PANEL DATA

Repeated observations over time of a sample of cross-sectional entities (i)

Ex. Panel Study of Income Dynamics (PSID)

economic info on about 6,000 families followed through time

why use panel data?

1. can capture dynamics (or lack thereof) of economic issues

ex. Ben-Porath (1973) married women have 50% labor force participation

is it heterogeneity: half never work, half always work

or is it dynamic: homogeneous population, each woman expects to work 50% of the time

2. often a useful way to implement instruments or as an alternative to an event study

ex. Nasdaq stocks are more volatile than NYSE stocks

is it the type of firm that lists, or is it the market structure?

panel data can be used to look at firms that switch

(presumably this controls for selection ~~bias~~ effects)

compare volatility before vs. after

3. can handle ~~heterogeneity and~~ omitted variables

ex. $y_{it} = \alpha + \beta x_{it} + \gamma z_{it} + u_{it}$

but z_{it} is unobserved and we suspect $\text{cov}(x_{it}, z_{it}) \neq 0$.

Are we in trouble? Not necessarily. If $z_{it} = z_i \ \forall t$ we can difference the data and get:

$$y_{it} - y_{i,t-1} = \beta(x_{it} - x_{i,t-1}) + u_{it} - u_{i,t-1}$$

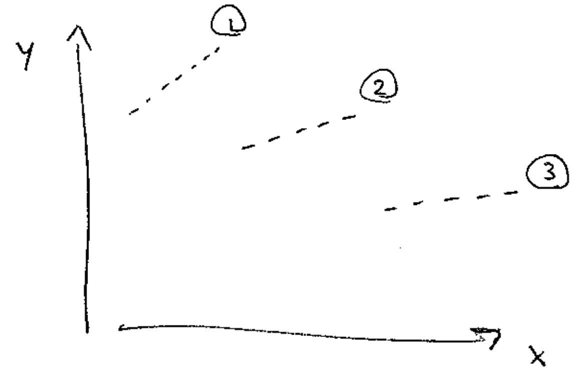
and we'll get consistent estimates of β.

(Note that differencing is not usually our best route.)

4. can handle heterogeneity

ex. $y_{it} = \alpha_i + \beta_i x_{it} + u_{it}$

Suppose the data look like this:



but if we pool all the data, we'll estimate $\hat{\beta} < 0$

Basic framework:

$$y_{it} = \underset{\sim}{\beta}' \underset{\sim}{x}_{it} + \underset{\sim}{\alpha}' \underset{\sim}{z}_i + \varepsilon_{it} \qquad \begin{array}{l} i = 1, \ldots, N \\ t = 1, \ldots, T \end{array}$$

where $\underset{\sim}{x}_{it}$ does <u>not</u> include a constant term but $\underset{\sim}{z}_i$ does

z_i can be thought of as 'individual or group-specific variables, observed or not

e.g., race, sex, SAT score, ability, effort, preferences

Note that we're not allowing the general heterogeneity discussed above for now.

Suppose now that we can't observe the z_i. (If we can observe all the z_i, there's no problem at all.)

To the econometrician, the model is:

$$y_{it} = \alpha_i + \underset{\sim}{\beta}' \underset{\sim}{x}_{it} + \varepsilon_{it} \qquad \boxed{\text{"FIXED EFFECTS MODEL"}}$$

(If we think that $\alpha_i = \alpha_j \; \forall i, j$, we can just estimate the <u>pooled</u> regression using all observations pooled together.)

How can we estimate this? Stack everything into a system:

$$\underset{NT \times 1}{\begin{bmatrix} \overset{T\times 1}{y_1} \\ \vdots \\ \underset{\sim}{y_n} \end{bmatrix}} = \underset{NT \times k}{\begin{bmatrix} \overset{T\times k}{X_1} \\ \vdots \\ X_n \end{bmatrix}} \underset{k \times 1}{\underset{\sim}{\beta}} + \underset{NT \times N}{\begin{bmatrix} \overset{T\times 1}{\underset{\sim}{1}} & & 0 \\ & \underset{\sim}{1} & \\ 0 & & \ddots \; \underset{\sim}{1} \end{bmatrix}} \underset{N \times 1}{\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}} + \underset{NT \times 1}{\begin{bmatrix} \overset{T\times 1}{\varepsilon_1} \\ \vdots \\ \underset{\sim}{\varepsilon_n} \end{bmatrix}}$$

or equivalently $\quad y = [\, X \;\; d_1 \; d_2 \cdots d_n \,] \begin{bmatrix} \beta \\ \alpha \end{bmatrix} + \varepsilon \quad$ where d_i is a dummy variable for the i^{th} unit or even more compactly as $y = X\beta + D\alpha + \varepsilon$, $D = [d_1 \cdots d_n]$

