2017 China Economics Prize Award Acceptance Speech

Xiaohong Chen

Cowles Foundation, Yale University

2017 China Economics Prize Ceremony

November 18, 2017, Beijing

Thanks to the National Economics Foundation

- I sincerely thank President Xia and the National Economics Foundation for establishing the China Economics Prize in 2016.
- I commend the Foundation for its noble deeds to "encourage theoretical innovation and economic science prosperity".
- I am deeply grateful to the 2017 Prize Committee for recognizing the works by Professor Gregory Chow and me on econometrics.

Gregory and My Generation of Chinese Economists

- I am extremely honored and humbled to share this year's China Economics Prize with Professor Gregory Chow.
- Gregory is a true giant among all economists, Chinese or not.
- Gregory also played fundamental roles in China economic reforms and modernization of economics education in China.
- Some Chinese economists were "Chow testers"; Many more (including me) were from Gregory's Economics Training program from 1985-96.

- I view my share of the prize as a salute to my generation of Chinese econometricians and Chinese female economists.
- I hope this award inspires young Chinese economists to expand true research frontiers of econometrics.

For My Education, I Am Thankful To

- the end of the Cultural Revolution in 1976; the start of China national college-entry exams in 1977; the start of China economic reform since December 78.

- All my teachers inside and outside China for their love and devotion to their teaching.
  - BS in mathematics from Wuhan University, 82-86.
  - Economics Training Program (initiated by Gregory Chow and President Huang Da of Renmin Univ.), class 86-87.
  - MA in economics from Univ. of Western Ontario, 87-88.
  - PhD in economics from UC San Diego, 88-93.

For My Professional Career, I Am Grateful To

- my **senior mentors**
  - at graduate school: Roger Gordon, Max Stinchcombe, the late Halbert White (my PhD advisor);
  - at Chicago: Lars Peter Hansen, Thomas Sargent (also at NYU);
  - at LSE: Peter Robinson;
  - at Princeton (I was a visitor): Gregory Chow, Bo Honore.
- my **long-term coauthors/friends**: Chunrong Ai, Yanqin Fan, Han Hong, Tong Li, Oliver Linton, Yixiao Sun, Elie Tamer, Zhijie Xiao,...
- my **colleagues/coauthors/friends**: Don Andrews, Richard Blundell, Javier Hidalgo, Cheng Hsiao, Sydney Ludvigson, Rosa Matzkin, Whitney Newey, Jim Powell, Ed Vytlacil, Chenggang Xu,...
- my **students/coauthors**: Tim Christensen, Zhipeng Liao, Demian Pouzo, Alex Torgovitsky, Yanping Yi,...
- my parents, my husband, my sister's and brother's families.

Semi-nonparametric Models and Sieve Methods

- Semi-nonparametric models are families of probability distributions that are indexed by both finite and infinite dimensional unknown parameters of interests.
- The method of sieves estimates semi-nonparametric models by optimizing an empirical criterion over a sequence of approximating parameter spaces (i.e., sieves).
- Two general classes of criteria: all the existing criteria for estimating nonlinear parametric models (e.g., Newey and McFaddern, 94) are valid.
  - M: maximum likelihood (ML), quasi ML, quantile regression, (nonlinear) least square, etc.
  - Minimum Distance (MD), Generalized Method of Moment (GMM), GEL, etc.
- Two general classes of sieves (e.g., Chen, 07)
  - Finite-dimensional linear or nonlinear sieves: Polynomials, Hermite polynomials, Fourier series, Splines, Wavelets, Artificial Neural Networks (**ANN**), Radial-basis, Ridgelets, etc.
  - Infinite-dimensional linear or nonlinear sieves with various penalty constraints.

More on the Method of Sieves

- With different choices of criteria and sieves, the method of sieves is very flexible in estimating complicated semi-nonparametric models, allowing for endogeneity, latent heterogeneity or latent state dependence.
  - Choice of criterion: depends on the model and the parameters of interests. E.g., M criterion for nonparametric regressions without endogeneity, and MD (or GMM) criterion for models with endogeneity
  - Choice of sieves: depends on prior information imposed on the unknown functions by the model, such as dimensions of the potential covariates entering the unknown functions.
- Easier to impose shape (monotonicity, concavity), additivity, non-negativity and other restrictions on unknown functions.
- Easy to compute. Once when the unknown functions are approximated by finite dimensional sieves, the implementation is the same as any parametric nonlinear extremum estimation.

