Partial Mean Processes with Generated Regressors:
Continuous Treatment Effects and Nonseparable Models

Ying-Ying Lee
University of Wisconsin-Madison

November, 2012
Partial Mean Process

\[
\{ \ t \rightarrow E\left[ F_{Y|TV}(y|t, V) \right] : y \in \mathcal{Y} \ \}\n\]

- \( E\left[ F_{Y|TV}(y|t, V) \right] = E\left[ E\left[ 1\{Y \leq y\} \mid T = t, V \right] \right] \): CDF of outcome for a fixed value of the endogenous/treatment variables \( T = t \)
  - \( E[E[Y|T = t, V]] \) partial mean in Newey (1994a)

- Distributional features by Hadamard-differentiable functionals: mean, quantiles, Gini coefficient, Lorenz curve, etc.

- Conditional Independence Assumption \( T \perp Y(t)|V(S), \forall t \)
  
  Matzkin (2007), “A control function is a function of observable variables \( V(S) \) such that conditioning on its value purges any statistical dependence that may exist between the observable \( T \) and unobservable explanatory variables in an original model.”
Partial Mean Process

\[
\left\{ t \rightarrow E \left[ F_{Y|TV}(y|t, V) \right] : y \in \mathcal{Y} \right\}
\]

- \( E \left[ F_{Y|TV}(y|t, V) \right] = E \left[ E[\mathbf{1}_{Y \leq y}|T = t, V] \right] \): CDF of outcome for a fixed value of the endogenous/treatment variables \( T = t \)
  - \( E[E[Y|T = t, V]] \) partial mean in Newey (1994a)

- Distributional features by Hadamard-differentiable functionals: mean, quantiles, Gini coefficient, Lorenz curve, etc.

- Conditional Independence Assumption \( T \perp Y(t)|V(S), \forall t \)

Matzkin (2007), “A control function is a function of observable variables \( V(S) \) such that conditioning on its value purges any statistical dependence that may exist between the observable \( T \) and unobservable explanatory variables in an original model.”
Partial Mean Process

\[
\left\{ \ t \to E\left[ F_{Y|TV}(y|t, V) \right] : y \in \mathcal{Y} \ \right\}
\]

- \( E\left[ F_{Y|TV}(y|t, V) \right] = E\left[ E[1\{Y \leq y\}|T = t, V] \right] \): CDF of outcome for a fixed value of the endogenous/treatment variables \( T = t \)
  - \( E[E[Y|T = t, V]] \) partial mean in Newey (1994a)

- Distributional features by Hadamard-differentiable functionals: mean, quantiles, Gini coefficient, Lorenz curve, etc.

- Conditional Independence Assumption \( T \perp Y(t)|V(S), \forall t \)
  
  Matzkin (2007), "A control function is a function of observable variables \( V(S) \) such that conditioning on its value purges any statistical dependence that may exist between the observable \( T \) and unobservable explanatory variables in an original model."
Continuous Treatment Effects

▶ Example: Program Evaluation

\[ Y_i = Y_i(T_i) : \text{Outcome variable (Income)} \]
\[ T_i : \text{Treatment (Length of exposure)} \]

\[ \{ Y_i(t) \}_{t \in T} : \text{Potential outcome} \]
(Potential income corresponding to length in the program \( t \))

▶ \( Y_i(t) \) is latent if \( T_i \neq t \)

Nonseparable models

▶ Example: Demand Analysis (Engel curve)

\[ Y_i = \phi(T_i, \epsilon_i) : \text{Share of expenditure on food} \]
\[ T_i : \text{log of total expenditure} \]

For any \( t \in T \), \( F_{Y(t)}(y) = E[1_{\{Y(t) \leq y\}}] \), a process indexed by \( y \).
Continuous Treatment Effects

▶ Example: Program Evaluation

\[ Y_i = Y_i(T_i): \text{Outcome variable (Income)} \]

\[ T_i: \text{Treatment (Length of exposure)} \]

\( \{Y_i(t)\}_{t \in T}: \text{Potential outcome} \)

(Potential income corresponding to length in the program \( t \))

▶ \( Y_i(t) \) is latent if \( T_i \neq t \)

Nonseparable models

▶ Example: Demand Analysis (Engel curve)

\[ Y_i = \phi(T_i, \epsilon_i): \text{Share of expenditure on food} \]

\[ T_i: \text{log of total expenditure} \]

For any \( t \in T \), \( F_Y(t)(y) = E[1_{\{Y(t) \leq y\}}], \) a process indexed by \( y \).
Distributional features $\Gamma(F_{Y(t)})$ by Hadamard-differentiable functionals

