D Estimating the parameters of the incentive function

(to be published on the web only)

Our exercise to estimate the parameters of our flow-performance function is standard with the only main difference that we transform the measures of flows and performance, so they correspond more closely to our theoretical framework.[30] Mutual fund holdings data are obtained from the CDA/Spectrum database by Thompson Financial. Mutual funds's total net assets, net monthly returns, expense ratios and other fund characteristics are obtained from the CRSP mutual fund database. For net returns and expense ratios we use the TNA-weighted averages across all share classes. Quarterly market returns are obtained from Kenneth French's website. Although our data is quarterly or monthly, we annualize our variables for our regressions. We use data between 1981-2009. To select domestic equity funds we follow Barber, Odean and Zheng (2005). In particular, we select funds with the following IOC objectives: aggressive growth, growth and income, long-term growth, or total return. If these objectives are missing, we select funds with the following Strategic Insight's fund objectives: aggressive growth, growth and income, growth, income growth, or small company growth. If both IOC and Strategic Insight's objectives are missing, we select funds with the following Weisenberger fund types: AAL, AGG, G, G-I, G-I-S, G-S, G-S-I, GCI, GRI, GRO, I-G, I-G-S, I-S, I-S-G, MCG, SCG, or TR. We drop funds where all of these variables are missing. Furthermore, we drop funds with total net assets less than 5 million dollars.[31]

Recall that we want to estimate the parameters corresponding to n_1,n_2,k in our theoretical specification

$$\ln \frac{w^M_{t+1}}{w^M_{t+1,-}} = \ln \frac{w^M_{t+1}}{\rho_{t+1}\left(\alpha^M_t\right)\left(1-\psi^M_t\right) w^M_t} =$$
$$= \ln \Gamma_t Z_B + (n_B - 1)\ln \frac{\rho_{t+1}\left(\alpha^M_t\right)}{\frac{q_{t+1}+\delta_{t+1}}{q_t}} + \left[(n_A - n_B)\, 1_{\upsilon_t \geq \kappa}\right]\left(\ln \frac{\rho_{t+1}\left(\alpha^M_t\right)}{\frac{q_{t+1}+\delta_{t+1}}{q_t}} - \kappa\right).$$

For this we rewrite this equation as

$$FL_t = FE_t + (n_B - 1)\, R^e_t + (n_A - n_B)\, 1_{R^e \geq k}\left(R^e - k\right) \tag{59}$$

[30]We are very grateful to Dong Lou for the compilation of the appropriate datasets and Eszter Nagy for research assistance. All codes are available from the authors on request.

[31]As robustness checks we repeat our excercise on samples restricted to (1) funds with assets not less than 1M instead of 5, (2) funds older than 2 years, (3) funds younger than 10 years, (4) pikcing only growth and aggressive growth funds. The results are virtually unchanged except with the difference that considering only young funds make the flow-performance relationship stronger (the estimated values for $(n_A - n_B)$ are around 1). In this sense, our estimates are on the conservative side.

where

$$
\begin{aligned}
FL_t &= \ln\left(\frac{TNA_t}{TNA_{t-1}R^{w/o}} - MRG_t\right) \\
R_t^e &= \ln R_t^w - \ln R_t^M \\
k &= \ln \kappa
\end{aligned}
$$

with $R^{w/o}$,R^w are the fund's gross return net of fees and including fees, respectively, R^M is the gross market return, MRG_{t-1} is the increase in TNA due to fund mergers in year t and FE_t are time fixed effects. We fix a k and estimate (59) as a fixed-effect panel regression by clustering standard errors on the fund level and including the age of the fund and turnover of the fund as controls. We reestimate this regression for different values of k in the range $[-0.1, 0.2]$, record the parameter estimates, $p-$values implied by the t-statistics and within-R^2 measures.

The following figure shows the results.

