41903: Group-Based Inference
Clustering and Fama-MacBeth

Notes 4
Many real-world data sets plausibly exhibit complicated dependence structures.

Consider “aggregate” panel data model:

\[ y_{it} = x_{it}' \beta + \alpha_i + \delta_t + \epsilon_{it} \]

- \( i \) represents something like firm, state, country
- seems likely that \( \epsilon_{it} \) will have a complicated structure
- e.g. consider a productivity shock to IBM today
  - likely related to future (and past) and past productivity shocks at IBM
  - plausibly spills over to related firms (e.g. Google) today
  - plausibly spills over to related firms in the future (and related to them in the past)
- Likely that \( E[\epsilon_{it}\epsilon_{jt}] \neq 0 \) even for \( i \neq j \) and \( s \neq t \)
- Failing to account for this dependence will lead to invalid inference for parameters of interest
Group-based inference approaches popular methods to estimating standard errors robust to non-identically distributed and dependent data.

Group-based inference procedures partition data into approximately uncorrelated groups

- clustered standard errors
  - really popularized in panel data application by Bertrand, Duflo, and Mullainathan (2004)
- Fama-MacBeth (1973)
  - popular in finance but much more generally applicable

Both approaches partition data into \( g = 1, \ldots, G \) groups

- \( I_g \) denotes the indices of observations in group \( g \)
- \( N_g \) observations per group
Illustrate mechanics of clustering and Fama-MacBeth in linear model:

\[ y_i = x_i' \beta + \varepsilon_i; \quad E[x_i \varepsilon_i] = 0 \]

- \( E[x_i x_j' \varepsilon_i \varepsilon_j] \neq 0 \) when \( i \) and \( j \) both belong to \( I_g \)
- \( E[x_i x_j' \varepsilon_i \varepsilon_j] = 0 \) when \( i \in I_g \) and \( j \in I_h \) with \( g \neq h \)
Recall $\hat{\beta} \sim N(\beta, V/n)$ with $V = Q_n^{-1} \Omega_n Q_n^{-1}$ for

$$Q_n = \frac{1}{n} \sum_{i=1}^{n} E[\mathbf{x}_i \mathbf{x}'_i] \approx \frac{1}{n} \sum_{i=1}^{n} [\mathbf{x}_i \mathbf{x}'_i] = \hat{Q}$$

and

$$\Omega_n = \text{Var}[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{\epsilon}_i] = \frac{1}{n} \sum_{g=1}^{G} \sum_{i \in \mathcal{I}_g} \sum_{j \in \mathcal{I}_g} E[\mathbf{x}_i \mathbf{x}'_j \mathbf{\epsilon}_i \mathbf{\epsilon}_j] = \frac{1}{n} \sum_{g=1}^{G} E[(X'_g E_g)(E'_g X_g)]$$

for $X_g$ the $N_g \times p$ matrix formed by stacking observations $\mathbf{x}_i$ for $i \in \mathcal{I}_g$ and $E_g$ the $N_g \times 1$ vector formed by stacking $\mathbf{\epsilon}_i$ for $i \in \mathcal{I}_g$

Clustered standard errors based on estimating $\Omega_n$ with

$$\hat{\Omega} = \frac{1}{n} \sum_{g=1}^{G} (X'_g \hat{E}_g)(\hat{E}'_g X_g) = \frac{1}{G} \sum_{g=1}^{G} \frac{N_g}{\bar{N}_g} (X'_g \hat{E}_g / \sqrt{N_g})(\hat{E}'_g X_g / \sqrt{N_g})$$

for $\hat{E}_g$ the $N_g \times 1$ vector formed by stacking sample residuals $\hat{\mathbf{\epsilon}}_i$ for $i \in \mathcal{I}_g$
Fama-MacBeth (FM) Mechanics

FM subset data into the $G$ different groups:

- $X_g$ the $N_g \times p$ matrix formed by stacking observations $x_i$ for $i \in \mathcal{I}_g$
- $Y_g$ the $N_g \times 1$ vector formed by stacking observations $y_i$ for $i \in \mathcal{I}_g$
- $\widehat{\beta}_g = (X'_g X_g)^{-1}(X'_g Y_g)$ - Estimator of parameter of interest within group $g$