Ex 1a: Shape-invariant System of Endogenous Engel Curves

- Blundell et al. (03): a system of Engel curves satisfying household utility optimization in a given year:

$$Y_{1\ell i} = h_{1\ell}(Y_{2i} - h_0(X_{1i})) + h_{2\ell}(X_{1i}) + \varepsilon_{\ell i}, \quad \ell = 1, ..., N,$$

  $Y_{1\ell i}$ is household i's budget share on goods ℓ, Y_{2i} is i's log-total non-durable expenditure–endogenous in that $E[\varepsilon_{\ell i}|Y_{2i}] \neq 0$.

- Blundell et al. (07): a semi-nonparametric mean IV

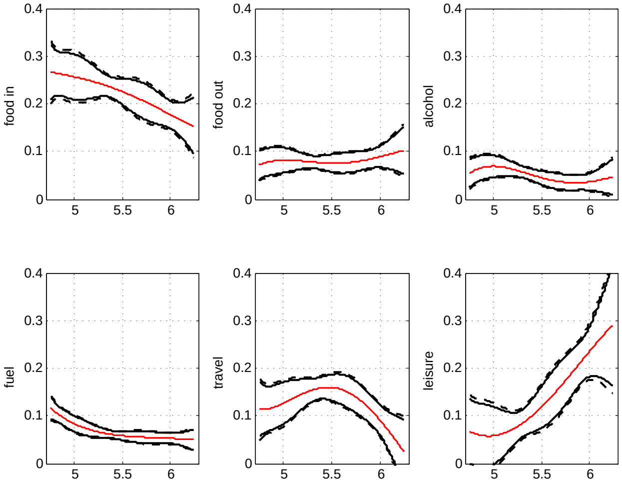
$$E[Y_{1\ell i} - \{h_{1\ell}(Y_{2i} - g(X_{1i}'\beta_1)) + X_{1i}'\beta_{2\ell}\}|X_{1i}, X_{2i}] = 0,$$

- Chen-Pouzo (09): a semi-nonparametric quantile IV

$$E[1\left(Y_{1\ell i} \leq h_{1\ell}(Y_{2i} - g(X_{1i}'\beta_1)) + X_{1i}'\beta_{2\ell}\right)|X_{1i}, X_{2i}] = \gamma \in (0, 1).$$

- Both are estimated via **sieve minimum distance (MD)**.

Figure: Engel Curve Estimation



Estimated Engel curves (red line) with bootstrap uniform confidence bands (solid black lines are 90%, dashed black lines are 95%). The x-axis is log total household expenditure, the y-axis is household budget share. Source: Chen-Christensen.

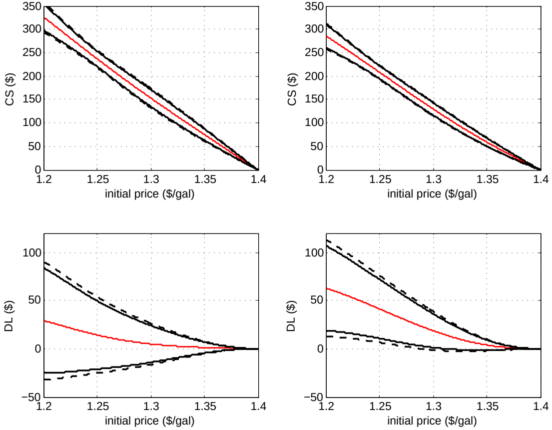
Ex 1b: Welfare Functionals of Endogenous Demand Curves

