Econometrics of AKM

$$Y_{it} = \alpha_i + \psi_{j(i,t)} + X_{it}'\xi + \varepsilon_{it}$$

where $j\left(i,t\right) \in \{1, ..., J\}$ gives identity of current employer.

Matrix representation:

$$Y = D\alpha + F\psi + X\xi + \varepsilon$$

- Isomorphic to standard panel model but with J treatments.
- Treat $Z = (D, F, X)$ as fixed (i.e. all expectations conditional on Z)

Identification:

- Exogeneity: $\mathbb{E}\left[\varepsilon\right] = 0$ (plausible?)
- Rank condition: need at least one restriction on the $\{\psi_j\}_{j=1}^{J}$ within each "connected set" of firms

Variance decomposition

Target parameter: size weighted variance of firm effects

$$\theta_\psi = \sum_{j=1}^{J} s_j \left(\psi_j - \bar{\psi}\right)^2,$$

where s_j is firm j's employment share and $\bar{\psi} = \sum_{j=1}^{J} s_j \psi_j$.

Customary to use OLS estimates $\hat{\psi}$ to compute "plug-in" estimates of variance components, e.g.:

$$\begin{aligned}\hat{\theta}_\psi &= \sum_{j=1}^{J} s_j \left(\hat{\psi}_j - \hat{\bar{\psi}}\right)^2 \\ &= \sum_{j=1}^{J} s_j \left(\hat{\psi}_j\right)^2 - \left(\hat{\bar{\psi}}\right)^2\end{aligned}$$

Bias in the square

OLS is unbiased

$$\mathbb{E}\left[\hat{\psi}_j\right] = \psi_j$$

But the square of an unbiased estimator is upward biased

$$\begin{aligned}
\mathbb{E}\left[\left(\hat{\psi}_j\right)^2\right] &= \mathbb{E}\left[\left(\hat{\psi}_j - \psi_j + \psi_j\right)^2\right] \\
&= \mathbb{E}\left[\left(\hat{\psi}_j - \psi_j\right)^2\right] + 2\mathbb{E}\left[\hat{\psi}_j - \psi_j\right]\psi_j + \psi_j^2 \\
&= \psi_j^2 + \underbrace{\mathbb{V}\left[\hat{\psi}_j\right]}_{\text{bias}}
\end{aligned}$$

Bias of plugin

By same argument plug-in estimator is biased

$$\begin{aligned}
\mathbb{E}\left[\hat{\theta}_{\psi}\right] &= \sum_{j=1}^{J} s_j \mathbb{E}\left[\left(\hat{\psi}_j\right)^2\right] - \mathbb{E}\left[\left(\hat{\bar{\psi}}\right)^2\right] \\
&= \sum_{j=1}^{J} s_j \left\{\psi_j^2 + \mathbb{V}\left[\hat{\psi}_j\right]\right\} - (\bar{\psi})^2 - \mathbb{V}\left[\hat{\bar{\psi}}\right] \\
&= \theta_{\psi} + \underbrace{\sum_{j=1}^{J} s_j \mathbb{V}\left[\hat{\psi}_j\right] - \mathbb{V}\left[\hat{\bar{\psi}}\right]}_{\text{bias}}
\end{aligned}$$

$\mathbb{V}\left[\hat{\bar{\psi}}\right]$ term typically negligible when J is large..

Correcting the bias

Bias is weighted average of squared standard errors on firm effects:

$$\mathbb{E}\left[\hat{\theta}_\psi - \theta_\psi\right] \approx \sum_{j=1}^{J} s_j \mathbb{V}\left[\hat{\psi}_j\right]$$

Correcting the bias

Bias is weighted average of squared standard errors on firm effects:

$$\mathbb{E}\left[\hat{\theta}_\psi - \theta_\psi\right] \approx \sum_{j=1}^{J} s_j \mathbb{V}\left[\hat{\psi}_j\right]$$

Can't we just do Krueger-Summers style correction based on conventional het-consistent ("robust") standard errors $\hat{\mathbb{V}}_{HC}\left[\hat{\psi}_j\right]$?

- No, because HC standard errors break down (are inconsistent) when # of regressors grow in proportion to sample size.
- Same problem for bootstrap (Bickel and Freedman, 1983)
- To handle high dimensionality: swap usual het-consistent estimators $\hat{\mathbb{V}}_{HC}\left[\hat{\psi}_j\right]$ for het-unbiased estimators $\hat{\mathbb{V}}_{HU}\left[\hat{\psi}_j\right]$. Noise averages out across estimates.