- Mean $E[Y(t)]$
  - Average structural function: Blundell & Powell (2000)
  - $\frac{\partial}{\partial s} E[Y(s)|T = t]|_{s=t}$: treatment effect on the treated in Florens, Heckman, Meghir, & Vytlacil (2008) or local average response in Altonji & Matzkin (2005)

- $F^{-1}_{Y(t)}(\tau)$ Quantile structural function Imbens & Newey (2009)

- Inequality measures: Gini coefficient, Lorenz curve
Distributional features $\Gamma(F_{Y(t)})$ by Hadamard-differentiable functionals

- **Mean $E[Y(t)]$**
  - Average structural function: Blundell & Powell (2000)
  - $\frac{\partial}{\partial s} E[Y(s)|T=t]_{s=t}$: treatment effect on the treated in Florens, Heckman, Meghir, & Vytlacil (2008) or local average response in Altonji & Matzkin (2005)

- $F_{Y(t)}^{-1}(\tau)$ Quantile structural function Imbens & Newey (2009)

- Inequality measures: Gini coefficient, Lorenz curve
Distributional features $\Gamma(F_{Y(t)})$ by Hadamard-differentiable functionals

- **Mean $E[Y(t)]$**
  - Average structural function: Blundell & Powell (2000)
  - $\frac{\partial}{\partial s} E[Y(s)|T = t]|_{s=t}$: treatment effect on the treated in Florens, Heckman, Meghir, & Vytlacil (2008) or local average response in Altonji & Matzkin (2005)

- $F_{Y(t)}^{-1}(\tau)$ Quantile structural function Imbens & Newey (2009)

- Inequality measures: Gini coefficient, Lorenz curve
Distributional features $\Gamma(F_{Y(t)})$ by Hadamard-differentiable functionals

- **Mean** $E[Y(t)]$
  - **Average structural function**: Blundell & Powell (2000)
  - $\frac{\partial}{\partial s} E[Y(s)|T = t]|_{s=t}$: treatment effect on the treated in Florens, Heckman, Meghir, & Vytlacil (2008) or local average response in Altonji & Matzkin (2005)

- $F_{Y(t)}^{-1}(\tau)$ Quantile structural function Imbens & Newey (2009)

- **Inequality measures**: Gini coefficient, Lorenz curve
Nonparametric Estimation

$$E\left[E\left[1_{\{Y \leq y\}} \mid T = t, V(S)\right]\right]$$

Step 1 (Estimate $V(S)$) \[ \sup_s |\hat{V}(s) - V(s)| = o_p(n^{-\delta}) \]

Step 2 (Regression) Nonparametric regression on $\hat{V}(S)$

$$\hat{F}_{Y|T\hat{V}}(y \mid t, v) = \frac{1}{n} \sum_{j=1}^{n} 1_{\{Y_j \leq y\}} K_h(T_j - t) K_h(\hat{V}(S_j) - v) \frac{\hat{f}_{T\hat{V}}(t, v)}{\hat{f}_{T\hat{V}}(t, v)}$$

Step 3 (Partial Sum) Fixing $T$ at $t$,

$$\frac{1}{n} \sum_{i=1}^{n} \hat{F}_{Y|T\hat{V}}(y \mid t, \hat{V}(S_i))$$

Roles of the Generated Regressor $\hat{V}(S_i)$

1. Arguments for the outer expectation - Step 3
2. Regressors for the conditional regression - Step 2
Nonparametric Estimation

\[
E \left[ E \left[ \mathbf{1}_{Y \leq y} \mid T = t, V(S) \right] \right]
\]

**Step 1** (Estimate \( V(S) \)) \( \sup_s |\hat{V}(s) - V(s)| = o_p(n^{-\delta}) \)

**Step 2** (Regression) Nonparametric regression on \( \hat{V}(S) \)

\[
\hat{F}_{Y \mid T \hat{V}}(y \mid t, \nu) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}_{Y_j \leq y} K_h(T_j - t) K_h(\hat{V}(S_j) - \nu) \hat{f}_{T \hat{V}}(t, \nu)
\]

**Step 3** (Partial Sum) Fixing \( T \) at \( t \),

\[
\frac{1}{n} \sum_{i=1}^{n} \hat{F}_{Y \mid T \hat{V}}(y \mid t, \hat{V}(S_i))
\]

Roles of the Generated Regressor \( \hat{V}(S_i) \)

1. **Arguments** for the outer expectation - **Step 3**
2. **Regressor**s for the conditional regression - **Step 2**
$E \left[ 1_{\{Y \leq y\}} \left| T = t, V(S) \right. \right]$ Nonparametric Regression with Generated Regressors

