A Framework for the Analysis of Dynamic Treatment Effects: Grade Retention and Test Scores

J. Cooley S. Navarro Y. Takahashi

University of Wisconsin-Madison

North American Summer Meeting of the Econometric Society

06/06/09
The decision to retain students is inherently dynamic and the treatment effect may vary over time.

Existing literature treats the problem as a binary static problem.

You learn to use a hammer: everything looks like a nail.

Look at the treatment literature in general: static (mostly binary) models

Look at the problems they (we) study: mostly dynamic
The decision to retain students is inherently dynamic and the treatment effect may vary over time.

- Existing literature treats the problem as a binary static problem.
- You learn to use a hammer: everything looks like a nail.
- Look at the treatment literature in general: static (mostly binary) models
- Look at the problems they (we) study: mostly dynamic
The decision to retain students is inherently dynamic and the treatment effect may vary over time.

- Existing literature treats the problem as a binary static problem.
- You learn to use a hammer: everything looks like a nail.
- Look at the treatment literature in general: static (mostly binary) models
- Look at the problems they (we) study: mostly dynamic
The decision to retain students is inherently dynamic and the treatment effect may vary over time.

- Existing literature treats the problem as a binary static problem.
- You learn to use a hammer: everything looks like a nail.
- Look at the treatment literature in general: **static (mostly binary) models**
- Look at the problems they (we) study: **mostly dynamic**
The decision to retain students is inherently dynamic and the treatment effect may vary over time.

- Existing literature treats the problem as a binary static problem.
- You learn to use a hammer: everything looks like a nail.
- Look at the treatment literature in general: static (mostly binary) models
- Look at the problems they (we) study: mostly dynamic
Contributions

1. Methodological Contributions:
   - Framework for analysis of **dynamic treatment effects**:
     - “Effect” varies based on the time at which treatment is received
     - Account for dynamic selection into treatment
   - Identification strategy that can be applied broadly and recovers missing counterfactuals that cannot be recovered with standard (and not so standard) methods including Randomization

2. Policy Contributions:
   - Does effect of retention vary across grades and over time?
   - How do effects vary by student types? (not restricted to focus on marginal types)
   - Do certain policies do a better job of retaining the “right” students?
Contributions

1. Methodological Contributions:
   - Framework for analysis of **dynamic treatment effects**:
     - “Effect” varies based on the time at which treatment is received
     - Account for dynamic selection into treatment
   - Identification strategy that can be applied broadly and recovers missing counterfactuals that cannot be recovered with standard (and not so standard) methods including Randomization

2. Policy Contributions:
   - Does effect of retention vary across grades and over time?
   - How do effects vary by student types? (not restricted to focus on marginal types)
   - Do certain policies do a better job of retaining the “right” students?
Contributions

1. Methodological Contributions:
   - Framework for analysis of dynamic treatment effects:
     - “Effect” varies based on the time at which treatment is received
     - Account for dynamic selection into treatment
   - Identification strategy that can be applied broadly and recovers missing counterfactuals that cannot be recovered with standard (and not so standard) methods including Randomization

2. Policy Contributions:
   - Does effect of retention vary across grades and over time?
   - How do effects vary by student types? (not restricted to focus on marginal types)
   - Do certain policies do a better job of retaining the “right” students?
1 Methodological Contributions:
   - Framework for analysis of **dynamic treatment effects**:
     - “Effect” varies based on the time at which treatment is received
     - Account for dynamic selection into treatment
   - Identification strategy that can be applied broadly and recovers missing counterfactuals that cannot be recovered with standard (and not so standard) methods including Randomization

2 Policy Contributions:
   - Does effect of retention vary across grades and over time?
   - How do effects vary by student types? (not restricted to focus on marginal types)
   - Do certain policies do a better job of retaining the “right” students?
Contributions

1** Methodological Contributions:**

- Framework for analysis of **dynamic treatment effects**:
  - “Effect” varies based on the time at which treatment is received
  - Account for dynamic selection into treatment

- Identification strategy that can be applied broadly and recovers missing counterfactuals that cannot be recovered with standard (and not so standard) methods including Randomization

2** Policy Contributions:**

- Does effect of retention vary across grades and over time?
- How do effects vary by student types? (not restricted to focus on marginal types)
- Do certain policies do a better job of retaining the “right” students?
Contributions

1. **Methodological Contributions:**
   - Framework for analysis of dynamic treatment effects:
     - “Effect” varies based on the time at which treatment is received
     - Account for dynamic selection into treatment
   - Identification strategy that can be applied broadly and recovers missing counterfactuals that cannot be recovered with standard (and not so standard) methods including Randomization

2. **Policy Contributions:**
   - Does effect of retention vary across grades and over time?
   - How do effects vary by student types? (not restricted to focus on marginal types)
   - Do certain policies do a better job of retaining the “right” students?
Methodological Contributions:

- Framework for analysis of **dynamic treatment effects**:
  - “Effect” varies based on the time at which treatment is received
  - Account for dynamic selection into treatment

- Identification strategy that can be applied broadly and recovers missing counterfactuals that cannot be recovered with standard (and not so standard) methods including Randomization

Policy Contributions:

- Does effect of retention vary across grades and over time?
- How do effects vary by student types? (not restricted to focus on marginal types)
- Do certain policies do a better job of retaining the “right” students?
Contributions

1. Methodological Contributions:
   - Framework for analysis of **dynamic treatment effects**:
     - “Effect” varies based on the time at which treatment is received
     - Account for dynamic selection into treatment
   - Identification strategy that can be applied broadly and recovers missing counterfactuals that cannot be recovered with standard (and not so standard) methods including Randomization

2. Policy Contributions:
   - Does effect of retention vary across grades and over time?
   - How do effects vary by student types? (not restricted to focus on marginal types)
   - Do certain policies do a better job of retaining the “right” students?
Contributions