Bias correction: homoscedastic case

Andrews et al (2008): bias correct assuming $\mathbb{V}[\varepsilon] = I\sigma^2$

$$\mathbb{V}\left[\hat{\psi}\right] = \left(\tilde{F}'\tilde{F}\right)^{-1}\sigma^2$$

where $\tilde{F}$ is residualized version of F (against D and X).

- Estimate $\mathbb{V}\left[\hat{\psi}\right]$ using DoF adjusted regression MSE

$$\hat{\sigma}^2 = \frac{SSR}{n - \dim(Z)}$$

- But homoscedasticity is a strong assumption
  - Can't be correct if outcome is bounded
  - And in the case of log wages there is ample evidence that error variance differs by gender / experience (e.g., Lemieux, 2006)

Bias correction: heteroscedasticity

Index each person-year observation by $\ell = \ell(i, t)$

- Suppose errors $\{\varepsilon_\ell\}$ are mutually independent
- But potentially heteroscedastic with variances $\sigma_\ell^2 = \mathbb{V}[\varepsilon_\ell]$

Yields familiar "sandwich" variance expression (White, 1980)

$$\mathbb{V}\left[\hat{\psi}\right] = \left(\tilde{F}'\tilde{F}\right)^{-1}\left(\tilde{F}'\Omega\tilde{F}\right)\left(\tilde{F}'\tilde{F}\right)^{-1}$$

where $\Omega = diag\left(\sigma_1^2, ..., \sigma_n^2\right)$.

Estimation challenge: How to get the error variances $\left\{\sigma_\ell^2\right\}_{\ell=1}^n$?

Kline, Saggio, and Sølvsten (2020)

Write AKM as high-dimensional regression:

$$Y_\ell = Z_\ell'\beta + \varepsilon_\ell, \quad \text{for } \ell = 1, ..., n.$$

- Let $\hat{\beta}_{-\ell}$ denote the OLS estimator of β obtained after leaving out obs ℓ. (Requires leave-out connectedness)
- "Cross-fit" estimator of σ_ℓ^2 is *unbiased*:

$$\hat{\sigma}_\ell^2 \quad = \quad Y_\ell \quad \underbrace{\left(Y_\ell - Z_\ell'\hat{\beta}_{-\ell}\right)}_{\text{leave-out prediction error}}$$

Cross-fitting

"Cross-fit" estimator of σ_ℓ^2 is *unbiased*:

$$\begin{aligned}
\hat{\sigma}_\ell^2 &= Y_\ell \underbrace{\left(Y_\ell - Z_\ell' \hat{\beta}_{-\ell}\right)}_{\text{leave-out prediction error}} \\
&= \left(\varepsilon_\ell + Z_\ell' \beta\right)\left(\varepsilon_\ell + Z_\ell'\left(\beta - \hat{\beta}_{-\ell}\right)\right)
\end{aligned}$$

Intuition: leave-out breaks corr between $\hat{\beta}$ and ε_ℓ

$$\begin{aligned}
\mathbb{E}\left[\varepsilon_\ell\left(\beta - \hat{\beta}_{-\ell}\right)\right] &= \mathbb{E}\left[\varepsilon_\ell\left(\sum_{l \neq \ell} Z_l Z_l'\right)^{-1} \sum_{l \neq \ell} Z_l \varepsilon_l\right] \\
&= \left(\sum_{l \neq \ell} Z_l Z_l'\right)^{-1} \sum_{l \neq \ell} Z_l \underbrace{\mathbb{E}\left[\varepsilon_\ell \varepsilon_l\right]}_{=0}
\end{aligned}$$

Bias correction

Proxy Ω with $\hat{\Omega} = diag \left\{ \hat{\sigma}_\ell^2 \right\}_{\ell=1}^n$ to get unbiased variance estimates

$$\hat{\mathbb{V}}_{HU}\left[\hat{\psi}\right] = \left(\tilde{F}'\tilde{F}\right)^{-1}\left(\tilde{F}'\hat{\Omega}\tilde{F}\right)\left(\tilde{F}'\tilde{F}\right)^{-1}$$

Bias corrected estimator of θ_ψ is:

$$\hat{\theta}_{\psi,HU} = \underbrace{\hat{\theta}_\psi}_{\text{plugin}} - \underbrace{\sum_{j=1}^{J} s_j \hat{\mathbb{V}}_{HU}\left[\hat{\psi}_j\right]}_{\text{average squared stderr}} + \underbrace{\hat{\mathbb{V}}_{HU}\left[\hat{\bar{\psi}}\right]}_{\text{stderr of mean}}$$

Generalization

What about other variances and covariances?