This model is called the LSDV (least squares dummy variable) model

We can estimate this with OLS, estimating $k + N$ parameters

No problem if N is small, but infeasible if N is in the thousands.
(inverting the design matrix is hard if it's $10{,}000 \times 10{,}000$)

Fortunately, there's a way to proceed. Go back to the original:

$$y_{it} = \alpha_i + \beta' \underset{\sim}{x}_{it} + \varepsilon_{it}$$

For each group i,

$$\bar{y}_i = \alpha_i + \underset{\sim}{\beta}' \underset{\sim}{\bar{x}}_i + \bar{\varepsilon}_i \qquad \text{(where the avgs are over time)}$$

and so

$$y_{it} - \bar{y}_i = \beta'(x_{it} - \bar{x}_i) + \varepsilon_{it} - \bar{\varepsilon}_i$$

If our original specification satisfies the ideal conditions, so does this; and $\hat{\beta}^{\text{within}}$ (the <u>within-group estimator</u>, so named because it makes use of TS variation in each group) is BLUE

We can use the group means equation to recover $\hat{\alpha}_i$ if this is of interest

$$\hat{\alpha}_i = \bar{y}_i - \hat{\beta}^{\text{within}\,\prime} \bar{x}_i$$

why is this easier? To calc $\hat{\beta}^{\text{within}}$, we just have to invert a $K \times K$ matrix

But <u>be careful with your</u> d.f.'s.

In the LSDV formulation

$$s^2 = \frac{e'e}{NT - N - K}$$

If you subtract group means, your regression software doesn't notice and instead calc's

$$s^2 = \frac{e'e}{NT - K} \quad \text{wrong!}$$

We can test whether $\alpha_i = \alpha_j \;\forall i,j$ using an F-test.

Under the null, we simply estimate a pooled model and using our std. results

$$\text{under } H_0: \quad \frac{\left(R^2_{LSDV} - R^2_{POOLED}\right)/(N-1)}{\left(1 - R^2_{LSDV}\right)/(NT-N-K)} \sim F(N-1,\, NT-N-K)$$

Within vs. between-groups estimators

If we have a pooled linear model, we can estimate it 3 ways.

$\hat{\beta}_{total}$: original: $y_{it} = \beta' x_{it} + \alpha + \varepsilon_{it}$

$\hat{\beta}_{within}$ deviations from group means: $y_{it} - \bar{y}_i = \beta'(x_{it} - \bar{x}_i) + \varepsilon_{it} - \bar{\varepsilon}_i$

$\hat{\beta}_{btw}$ group means: $\bar{y}_i = \beta' \bar{x}_i + \alpha + \bar{\varepsilon}_i$

called between-groups estimator because it focuses on C-S differences in means

$$\hat{\beta}_{total} = \left(S^{total}_{xx}\right)^{-1}\left(S^{total}_{xy}\right) \qquad S^{total}_{xx} = \sum_i \sum_t (x_{it} - \bar{\bar{x}})(x_{it} - \bar{\bar{x}})'$$

$$S^{total}_{xy} = \sum_i \sum_t (x_{it} - \bar{\bar{x}})(y_{it} - \bar{\bar{y}})$$

where $\bar{\bar{x}}$ and $\bar{\bar{y}}$ are the overall means (avg'd over both N and T)