- ith household quantity demand on gasoline: $Q_i = h_0(P_i, Y_i) + \epsilon_i$, P_i price and Y_i income of ith household. P_i is endogenous. Consumer surplus (CS) and deadweight loss (DL) are welfare measures for impact of price (i.e, tax) changes on consumers. Hausman (81) shows CS from a price change from p^0 to p^1 at income level y, denoted $S_y(p^0)$, solves

$$\frac{\partial S_y(p(u))}{\partial u} = -h_0(p(u), y - S_y(p(u)))\frac{dp(u)}{du}, \quad S_y(p(1)) = 0,$$

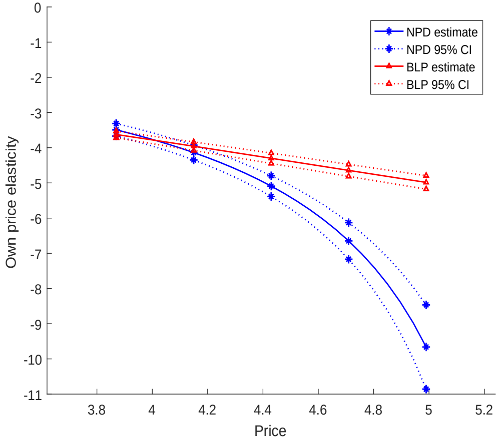
  $p : [0,1] \to R$ is a continuously differentiable path with $p(0) = p^0$ and $p(1) = p^1$.
- Chen-Christensen (forthcoming) estimates the CS and DL welfare impacts of endogenous gasoline price increase via sieve NPIV, using data set of Blundell et al. (12).
- Structural demand in IO: Compiani (yale phD JMP, 17) uses sieve NPIV to estimate endogenous demand nonparametrically, and aims at price elasticity of demand.

Figure: Welfare Impact from Gasoline Price Increase



CS and DL from a price increase to \$1.40/gal (solid red line) and the bootstrap UCBs (solid black lines are 90%, dashed black lines are 95%) when demand is estimated via sieve NPIV. Left panels are for household income of \$72,500; right panels are for household income of \$42,500. Source: Chen-Christensen (forthcoming)

Figure: Organic Strawberries: Own-price Elasticity Function



Own-price elasticity function estimated via BLP vs sieve NPIV. BLP estimates relies on parametric assumptions, while sieve NPIV uses Chen-Christensen for inference. BLP underestimates price elasticity. Source: Compiani (17, yale JMP)

Ex 2a: Semiparametric Asset Pricing Models

- Consumption based asset pricing models:

$$E(M_{t+1}R_{j,t+1} - 1|\mathcal{I}_t) = 0, \quad j = 1, ..., N,$$

$M_{t+1} = \frac{\partial U/\partial C_{t+1}}{\partial U/\partial C_t}$ is IMRS (intertemporal marginal rate of substitution in consumption), or pricing kernel or SDF.

- Hansen-Singleton (82): $U = \sum_{t=0}^{\infty} \delta^t \left[(C_t^{1-\gamma} - 1)/(1-\gamma) \right]$, $M_{t+1} = \delta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma}$. GMM with *unconditional* moment restrictions

$$E\left(\left[\delta \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} R_{j,t+1} - 1 \right] \mathbf{Z}_t \right) = 0, \quad j = 1, ..., N,$$

- Many worry misspecification of time separable utility in consumption.

Ex 2b: Nonlinear Habit-based Asset Pricing Models

- Chen-Ludvigson (09): $U = \sum_{t=0}^{\infty} \delta^t \left[\left(\left(C_t - H_t \right)^{1-\gamma} - 1 \right) / \left(1 - \gamma \right) \right]$, here $H_t = C_t g\left(c_t^* \right)$ is unknown habit level, $0 \leq g < 1$, g nondecreasing in first argument of $c_t^* = \left(\frac{C_{t-1}}{C_t}, \ldots, \frac{C_{t-L}}{C_t} \right)$. $M_{t+1} = \frac{\partial U / \partial C_{t+1}}{\partial U / \partial C_t}$. For external habit, $\partial U / \partial C_t = C_t^{-\gamma} \left(1 - g\left(c_t^* \right) \right)^{-\gamma}$; For internal habit,

$$\frac{\partial U}{\partial C_t} = C_t^{-\gamma} \left[\left(1 - g\left(c_t^* \right) \right)^{-\gamma} - E_t \{ \sum_{j=0}^{L} \delta^j \left(\frac{C_{t+j}}{C_t} \right)^{-\gamma} \left(1 - g(c_{t+j}^*) \right)^{-\gamma} \frac{\partial H_{t+j}}{\partial C_t} \} \right].$$

- welfare implication is different under internal vs external habits. One needs nonlinear habit $g()$ to identify external vs internal habits.