1. Methodological Contributions:
   - Framework for analysis of **dynamic treatment effects**:
     - “Effect” varies based on the time at which treatment is received
     - Account for dynamic selection into treatment
   - Identification strategy that can be applied broadly and recovers missing counterfactuals that cannot be recovered with standard (and not so standard) methods including *Randomization*

2. Policy Contributions:
   - Does effect of retention vary across grades and over time?
   - How do effects vary by student types? (not restricted to focus on marginal types)
   - Do certain policies do a better job of retaining the “right” students?
Framework (1)

- $P \in \mathcal{P} = \{1, 2, ..., \bar{P}\}$ – index calendar time (period)
- $i = 1, ..., I$ – index the individual.
- $T$ – random variable that indicates the time at which treatment is received
  - $T = 0$ for the “never” treated state
  - $T \in \mathcal{T} \subseteq \mathcal{P}$

- Outcome of interest (e.g., test scores):
  \[
  Y_i(P, T) = \Phi(P, T) + \epsilon_i(P, T)
  \]

- Impose (Abbring and Van den Berg, 2003)
  \[
  Y_i(P, T) = Y_i(P, 0) = Y_i(P) \implies \Phi(P, T) = 0 \text{ and } \epsilon_i(P, T) = \epsilon_i(P) \text{ for } T \geq P
  \]
Framework (1)

- $P \in \mathcal{P} = \{1, 2, ..., \bar{P}\}$ – index calendar time (period)
- $i = 1, ..., l$ – index the individual.
- $T$ – random variable that indicates the time at which treatment is received
  - $T = 0$ for the “never” treated state
  - $T \in \mathcal{T} \subseteq \mathcal{P}$
- Outcome of interest (e.g., test scores):
  \[
  Y_i(P, T) = \Phi(P, T) + \epsilon_i(P, T)
  \]
  - Impose (Abbring and Van den Berg, 2003)
    \[
    Y_i(P, T) = Y_i(P, 0) = Y_i(P) \implies \Phi(P, T) = 0 \quad \text{and} \quad \epsilon_i(P, T) = \epsilon_i(P) \quad \text{for} \quad T \geq P
    \]
Framework (2)

- Outcome of interest (e.g., test scores):
  \[ Y_i(P, T) = \Phi(P, T) + \epsilon_i(P, T) \]

- Determining Treatment:
  \[ V_i(T) = \lambda(T) + U_i(T) \]
  \[ D_i(T) = 1 \left( V_i(T) > 0 \mid \{ V_i(\tau) < 0 \}_{\tau=1}^{T-1} \right) \]
  \[ = 1 \left( V_i(T) > 0 \mid \{ D_i(\tau) = 0 \}_{\tau=1}^{T-1} \right) \]

- Observed outcome at \( P \):
  \[ Y_i(P) = \Phi(P, 0) + \epsilon(P, 0) + \sum_{\tau=1}^{\min\{P, \bar{T}\}} D_i(\tau) (\Phi(P, \tau) - \Phi(P, 0)). \]
  \[ + \sum_{\tau=1}^{\min\{P, \bar{T}\}} D_i(\tau) (\epsilon_i(P, \tau) - \epsilon(P, 0)) \]
Framework (2)

- **Outcome of interest (e.g., test scores):**

\[ Y_i(P, T) = \Phi(P, T) + \epsilon_i(P, T) \]

- **Determining Treatment:**

\[ V_i(T) = \lambda(T) + U_i(T) \]

\[ D_i(T) = 1 \left( V_i(T) > 0 \mid \{V_i(\tau) < 0\}_{\tau=1}^{T-1} \right) \]

\[ = 1 \left( V_i(T) > 0 \mid \{D_i(\tau) = 0\}_{\tau=1}^{T-1} \right) \]

- **Observed outcome at \( P \):**

\[ Y_i(P) = \Phi(P, 0) + \epsilon(P, 0) + \sum_{\tau=1}^{\min\{P, \bar{T} \}} D_i(\tau) (\Phi(P, \tau) - \Phi(P, 0)) + \sum_{\tau=1}^{\min\{P, \bar{T} \}} D_i(\tau) (\epsilon_i(P, \tau) - \epsilon(P, 0)) \]
Framework (2)

- **Outcome of interest (e.g., test scores):**
  \[ Y_i(P, T) = \Phi(P, T) + \epsilon_i(P, T) \]

- **Determining Treatment:**
  \[ V_i(T) = \lambda(T) + U_i(T) \]

  \[ D_i(T) = 1 \left( V_i(T) > 0 \mid \{ V_i(\tau) < 0 \}_{\tau=1}^{T-1} \right) \]

  \[ = 1 \left( V_i(T) > 0 \mid \{ D_i(\tau) = 0 \}_{\tau=1}^{T-1} \right) \]

- **Observed outcome at** \( P \):
  \[ Y_i(P) = \Phi(P, 0) + \varepsilon(P, 0) + \sum_{\tau=1}^{\min\{P, \bar{T}\}} D_i(\tau) (\Phi(P, \tau) - \Phi(P, 0)). \]

  \[ + \sum_{\tau=1}^{\min\{P, \bar{T}\}} D_i(\tau) (\epsilon_i(P, \tau) - \epsilon(P, 0)) \]
Defining Treatment Effects

- Many possible individual effects. Even more population average effects.

- On average do students perform better if retained in Kindergarten vs. 1st grade?

\[
ATE(P, t, t') = E(Y(P, t) - Y(P, t'))
\]

- Do students who are retained in Kindergarten perform better on average the next year (or 3 years later, etc..) than if they had not been retained?