- Mammen, Rothe, & Schienle (2012a, 2012b): $E[Y|V(S) = v]$
- Escanciano, Jacho-Chavez, & Lewbel (2012):
  \[
  \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left( Y_i - \hat{E}[Y|V(S) = v_i] \right) W(X_i) \hat{t}(V, v_i)
  \]
- Song (2008): $E[\alpha(Y)|V(S) = v]$
- Hahn & Ridder (2012): $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{E}[Y|V(S) = \hat{V}(S_i)]$
Theorem - Stochastic Expansion (Theorem 4.2)

\[
\sqrt{nh} \left( \frac{1}{n} \sum_{i=1}^{n} \hat{F}_{Y|TV}(y|t, \hat{V}(S_i)) \right) - E\left[ F_{Y|TV}(y|t, V) \right] = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{Y}(y) + \sqrt{nh} E\left[ (\hat{V}(S) - V(S))' A(y, S) \right] + \sqrt{nh}R_n
\]

uniformly in \( y \in \mathcal{Y} \), where the influence function for the true regressor \( V(S) \)

\[
\psi_{Y}(y) \equiv \left( 1_{\{Y_i \leq y\}} - F_{Y|TV}(y|t, V(S_i)) \right) \cdot \frac{1}{f_{T|V}(t|V(S_i))} \frac{1}{\sqrt{h}} K\left( \frac{T_i - t}{h} \right)
\]

\[
A(y, S) \equiv \nabla_V F_{Y|TV}(y|t, V(S)) + \frac{f_{T|S}(t|S)}{f_{T|V}(t|V(S))} \left( - \nabla_V F_{Y|TV}(y|t, V(S)) \right)
\]

\[
+ \frac{\nabla_V f_{T|V}(t|V(S))}{f_{T|V}(t|V(S))} \left( F_{Y|TV}(y|t, V(S)) - F_{Y|TS}(y|t, S) \right)
\]
Example 1 Unconfoundedness: $T \perp Y(t) | X$

- $E[Y(t)] = E[E[Y|T = t, X]]$

Example 2 Generalized Propensity Score $T \perp Y(t) | f_{T|X}(t|X)$

- $F_{Y(t)}(y) = E\left[ E\left[ 1_{\{Y \leq y\}} \bigg| T = t, f_{T|X}(t|X) \right] \right] \quad \text{Hirano & Imbens (2004)}$
Example 2: Generalized Propensity Score

If \( \hat{V}(X) = \hat{f}_{T|X}(t|X) \) uses the same kernel and bandwidth as the 2nd-stage regression, then

\[
E \left[ \left( \hat{f}_{T|X}(t|X) - f_{T|X}(t|X) \right)^\prime A(y, S) \right] = O((nh)^{-1/2})
\]

Corollary 4.2 (Weak convergence)

\[
\sqrt{nh} \left( \frac{1}{n} \sum_{i=1}^{n} \hat{F}_{Y|T\hat{V}}(\cdot|t, \hat{V}(X_i)) - F_{Y(t)}(\cdot) \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{ti}(\cdot) + o_p(1) \Rightarrow X^X(\cdot)
\]

a Gaussian process with the covariance \( Cov^X(y_1, y_2) \)

\[
= E \left[ \left( F_{Y|T X}(\min\{y_1, y_2\}|t, X) - F_{Y|T X}(y_1|t, X)F_{Y|T X}(y_2|t, X) \right) \frac{1}{f_{T|X}(t|X)} \right] \int K^2(v)dv
\]

\[\blacksquare\]
Example 2: Generalized Propensity Score (Remark 4.3)

1. Regression on the nonparametric estimated GPS is first-order asymptotically equivalent to regressing on $X$.
   ▶ no efficiency gain in using the GPS

2. The estimator based on the regression on the true GPS is less efficient than the estimator using the nonparametrically estimated GPS or $X$.
   ▶ knowledge of the GPS $f_{T|X}$, sample-specific information is lost for true GPS

Example 3: Control Variables

Nonseparable model \( Y = \phi(T, X, \epsilon) \)

\[ T = g(Z, e) \quad \text{The instruments } Z \perp (\epsilon, e) \]

1. Imbens & Newey (2009) \( V(S) = V(T, Z) = F_{T|Z}(T|Z) \)

2. Newey, Powell, & Vella (1999) \( T = g(Z) + e, \text{ where } E[e|Z] = 0 \)

\[ V(S) = V(T, Z) = T - E[T|Z] \]

\[ \Rightarrow E[(\hat{V}(S) - V(S))'A(y, S)] = o_p(n^{-1/2}) \quad \text{first-order ignorable} \]