FM Estimator:

$$\bar{\beta} = \frac{1}{G} \sum_{g=1}^{G} \widehat{\beta}_g$$

FM Variance estimator:

$$\widehat{V}_{FM} = \frac{1}{G} \left[ \frac{1}{G-1} \sum_{g=1}^{G} (\widehat{\beta}_g - \bar{\beta})(\widehat{\beta}_g - \bar{\beta})' \right]$$

I.e. the $\widehat{\beta}_g$ are just treated as observations. The estimator is their sample mean. The estimated variance is just the estimator of the variance of this sample mean.
Traditional analysis considers large G asymptotics ($G \to \infty$) have
\( \hat{\Omega} - \Omega_n \xrightarrow{p} 0 \):

- \( \hat{E}_g = E_g - X_g(\hat{\beta} - \beta) \Rightarrow (X'_g \hat{E}_g)(\hat{E}'_g X_g) = (X'_g E_g)(E'_g X_g) - (X'_g E_g)((\hat{\beta} - \beta)X'_g X_g) - (X'_g X_g(\hat{\beta} - \beta))(E'_g X_g) + (X'_g X_g(\hat{\beta} - \beta))((\hat{\beta} - \beta)X'_g X_g) \)

- \( \frac{G}{n} \sum_{g=1}^{G} (X'_g E_g)(E'_g X_g) - \Omega_n \xrightarrow{p} 0 \) by LLN

- \( \text{vec} \left( \frac{1}{G} \sum_{g=1}^{G} (X'_g E_g)((\hat{\beta} - \beta)X'_g X_g) \right) = \frac{1}{G} \sum_{g=1}^{G} (X'_g X_g \otimes X'_g E_g) (\hat{\beta} - \beta) = O_p(1) o_p(1) \) by LLN and consistency of \( \hat{\beta} \)

- Other terms handled similarly

- Applies more generally than OLS by replacing \( x_i \hat{\varepsilon}_i \) by \( s_i(\hat{\theta}) \) (the element of the score vector for estimating \( \theta \) for observation \( i \))
Comments on Large G Clustered s.e.

Comments:

- Consistency of s.e. estimator obtains with essentially no restrictions on within group covariance structure
  - Essentially arbitrary heteroskedasticity and dependence allowed
- $\hat{\Omega}$ p.s.d. by construction
- Leverages $G \to \infty$ - need a large number of groups that can be taken to be “approximately independent”
  - Appropriate notion of “approximately independent” does not require there be literally no dependence across groups
  - Hansen (2007) verifies that no further restrictions are required on $N_g$ when $G \to \infty$ if groups are literally independent
  - Bester, Conley, and Hansen (2011) shows that $N_g \to \infty$ and $G \to \infty$ required if there is “small correlation” across groups
- Similar properties can be established for FM
Suppose you want to estimate

\[ \Omega_n = \text{Var} \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \eta_i \right] \]

where \( \text{E}[\eta_i] = 0 \), \( \text{Var}[\eta_i] = \sigma^2 \), \( \text{Cov}(\eta_i, \eta_{i-1}) = \theta \)

and all other covariances are 0.

We have

\[ \Omega_n = \sigma^2 + 2\theta \frac{n-1}{n} \]
Consider clustering estimators with $G_n$ groups of size $N_G$ (ignore integer problems)

$$\hat{\Omega}_{G_n} = \frac{1}{n} \sum_{g=1}^{G_n} g N_G \sum_{s=(g-1)N_G+1}^{g N_G} \sum_{t=(g-1)N_G+1}^{g N_G} \eta_s \eta_t$$

\[\] \[\]

- $E[\hat{\Omega}_{G_n}] = \frac{1}{n} G_n (N_G \sigma^2 + 2(N_G - 1) \theta) = \sigma^2 + 2 \theta \frac{n-G_n}{n}$
  - Bias = $-2 \theta \frac{G_n-1}{n}$
  - no asymptotic bias requires $\frac{G_n}{n} = \frac{1}{N_G} \to 0$