- For students who are retained in 1st grade, would they have been better off if retained in Kindergarten instead?

\[
TT(P, t, t', t'') = E(Y(P, t) - Y(P, t') | T = t'')
\]
Defining Treatment Effects

- Many possible individual effects. Even more population average effects.
- On average do students perform better if retained in Kindergarten vs. 1st grade?

\[ ATE(P, t, t') = E(Y(P, t) - Y(P, t')) \]

- Do students who are retained in Kindergarten perform better on average the next year (or 3 years later, etc..) than if they had not been retained?
- For students who are retained in 1st grade, would they have been better off if retained in Kindergarten instead?

\[ TT(P, t, t', t'') = E(Y(P, t) - Y(P, t') | T = t'') \]
Defining Treatment Effects

- Many possible individual effects. Even more population average effects.

- On average do students perform better if retained in Kindergarten vs. 1st grade?

  \[ ATE (P, t, t') = E (Y (P, t) - Y (P, t')) \]

- Do students who are retained in Kindergarten perform better on average the next year (or 3 years later, etc.) than if they had not been retained?

- For students who are retained in 1st grade, would they have been better off if retained in Kindergarten instead?

  \[ TT (P, t, t', t'') = E (Y (P, t) - Y (P, t') | T = t'') \]
Consider a 3 period model where treatment can be received at periods 1 or 2 (i.e., $T = 0, 1, 2$).

Essential Heterogeneity: $D_i(T)$ correlated with $\epsilon_i(3,T) - \epsilon_i(3,T')$ (e.g., less "able" students are more likely to be retained and also experience smaller gains in the absence of retention)

\[
Y_i(3) = \Phi(3,T) + \epsilon_i(3,T)
\]

\[
Y_i(3,T) = \Phi(3,T) + \epsilon_i(3,T)
\]
Identification Problem

- Consider a 3 period model where treatment can be received at periods 1 or 2 (i.e., $T = 0, 1, 2$)

- Outcome in period 3:

$$Y_i(3, T) = \Phi(3, T) + \epsilon_i(3, T)$$

- Observed outcome:

$$Y_i(3) = \Phi(3, 0) + \epsilon_i(3, 0) + D_i(1) [\Phi(3, 1) - \Phi(3, 0)] + D_i(1) [\epsilon_i(3, 1) - \epsilon_i(3, 0)] + D_i(2) [\Phi(3, 2) - \Phi(3, 0)] + D_i(2) [\epsilon_i(3, 2) - \epsilon_i(3, 0)].$$

- **Essential Heterogeneity:** $D_i(T)$ correlated with $\epsilon_i(3, T) - \epsilon_i(3, T')$ (e.g., less “able” students are more likely to be retained and also experience smaller gains in the absence of retention)
Consider a 3 period model where treatment can be received at periods 1 or 2 (i.e., $T = 0, 1, 2$)

Outcome in period 3:

$$Y_i(3, T) = \Phi(3, T) + \epsilon_i(3, T)$$

Observed outcome:

$$Y_i(3) = \Phi(3, 0) + \epsilon_i(3, 0)$$

$$+ D_i(1)[\Phi(3, 1) - \Phi(3, 0)] + D_i(1)[\epsilon_i(3, 1) - \epsilon_i(3, 0)]$$

$$+ D_i(2)[\Phi(3, 2) - \Phi(3, 0)] + D_i(2)[\epsilon_i(3, 2) - \epsilon_i(3, 0)].$$

**Essential Heterogeneity:** $D_i(T)$ correlated with $\epsilon_i(3, T) - \epsilon_i(3, T')$ (e.g., less “able” students are more likely to be retained and also experience smaller gains in the absence of retention)
Identification Problem

- Consider a 3 period model where treatment can be received at periods 1 or 2 (i.e., $T = 0, 1, 2$)

- Outcome in period 3:
  \[ Y_i(3, T) = \Phi(3, T) + \epsilon_i(3, T) \]

- Observed outcome:
  \[
  Y_i(3) = \Phi(3, 0) + \epsilon_i(3, 0) \\
  + D_i(1)[\Phi(3, 1) - \Phi(3, 0)] + D_i(1)[\epsilon_i(3, 1) - \epsilon_i(3, 0)] \\
  + D_i(2)[\Phi(3, 2) - \Phi(3, 0)] + D_i(2)[\epsilon_i(3, 2) - \epsilon_i(3, 0)].
  \]

- **Essential Heterogeneity**: $D_i(T)$ correlated with $\epsilon_i(3, T) - \epsilon_i(3, T')$ (e.g., less “able” students are more likely to be retained and also experience smaller gains in the absence of retention)
Experimental Data

- What if we could run the experiment we wanted to account for selection?

- Suppose our outcome of interest is period 3 test scores.

- Students can be retained in $K$ or $1^{st}$ grade.

- Show you that we will lose information by doing randomization

- Paper: problems and potential solutions for instrumental variables, LATE, RD, control functions.
Experimental Data

- What if we could run the experiment we wanted to account for selection?

- Suppose our outcome of interest is period 3 test scores.

- Students can be retained in $K$ or $1^{st}$ grade.

- Show you that we will loose information by doing randomization

- Paper: problems and potential solutions for instrumental variables, LATE, RD, control functions.
What if we could run the experiment we wanted to account for selection?

Suppose our outcome of interest is period 3 test scores.

Students can be retained in K or 1st grade.

Show you that we will loose information by doing randomization

Paper: problems and potential solutions for instrumental variables, LATE, RD, control functions.
Experimental Data

- What if we could run the experiment we wanted to account for selection?

- Suppose our outcome of interest is period 3 test scores.

- Students can be retained in K or 1st grade.

- Show you that we will **loose information** by doing randomization

- Paper: problems and potential solutions for instrumental variables, LATE, RD, control functions.
What if we could run the experiment we wanted to account for selection?

Suppose our outcome of interest is period 3 test scores.

Students can be retained in $K$ or 1$^{st}$ grade.