Corollary 4.1 (Weak convergence)

\[ \sqrt{n}h \left( \frac{1}{n} \sum_{i=1}^{n} \hat{F}_{Y|T}(\cdot|t, \hat{V}(S_i)) - F_{Y(t)}(\cdot) \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_t V(\cdot) + o_p(1) \Rightarrow \mathcal{X}^V(\cdot) \]

a Gaussian process with the covariance \( \text{Cov}^V(y_1, y_2) \)

\[ = E \left[ \left( F_{Y|TV}(\min\{y_1, y_2\}|t, V) - F_{Y|TV}(y_1|t, V)F_{Y|TV}(y_2|t, V) \right) \frac{1}{f_{T|V}(t|V)} \right] \int K^2(v)dv \]
Inference by a Multiplier Method

Theorem - Multiplier CLT (Theorem 5.2)

1. Estimate the influence function

\[
\hat{\psi}_{ti}(y) \equiv \left(1_{\{Y_i \leq y\}} - \hat{F}_{Y|TV}(y|t, V_i)\right) \cdot \frac{1}{\sqrt{h}} K\left(\frac{T_i - t}{h}\right) \frac{1}{\hat{f}_{T|V}(t|V_i)}
\]

2. Draw \( U_i \) be i.i.d. \( \mathcal{N}(0, 1) \), independent of the data.

\[
\mathcal{X}_M(\cdot) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i \cdot \hat{\psi}_{ti}(\cdot) \Rightarrow \mathcal{X}(\cdot)
\]

conditional on sample path with probability approaching 1.

\[\blacksquare\]

▶ Donald, Hsu, & Barrett (2012) for the conditional CDF process
Functional Delta Method (Theorem 5.1)

Let $\Gamma$ be a **Hadamard-differentiable** functional with derivative $\Gamma'$.

$$\sqrt{nh} \left( \Gamma(\hat{F}_{Y(t)}) - \Gamma(F_{Y(t)}) \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Gamma'(\psi_{ti}) + o_p(1) \Rightarrow \Gamma'(X)$$

a Gaussian process with mean zero and covariance defined by the limit of the second moment of $\Gamma'(\psi_{ti})$.

\[\square\]

- Inference by the Multiplier method:

$$\Gamma'(X_M) \equiv \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i \cdot \Gamma'(\hat{\psi}_{ti}) \Rightarrow \Gamma'(X)$$

conditional on sample path with probability approaching 1.
Counterfactual Distribution $W(X)$

\[
\left\{ \begin{array}{c}
t \to E \left[ E \left[ \mathbf{1}_{Y \leq y} \right| T = t, V(S) \right] \cdot W(S) \right] : y \in \mathcal{Y} \\
\end{array} \right.
\]

Consider a counterfactual distribution $F^*_X$.

\[
F^*_Y(t)(y) = E^* \left[ \mathbf{1}_{Y(t) \leq y} \right] = \int F_{Y|TX}(y|t, X) dF^*_X(x)
\]

\[
= \int F_{Y|TX}(y|t, X) dF^*_X(x) = E \left[ F_{Y|TX}(y|t, X) \cdot W(X) \right], \quad W(X) = \frac{f^*_X(X)}{f_X(X)}
\]

- Overall CDF $F_Y(t)(y)$: $F^*_X = F_X$
- CDF on the treated $F_{Y(t)|T}(y|\bar{t}) = E \left[ \mathbf{1}_{Y(t) \leq y} \right| T = \bar{t}]$:

\[
F^*_X = F_{X|T=\bar{t}} \quad \text{and} \quad W(X) = f_{T|X}(\bar{t}|X)/f_X(X)
\]

*DiNardo, Fortin, and Lemieux (1996) and Chernozhukov et. al. (2012)*
Conclusions

\[
\left\{ \begin{array}{c}
t \to F_{Y(t)}(y) = E \left[ E \left[ \mathbf{1}_{Y \leq y} \mid T = t, V(S) \right] \cdot W(S) \right] : y \in \mathcal{Y} \\
\end{array} \right.
\]

- Limit theory for nonparametrically estimating a **partial mean process** with generated regressors
  1. Control function in triangular simultaneous equations models
  2. Generalized Propensity Score \( f_{T \mid X}(t \mid X) \)

- **Distributional** impacts of **continuous** treatments

- Application:
  - Continuous Treatment Effect: program evaluation of a Conditional Cash Transfer program in Colombia (joint with Juan Villa)
  - Nonseparable model: Engel curve

*Thank you.*