# Online appendix to
## *Non-Standard Errors*

The online appendix contains the following additional material:

- Table OA.1: Summary statistics stage-2 dispersion in RT-results.

- Table OA.2: Summary statistics stage-3 dispersion in RT-results.

- Table OA.3: Summary statistics stage-4 dispersion in RT-results.

- Table OA.4: Principal component analysis team quality.

- Table OA.5: Stage-1 quantile regressions with all team quality variables.

- Table OA.6 Dispersion in research team beliefs.

- Table OA.7: Statistical significance under multiple testing.

- Figure OA.1: Distribution of the median/mean estimate.

- Figure OA.2: Countries of #fincap community.

- Figure OA.3: Dispersion estimates across all feedback stages.

- Figure OA.4: Multiverse fit.

- Figure OA.5: Distribution decisions at the forks.

- Figure OA.6: Details on the data sample.

- Figure OA.7: Sign-up sheet research team.

- Figure OA.8: Non-disclosure agreement research teams.

- Figure OA.9: Non-disclosure agreement peer evaluators.

- Figure OA.10: Registration as a team member.

- Figure OA.11: Instruction sheet for research teams.

- Figure OA.12: Form to submit stage 1 results.

- Figure OA.13: Form to submit stage 2 results.

- Figure OA.14: Form to submit stage 3 results.

- Figure OA.15: Belief elicitation sheet.

- Figure OA.16: Instruction sheet peer evaluators.

- Figure OA.17: The form that RTs completed for multiverse analysis.

**Table OA.1: Summary statistics stage-2 dispersion in RT-results**

This table mirrors panel (c) in Table 1, except this dispersion pertains
to stage-2 results instead of stage 1 results.

| | RT-H1 Efficiency | RT-H2 RSpread | RT-H3 Client Volume | RT-H4 Client RSpread | RT-H5 Client MOrders | RT-H6 Client GTR |
|---|---|---|---|---|---|---|
| *Estimate (yearly change, %)* | | | | | | |
| Mean | 451.2 | -1,122.2 | -3.6 | -38,254.7 | 0.4 | -37.1 |
| SD | 5,817.1 | 14,531.4 | 9.2 | 490,025.9 | 7.5 | 264.7 |
| Min | -291.3 | -186,074.5 | -117.5 | -6,275,383.0 | -30.8 | -3,024.9 |
| Q(0.10) | -13.1 | -7.7 | -3.8 | -5.8 | -1.8 | -83.4 |
| Q(0.25) | -4.4 | -4.7 | -3.7 | -2.7 | -0.6 | -9.4 |
| Median | -1.2 | -0.9 | -3.3 | 0.0 | -0.0 | -0.0 |
| Q(0.75) | 0.3 | 2.5 | -2.1 | 3.5 | 0.2 | 2.1 |
| Q(0.90) | 3.4 | 12.6 | -0.3 | 14.2 | 1.1 | 33.7 |
| IQR (i.e., NSE) | 4.7 | 7.1 | 1.7 | 6.2 | 0.8 | 11.6 |
| IDR | 16.5 | 20.2 | 3.5 | 20.0 | 2.8 | 117.1 |
| Max | 74,491.1 | 1,098.0 | 4.8 | 870.2 | 86.1 | 486.5 |
| *Standard error* | | | | | | |
| Mean | 462.2 | 1,166.4 | 3.5 | 38,279.4 | 2.0 | 86.3 |
| SD | 5,811.0 | 14,710.6 | 29.6 | 489,931.2 | 8.0 | 308.4 |
| Min | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Q(0.10) | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Q(0.25) | 0.4 | 0.5 | 0.1 | 0.7 | 0.1 | 0.3 |
| Median | 1.4 | 3.3 | 0.4 | 3.0 | 0.5 | 4.9 |
| Q(0.75) | 5.8 | 8.4 | 1.9 | 9.0 | 1.4 | 45.5 |
| Q(0.90) | 29.0 | 28.5 | 2.0 | 24.7 | 2.8 | 160.7 |
| IQR | 5.4 | 7.8 | 1.8 | 8.3 | 1.2 | 45.2 |
| IDR | 28.9 | 28.4 | 2.0 | 24.7 | 2.8 | 160.7 |
| Max | 74,425.5 | 188,404.1 | 378.8 | 6,274,203.0 | 86.1 | 2,740.2 |
| *t-value* | | | | | | |
| Mean | -4.8 | 24.9 | -99.4 | 28.8 | -3.5 | -0.1 |
| SD | 28.4 | 434.5 | 627.1 | 400.5 | 73.8 | 4.4 |
| Min | -322.3 | -907.7 | -7,208.7 | -160.0 | -876.2 | -38.3 |
| Q(0.10) | -7.3 | -8.4 | -45.4 | -5.0 | -3.7 | -2.0 |
| Q(0.25) | -2.3 | -2.5 | -17.4 | -1.2 | -0.8 | -0.8 |
| Median | -0.8 | -0.4 | -3.5 | 0.0 | -0.1 | -0.0 |
| Q(0.75) | 0.5 | 0.7 | -1.7 | 1.0 | 1.0 | 1.0 |
| Q(0.90) | 2.3 | 1.6 | -0.4 | 2.5 | 3.1 | 1.4 |
| IQR | 2.8 | 3.3 | 15.7 | 2.3 | 1.7 | 1.8 |
| IDR | 9.6 | 10.0 | 45.0 | 7.5 | 6.8 | 3.4 |
| Max | 30.8 | 5,479.5 | 56.1 | 5,120.8 | 318.8 | 25.2 |