- KSS consider more general (co-)variance components

$$\theta = \beta' A \beta$$

where A is user specified matrix.

- General bias correction formula:

$$\hat{\theta}_{HU} = \hat{\theta} - \sum_{\ell=1}^{n} B_{\ell\ell} \hat{\sigma}_\ell^2$$

where $B_{\ell\ell} = Z_\ell' \left(\sum_{l=1}^n Z_l Z_l'\right)^{-1} A \left(\sum_{l=1}^n Z_l Z_l'\right)^{-1} Z_\ell$ gives influence of ε_ℓ^2 on $\hat{\theta}$. Mathematical intuition:

$$\hat{\theta} = \theta + \sum_{\ell=1}^{n} B_{\ell\ell} \varepsilon_\ell^2 + o_p(1).$$

Computation

A useful trick:

$$\begin{aligned}\hat{\sigma}_\ell^2 &= Y_\ell\left(Y_\ell - Z_\ell'\hat{\beta}_{-\ell}\right)\\ &= Y_\ell\frac{\left(Y_\ell - Z_\ell'\hat{\beta}\right)}{1-P_{\ell\ell}}\end{aligned}$$

where $\{P_{\ell\ell}\}$ are the diagonal elements of $P = Z\left(Z'Z\right)^{-1}Z'$.

- Note: only need to compute $\hat{\beta}$ once!
- In large problems can stochastically approximate $\{B_{\ell\ell}, P_{\ell\ell}\}$ (CHK size application in $<$1hr)
- Code / executables available at GitHub repository

Application to Italian data

Administrative records from Italian province of Veneto

Compare plug-in (AKM), homoscedasticity-only (HO) estimator of Andrews (2008), and KSS

Base sample: two wage observations per worker

- With a single wage change per worker we can ignore serial correlation / clustering when computing firm effect variances
- Allows us to focus on importance of heteroscedasticity, but throws away some of the data
- Analyzing 6 year panel via leave-worker-out yields similar results

Split by age: older workers move less $\Rightarrow$ more bias

Bias correction to variance of firm effs

Homoscedastic correction about half way between naive plug-in and KSS

VARIANCE DECOMPOSITION[a]

	Pooled	Younger Workers	Older Workers
Variance of Firm Effects			
Plug in (PI)	$\boxed{0.0358}$	$\boxed{0.0368}$	$\boxed{0.0415}$
Homoscedasticity Only (HO)	$\boxed{0.0295}$	$\boxed{0.0270}$	$\boxed{0.0350}$
Leave Out (KSS)	$\boxed{0.0240}$	$\boxed{0.0218}$	$\boxed{0.0204}$
Variance of Person Effects			
Plug in (PI)	0.1321	0.0843	0.2180
Homoscedasticity Only (HO)	0.1173	0.0647	0.2046
Leave Out (KSS)	0.1119	0.0596	0.1910
Covariance of Firm, Person Effects			
Plug in (PI)	0.0039	−0.0058	−0.0032
Homoscedasticity Only (HO)	0.0097	0.0030	0.0040
Leave Out (KSS)	0.0147	0.0075	0.0171
Correlation of Firm, Person Effects			
Plug in (PI)	0.0565	−0.1040	−0.0334
Homoscedasticity Only (HO)	0.1649	0.0726	0.0475
Leave Out (KSS)	0.2830	0.2092	0.2744
Coefficient of Determination (R^2)			
Plug in (PI)	0.9546	0.9183	0.9774
Homoscedasticity Only (HO)	0.9029	0.8184	0.9524
Leave Out (KSS)	0.8976	0.8091	0.9489

Large bias in correlation coefficient

Flips sign in age-specific samples!

Variance Decomposition[a]

	Pooled	Younger Workers	Older Workers
Variance of Firm Effects			
Plug in (PI)	0.0358	0.0368	0.0415
Homoscedasticity Only (HO)	0.0295	0.0270	0.0350
Leave Out (KSS)	0.0240	0.0218	0.0204
Variance of Person Effects			
Plug in (PI)	0.1321	0.0843	0.2180
Homoscedasticity Only (HO)	0.1173	0.0647	0.2046
Leave Out (KSS)	0.1119	0.0596	0.1910
Covariance of Firm, Person Effects			
Plug in (PI)	0.0039	−0.0058	−0.0032
Homoscedasticity Only (HO)	0.0097	0.0030	0.0040
Leave Out (KSS)	0.0147	0.0075	0.0171
Correlation of Firm, Person Effects			
Plug in (PI)	0.0565	−0.1040	−0.0334
Homoscedasticity Only (HO)	0.1649	0.0726	0.0475
Leave Out (KSS)	0.2830	0.2092	0.2744
Coefficient of Determination (R^2)			
Plug in (PI)	0.9546	0.9183	0.9774
Homoscedasticity Only (HO)	0.9029	0.8184	0.9524
Leave Out (KSS)	0.8976	0.8091	0.9489

Small decrease in total explanatory power of model

Note: HO estimate is familiar "adjusted" R^2, which seems to exhibit negligible bias.