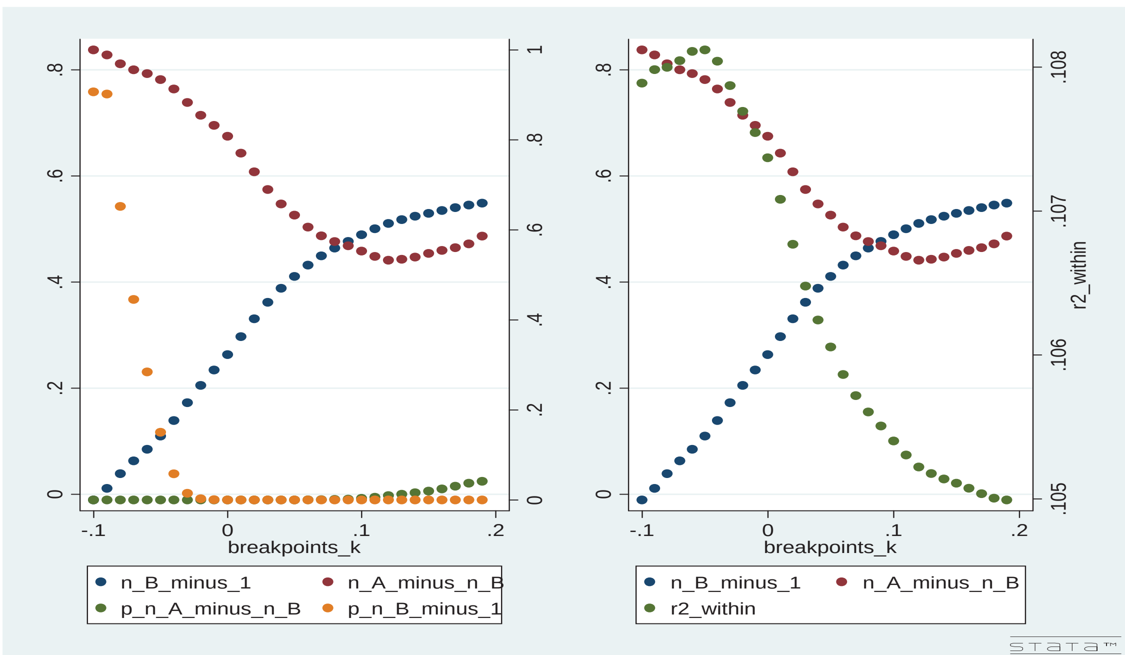

Figure 7: The left panel shows the estimated values of the coefficients, $(n_B - 1), (n_A - n_B)$ and the corresponding $p-$values as a function of k, the log of the breakpoint, while the right panel shows the coefficients and the within-R^2 measure as a function of k.

The left panel shows the estimated values of the coefficients, $(n_B - 1), (n_A - n_B)$ and the corresponding $p-$values as a function of k, while the right panel shows the coefficients and the within-R^2 measure as a function of k. Considering only the theoretically plausable range of $k > 0$, it is apparent that estimates for $(n_A - n_B)$ are within the range $[0.4, 0.7]$ while estimates for $(n_B - 1)$

are within the range $[0.25, 0.6]$ As long as $k \in (0, 0.11)$ both coefficients are significant on the 1% level. Looking at the right axis on the right panel shows that although the relevant R^2 is monotonically decreasing for positive k values, the decrease is only on the level of the third decimal. That, is based on R^2 any k in the given range perform very similarly. Considering these results, we conclude that estimates corresponding to the range $k \in (0, 0.11)$ are all valid starting points for our numerical example. We pick $k = \ln(1 + 0.05)$ arbitrarily, but consider perturbations around this value numerically.

The exact test statistics to the regression with our choice of $k = \ln(1 + 0.05)$ are in the following table.

```
Fixed-effects (within) regression               Number of obs      =     13115
Group variable: wficn                           Number of groups   =      1896

R-sq:  within  = 0.1061                         Obs per group: min =         1
       between = 0.0195                                        avg =       6.9
       overall = 0.0343                                        max =        21

                                                F(30,1895)         =     37.99
corr(u_i, Xb)  = -0.6237                        Prob > F           =    0.0000

                               (Std. Err. adjusted for 1896 clusters in wficn)
------------------------------------------------------------------------------
             |               Robust
 logpercflow |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       n_B-1 |   .4104857   .0561357     7.31   0.000     .3003916    .5205799
     n_A-n_B |   .5260528   .1336498     3.94   0.000     .2639365    .7881691
         age |  -.0154047   .0133265    -1.16   0.248    -.0415408    .0107314
    turnover |   .0064846   .0380249     0.17   0.865    -.0680905    .0810597
       timeFE| yes
       _cons |   .2705087   .3229564     0.84   0.402    -.3628788    .9038962
-------------+----------------------------------------------------------------
     sigma_u |  .30503114
     sigma_e |  .45128025
         rho |  .31359848   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```