Online Appendix to "The Granular Origins of Aggregate Fluctuations"

Xavier Gabaix

August 13, 2009

Attenuation bias in the granular residual

I analyze the properties of the granular residual. The conclusion is that it suffers from attenuation bias, but the bias goes to 0 as the number of firm K becomes large. I did not take a very large number of firms in the analysis, because this introduced new difficulties – the homogeneity assumption (55) is likely to be a less good approximation.

I consider first a one-factor model (no industry shocks). For firm i:

$$g_{it} = a_t + \varepsilon_{it} \tag{55}$$

where a_t is a common shock, and ε_{it} is an idiosyncratic shock. The granular residual is:

$$\Gamma_K = \frac{\sum_{i=1}^{K} S_i \left(g_i - \overline{g}\right)}{\sum_{i=1}^{K} S_t} \tag{56}$$

while the econometrician would like to know the "ideal" granular residual – a weighted mean of the idiosyncratic shocks of the top K firms.

$$\Gamma_K^* = \frac{\sum_{i=1}^{K} S_i \varepsilon_i}{\sum_{i=1}^{K} S_t} \tag{57}$$

(the specific choice of the denominator does not matter here, as I investigate the R^2's, and R^2 do not change when one multiplies some variables by a constant).

GDP growth follows, as in the model of section 3:

$$y_t = \phi \Gamma_{Kt}^* + u_t \tag{58}$$

where u_t is a disturbance orthogonal to $\left(\varepsilon_{it}\right)_{i=1...K}$. One would like to know how much R^2 of the idiosyncratic shocks of the top K firms explain, i.e. the R^2 of the *ideal* granular residual:

$$R^2_{\Gamma_K^*} = \frac{cov\left(y_t, \Gamma_{Kt}^*\right)^2}{var\left(y_t\right) var\left(\Gamma_{Kt}^*\right)} \tag{59}$$

The empirical analysis only gives the R^2 of the granular residual Γ :

$$R^2_{\Gamma_K} = \frac{cov\left(y_t, \Gamma_{Kt}\right)^2}{var\left(y_t\right) var\left(\Gamma_{Kt}\right)} \tag{60}$$

Econometrically, the situation is tricky, because economically, a_t is correlated with Γ^*_K.

A quantity of interest is the Herfindahl of the top K firms:

$$H_K = \frac{\sum_{i=1}^K S_i^2}{\left(\sum_{i=1}^K S_i\right)^2} \tag{61}$$

By the Cauchy-Schwartz inequality, $KH_K \geq 1$.

Lemma 2 *The R^2 of the granular residual is a downward biased estimate of the R^2 of the ideal granular residual, by a factor $1 - \frac{1}{KH_K}$.*

$$R^2_{\Gamma_K} = R^2_{\Gamma^*_K}\left(1 - \frac{1}{KH_K}\right)$$

Proof. We first observe that, by rescaling, it is enough to analyze the case where $\sigma_\varepsilon = 1$. I call $\sum_{i=1}^K S_i = s$. Call $\overline{X} = K^{-1}\sum_{i=1}^K X_i$ for a variable X. Then: $\Gamma = \sum_{i=1}^K \left(S_i - s/K\right)\varepsilon_i$, which gives, dropping the K subscripts when there is no ambiguity:

$$\Gamma^*_t = \Gamma_t + \overline{\varepsilon_t} \text{ with } cov\left(\Gamma_t, \overline{\varepsilon_t}\right) = 0$$

which means that Γ is a noisy proxy for Γ^*. Also

$$cov\left(\Gamma^*_t, \Gamma_t\right) = var\Gamma_t = \left(\sum_{i=1}^K S_i\right)^2 \left(H - \frac{1}{K}\right)$$
$$var\Gamma^*_t = \left(\sum_{i=1}^K S_i\right)^2 H$$

and

$$R^2_\Gamma = \frac{cov\left(y, \Gamma_t\right)^2}{vary \cdot var\left(\Gamma_t\right)} = \frac{\phi^2 cov\left(\Gamma_t, \Gamma^*_t\right)^2}{vary \cdot var\left(\Gamma_t\right)} = \frac{\phi^2\left(var\Gamma^*\right)^2\left(1 - \frac{1}{HK}\right)^2}{vary \cdot var\left(\Gamma_t\right)} = \frac{cov\left(y, \Gamma^*_t\right)^2}{vary \cdot var\Gamma^*}\left(1 - \frac{1}{HK}\right)^2 \frac{var\Gamma^*}{var\Gamma}$$
$$= R^2_{\Gamma^*}\left(1 - \frac{1}{HK}\right)^2 \frac{H}{H - \frac{1}{K}} = R^2_{\Gamma^*}\left(1 - \frac{1}{HK}\right).$$

■

Empirically, for the $K = 100$, firms,$\left(1 - \frac{1}{KH_K}\right) = 2/3$. Hence if empirically the $R^2_{\Gamma_K} = 1/3$, the R^2 of the ideal granular residual is $R^2_{\Gamma^*_K} = 1/2$. This bias is an attenuation bias, as the granular residual is a noisy proxy for the ideal granular residual.

If the distribution is very concentrated, then $H_K \gg 1/K$. Formally, the proof of Proposition 2 shows that if the Pareto exponent of the distribution is $1 \leq \zeta < 2$, $KH_K \sim K^{2-2/\zeta}$, so $\lim_{K\to\infty} (KH_K)^{-1} = 0$, and as $K \to \infty$, $R^2_{Y,\Gamma_K}/R^2_{Y,\Gamma^*_K} \to 1$. This is the sense in which, for large K, the granular residual identifies the explanatory power of the ideal granular residual.

The same reasoning applies, with messier expressions, with industry specific shocks, model: $g_{it} = a_t + a_{I_i} + \varepsilon_{it}$. The R^2 of the Γ^{ind} is a downward estimate R^2 of the ideal granular residual $\Gamma^{*,ind}$. The bias goes to 0 as the number of firms becomes large.