Show you that we will lose information by doing randomization.

Paper: problems and potential solutions for instrumental variables, LATE, RD, control functions.
Experimental Data (2)

Table 1: Experimental Solution

<table>
<thead>
<tr>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>V(1)&lt;0</td>
<td>R=0</td>
<td>E(Y(3,0)</td>
</tr>
<tr>
<td>R=1</td>
<td>E(Y(3,1)</td>
<td>T=0)</td>
</tr>
<tr>
<td>R=2</td>
<td>E(Y(3,2)</td>
<td>T=0)</td>
</tr>
<tr>
<td>V(2)&lt;0</td>
<td>R=0</td>
<td>E(Y(3,0)</td>
</tr>
<tr>
<td>R=1</td>
<td>E(Y(3,1)</td>
<td>T=2)</td>
</tr>
<tr>
<td>R=2</td>
<td>E(Y(3,2)</td>
<td>T=2)</td>
</tr>
<tr>
<td>V(2)&gt;0</td>
<td>R=0</td>
<td>E(Y(3,0)</td>
</tr>
<tr>
<td>R=1</td>
<td>E(Y(3,1)</td>
<td>T=1)</td>
</tr>
<tr>
<td>R=2</td>
<td>E(Y(3,2)</td>
<td>T=1)</td>
</tr>
</tbody>
</table>

V(1)>0

R=0
R=1
R=2

Table 1: Experimental Solution
Factor Structure

- Assume residual in selection equation and treatment equation include vector of mutually independent factors $\theta_i$:
  \[
  \epsilon_i(P, T) = \theta_i \alpha(P, T) + \epsilon_i(P) \\
  U_i(T) = \theta_i \rho(T) + \nu_i(T)
  \]

- Assume $\epsilon_i(P) \perp \perp \epsilon_i(P')$ for all $P \neq P'$ and $\nu_i(T) \perp \perp \nu_i(T') \perp \perp \epsilon_i(P)$ for all $T \neq T'$

- Dimension reduction technique – only need to recover loadings $\alpha(P, T), \rho(T)$ and marginals of $\theta, \epsilon, \nu$ rather than all possible joints.

- More importantly: aids in interpretation of results (low-dimensional common set of cases: Goldberger).
Factor Structure

- Assume residual in selection equation and treatment equation include vector of mutually independent factors $\theta_i$:

$$\epsilon_i(P, T) = \theta_i \alpha(P, T) + \epsilon_i(P)$$

$$U_i(T) = \theta_i \rho(T) + \nu_i(T)$$

- Assume $\epsilon_i(P) \perp \perp \epsilon_i(P')$ for all $P \neq P'$ and $\nu_i(T) \perp \perp \nu_i(T') \perp \perp \epsilon_i(P)$ for all $T \neq T'$

- Dimension reduction technique – only need to recover loadings $\alpha(P, T), \rho(T)$ and marginals of $\theta, \epsilon, \nu$ rather than all possible joints.

- More importantly: aids in interpretation of results (low-dimensional common set of cases: Goldberger).
Assume residual in selection equation and treatment equation include vector of mutually independent factors $\theta_i$:

$$
\begin{align*}
\epsilon_i(P, T) &= \theta_i \alpha(P, T) + \epsilon_i(P) \\
U_i(T) &= \theta_i \rho(T) + \nu_i(T)
\end{align*}
$$

Assume $\epsilon_i(P) \perp \perp \epsilon_i(P')$ for all $P \neq P'$ and $\nu_i(T) \perp \perp \nu_i(T') \perp \perp \epsilon_i(P)$ for all $T \neq T'$.

Dimension reduction technique – only need to recover loadings $\alpha(P, T), \rho(T)$ and marginals of $\theta, \epsilon, \nu$ rather than all possible joints.

More importantly: aids in interpretation of results (low-dimensional common set of cases: Goldberger).
Factor Structure (2)

- Back to our 3 period example. Period 3 outcome:

\[ Y(3) = \Phi(3, 0) + \theta_i \alpha(3, 0) + \varepsilon(3) \]

\[ + D_i(1) [\Phi(3, 1) - \Phi(3, 0)] + D_i(2) [\Phi(3, 2) - \Phi(3, 0)] \]

\[ + D_i(1) \theta_i [\alpha(3, 1) - \alpha(3, 0)] + D_i(2) \theta_i [\alpha(3, 2) - \alpha(3, 0)] \]

- Selection equation:

\[ V_i(T) = \lambda(T) + \theta_i \rho(T) + \nu_i(T) \]

- Essential heterogeneity: \( \alpha(3, T) \neq \alpha(3, 0) \)

- Solution: Recover and control for \( \theta_i \)
Back to our 3 period example. Period 3 outcome:

\[ Y(3) = \Phi(3, 0) + \theta_i \alpha(3, 0) + \varepsilon(3) \]

\[ + D_i(1) [\Phi(3, 1) - \Phi(3, 0)] + D_i(2) [\Phi(3, 2) - \Phi(3, 0)] \]

\[ + D_i(1) \theta_i [\alpha(3, 1) - \alpha(3, 0)] + D_i(2) \theta_i [\alpha(3, 2) - \alpha(3, 0)] \]

Selection equation:

\[ V_i(T) = \lambda(T) + \theta_i \rho(T) + \upsilon_i(T) \]

Essential heterogeneity: \( \alpha(3, T) \neq \alpha(3, 0) \)

Solution: Recover and control for \( \theta_i \)
Factor Structure (2)

- Back to our 3 period example. Period 3 outcome:

\[ Y(3) = \Phi(3, 0) + \theta_i \alpha(3, 0) + \varepsilon(3) \]

\[ + D_i(1) [\Phi(3, 1) - \Phi(3, 0)] + D_i(2) [\Phi(3, 2) - \Phi(3, 0)] \]

\[ + D_i(1) \theta_i [\alpha(3, 1) - \alpha(3, 0)] + D_i(2) \theta_i [\alpha(3, 2) - \alpha(3, 0)] \]

- Selection equation:

\[ V_i(T) = \lambda(T) + \theta_i \rho(T) + \nu_i(T) \]

- Essential heterogeneity: \( \alpha(3, T) \neq \alpha(3, 0) \)

- Solution: Recover and control for \( \theta_i \)
The ABC of Ability

- Ability is multidimensional. 3 types of ability: \( \theta_i = (A_i, B_i, C_i) \).

