independence Assumption – selection is ignorable after controlling a set of observable confounders $\mathbf{w}_i = \{x_i \cup z_i\}$

$$d_i \perp (y_{1i}^*, y_{0i}^*) \,/\mathbf{w}_i$$

then

$$\begin{aligned} E\left(y_{1i}^*/d_i, \mathbf{w}_i\right) &= E\left(y_{1i}^*/\mathbf{w}_i\right), \\ E\left(y_{0i}^*/d_i, \mathbf{w}_i\right) &= E\left(y_{0i}^*/\mathbf{w}_i\right), \end{aligned}$$

$$\begin{aligned} ATE &= E\left(y_{1i}^* - y_{0i}^*/\mathbf{w}_i = \mathbf{w}\right) \\ &= E\left(y_i/d_i = 1, \mathbf{w}_i = \mathbf{w}\right) - E\left(y_i/d_i = 0, \mathbf{w}_i = \mathbf{w}\right) \\ &= E\left(y_{1i}^*/\mathbf{w}_i = \mathbf{w}\right) - E\left(y_{0i}^*/\mathbf{w}_i = \mathbf{w}\right) \end{aligned}$$

$$
\begin{aligned}
ATE &= E_{p(\mathbf{w})}\left\{E\left(y_i/p\left(\mathbf{w}\right), d_i = 1\right) - E\left(y_i/p\left(\mathbf{w}\right), d_i = 0\right)\right\} \\
&= E_{p(\mathbf{w})}\left\{E\left(y_{1i}^{*}/p\left(\mathbf{w}\right)\right) - E\left(y_{0i}^{*}/p\left(\mathbf{w}\right)\right)\right\} \\
&= E\left(y_{1i}^{*} - y_{0i}^{*}\right)
\end{aligned}
$$

$$
TT = E_{p(\mathbf{w})/d_i=1}\left\{E\left(y_i/p\left(\mathbf{w}\right), d_i = 1\right) - E\left(y_i/p\left(\mathbf{w}\right), d_i = 0\right)\right\}
$$

Decriminalization and Marijuana Smoking Prevalence: Evidence from Australia

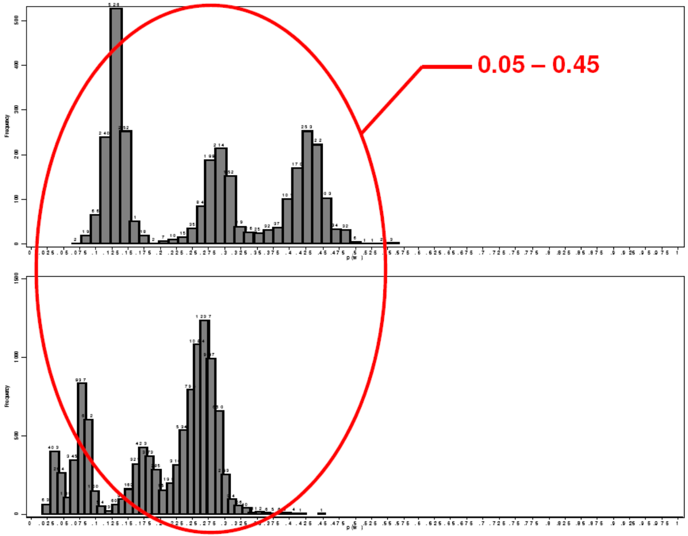
Kannika Damrongplasit, Cheng Hsiao and Xueyan Zhao

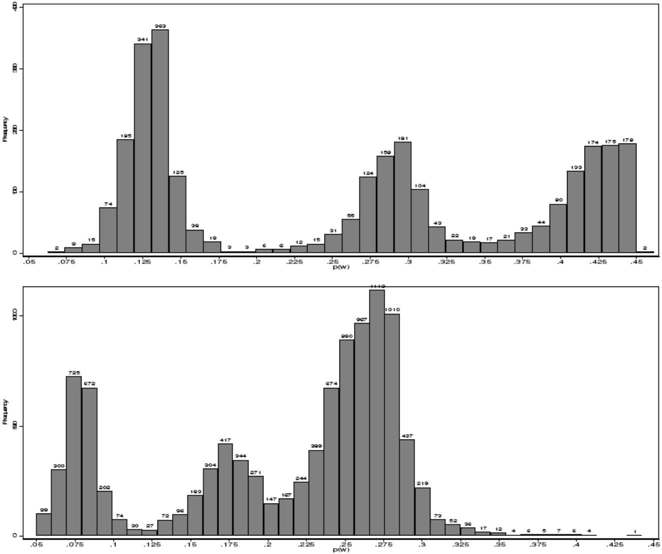
What is Decriminalization Policy?