- Chen-Ludvigson (09): **Sieve minimum distance** (SMD) with *conditional* moment restrictions:

$$E\left(M_{t+1}R_{j,t+1}-1\middle|\mathbf{w}_t\right)=0,\ j=1,...,N,\quad \mathbf{w}_t\subset\mathcal{I}_t,$$

  with $\mathbf{w}_t=\left[\widehat{cay}_t, RREL_t, SPEX_t, \frac{C_t}{C_{t-1}}\right]'$ in empirical work.
- **ANN sieve** to approximate unknown habit $g()$ with $L=3,4$ (quarters).

- Using quarterly data, some empirical findings are: (1) estimated *habit is nonlinear*; (2) *internal habit fits data significantly better* than external habit; (3) estimated δ,γ are sensible; (4) estimated habit generated SDF performs well in explaining cross-sectional stock returns; (5) more findings about pricing errors, and model comparison in terms of HJ pricing errors.

Brief Summary of My Past Works on Sieve MD Estimation

For semi-nonparametric conditional moment models: $E[\rho(Y, \beta_0, h_0(\cdot))|X] = 0$, where the unknown functions $h()$ could depend on endogenous variable,

- Ai-Chen (03): propose sieve MD estimation of $(\beta, h(\cdot))$, root-n asymp. normality and efficient estimation, and sieve Wald stat for β.
- Ai-Chen (07, 12): modified sieve MD for root-n normality of β in general misspecified models with different conditioning sets (07), or efficient estimation of β in sequential moment restrictions (12).
- Blundell-Chen-Kristensen (07): nonparametric mean squared error convergence rate of sieve NPIV estimation of $h()$ in system of endogenous Engel curves (Ex 1a).
- Chen-Pouzo (09, 12, 15): Large sample properties (consistency, rate, limiting distribution) of sieve MD, sieve Wald and QLR tests for possibly non-smooth generalized residuals $\rho()$, e.g., quantile IV.
- Chen-Christensen (15): data-driven optimal sup-norm rates and uniform confidence bands of nonlinear welfare functionals of sieve NPIV of h, applications to endogenous demand estimation (Ex 1b).

Ex 3: Duration Model with Unobserved Heterogeneity

- Heckman and Singer (84): data $\{T_i, X_i\}_{i=1}^n$ i.i.d. from

$$p(T|X, \beta_0, h_0) = \int_{\mathcal{U}} g(T|X, u, \beta_0) f_U(u) du,$$

- $g(T|X, u, \beta_0)$: the density of duration T conditional on unobserved heterogeneity U and observed X. E.g., a Weibull density is

$$g(T|X, u, \beta_0) = \beta_{0,1} T^{\beta_{0,1}-1} \exp\left[\beta_{0,2}' X + u - T^{\beta_{0,1}} \exp\left(\beta_{0,2}' X + u\right)\right].$$

- U is indep. of X. Misspecifying density $f_U(u) \equiv h_0^2(u)$ leads to inconsistent estimation of β_0.
- **Semiparametric mixture** models are widely used.

- Let $\alpha_0 = (\beta_0, h_0) \in B \times \mathcal{H}$, which can be estimated by **sieve MLE**:

$$\widehat{\alpha}_n = \arg \max_{\beta \in B,\ h \in \mathcal{H}_n} \sum_{i=1}^{n} \log\{ \int_{\mathcal{U}} g(T_i | X_i, u, \beta) h^2(u) du \}$$

  where $\mathcal{H}_n$ is a sieve space that becomes dense in $\mathcal{H}$ as $n \to \infty$.
- Heckman and Singer (84): first-order spline sieve for $\mathcal{H}_n$; consistency.
- Chen and Liao (14): Hermite polynomial sieve for $\mathcal{H}_n$; sieve LR based inference, with an application to Chinese data.
- Chen, Tamer and Torgovitsky (11): bootstrap sieve LR inference robust to potential loss of point-identification.
- The theory is directly applicable to random coeff., measurement error, etc.