- \( j^{th} \) demeaned cognitive test (\( N_c \geq 2 \)):
  \[ \zeta_{i,j} = A_i \alpha_{\zeta,j} + C_i \pi_{\zeta,j} + \varepsilon_{\zeta,j}. \]

- \( j^{th} \) demeaned behavioral test (\( N_b \geq 2 \)):
  \[ \beta_{i,j} = A_i \alpha_{\beta,j} + B_i \phi_{\beta,j} + \varepsilon_{\beta,j}. \]

- **Loadings on general ability** (\( A_i \)) recovered from cross moments *between* cognitive and behavioral measures.

- **Loadings on cognitive ability** (\( C_i \)) from cross-moments *within* cognitive markers. Similarly for **loadings on behavioral ability** from behavioral measures.

- Also, allow for **correlated (permanent) shocks** by taking cross-moments over time.
The ABC of Ability

- Ability is multidimensional. 3 types of ability: \( \theta_i = (A_i, B_i, C_i) \).

- \( j^{th} \) demeaned cognitive test (\( N_c \geq 2 \)):
  \[
  \zeta_{i,j} = A_i \alpha \zeta_{i,j} + C_i \pi \zeta_{i,j} + \varepsilon_{\zeta,j}.
  \]

- \( j^{th} \) demeaned behavioral test (\( N_b \geq 2 \)):
  \[
  \beta_{i,j} = A_i \alpha \beta_{i,j} + B_i \phi \beta_{i,j} + \varepsilon_{\beta,j}.
  \]

- **Loadings on general ability** (\( A_i \)) recovered from cross moments *between* cognitive and behavioral measures.

- **Loadings on cognitive ability** (\( C_i \)) from cross-moments *within* cognitive markers. Similarly for **loadings on behavioral ability** from behavioral measures.

- Also, allow for **correlated (permanent) shocks** by taking cross-moments over time.
The ABC of Ability

- Ability is multidimensional. 3 types of ability: 
  \[ \theta_i = (A_i, B_i, C_i). \]
- \( j^{th} \) demeaned cognitive test \( (N_c \geq 2) \):
  \[ \zeta_{i,j} = A_i \alpha_{\zeta,j} + C_i \pi_{\zeta,j} + \varepsilon_{\zeta,j}. \]
- \( j^{th} \) demeaned behavioral test \( (N_b \geq 2) \):
  \[ \beta_{i,j} = A_i \alpha_{\beta,j} + B_i \phi_{\beta,j} + \varepsilon_{\beta,j}. \]

- **Loadings on general ability** \( (A_i) \) recovered from cross moments *between* cognitive and behavioral measures.
- **Loadings on cognitive ability** \( (C_i) \) from cross-moments *within* cognitive markers. Similarly for **loadings on behavioral ability** from behavioral measures.
- Also, allow for **correlated (permanent) shocks** by taking cross-moments over time.
The ABC of Ability

- Ability is multidimensional. 3 types of ability: $\theta_i = (A_i, B_i, C_i)$.

- $j^{th}$ demeaned cognitive test ($N_c \geq 2$):
  \[ \zeta_{i,j} = A_i \alpha_{\zeta,j} + C_i \pi_{\zeta,j} + \varepsilon_{\zeta,j}. \]

- $j^{th}$ demeaned behavioral test ($N_b \geq 2$):
  \[ \beta_{i,j} = A_i \alpha_{\beta,j} + B_i \phi_{\beta,j} + \varepsilon_{\beta,j}. \]

- **Loadings on general ability** ($A_i$) recovered from cross moments *between* cognitive and behavioral measures.

- **Loadings on cognitive ability** ($C_i$) from cross-moments *within* cognitive markers. Similarly for **loadings on behavioral ability** from behavioral measures.

- Also, allow for **correlated (permanent) shocks** by taking cross-moments over time.
Data

- Early Childhood Longitudinal Study of Kindergartners
  - Follows students from K to 5th grade (in absence of retention)
  - Retention definitions:
    - Kindergarten retention
    - Early retention—retained in 1st or 2nd grade
    - Late retention—in 3rd or 4th grade
  - Cognitive tests—IRT scores for reading, math, “general” or science tests
  - Behavioral tests—Social Rating Scale (SRS) evaluated by teachers (approaches to learning, self control and interpersonal skills components)
  - Include rich set of controls
Data

- Early Childhood Longitudinal Study of Kindergartners
  - Follows students from K to 5th grade (in absence of retention)
  - Retention definitions:
    - Kindergarten retention
    - Early retention—retained in 1st or 2nd grade
    - Late retention—in 3rd or 4th grade
  - Cognitive tests—IRT scores for reading, math, “general” or science tests
  - Behavioral tests—Social Rating Scale (SRS) evaluated by teachers (approaches to learning, self control and interpersonal skills components)
  - Include rich set of controls
Data

- Early Childhood Longitudinal Study of Kindergartners
- Follows students from K to 5th grade (in absence of retention)

- Retention definitions:
  - Kindergarten retention
  - Early retention—retained in 1st or 2nd grade
  - Late retention—in 3rd or 4th grade