- Let $\omega_{st} = \eta_s \eta_t - E[\eta_s \eta_t]$. $\text{Var}[\hat{\Omega}_{G_n}] = \frac{1}{n^2} \sum_{g=1}^{G_n} g N_G \sum_{q=(g-1)N_G+1}^{g N_G} g N_G \sum_{r=(g-1)N_G+1}^{g N_G} \sum_{h=1}^{G_n} g N_G \sum_{s=(g-1)N_G+1}^{g N_G} \sum_{t=(g-1)N_G+1}^{g N_G} E[\omega_{qr} \omega_{st}]$
  - order $G_n N_G^2$ non-zero terms even if observations were independent
  - Using this optimistic rate gives $\text{Var}[\hat{\Omega}_{G_n}] = C \frac{G_n N_G^2}{n^2} = C \frac{N_G}{n}$ for some $C$
  - No asymptotic variance requires $\frac{N_G}{n} = \frac{1}{G} \to 0$

Need $G \to \infty$ for consistency. With literally no dependence across groups, can get away with no restrictions on group size. Otherwise, need large groups.
Two-Way Clustering

Cameron, Gelbach, and Miller (2011) propose a multiway clustering estimator.

Recall general HAC estimator for LS coefficient estimator:

$$\hat{\Omega} = \frac{1}{n} \sum_i \sum_j W(d(i,j)) x_i x_j' e_i e_j$$

- Cluster estimator special case with $W(d(i,j)) = \mathbb{1}(g(i) = g(j))$ where $g(i)$ returns the group membership of observation $i$.
- Consider two grouping variables (e.g. state and year). Let $g(i)$ denote group membership in first dimension and $h(i)$ denote group membership in second dimension. Two-way clustering is HAC with $W(d(i,j)) = \mathbb{1}(g(i) = g(j) \text{ or } h(i) = h(j))$.
  - Can obviously be generalized to more than two grouping variables.
  - Simple computational trick - $\hat{\Omega}_{TW} = \hat{\Omega}_{\text{Cluster}, g} + \hat{\Omega}_{\text{Cluster}, h} - \hat{\Omega}_{\text{Cluster}, g \times h}$.
  - Good properties will require minimum number of groups along any dimension to be very large.
  - Not necessarily p.s.d. (may return negative variance estimates) - in practice, common “fix” is to add a large enough p.d. matrix to the estimate to maintain positivity.
  - Rules out very sensible models of correlation across observations.

Group-Based Inference
Using a Small Number of Large Groups

Takeaways from previous section:

- Using many groups keeps variance of standard error estimator small
- Using large groups keeps bias of standard error estimator small
- These two forces work against each other - May lead to poor performance in finite samples

Revisit testing problem for scalar parameter of interest ($\hat{\beta}$) with estimated standard error ($\hat{V}$)

- Assume $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} A = \sqrt{V}Z$ where $Z \sim N(0, 1)$
- Assume $\hat{V} \xrightarrow{d} VW$ for some random variable $W$
  - Traditional approach that relies on consistency takes $W = 1$
- $t = \frac{\sqrt{n}(\hat{\beta} - \beta)}{\sqrt{\hat{V}}} \xrightarrow{d} \frac{A}{\sqrt{VW}} = \frac{Z}{\sqrt{W}}$
  - Can use this as long as distribution of $\frac{Z}{\sqrt{W}}$ can be obtained
  - Does not require consistent variance estimation - only a notion of low bias so $V$ cancels
- We know how to keep bias small - Use a small $G$ (and resulting large groups)
Clustered Standard Errors with Small $G$

Bester, Conley, and Hansen (2011) consider clustering with general dependence using small $G$

Most useful approach relies on

- $N_g/N_h \to 1$ for all $g, h$
  - Given this, simplify and assume $N_g = N \forall g$

\[
\begin{align*}
\frac{1}{N} & \begin{pmatrix} x_1' x_1 \\ \vdots \\ x_G' x_G \end{pmatrix} \xrightarrow{p} \begin{pmatrix} Q_1 \\ \vdots \\ Q_G \end{pmatrix} \\
\frac{1}{\sqrt{N}} & \begin{pmatrix} x_1' \varepsilon_1 \\ \vdots \\ x_G' \varepsilon_G \end{pmatrix} \xrightarrow{d} \begin{pmatrix} B_1 \\ \vdots \\ B_G \end{pmatrix}
\end{align*}
\]

with $Q_g = Q$ for all $g$

\[
\begin{pmatrix} B_1 \\ \vdots \\ B_G \end{pmatrix} \sim N \left( \begin{pmatrix} \Omega_1 & 0 & \cdots & 0 \\ 0 & \Omega_2 & 0 & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \Omega_G \end{pmatrix} \right)
\]