Similarly

$$\hat{\beta}_{within} = \left(S^{within}_{xx}\right)^{-1}\left(S^{within}_{xy}\right) \qquad S^{within}_{xx} = \sum_i \sum_t (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)'$$

$$S^{within}_{xy} = \sum_i \sum_t (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i)$$

$$S^{btw}_{xx} = \sum_i T(\bar{x}_i - \bar{\bar{x}})(\bar{x}_i - \bar{\bar{x}})'$$

$$\hat{\beta}_{btw} = \left(S^{btw}_{xx}\right)^{-1}\left(S^{btw}_{xy}\right)$$

$$S^{btw}_{xy} = \sum_i T(x_i - \bar{\bar{x}})(\bar{y}_i - \bar{\bar{y}})$$

From this, it should be clear that

$$S_{xx}^{total} = S_{xx}^{within} + S_{xx}^{btw}$$

$$S_{xy}^{total} = S_{xy}^{within} + S_{xy}^{btw}$$

So $$\hat{\beta}_{total} = \left(S_{xx}^{within} + S_{xx}^{btw}\right)^{-1}\left(S_{xy}^{within} + S_{xy}^{btw}\right)$$

Sub in $$S_{xy}^{within} = S_{xx}^{within}\hat{\beta}_{within} \quad \text{and} \quad S_{xy}^{btw} = S_{xx}^{btw}\hat{\beta}_{btw}$$

to get

$$\hat{\beta}_{total} = \left(S_{xx}^{within} + S_{xx}^{btw}\right)^{-1}\left(S_{xx}^{within}\hat{\beta}_{within} + S_{xx}^{btw}\hat{\beta}_{btw}\right)$$

$$= F\,\hat{\beta}_{within} + (I - F)\,\hat{\beta}_{btw}$$

So the OLS estimator is a matrix wtd avg. of the btw and within estimators

In the pooled model, we don't want to throw out the btw data – it's not optimal to just use the within-group estimator, because there's additional c-s information that we want to capture

but this isn't true for the fixed-effect model, where the N group means are being "used up" to estimate N group intercepts

<u>Two-way fixed effects</u>

LSDV model can be extended to include a time effect as well

$$y_{it} = \underset{\sim}{\beta}'\underset{\sim}{x}_{it} + \alpha_i + \gamma_t + \varepsilon_{it}$$

(be sure to only add T−1 time dummies – why?)

If $K + N + T$ parameters are too much for your computer, there are ways to estimate the slope parms using deviations from appropriate means and then back into $\hat{\alpha}_i$ and $\hat{\gamma}_t$ (See Greene 13.3.3)

but usually your package won't have trouble doing this

<u>Balanced vs. unbalanced panels</u>

Missing data are common in panels, especially those that use surveys

Or different groups have different length time series due to mergers, delistings

Either way, we get an <u>unbalanced</u> panel

But the effects are minimal for fixed effects:

new sample size $\sum_i T_i$ for calculating s^2, F-stat, etc.

group means based on T_i rather than t

And it's even easier using statistical software

if there's no time dummy, nothing changes from the LSDV specification

with a time ~~dumm~~ effect, create a time dummy for each date that appears in the full data set, and estimate an LSDV model

<u>Ex.</u> short selling across countries and over time

Bris, Goetzmann, Zhu (2005)

Fixed vs. random effects

The major disadvantage of a fixed-effects model is the num. of parms. estimated

If we are willing to make some stronger assumptions, we can estimate many fewer parms.

Rewrite our standard model w/ heterogeneous intercepts:

$$y_{it} = \alpha + \beta' \underset{\sim}{x}_{it} + u_i + \varepsilon_{it}$$

Instead of estimating u_i for each firm/unit/individual, now we'll assume the u_i are random and indep of everything else in the model

$$E(\varepsilon_{it} \mid X) = E(u_i \mid X) = 0$$

$$E(\varepsilon_{it}^2 \mid X) = \sigma_\varepsilon^2 \qquad E(u_i^2 \mid X) = \sigma_u^2$$

$$E(\varepsilon_{it} u_j \mid X) = 0 \quad \forall i, t, j$$

$$E(\varepsilon_{it} \varepsilon_{js} \mid X) = 0 \quad \text{if } t \neq s \text{ or } i \neq j$$

$$E(u_i u_j \mid X) = 0 \quad \forall i \neq j$$

This is called a random effects model.

It turns out that the GLS estimator given this structure is still a wtd. avg. of the btw and within estimators

$$\hat{\beta}_{GLS} = \hat{F}_{within} \hat{\beta}_{within} + (I - \hat{F}_{within}) \hat{\beta}_{btw}$$

where $\hat{F}_{within} = \left[S_{xx}^{within} + \lambda S_{xx}^{btw} \right]^{-1} S_{xx}^{within}$, $\lambda = \dfrac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_u^2}$

This differs from the OLS pooled estimator to the extent that $\lambda \neq 1$.