- Cognitive tests—IRT scores for reading, math, “general” or science tests

- Behavioral tests—Social Rating Scale (SRS) evaluated by teachers (approaches to learning, self control and interpersonal skills components)

- Include rich set of controls
Data

- Early Childhood Longitudinal Study of Kindergartners
- Follows students from K to 5th grade (in absence of retention)
- Retention definitions:
  - Kindergarten retention
  - Early retention—retained in 1st or 2nd grade
  - Late retention—in 3rd or 4th grade
- Cognitive tests—IRT scores for reading, math, “general” or science tests
- Behavioral tests—Social Rating Scale (SRS) evaluated by teachers (approaches to learning, self control and interpersonal skills components)
- Include rich set of controls
Early Childhood Longitudinal Study of Kindergartners

Follows students from K to 5th grade (in absence of retention)


Retention definitions:
- Kindergarten retention
- Early retention—retained in 1st or 2nd grade
- Late retention—in 3rd or 4th grade

Cognitive tests—IRT scores for reading, math, “general” or science tests

Behavioral tests—Social Rating Scale (SRS) evaluated by teachers (approaches to learning, self control and interpersonal skills components)

Include rich set of controls
Data

- Early Childhood Longitudinal Study of Kindergartners
  - Follows students from K to 5th grade (in absence of retention)
- Retention definitions:
  - Kindergarten retention
  - Early retention—retained in 1st or 2nd grade
  - Late retention—in 3rd or 4th grade
- Cognitive tests—IRT scores for reading, math, “general” or science tests
- Behavioral tests—Social Rating Scale (SRS) evaluated by teachers (approaches to learning, self control and interpersonal skills components)
- Include rich set of controls
Data

- Early Childhood Longitudinal Study of Kindergartners
- Follows students from K to 5th grade (in absence of retention)

- Retention definitions:
  - Kindergarten retention
  - Early retention—retained in 1st or 2nd grade
  - Late retention—in 3rd or 4th grade

- Cognitive tests—IRT scores for reading, math, “general” or science tests

- Behavioral tests—Social Rating Scale (SRS) evaluated by teachers (approaches to learning, self control and interpersonal skills components)

- Include rich set of controls
Evaluating Grade Retention

- What is effect of grade retention?

- Impossible to hold both age and grade fixed

- Let $S$ denote test scores, $R$ retention

\[ S = \text{Age} \gamma_a + \text{Grade} \gamma_g + R \gamma_R + \varepsilon. \]

- Hold grade fixed $\Rightarrow \gamma_a + \gamma_R$

- Hold age fixed $\Rightarrow \gamma_g + \gamma_R$

- Estimate the effect holding age fixed because we use IRT scores
Evaluating Grade Retention

- What is effect of grade retention?

- Impossible to hold both age and grade fixed

- Let $S$ denote test scores, $R$ retention

  $$S = \text{Age}\gamma_a + \text{Grade}\gamma_g + R\gamma_R + \varepsilon.$$  

- Hold grade fixed $\Rightarrow \gamma_a + \gamma_R$

- Hold age fixed $\Rightarrow \gamma_g + \gamma_R$

- Estimate the effect holding age fixed because we use IRT scores
Evaluating Grade Retention

- What is effect of grade retention?
- Impossible to hold both age and grade fixed
- Let $S$ denote test scores, $R$ retention

$$S = \text{Age} \gamma_a + \text{Grade} \gamma_g + R \gamma_R + \varepsilon.$$ 

- Hold grade fixed $\Rightarrow \gamma_a + \gamma_R$
- Hold age fixed $\Rightarrow \gamma_g + \gamma_R$
- Estimate the effect holding age fixed because we use IRT scores
Evaluating Grade Retention

- What is effect of grade retention?
- Impossible to hold both age and grade fixed
- Let \( S \) denote test scores, \( R \) retention
  \[
  S = \text{Age} \gamma_a + \text{Grade} \gamma_g + R \gamma_R + \varepsilon.
  \]
- Hold grade fixed \( \Rightarrow \gamma_a + \gamma_R \)
- Hold age fixed \( \Rightarrow \gamma_g + \gamma_R \)
- Estimate the effect holding age fixed because we use IRT scores
Evaluating Grade Retention

- What is effect of grade retention?

- Impossible to hold both age and grade fixed

- Let $S$ denote test scores, $R$ retention
  
  $$S = \text{Age} \gamma_a + \text{Grade} \gamma_g + R \gamma_R + \varepsilon.$$

- Hold grade fixed $\Rightarrow \gamma_a + \gamma_R$

- Hold age fixed $\Rightarrow \gamma_g + \gamma_R$

- Estimate the effect holding age fixed because we use IRT scores
Evaluating Grade Retention

- What is effect of grade retention?
- Impossible to hold both age and grade fixed
- Let $S$ denote test scores, $R$ retention
  \[ S = \text{Age} \gamma_a + \text{Grade} \gamma_g + R \gamma_R + \varepsilon. \]
- Hold grade fixed $\Rightarrow \gamma_a + \gamma_R$
- Hold age fixed $\Rightarrow \gamma_g + \gamma_R$
- Estimate the effect holding age fixed because we use IRT scores
Baseline Estimates: Math

Table 3: Evidence for Dynamic Selection and Treatment Effect (Math Score)