- For inference about scalar parameter, can allow $\Omega_g$ to differ across groups

Group-Based Inference
1. Properties of OLS

\[
\sqrt{n}(\hat{\beta} - \beta) = \left( \frac{1}{G} \sum_g (x'_g x_g / N) \right)^{-1} \left( \frac{1}{\sqrt{G}} \sum_g (x'_g \varepsilon_g / \sqrt{N}) \right) \]

\[
\xrightarrow{d} \left( \frac{1}{G} \sum_g Q_g \right)^{-1} \left( \frac{1}{\sqrt{G}} \sum_g B_g \right) \]

\[
= Q^{-1} \frac{1}{\sqrt{G}} \Lambda \sum_g Z_g \]

where \( \Omega = \Lambda \Lambda' \) and \( Z_g \sim N(0, I) \) iid across \( g \)

2. Properties of \( \hat{\Omega} \): (Let \( Q = \sum_g Q_g \) and \( S = \sum_g \Lambda_g Z_g \))

\[
\hat{\Omega} \xrightarrow{d} W = \frac{1}{G} \sum_g \left( \Lambda_g Z_g Z'_g \Lambda_g' - Q_g Q^{-1} S Z'_g \Lambda_g' \right) \]

\[
- \Lambda_g Z_g S' Q^{-1} Q_g + Q_g Q^{-1} S S' Q^{-1} Q_g \right) \]

\[
= \frac{1}{G} \Lambda \left[ \sum_g Z_g Z'_g - \frac{1}{G} \left( \sum_g Z_g \right) \left( \sum_g Z'_g \right) \right] \Lambda' \]
3. \( \hat{V} = \left( \frac{1}{G} \sum_g (x'_g x_g / N) \right)^{-1} \hat{\Omega} \left( \frac{1}{G} \sum_g (x'_g x_g / N) \right)^{-1} \xrightarrow{d} Q^{-1} W Q^{-1} \)

For simplicity, look at Wald test of \( H_0 : \beta = \beta_0 \):

\[
\hat{F} = \sqrt{n}(\hat{\beta} - \beta)' \hat{V}^{-1} \sqrt{n}(\hat{\beta} - \beta)
\]

\[
\xrightarrow{d} \left( \frac{1}{\sqrt{G}} \sum_g Z_g' \right) \Lambda' Q^{-1} Q \left[ \frac{1}{G} \Lambda \left[ \sum_g Z_g Z_g' - \frac{1}{G} \left( \sum_g Z_g \right) \left( \sum_g Z_g' \right) \right] \Lambda' \right]^{-1} 
\times QQ^{-1} \Lambda \left( \frac{1}{\sqrt{G}} \sum_g Z_g' \right)
\]

\[
= \left( \frac{1}{\sqrt{G}} \sum_g Z_g' \right) \left[ \frac{1}{G} \Lambda \left[ \sum_g Z_g Z_g' - \frac{1}{G} \left( \sum_g Z_g \right) \left( \sum_g Z_g' \right) \right] \right]^{-1} \left( \frac{1}{\sqrt{G}} \sum_g Z_g' \right)
\]

\( = F \)

- The asymptotic distribution is pivotal
- \( F \sim \frac{Gp}{G-p} F_{p,G-p} \) (e.g. Rao (2002), *Linear Statistical Inference and its Application*, Chapter 8b)
Main BCH Results: Under conditions above and for testing hypotheses of the form $R\beta = c$

1. $\hat{F} \overset{d}{\rightarrow} \frac{Gq}{G-q} F_{q,G-q}$ where $c$ is a $q \times 1$ vector