Decriminalized States	Non-decriminalized states
South Australia 1987	New South Wales
ACT 1992	Queensland
Northern Territory 1996	Victoria
Western Australia 2003	Tasmania

Sources of Data

1. 2001 National Drug Strategy Household survey (NDSHS)
    1. Nationally representative survey of non-institutionalized civilian population aged 14 and above
    2. 14008 resulting samples after delete missing data
    3. Treatment group = 2968, Control group = 11040
2. Australian Illicit Drug Report
3. Australia Bureau of Statistics


0.05 – 0.45
Frequency
p(w)



p(w)
.05 .075 .1 .125 .15 .175 .2 .225 .25 .275 .3 .325 .35 .375 .4 .425 .45
p(w)
.05 .075 .1 .125 .15 .175 .2 .225 .25 .275 .3 .325 .35 .375 .4 .425 .45

Range of Estimated Propensity Score	Number of Treatment observations	Number of Control observations	ATE	ATET
0.05 – 0.45 Length of interval 0.025	2810	10301	0.092*** (0.026)	-0.032 (0.053)
0.05 – 0.45 With STATA interval	2810	10301	0.096*** (0.025)	-0.069* (0.038)
0.075 – 0.425 Length of interval 0.025	2432	9509	0.074*** (0.018)	-0.056** (0.025)
0.075 – 0.425 With STATA interval	2432	9509	0.086*** (0.020)	-0.056** (0.025)
0.05 – 0.4 Length of interval 0.05	2116	10295	0.098*** (0.017)	-0.024† (0.014)
0.05 – 0.4 With STATA interval	2116	10295	0.112*** (0.026)	-0.026** (0.012)
0.1 – 0.35 Length of interval 0.05	1909	8263	0.067*** (0.016)	-0.021† (0.015)
0.1 – 0.35 With STATA interval	1909	8263	0.059*** (0.020)	-0.022† (0.015)

Endogenous Probit Switching Model
(Model i)

$$\begin{aligned} y_{1i}^* &= \alpha_1 + \beta_1 x_i + \varepsilon_{1i} \\ y_{0i}^* &= \alpha_0 + \beta_0 x_i + \varepsilon_{0i} \\ d_i^* &= \gamma_1 x_i + \gamma_2 z_i + \nu_i \end{aligned}$$

$$\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{0i} \\ \nu_i \end{pmatrix} \backsim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{10} & \rho_{1\nu} \\ \rho_{10} & 1 & \rho_{0\nu} \\ \rho_{1\nu} & \rho_{0\nu} & 1 \end{pmatrix} \right)$$

Also, we consider a restricted probit switching model (Model ii) where $\rho_{1\nu} = \rho_{0\nu} \neq 0$

Coefficient Estimates for Marijuana Smoking Equation (Continue)

Variable	Unrestricted Switching: Treatment Model (i)	Unrestricted Switching: Control	Restricted Switching: Treatment Model (ii)	Restricted Switching: Control	Two-Part: Treatment Model (iii)	Two-Part: Control	Bivariate Probit Model (iv)	Binary Probit Model (v)
Working Status	0.307† (0.188)	0.265*** (0.080)	0.318* (0.186)	0.256*** (0.081)	0.306† (0.187)	0.249*** (0.081)	0.257*** (0.075)	0.256*** (0.074)
Aboriginal	-0.308† (0.222)	0.377*** (0.123)	-0.356† (0.219)	0.421*** (0.124)	-0.304† (0.218)	0.438*** (0.124)	0.281*** (0.107)	0.282*** (0.107)
Constant	29.475*** (5.022)	-0.016 (0.475)	31.052*** (4.836)	-0.412 (0.471)	29.338*** (4.832)	-0.650 (0.457)	-0.454 (0.455)	-0.468 (0.429)
$\rho_{1v} = \rho_{0v}$			-0.148* (0.083)				-0.007 (0.077)	
ρ_{1v}	-0.011 (0.109)							
ρ_{0v}		-0.465*** (0.179)						
Log Likelihood	-11928.673		-11931.323		-1183.790	-3997.998	-11969.155	-5218.030
Average Treatment Effect (ATE)	**0.162**		**0.183**		**0.137**		**0.040**	**0.037**

	Non-parametric Approach	Parametric Approach
Advantages	- Do not impose any distributional assumption	-Take account of both selection on observables and unobservables - Can estimate the impact of other explanatory variables on smoking outcome in addition to the effect of decriminalization policy
Disadvantages	- Conditional independence assumption is the maintained hypothesis - Only take account of selection on observables	- Need to impose both functional form and distributional assumptions

One cannot build a castle on the sand. Poor data contribute to misperceptions about the economy. For instance, L. Ivanic, K.J. Fox and W.E. Diewert, "Scanner Data, Time Aggregation and the Construction of Price Indexes" finds that time aggregation choices (the choice of a weekly, monthly, or quarterly unit value concept for prices) have a considerable impact on estimates of price change. When chained indexes are used, the difference in price change estimates can be huge, ranging from -1.42% to -25.78% for a superlative (Fisher) index and an incredible 17.22% to 9.548% for a non-superlative (Laspeyres) index.

Measurement of data guided by economic theory?

Heckman (2001) warns:

"There are, however, some risks in uncritically adopting the methods and the mind set of the statisticians... A theorem proof format is poorly suited for analyzing economic data which requires skills of synthesis, interpretation and empirical imagination. Command of statistical methods is only a part and sometimes a very small part, of what is required to do a first-class empirical research."

Heckman (2001) warns:

"If knowledge transfer from mathematical statistics continues as the mainstream activity of theoretical econometrics, it will increasingly be perceived as irrelevant to economics and empirical work, and will be perceived as a branch of statistics."

I am very impressed by some of the leading theoretical statisticians and econometricians by their mastery of powerful mathematical techniques and their eagerness to apply the methods to important practical problem.