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Kindergarten Math Score*</th>
<th>Math Score for 2003-04 School Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retained in Kindergarten</td>
<td>-0.19*</td>
<td>-0.18*</td>
</tr>
<tr>
<td>Retained Early (1st or 2nd grade)</td>
<td>-0.29*</td>
<td>-0.24*</td>
</tr>
<tr>
<td>Retained Late (3rd or 4th grade)</td>
<td>-0.20*</td>
<td>-0.18*</td>
</tr>
<tr>
<td>Child's Characteristics</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Family Characteristics</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>School Characteristics</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Age and Age Squared</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Kindergarten Cognitive Tests</td>
<td>--</td>
<td>No</td>
</tr>
<tr>
<td>Kindergarten Behavioral Measures</td>
<td>--</td>
<td>No</td>
</tr>
<tr>
<td>No. of Observations</td>
<td>7771</td>
<td>3964</td>
</tr>
</tbody>
</table>

| P-value for KI = EA = LA*               | 0.000                    | 0.058                            |
| P-value for KI = EA                    | 0.000                    | 0.062                            |
| P-value for EA = LA                    | 0.002                    | 0.039                            |
| P-value for KI = LA                    | 0.730                    | 0.945                            |
| R squared                              | 0.354                    | 0.316                            |

* Statistically significant at 5% level

# 1998-99 School Year

Note: If the p value is small compared to the critical value, we reject the hypothesis of equality of coefficients. P values less than 0.05 are colored with yellow. Yes/No indicates if each group of variables is included as controls.
Estimation

- Use school retention policies as instruments in the sense that affect retention decision but do not affect outcomes:
  - School Policies: (1) allows children to be retained in any grade, (2) to be retained because of immaturity, (3) to be retained at the parents request and (4) to be retained without parental authorization

- Restrict $X'$s to have same marginal effect over time

- Estimate via maximum likelihood allowing all distributions to be mixtures of normals (show it is needed)
## Model Fit: Tests

### Table 5: Predicted and Actual Means and Standard Deviations of Kindergarten (1998-99 School Year) Test Scores/Measures

<table>
<thead>
<tr>
<th>Test / Measure</th>
<th>Predicted Mean</th>
<th>Predicted Standard Deviation</th>
<th>Actual Mean</th>
<th>Actual Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Test</td>
<td>3.067</td>
<td>0.352</td>
<td>3.085</td>
<td>0.347</td>
</tr>
<tr>
<td>Reading Test</td>
<td>3.357</td>
<td>0.291</td>
<td>3.362</td>
<td>0.281</td>
</tr>
<tr>
<td>Math Test</td>
<td>3.093</td>
<td>0.367</td>
<td>3.097</td>
<td>0.361</td>
</tr>
<tr>
<td>Approach to Learning</td>
<td>-0.019</td>
<td>0.998</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Self-Control</td>
<td>-0.006</td>
<td>1.007</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Interpersonal Skills</td>
<td>-0.005</td>
<td>1.012</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Note: Behavioral measures are standardized to have mean zero and variance equal to one. These figures are calculated from 500000 simulations based on the estimated model.
### Table 6: Predicted and Actual Retention Probabilities (Conditional on Survival)*

<table>
<thead>
<tr>
<th></th>
<th>Data</th>
<th>Model Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retained in Kindergarten</td>
<td>5.06%</td>
<td>5.14%</td>
</tr>
<tr>
<td>Retained Early (1st or 2nd grade)</td>
<td>3.64%</td>
<td>3.62%</td>
</tr>
<tr>
<td>Retained Late (3rd or 4th grade)</td>
<td>1.07%</td>
<td>1.03%</td>
</tr>
</tbody>
</table>

*Note: The table calculates the probability of retention at t, conditional on not having been retained before t.*
Let $f(A)$ denote the probability density function of general ability. We assume that $f(A)$ is a mixture of normals. Let $T=0,1,2,3$ denote retention status: not retained, retained in kindergarten, retained early (1 or 2 grade) and retained late (3 or 4). The graph shows $f(A|T=t)$ for each retention status.

Figure 1
Density of General Ability Conditional on Retention Status
Let \( f(B) \) denote the probability density function of behavioral ability. We assume that \( f(B) \) is a mixture of normals. Let \( T = 0, 1, 2, 3 \) denote retention status: not retained, retained in kindergarten, retained early (1 or 2 grade) and retained late (3 or 4). The graph shows \( f(B | T = t) \) for each retention status.
Let $f(C)$ denote the probability density function of cognitive ability. We assume that $f(C)$ is a mixture of normals. Let $T=0,1,2,3$ denote retention status: not retained, retained in kindergarten, retained early (1 or 2 grade), and retained late (3 or 4). The graph shows $f(C|T=t)$ for each retention status.
## Average Reading Test Scores 2003/04

### Table 9: Average Reading Test Scores by Potential and Actual Retention Status: 2003-04 School Year

<table>
<thead>
<tr>
<th>Potential Retention Status</th>
<th>Actual Retention Status</th>
<th>A kid who is actually (i.e. conditional on retention status being:)</th>
<th>Unconditional</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not Retained</td>
<td>Retained in Kindergarten</td>
<td>Retained Early</td>
</tr>
<tr>
<td>Not Retained</td>
<td></td>
<td>4.966</td>
<td>4.829</td>
</tr>
<tr>
<td>Retained in Kindergarten</td>
<td></td>
<td>5.052</td>
<td>4.839</td>
</tr>
<tr>
<td>Retained Early</td>
<td></td>
<td>4.991</td>
<td>4.788</td>
</tr>
<tr>
<td>Retained Late</td>
<td></td>
<td>4.941</td>
<td>4.838</td>
</tr>
</tbody>
</table>

**Factual Results**

Note: Let $T = 0, 1, 2, 3$ represent the actual retention status of a kid: never retained, retained in kindergarten, retained early (at grade 1 or 2), or retained late (at grade 3 or 4), respectively. Let $S(i)$ be the potential test score at 2003-04 school year if the kid were retained at time $i=0, 1, 2, 3$. The row $i$, column $j$ element of this table calculates $E[S(i) | D=j]$. For example, a kid who was actually not retained would get 4.991 on average if the kid were retained at 1 or 2 grade instead. When calculating them, we keep kid's age fixed at 11.
### Table 11: Average Reading Test Score Gain by Retention Status: 2003-04 School Year

| Average Gain                           | A kid who is actually
<table>
<thead>
<tr>
<th></th>
<th>(i.e. conditional on the retention status being:)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not Retained</td>
</tr>
<tr>
<td>Retained in Kindergarten vs Not Retained</td>
<td>0.085 (0.0038)</td>
</tr>
<tr>
<td>Retained Early vs Not Retained</td>
<td>0.024 (0.0023)</td>
</tr>
<tr>
<td>Retained Late vs Not Retained</td>
<td>-0.025 (0.0023)</td>
</tr>
</tbody>
</table>

Figures in parentheses are standard errors which were obtained via 200 bootstrap replications.