**Table OA.2: Summary statistics stage-3 dispersion in RT-results**

This table mirrors panel (c) in Table 1, except this dispersion pertains to stage-3 results instead of stage 1 results.

| | RT-H1 Efficiency | RT-H2 RSpread | RT-H3 Client Volume | RT-H4 Client RSpread | RT-H5 Client MOrders | RT-H6 Client GTR |
|---|---|---|---|---|---|---|
| *Estimate (yearly change, %)* | | | | | | |
| Mean | 453.4 | -1,130.6 | -3.1 | -38,263.4 | -2.4 | -2.4 |
| SD | 5,816.9 | 14,530.5 | 10.7 | 490,025.2 | 36.2 | 105.9 |
| Min | -70.5 | -186,074.0 | -117.5 | -6,275,383.0 | -452.9 | -898.7 |
| Q(0.10) | -6.7 | -8.0 | -3.9 | -7.1 | -1.7 | -15.1 |
| Q(0.25) | -3.2 | -5.7 | -3.8 | -3.4 | -0.6 | -0.5 |
| Median | -1.0 | -1.8 | -3.3 | -0.3 | -0.0 | 0.0 |
| Q(0.75) | -0.0 | 0.0 | -1.3 | 0.8 | 0.2 | 1.4 |
| Q(0.90) | 2.2 | 5.5 | -0.4 | 5.3 | 0.9 | 12.8 |
| IQR (i.e., NSE) | 3.2 | 5.8 | 2.4 | 4.1 | 0.7 | 1.8 |
| IDR | 8.9 | 13.6 | 3.5 | 12.4 | 2.6 | 28.0 |
| Max | 74,491.1 | 1,098.0 | 66.7 | 302.4 | 86.1 | 486.5 |
| *Standard error* | | | | | | |
| Mean | 458.6 | 1,156.2 | 3.0 | 38,264.0 | 4.2 | 40.5 |
| SD | 5,811.3 | 14,711.3 | 29.6 | 489,932.4 | 36.8 | 149.5 |
| Min | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Q(0.10) | 0.1 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 |
| Q(0.25) | 0.3 | 0.4 | 0.1 | 0.4 | 0.1 | 0.0 |
| Median | 0.6 | 1.1 | 0.2 | 1.2 | 0.3 | 1.6 |
| Q(0.75) | 2.1 | 3.8 | 0.7 | 3.4 | 0.7 | 8.0 |
| Q(0.90) | 7.9 | 10.0 | 2.0 | 10.1 | 1.9 | 58.5 |
| IQR | 1.8 | 3.5 | 0.6 | 3.0 | 0.6 | 8.0 |
| IDR | 7.8 | 9.9 | 1.9 | 10.1 | 1.9 | 58.5 |
| Max | 74,425.5 | 188,404.0 | 378.8 | 6,274,203.0 | 463.7 | 1,149.3 |
| *t-value* | | | | | | |
| Mean | -3.7 | 25.2 | -56.5 | 29.8 | -3.3 | 0.3 |
| SD | 12.5 | 434.4 | 363.3 | 400.3 | 73.2 | 3.3 |
| Min | -131.7 | -908.3 | -3,800.0 | -160.0 | -876.2 | -10.2 |
| Q(0.10) | -7.9 | -8.8 | -36.5 | -5.5 | -4.2 | -1.6 |
| Q(0.25) | -3.8 | -5.1 | -28.5 | -2.9 | -1.5 | -0.6 |
| Median | -1.8 | -1.2 | -11.1 | -0.3 | -0.0 | 0.1 |
| Q(0.75) | 0.1 | 0.3 | -3.1 | 0.9 | 1.0 | 1.0 |
| Q(0.90) | 2.2 | 1.6 | -1.4 | 2.5 | 3.5 | 1.6 |
| IQR | 3.8 | 5.4 | 25.4 | 3.9 | 2.5 | 1.6 |
| IDR | 10.0 | 10.5 | 35.0 | 8.0 | 7.7 | 3.2 |
| Max | 11.9 | 5,479.5 | 56.1 | 5,120.8 | 318.8 | 25.2 |