2. If $q = 1$, the usual $t$-statistic, $\hat{t} \overset{d}{\rightarrow} \sqrt{\frac{G-1}{G}} t_{G-1}$
   
   - Note that using $\tilde{\Omega} = \frac{G}{G-1} \hat{\Omega}$ in forming the t-statistic gives $\tilde{t} \overset{d}{\rightarrow} t_{G-1}$ - This is the scaling used in Stata
   - When using clustered standard errors, use critical values from a $t$ or $F$ distribution (rather than $N(0,1)$ or $\chi^2$) and rescale - super easy

Main drawback: Strong homogeneity conditions employed
Ibragimov and Müller (2010) consider FM with a small number of groups

Approach only allows testing a scalar hypothesis of the form $H_0 : \beta_{0,j} = c$ for some $j$

- Extended to two-sample t-tests in Ibragimov and Müller (2016)

Let $\hat{\theta}_g$ be the estimator of the parameter of interest obtained within the $g^{th}$ subgroup, and let $n = G\tilde{N}$ where $\tilde{N} = \frac{1}{G} \sum_g N_g$. IM really need one condition:

- $(\sqrt{\tilde{N}}(\hat{\theta}_1 - \theta), ..., \sqrt{\tilde{N}}(\hat{\theta}_G - \theta))' \xrightarrow{d} (b_1, ..., b_G)' \sim N(0, \Sigma)$ where $\Sigma$ is diagonal

- Note - no homogeneity imposed
1. t-statistic:

\[ t = \frac{\bar{\theta} - \theta}{\sqrt{V_{FM}}} = \frac{\sqrt{n}(\bar{\theta} - \theta)}{\sqrt{nV_{FM}}} \]

2. FM Estimator:

\[
\sqrt{n}(\bar{\theta} - \theta) = \frac{1}{\sqrt{G}} \sum_{g} \sqrt{\bar{N}}(\hat{\theta}_g - \theta)
\]

\[
\rightarrow \quad \frac{1}{\sqrt{G}} \sum_{g} b_g
\]
3. FM Variance Estimator:

\[ n\hat{V}_{FM} = \frac{n}{G} \left[ \frac{1}{G-1} \sum_{g=1}^{G} (\hat{\theta}_g - \bar{\theta})(\hat{\theta}_g - \bar{\theta})' \right] \]

\[ = \frac{1}{G-1} \sum_{g} \left( \sqrt{\bar{N}}(\hat{\theta}_g - \bar{\theta}) \right)^2 \]

\[ = \frac{1}{G-1} \sum_{g} \left( \sqrt{\bar{N}}(\hat{\theta}_g - \theta) - \frac{1}{G} \sum_{h} \sqrt{\bar{N}}(\hat{\theta}_h - \theta) \right)^2 \]

\[ \overset{d}{\rightarrow} \frac{1}{G-1} \sum_{g} (b_g - \bar{b})^2 = s_G^2 \]

Putting 1., 2., and 3. together gives

\[ t \overset{d}{\rightarrow} t^* = \frac{1}{\sqrt{G}} \sum_{g} b_g \]

\[ s_G / \sqrt{G} \]

i.e. the FM t-statistic is asymptotically the same as a t-statistic for testing \( H_0 : \mu = 0 \) using the sample mean estimated from \( g \) independent normals with different variances using the usual standard error estimator.
How does this help? Variances are not homogeneous, so $t^*$ does not follow a t-distribution.

Simplified statement of theorem from Bakirov and Székely (2005):

Let $cv_{G-1}(\alpha)$ be the usual two-sided critical value from a $t_{G-1}$ random variable. If $\alpha < .083$, then for all $G \geq 2$,

$$\sup_{\sigma_1^2, \sigma_2^2, \ldots, \sigma_G^2} P(|t^*| > cv_{G-1}(\alpha)|H_0) = P(|t_{G-1}| > cv_{G-1}(\alpha)) = \alpha$$

The $t$-test for a sample mean assuming common variances remains valid for sizes less than 8% or so regardless of heterogeneity in the variances. The FM-based $t$-test is asymptotically equivalent to a $t$-test for a sample mean so remains valid with essentially no homogeneity restrictions.
Robustness of FM approach relies on

- test of single hypothesis
- being willing to live with a conservative test

Canay, Romano, and Shaikh (2014) consider a randomization test procedure that addresses both concerns.