VARIANCE DECOMPOSITION[a]

	Pooled	Younger Workers	Older Workers
Variance of Firm Effects			
Plug in (PI)	0.0358	0.0368	0.0415
Homoscedasticity Only (HO)	0.0295	0.0270	0.0350
Leave Out (KSS)	0.0240	0.0218	0.0204
Variance of Person Effects			
Plug in (PI)	0.1321	0.0843	0.2180
Homoscedasticity Only (HO)	0.1173	0.0647	0.2046
Leave Out (KSS)	0.1119	0.0596	0.1910
Covariance of Firm, Person Effects			
Plug in (PI)	0.0039	−0.0058	−0.0032
Homoscedasticity Only (HO)	0.0097	0.0030	0.0040
Leave Out (KSS)	0.0147	0.0075	0.0171
Correlation of Firm, Person Effects			
Plug in (PI)	0.0565	−0.1040	−0.0334
Homoscedasticity Only (HO)	0.1649	0.0726	0.0475
Leave Out (KSS)	0.2830	0.2092	0.2744
Coefficient of Determination (R^2)			
Plug in (PI)	0.9546	0.9183	0.9774
Homoscedasticity Only (HO)	0.9029	0.8184	0.9524
Leave Out (KSS)	0.8976	0.8091	0.9489

Estimates from 6 year panel nearly identical after accounting for serial correlation

TABLE A.I

VARIANCE OF FIRM EFFECTS UNDER DIFFERENT LEAVE-OUT STRATEGIES[a]

	Pooled	Younger Workers	Older Workers
Variance of Firm Effects			
Plug-in	0.0304	0.0303	0.0376
Leave Person-Year Out	0.0296	0.0302	0.0314
Leave Match Out	0.0243	0.0221	0.0265
Leave Worker Out	0.0241	0.0227	0.0270

Leaving match out yields same answer as leaving whole worker out
$\Rightarrow$ sufficient to "cluster" std err estimates $\hat{\mathbb{V}}_{HU}\left[\hat{\psi}_j\right]$ by match

Projecting fixed effects onto observables

- Common to project fixed effect estimates $\hat{\psi}$ onto covariates
- Problem: $\hat{\psi}$ are correlated with one another
- Dependence hinges on design because

$$\hat{\psi} = \psi + \underbrace{\left(\tilde{F}'\tilde{F}\right)^{-1}\tilde{F}'\varepsilon}_{\text{correlated noise}}$$

- Solution: use HU variance estimator

$$\begin{aligned}
\hat{\mathbb{V}}_{HU}\left[\hat{\psi}\right] &= \left(\tilde{F}'\tilde{F}\right)^{-1}\left(\tilde{F}'\hat{\Omega}\tilde{F}\right)\left(\tilde{F}'\tilde{F}\right)^{-1} \\
&= \left(\tilde{F}'\tilde{F}\right)^{-1}\left(\sum_{\ell=1}^{n}\tilde{f}_\ell\tilde{f}_\ell'\hat{\sigma}_\ell^2\right)\left(\tilde{F}'\tilde{F}\right)^{-1}
\end{aligned}$$

Connection to HC2

HC2 estimator (Mackinnon and White, 1985) is:

$$\hat{\mathbb{V}}_{HC2}\left[\hat{\psi}\right] = \left(\tilde{F}'\tilde{F}\right)^{-1}\left(\sum_{\ell=1}^{n}\tilde{f}_\ell\tilde{f}_\ell'\frac{\left(Y_\ell - Z_\ell'\hat{\beta}\right)^{2}}{1-P_{\ell\ell}}\right)\left(\tilde{F}'\tilde{F}\right)^{-1}$$

- HC2 is unbiased under *homo*-scedasticity but otherwise inconsistent when $dim\left(\tilde{F}\right) \propto n$.