Not used in finance very often, because these assumptions are usually quite suspect.

Nonspherical disturbances in panels

The models here are extensions of the classical assumptions, so we can calculate $\hat{\beta}_{OLS}$ and use a robust std error to treat heteroskedasticity.

In the standard LSDV formulation, we have

$$y = X\beta + D\alpha + \varepsilon = Zb + \varepsilon$$

and we can apply White std errors & recall that $\text{var}(\hat{b}) = (Z'Z)^{-1} Z'\widehat{\Omega} Z (Z'Z)^{-1}$

$$\Omega = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_{NT}^2 \end{bmatrix}$$

We can actually do this using the within-group estimator.

$$y_{it} - \bar{y}_i = \beta'(x_{it} - \bar{x}_i) + \varepsilon_{it} - \bar{\varepsilon}_i = \tilde{X}\beta + \tilde{\varepsilon}$$

still satisfies the required conditions, so we can calculate

$$\text{var}(\hat{\beta}) = (\tilde{X}'\tilde{X})^{-1} \tilde{X}'\widehat{\Omega}\tilde{X}(\tilde{X}'\tilde{X})^{-1}$$

and get a numerically identical estimate to the LSDV version

The bigger issue is how to proceed if there is cross-sectional correlation in the residuals, or within-firm correlation in the residuals

Then we need to calculated clustered std errors

Essentially, we assume $\begin{cases} \text{independence across clusters} \\ \text{arbitrary correlation across obs. within a cluster} \end{cases}$

These are called Rogers std. errors in the finance literature, named for a STATA employee who wrote and documented the Stata command

but they are also developed by Andrews (1991) Moulton (1990) Arellano (1987) Moulton (1986) Liang and Zeger (1986)

Clustered standard errors

Basically a generalization of White std. errors.

Suppose we assume independence across Q clusters, and let $\underset{m\times 1}{X_q}$, $\underset{m\times 1}{\varepsilon_q}$, and $\underset{m\times m}{\Omega_q}$ be the vector of RHS variables, errors, and the cov matrix for the q^{th} cluster w/ m elements

Recall that

$$\text{var}(\hat{\beta}) = (X'X)^{-1} X'\Omega X (X'X)^{-1}$$

Assuming independence across clusters simplifies this to

$$\text{var}(\hat{\beta}) = (X'X)^{-1} \sum_q (X_q' \Omega_q X_q)(X'X)^{-1}$$

and we can estimate this consistently using

$$\widehat{\text{avar}}(\hat{\beta}) = (X'X)^{-1} \sum_q (X_q' e_q e_q' X_q)(X'X)^{-1}$$

Because these are asymptotics, we need to have a lg. number of independent clusters for the result to go through — Rogers (1993) finds bias when Q is small (5-10).

In finance, we often assume contemporaneous correlations are non-zero, so $Q = T$.

In corporate, we sometimes assume within-firm correlations are non-zero, so $Q = N$.

(I tend to be skeptical of this, since it assumes $\text{cov}(\varepsilon_{it}, \varepsilon_{jt}) = 0$.

Let's see how clustered std errors do in the presence of a firm fixed effect.

Petersen (2006)

To fix ideas, we'll work w/ 1 RHS variable and everything demeaned.

If we estimate a pooled regression $y_{it} = \beta x_{it} + \varepsilon_{it}$

$$\hat{\beta}_{OLS} = \frac{S_{xy}}{S_{xx}} = \frac{\sum_i \sum_t x_{it} y_{it}}{\sum_i \sum_t x_{it} x_{it}} = \beta + \frac{\sum_i \sum_t x_{it} \varepsilon_{it}}{\sum_i \sum_t x_{it}^2}$$

If the errors are iid.

$$var(\hat{\beta}_{OLS}) = \frac{\sigma_\varepsilon^2}{NT\sigma_x^2}$$

Now let's assume there's a firm fixed effect, but we continue to estimate pooled.