Let $\hat{\theta}_g$ be the estimator of the parameter(s) of interest obtained within the $g^{th}$ subgroup. Will use

$$(\sqrt{N}(\hat{\theta}_1 - \theta)', ..., \sqrt{N}(\hat{\theta}_G - \theta)') \xrightarrow{d} (b'_1, ..., b'_G)' \sim N(0, \Sigma)$$

where $\Sigma$ is block diagonal - i.e. $E[b_g b'_h] = 0$ for $g \neq h$

- Note - Good properties of CRS can be obtained under weaker conditions
- Again no homogeneity imposed
Let

- $T(X)$ be a test-statistic that takes data $X$ and rejects if $T(X)$ is large (e.g. a t-statistic)
  - In our application, we will set $X = (\sqrt{N}(\hat{\theta}_1 - \theta), \ldots, \sqrt{N}(\hat{\theta}_G - \theta))'$ where $\theta$ is the value of the parameter under the null hypothesis
- $M = [-1, 1]^G$ be the set of sign-changes for $G$ variables
- $mX$ be an element from $M$ applied to $X$
  - Note that as long as the distribution of $X$ is symmetric, $mX$ has the same distribution as $X$

Implementation:

- Compute $T(mX)$ for all $m \in M$. (If $G$ is large, this will be infeasible - can instead randomly choose elements from $M$). Let $M_R$ be the set of $m$ used where $R$ denotes the number of elements in this set.
- Calculate p-value: $\hat{p} = \frac{1}{R} \sum_{m \in M_R} \mathbb{1}(T(mX) \geq T(X))$
- Reject hypothesized value at level $\alpha$ if $\hat{p} < \alpha$
Intuition

Basic idea for why this should work:

- Under $H_0$, distribution of $X = (\sqrt{N}(\hat{\theta}_1 - \theta)', ..., \sqrt{N}(\hat{\theta}_G - \theta)')'$ is symmetric and elements of $X$ are independent.
- Multiplying the elements $X$ by 1 or $-1$ does not change the distribution under $H_0$ but does change the values of the test-statistic.
- By looking at the different values of the test-statistic produced we can then get the distribution of our test-statistic.
Thoughts on these procedures:

- All are potentially useful and none uniformly dominates in simulation
  - CRS and IM require formally weaker conditions than clustering
  - CRS is not conservative under heterogeneity while IM is
  - For small $\alpha$, CRS will have no power unless a moderate number of groups are available
  - Clustering uses the full sample in forming its point estimator - FM (e.g. IM and CRS) uses the average of subsample estimators - finite-sample bias will behave as if it is from the smallest subsample size while variance will shrink as if parameter estimated from the full sample
  - Use clustering when finite-sample bias is a bigger concern than heterogeneity
  - Use CRS when a moderate number of clusters is available
  - Use IM when using very few clusters and heterogeneity is a concern

- All procedure rely on having formed large enough groups to “capture” important sources of dependence

- Be careful of “fixed effects” that cross group boundaries - e.g. including four-digit industry classification dummies and grouping by state may induce dependence across states
Example: Health Insurance Reform (Again)

Kaestner and Simon (ILLR 2002)

Individual level data from CPS March Supplement

Data:
- 36999 individuals in 51 states (including DC) and 10 years
- Outcome: \( \log(wage) \)
- Variables of Interest (Treatments):
  - \( fullr \) - "full" HI reform indicator
  - \( partialr \) - "partial" HI reform indicator
  - \( highcost \) - number of mandates to cover high cost procedures
  - \( women \) - number of mandates for covering largely female specific procedures
  - \( other \) - number of mandates covering other procedures
- Controls:
  - \( age \): dummies for age categories (\( a2 - a8 \))
  - \( gender \): dummy for male
  - \( education \): dummies for high school, some college, and college
  - \( maritalstatus \): dummies for married and divorced
  - \( race \): dummies for African American and other race (white excluded category)
  - \( children \): number of children < 6 and number of children between 6 and 18
  - state effects, year effects, state specific trends

Group-Based Inference
HI Reform Regression

Estimated model:

$$\log(wage)_{it} = \alpha_s + \kappa_t + \delta_s t + treatment'_{st}\gamma + x'_{it}\beta + \epsilon_{it}$$

where $treatment_{st}$ is a $5 \times 1$ vector of the state level treatment variables applying to individual $i$ at time $t$.