HU estimator is:

$$\hat{\mathbb{V}}_{HU}\left[\hat{\psi}\right] = \left(\tilde{F}'\tilde{F}\right)^{-1}\left(\sum_{\ell=1}^{n}\tilde{f}_\ell\tilde{f}_\ell'\frac{Y_\ell\left(Y_\ell - Z_\ell'\hat{\beta}\right)}{1-P_{\ell\ell}}\right)\left(\tilde{F}'\tilde{F}\right)^{-1}$$

- Unbiased under arbitrary heteroscedasticity.

Standard errors on projection

Projection of ψ onto W is linear combination:

$$\left(W'W\right)^{-1}W'\psi = v'\psi$$

- Estimator of variance of projection coefficients is

$$\hat{\mathbb{V}}_{HU}\left[v'\hat{\psi}\right] = v'\left(\tilde{F}'\tilde{F}\right)^{-1}\left(\tilde{F}'\hat{\Omega}\tilde{F}\right)\left(\tilde{F}'\tilde{F}\right)^{-1}v$$

- Suppose v is $J \times 1$ (i.e., single projection coefficient of interest)
- Provided v' doesn't place "too much" weight on any particular coefficient KSS show that:

$$\frac{v'\left(\hat{\psi}-\psi\right)}{\sqrt{\hat{\mathbb{V}}_{HU}\left[v'\hat{\psi}\right]}} \rightarrow N\left(0,1\right)$$

lincom_KSS: high-dim version of Stata "lincom" command

Naive "robust" std err order of magnitude too small!

Projecting Firm Effects Onto Covariates[a]

	(1)	(2)
Older Worker	0.0272	$\boxed{-0.0016}$
	(0.0009)	$\boxed{(0.0024)}$
	[0.0003]	$\boxed{[0.0001]}$
Log Firm Size		0.0276
		(0.0007)
		[0.0001]
Older Worker × Log Firm Size		0.0028
		(0.0005)
		[0.0002]
Predicted Gap in Firm Effects (Older vs. Younger Workers)	0.0272	0.0054
	(0.0009)	(0.0019)
	[0.0003]	[0.0008]
Number of Observations	1,319,972	1,319,972

[a]This table reports the coefficients from projections of firm effects onto worker and firm characteristics in the pooled leave-one-out sample. A constant is included in each model. Standard errors based on equation (7) reported in parentheses. Naive Eicker–White (HC1) standard errors shown in square brackets. "Predicted Gap in Firm Effects" reports the predicted difference in firm effects between older and younger workers according to either Column (1) or Column (2) evaluated at the median firm size of 12 workers.

Naive std error on old dummy off by a factor of 24 in Col 2! Leave out std error reveals that older workers no more likely to work at high paying firms after adjusting for firm size.

Testing high dimensional hypotheses about fixed effects

Do the firm effects for younger workers equal those faced by older workers?

$$H_0 : \psi_j^O = \psi_j^Y \quad \text{for } j = 1, ..., J$$

- $J = 8,578 \Rightarrow$ cannot rely on standard $\chi^2\,(8578)$ approximation to F-test
- Bootstrap also fails

KSS: test by estimating the variance component

$$\theta_{H_0} = \frac{1}{8578} \left(\psi^O - \psi^Y\right)' \left(\tilde{F}'\tilde{F}\right) \left(\psi^O - \psi^Y\right)$$

Intuition:

- If H_0 is true, we must have $\theta_{H_0} = 0$
- $\tilde{F}'\tilde{F}$ gives optimal (i.e. inverse variance) weighting of differences $\hat{\psi}^O - \hat{\psi}^Y$ under homoscedasticity

Testing high dimensional hypotheses about fixed effects

Do the firm effects for younger workers equal those faced by older workers?

$$H_0 : \psi_j^O = \psi_j^Y \quad \text{for } j = 1, ..., J$$

KSS: test by estimating the variance component

$$\theta_{H_0} = \frac{1}{8578} \left(\psi^O - \psi^Y\right)' \left(\tilde{F}'\tilde{F}\right) \left(\psi^O - \psi^Y\right)$$

Under H_0: $\hat{\theta}_{H_0}$converges to $\mathcal{N}\left(0, \mathbb{V}\left[\hat{\theta}_{H_0}\right]\right)$.

- Estimation of $\mathbb{V}\left[\hat{\theta}_{H_0}\right]$ explained in paper.
- Test statistic is simple t-stat $\hat{\theta}_{H_0}/\sqrt{\hat{\mathbb{V}}_{HU}\left[\hat{\theta}_{H_0}\right]}$

Firm effects highly correlated across age groups

But can decisively reject that they are exactly the same