Table: Duration Analysis of the Second Birth in China

AIC	Sieve MLE	Std	Gaussian CIs[5]	Chi-square CIs[5]
log*(Duration)*	3.3621	0.0975	(3.1710 3.5532)	(3.2720 3.4496)
Constant	-5.3175	0.2624	(-5.8317 -4.8032)	(-5.4516 -5.1844)
Gender of 1st kid[2]	0.0584	0.1256	(-0.1877 0.3045)	(-0.1182 0.2322)
Years of schooling	0.0088	0.0181	(-0.0267 0.0442)	(-0.0102 0.0275)
Bonus[3]	-0.2270	0.1331	(-0.4880 0.0339)	(-0.4164 -0.0418)
Household type[4]	0.7900	0.1769	(0.4433 1.1366)	(0.6427 0.9365)
BIC	Sieve MLE	Std	Gaussian CIs	Chi-square CIs
log*(Duration)*	3.2404	0.0718	(3.0090 3.2905)	(3.0642 3.2321)
Constant	-4.9435	0.2284	(-5.3232 -4.4281)	(-5.0064 -4.7476)
Gender of 1st kid	0.0397	0.0840	(-0.0993 0.2298)	(-0.1100 0.2309)
Years of schooling	0.0075	0.0157	(-0.0253 0.0364)	(-0.0129 0.0235)
Bonus	-0.2836	0.1118	(-0.4816 -0.0434)	(-0.4459 -0.0818)
Household type	0.7326	0.1591	(0.4291 1.0526)	(0.5974 0.8824)

1. Sample size n=694; 2. gender dummy variable equals 1 if the 1st kid is a girl and 0 otherwise; 3. bonus dummy variable equals 1 if there are subsidies awarded to the household accepting the one child policy; 4. household type is 1 if it is registered in rural area and 0 otherwise; 5. confidence intervals are constructed under 0.95. Source: Chen-Liao (14)

Ex 4: Semi-nonparametric GARCH + Residual Copula

- Geanakoplos (10): bad news is accompanied by increased uncertainty (volatility). "News impact curve".
- Engle (10): "risk assessment" is also important in understanding the financial crisis.
- To capture both news impact curve and tail joint dependence, consider a semi-nonparametric GARCH + residual copula model.
- We use daily data from the last 4 years to address both "news impact curve" and risk assessment" based on 3 series: mortgage-backed security (MBS), stock, and bond market returns.

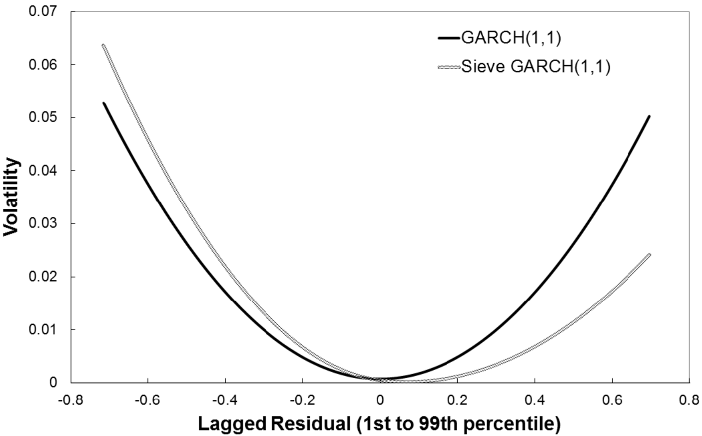
Chen-Fan (06) SCOMDY model: Excess returns on Barclays MBS index (S_t^e), excess market (daily Fama-French factor) returns (M_t^e), and excess returns on the Barclays bond index (B_t^e):