Note: Let $T = 0, 1, 2$, or 3 represent the actual retention status of a kid: never retained, retained in kindergarten, retained early (at grade 1 or 2), or retained late (at grade 3 or 4), respectively. Let $S(i)$ be the potential test score if the kid were retained at time $i=0, 1, 2, 3$. The row $i$, column $j$ element of this table calculates $E[S(i) - S(0) | D=j]$. For example, the test score of a kid who was actually not retained would increase by 0.024 if he were retained at 1 or 2 grade instead. When calculating these figures, we keep kid’s age fixed at 11.
Let $S(t,1)$ and $S(t,0)$ be the potential test scores at time $t$ if the kid were retained in kindergarten and if the kid were not retained, respectively. The graph shows $E[S(t,1)-S(t,0)]$ for $t=1,2,$ and $3$ for each test score.
Let $S(t,1)$ and $S(t,0)$ be the potential test scores at time $t$ if the kid were retained in kindergarten and if the kid were not retained, respectively. The graph shows $E[S(t,1)-S(t,0)]$ for $t=1,2,$ and 3 for each test score.
Treatment Effect by Type: Reading

ATE Over Achievement Quantiles in 2003/04

- Average Gain in Reading Score
- Lagged Achievement at Each Quantile

Retained in K vs. not retained: ●
Retained early vs. not retained: ▲
Retained late vs. not retained: ▼

-1 -0.2 -0.1 0 0.1 0.2
4.2 4.4 4.6 4.8 5 5.2

Dynamic Treatment Effects
J. Cooley, S. Navarro, Y. Takahashi

Motivation
The Framework
Identification
Data
Estimation
Results
Model Fit
Selection on Unobservables
Treatment Effects
Heterogeneity of Effects
Conclusion
Treatment Effect by Ability Type: Reading

ATE Over Achievement Quantiles in 2003/04

- Average Gain in Reading Score
- Avg. General Ability at Each Quantile
- Retained in K vs. not retained
- Retained early vs. not retained
- Retained late vs. not retained

ATE Over Achievement Quantiles in 2003/04

- Average Gain in Reading Score
- Avg. Cognitive Ability at Each Quantile
- Retained in K vs. not retained
- Retained early vs. not retained
- Retained late vs. not retained
Different student types are selected into retention at different grades (e.g., early retainees have lower cognitive and general ability)

Treatment effect of retention varies substantially by student types and over time

Focusing on 2003/04 (period 4 outcomes)

- Average treatment effect of retention is positive
- Treatment effect of early retention on early retainees is negative, whereas TT for retention in K or late on students is 0
- High “ability” students benefit from retention by period 4, low “ability” students have 0 to negative effects

Treatment effect becomes more positive as time since retention passes
Conclusion

- Different student types are selected into retention at different grades (e.g., early retainees have lower cognitive and general ability)

- Treatment effect of retention varies substantially by student types and over time

- Focusing on 2003/04 (period 4 outcomes)
  - Average treatment effect of retention is positive
  - Treatment effect of early retention on early retainees is negative, whereas TT for retention in K or late on students is 0
  - High “ability” students benefit from retention by period 4, low “ability” students have 0 to negative effects

- Treatment effect becomes more positive as time since retention passes
Conclusion

- Different student types are selected into retention at different grades (e.g., early retainees have lower cognitive and general ability)

- Treatment effect of retention varies substantially by student types and over time

- Focusing on 2003/04 (period 4 outcomes)
  - Average treatment effect of retention is positive
    - Treatment effect of early retention on early retainees is negative, whereas TT for retention in K or late on students is 0
    - High “ability” students benefit from retention by period 4, low “ability” students have 0 to negative effects
  - Treatment effect becomes more positive as time since retention passes
Consequence

- Different student types are selected into retention at different grades (e.g., early retainees have lower cognitive and general ability).

- Treatment effect of retention varies substantially by student types and over time.

- Focusing on 2003/04 (period 4 outcomes)
  - Average treatment effect of retention is positive.
  - Treatment effect of early retention on early retainees is negative, whereas TT for retention in K or late on students is 0.
  - High “ability” students benefit from retention by period 4, low “ability” students have 0 to negative effects.
  - Treatment effect becomes more positive as time since retention passes.
Conclusion

- Different student types are selected into retention at different grades (e.g., early retainees have lower cognitive and general ability)

- Treatment effect of retention varies substantially by student types and over time

- Focusing on 2003/04 (period 4 outcomes)
  - Average treatment effect of retention is positive
  - Treatment effect of early retention on early retainees is negative, whereas TT for retention in K or late on students is 0
  - High “ability” students benefit from retention by period 4, low “ability” students have 0 to negative effects

- Treatment effect becomes more positive as time since retention passes
Different student types are selected into retention at different grades (e.g., early retainees have lower cognitive and general ability)

Treatment effect of retention varies substantially by student types and over time

Focusing on 2003/04 (period 4 outcomes)
- Average treatment effect of retention is positive
- Treatment effect of early retention on early retainees is negative, whereas TT for retention in K or late on students is 0
- High “ability” students benefit from retention by period 4, low “ability” students have 0 to negative effects

Treatment effect becomes more positive as time since retention passes