How should we group?

- state $\times$ year?
- state?
- year?
- two-way state-year?
- 9 census-region?
- 4 census-region?
- ...
## HI Reform Regression: Results

### OLS Estimation Results:

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\beta}$</th>
<th>State</th>
<th>9 Regions</th>
<th>4 Regions</th>
<th>Year</th>
<th>State X Year</th>
<th>Two-Way</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Reform</td>
<td>-0.0332</td>
<td>0.0212</td>
<td>0.0208</td>
<td>0.0181</td>
<td>0.0191</td>
<td>0.0156</td>
<td>0.0238</td>
</tr>
<tr>
<td>Partial Reform</td>
<td>0.0036</td>
<td>0.0168</td>
<td>0.0156</td>
<td>0.0125</td>
<td>0.0183</td>
<td>0.0141</td>
<td>0.0205</td>
</tr>
<tr>
<td># High Cost</td>
<td>-0.0021</td>
<td>0.0072</td>
<td>0.0073</td>
<td>0.0072</td>
<td>0.0052</td>
<td>0.0078</td>
<td>0.0041</td>
</tr>
<tr>
<td># Women</td>
<td>0.0065</td>
<td>0.0075</td>
<td>0.0075</td>
<td>0.0099</td>
<td>0.0070</td>
<td>0.0075</td>
<td>0.0070</td>
</tr>
<tr>
<td># Other</td>
<td>0.0022</td>
<td>0.0073</td>
<td>0.0055</td>
<td>0.0061</td>
<td>0.0056</td>
<td>0.0055</td>
<td>0.70073</td>
</tr>
<tr>
<td>cv(5%)</td>
<td>2.01</td>
<td>2.31</td>
<td>3.18</td>
<td>2.26</td>
<td>1.96</td>
<td>2.26</td>
<td></td>
</tr>
</tbody>
</table>

### Fama-MacBeth Estimation Results:

<table>
<thead>
<tr>
<th></th>
<th>$\bar{\beta}$</th>
<th>s.e.</th>
<th>$\bar{\beta}_{9}\text{Regions}$</th>
<th>s.e.</th>
<th>$\bar{\beta}_{4}\text{Regions}$</th>
<th>s.e.</th>
<th>$\bar{\beta}_{\text{Year}}$</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Reform</td>
<td>0.0029</td>
<td>0.0429</td>
<td>-0.0167</td>
<td>0.0225</td>
<td>-0.0227</td>
<td>0.0204</td>
<td>0.0471</td>
<td>0.0592</td>
</tr>
<tr>
<td>Partial Reform</td>
<td>0.2503</td>
<td>0.1040</td>
<td>0.0048</td>
<td>0.0141</td>
<td>0.0086</td>
<td>0.0134</td>
<td>0.2177</td>
<td>0.1078</td>
</tr>
<tr>
<td># High Cost</td>
<td>-0.0864</td>
<td>0.0792</td>
<td>-0.0286</td>
<td>0.0235</td>
<td>-0.0042</td>
<td>0.0027</td>
<td>0.0084</td>
<td>0.0158</td>
</tr>
<tr>
<td># Women</td>
<td>0.0024</td>
<td>0.0378</td>
<td>0.0079</td>
<td>0.0115</td>
<td>0.0023</td>
<td>0.0166</td>
<td>0.0068</td>
<td>0.0216</td>
</tr>
<tr>
<td># Other</td>
<td>0.0298</td>
<td>0.0242</td>
<td>0.0047</td>
<td>0.0095</td>
<td>0.0077</td>
<td>0.0089</td>
<td>-0.0013</td>
<td>0.0130</td>
</tr>
<tr>
<td>cv(5%)</td>
<td>2.01</td>
<td></td>
<td>2.31</td>
<td></td>
<td>3.18</td>
<td></td>
<td>2.26</td>
<td></td>
</tr>
</tbody>
</table>

### Group-Based Inference