$$
\begin{aligned}
MBS\ \ Market &: \quad S_t^e = c_S + \rho_S S_{t-1}^e + \beta_S M_{t-1}^e + \sigma_{S,t}\varepsilon_{S,t} \\
Stock\ \ Market &: \quad M_t^e = c_M + \rho_M M_{t-1}^e + \sigma_{M,t}\varepsilon_{M,t} \\
Bonds\ \ Market &: \quad B_t^e = c_B + \rho_M B_{t-1}^e + \beta_B M_{t-1}^e + \sigma_{B,t}\varepsilon_{B,t} \\
Volatility &: \quad \sigma_{i,t}^2 = \omega_i + \theta_i \sigma_{i,t-1}^2 + h_i\left(\sigma_{i,t-1}\varepsilon_{i,t-1}\right), \ \ i \in \{S, M, B\},
\end{aligned}
$$

$E\left(\varepsilon_{i,t}\right) = 0$, $E\left(\varepsilon_{i,t}^2\right) = 1$ for $i \in \{S, M, B\}$. $\left(\varepsilon_{S,t}, \varepsilon_{M,t}, \varepsilon_{B,t}\right)'$ are indep. in time but jointly dist. as $F(\varepsilon) = C(F_1(\varepsilon_1), F_2(\varepsilon_2), F_3(\varepsilon_3); \Sigma, v)$, where $C(\cdot) : [0,1]^3 \to [0,1]$ is a Student's t-copula with unknown parameter (Σ, v), and unknown marginals $F_i(\cdot)$, $i \in \{S, M, B\}$.

- All three estimated “news impact curves” exhibit the same asymmetry: bad news increases volatility more than does good news. For mortgage-backed securities and stocks, some good news actually decreases volatility, as in Fostel and Geanakoplos (10). As in Linton and Mammen (05), most good news in the stock market does not have much effect on volatility.
- We find (i) shocks to bonds and shocks to mortgage-backed securities are highly correlated, (ii) shocks to mortgage-backed securities and shocks to stocks are moderately negatively correlated, and (iii) shocks to bonds and shocks to stocks are also moderately negatively correlated.
- With estimated semi-nonparametric GARCH and residual copula dependence parameters, we can easily calculate VaR for a portfolio comprised of mortgage-backed securities, stocks, and bonds.

Figure: MBS: News Impact Curve



For negative shocks Sieve-GARCH(1,1) predicts more volatility than standard GARCH(1,1) does, and for positive shocks Sieve-GARCH predicts less volatility than GARCH does. Source: Chen (2013)

Brief Summary of My Past Works on Sieve M Procedure

For sieve M estimation and inference,

- Chen-Shen (98): rates of convergence and limiting dists of general sieve M estimation for semi-nonparametric time series models. It contains ANN nonlinear sieve as an example.
- Chen-White (99): ANN nonlinear sieve estimation of conditional density, conditional quantile of functions of high-dimensional covariates.
- Chen-Liao-Sun (14): limiting dist. of possibly slower-than-root-n estimable functionals of linear sieve M-estimator, sieve Wald tests for weakly dependent time series models.
- Chen-Tamer-Torgovitsky (11): bootstrapping sieve LR inference for potentially partially identified semi-nonparametric likelihood models.
- Chen-Liao (15): sieve semiparametric two-step GMM for weakly dependent data, where the first step could be linear sieve M or sieve MD estimation.

Outlook: Discover More Unknowns

- Inference results such as uniform confidence bands or sieve Wald and QLR tests based on *nonlinear sieves* such as ANNs (or "deep learners") for slower-than-root-n functionals (such as unknown functions) are still missing.
- Need more research on simulation based sieve methods for semi-nonparametric dynamic models with nonlinear non-Gaussian latent structures (e.g., DSGE).
- Big Data era: more complex economic models, more possibilities for policy counterfactuals, more issues with: endogeneity, latent heterogeneity, latent dynamic state dependence, nonstationarity, measurement error, variables (or model) selections.
- Methods of penalized sieves are still applicable, but we need to study more carefully the interplay of stochastic optimization algorithms and econometric properties, and modify the existing modeling and inference theories.

THANK YOU!