$$\varepsilon_{it} = \gamma_i + \eta_{it}$$

Also assume

$$x_{it} = \mu_i + \nu_{it} \quad , \text{with } \gamma, \eta, \mu, \nu \text{ all indep of each other}$$

and correlation across observations for the same firm

$$corr(x_{it}, x_{js}) = \begin{cases} \rho_x = \frac{\sigma_\mu^2}{\sigma_x^2} & \forall\, i=j,\ t\neq s \\ 0 & \forall\, i \neq j \end{cases}$$

$$corr(\varepsilon_{it}, \varepsilon_{js}) = \begin{cases} \rho_\varepsilon = \frac{\sigma_\gamma^2}{\sigma_\varepsilon^2} & \forall\, i=j,\ t\neq s \\ 0 & \forall\, i \neq j \end{cases}$$

what is $\hat{\beta}_{OLS}$ under this setup?

$$var(\hat{\beta}_{OLS}) = E\left[\left(\sum_i\sum_t x_{it}\varepsilon_{it}\right)^2\left(\sum_i\sum_t x_{it}^2\right)^{-2}\right]$$

with independence across firms:

$$= E\left[\sum_i\left(\sum_t x_{it}\varepsilon_{it}\right)^2\left(\sum_i\sum_t x_{it}^2\right)^{-2}\right] \qquad (*)$$

$$= E\left[\sum_i\left(\sum_t x_{it}^2\varepsilon_{it}^2 + \sum_{t\neq s}\sum x_{it}x_{is}\varepsilon_{it}\varepsilon_{is}\right)\left(\sum_i\sum_t x_{it}^2\right)^{-2}\right]$$

$$= \left(NT\sigma_x^2\sigma_\varepsilon^2 + NT(T-1)\rho_x\sigma_x^2\rho_\varepsilon\sigma_\varepsilon^2\right)\left(NT\sigma_x^2\right)^{-2}$$

$$var(\hat{\beta}_{OLS}) = \frac{\sigma_\varepsilon^2}{NT\sigma_x^2}\left(1 + (T-1)\rho_x\rho_\varepsilon\right)$$

Note that in the presence of a firm fixed effect, the OLS variance is too low.

Intuition: consider the extreme case of perfect correlation across time ($\rho_x = 1$, $\rho_\varepsilon = 1$)

Another year of data won't help at all in reality

but OLS assumes we get another N observations (indep.), shrinking s.e.'s

Here, clustered standard errors are

$$\hat{var}(\hat{\beta}) = \frac{\sum_i \left(\sum_t x_{it} \varepsilon_{it} \right)^2}{\left(\sum_i \sum_t x_{it}^2 \right)^2}$$

and you can see these are correct, in that they match eqn (*) on the previous page

what happens if we run Fama-MacBeth regressions?

Recall, T cross-sectional regs, and $\hat{\beta}_{FM} = \frac{1}{T} \sum_t \hat{\beta}_t = \frac{1}{T} \sum_t \left(\frac{\sum_i x_{it} y_{it}}{\sum_i x_{it}^2} \right)$

with $\hat{var}(\hat{\beta}_{FM}) = \frac{1}{T} \sum_t \frac{(\hat{\beta}_t - \hat{\beta}_{FM})^2}{T-1}$

Recall that this assumes indep. over time, or equiv. $E(x_{it} \varepsilon_{it} \varepsilon_{is} x_{is}) = 0 \ \forall t \neq s$

This isn't true if there is a firm effect in the data, in which case

$$var(\hat{\beta}_{FM}) = \frac{1}{T^2} Var\left(\sum_t \hat{\beta}_t \right)$$

$$= \frac{Var(\hat{\beta}_t)}{T} + \frac{\sum\sum_{t \neq s} cov(\hat{\beta}_t, \hat{\beta}_s)}{T^2}$$

$$= \frac{var(\hat{\beta}_t)}{T} + \frac{T(T-1)}{T^2} cov(\hat{\beta}_t, \hat{\beta}_s)$$

with our firm effect

$$cov(\hat{\beta}_t, \hat{\beta}_s) = E\left[\left(\sum_i x_{it}^2 \right)^{-1} \left(\sum_i x_{it} \varepsilon_{it} \right) \left(\sum_i x_{is} \varepsilon_{is} \right) \left(\sum_i x_{is}^2 \right)^{-1} \right]$$

$$= (N \sigma_x^2)^{-2} E\left[\left(\sum_i x_{it} \varepsilon_{it} \right) \left(\sum_i x_{is} \varepsilon_{is} \right) \right]$$

$$= (N \sigma_x^2)^{-2} E\left[\sum_i x_{it} x_{is} \varepsilon_{it} \varepsilon_{is} \right] = (N \sigma_x^2)^{-2} N \rho_x \sigma_x^2 \rho_\varepsilon \sigma_\varepsilon^2$$

So
$$var(\hat\beta_{FM}) = \frac{var(\hat\beta_t)}{T} + \frac{T-1}{T}\left(\frac{\rho_x \rho_\varepsilon \sigma_\varepsilon^2}{N\sigma_x^2}\right)$$

using $var(\hat\beta_t) = \frac{\sigma_\varepsilon^2}{N\sigma_x^2}$ we get:

$$var(\hat\beta_{FM}) = \frac{\sigma_\varepsilon^2}{TN\sigma_x^2}\left(1 + (T-1)\rho_x\rho_\varepsilon\right)$$

and hey!, this is exactly the $var(\hat\beta_{OLS})$

Famous people have made this mistake, e.g. Fama-French (2001)

who use FM on $DIV_{it} = \alpha + \beta_1 (M/B)_{it} + \beta_2 (E/A)_{it} + \beta_3 \, SIZE_{it} + \varepsilon_{it}$

indicator variable set to 1 if firm pays a divd

(almost perfectly persistent!)

~~So... what should we do if there's a time effect instead?~~

Can we fix this correlation problem with Newey-West?

original NW designed to account for autocorrelation in a single time series

modified for use in panels to estimate only within-cluster autocorrelations

(which means a maximum lag of T-1)

Recall that the covariance at lag j ($\varepsilon_t \varepsilon_{t-j}$) is scaled by wt. $1 - \frac{j}{M+1}$ maximum lag

If $M = T-1$, then the middle part of the NW s.e.'s is

$$\sum_i \left(\sum_t x_{it}\varepsilon_{it}\right)^2 = \sum_i \left(\sum_t x_{it}^2 \varepsilon_{it}^2 + \sum_{t\neq s}\sum W(t-s)\, x_{it} x_{is} \varepsilon_{it}\varepsilon_{is}\right)$$

$$= \sum_i \left(\sum_t x_{it}\varepsilon_{it}^2 + 2\sum_{t=1}^{T}\sum_{j=1}^{T-t}\left(1-\frac{j}{T}\right) x_{it} x_{i,t-j}\, \varepsilon_{it}\varepsilon_{i,t-j}\right)$$

Note that this doesn't match the clustered std. errors because of the weighting function. It turns out that in a panel context with a permanent firm effect we don't want declining weights, so we should exclude the weighting function what if there is a time effect instead? i.e. there's contemporaneous C-S corr.

$$\mathrm{corr}(X_{it}, X_{jt}) \neq 0 \quad \text{and} \quad \mathrm{corr}(\varepsilon_{it}, \varepsilon_{jt}) \neq 0$$

This is much closer to most asset pricing exercises

Analytically, just exchange N and T everywhere above and proceed.

Similar results:

OLS std errors are too small

clustered std errors are fine (clustered by time) as long as enough clusters

(can be a problem if the panel is short)

now FM works fine (this is what it was designed for)

what if we think there's a time effect <u>and</u> a firm effect? eek!

one approach: add dummy variables for each time period

cluster by firm to calculate s.e.'s

this works well if the time effect is the same for all firms

another approach:

$$\mathrm{var}(\hat{\beta}_{\text{firm \& time}}) = \underbrace{\mathrm{var}(\hat{\beta}_{\text{firm}})}_{\substack{\text{captures unspecified} \\ \text{correlation within firm}}} + \underbrace{\mathrm{var}(\hat{\beta}_{\text{time}})}_{\substack{\text{captures unspecified} \\ \text{correlation in C-S}}} - \underbrace{\mathrm{var}(\hat{\beta}_{\text{white}})}_{\substack{\text{both other terms have the} \\ \text{same diagonal elements,} \\ \text{so we have to subtract one}}}$$

This seems to work well if there are enough clusters.