**Table OA.3: Summary statistics stage-4 dispersion in RT-results**

This table mirrors panel (c) in Table 1, except this dispersion pertains to stage-4 results instead of stage 1 results.

| | RT-H1 Efficiency | RT-H2 RSpread | RT-H3 Client Volume | RT-H4 Client RSpread | RT-H5 Client MOrders | RT-H6 Client GTR |
|---|---|---|---|---|---|---|
| *Estimate (yearly change, %)* | | | | | | |
| Mean | 453.5 | 1,138.6 | -1.8 | -38,263.2 | -2.9 | 4.9 |
| SD | 5,816.9 | 14,529.9 | 9.6 | 490,025.2 | 35.4 | 71.2 |
| Min | -70.5 | -90.1 | -6.9 | -6,275,383.0 | -452.9 | -360.7 |
| Q(0.10) | -6.2 | -6.9 | -3.8 | -5.8 | -1.3 | -5.0 |
| Q(0.25) | -2.8 | -4.4 | -3.8 | -2.0 | -0.5 | -0.2 |
| Median | -1.1 | -2.3 | -2.9 | -0.2 | 0.0 | 0.0 |
| Q(0.75) | -0.2 | -0.1 | -2.0 | 0.4 | 0.1 | 0.8 |
| Q(0.90) | 1.2 | 2.2 | -1.1 | 3.6 | 0.8 | 5.7 |
| IQR (i.e., NSE) | 2.6 | 4.3 | 1.7 | 2.4 | 0.6 | 1.1 |
| IDR | 7.4 | 9.1 | 2.7 | 9.5 | 2.1 | 10.8 |
| Max | 74,491.1 | 186,074.5 | 117.5 | 302.4 | 7.1 | 486.5 |
| *Standard error* | | | | | | |
| Mean | 457.9 | 1,155.3 | 3.0 | 38,261.5 | 3.5 | 24.7 |
| SD | 5,811.4 | 14,711.4 | 29.6 | 489,932.6 | 36.2 | 88.8 |
| Min | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Q(0.10) | 0.1 | 0.1 | 0.1 | 0.0 | 0.1 | 0.0 |
| Q(0.25) | 0.3 | 0.6 | 0.1 | 0.6 | 0.1 | 0.1 |
| Median | 0.5 | 1.2 | 0.3 | 1.3 | 0.3 | 2.0 |
| Q(0.75) | 1.5 | 3.0 | 0.5 | 2.7 | 0.6 | 5.2 |
| Q(0.90) | 5.2 | 7.0 | 1.8 | 5.5 | 1.2 | 46.6 |
| IQR | 1.2 | 2.4 | 0.4 | 2.1 | 0.4 | 5.1 |
| IDR | 5.1 | 7.0 | 1.8 | 5.5 | 1.1 | 46.6 |
| Max | 74,425.5 | 188,404.1 | 378.8 | 6,274,203.0 | 463.7 | 786.1 |
| *t-value* | | | | | | |
| Mean | -3.7 | 25.1 | -54.7 | 29.7 | -3.6 | 0.6 |
| SD | 12.1 | 434.4 | 363.4 | 400.2 | 73.0 | 6.4 |
| Min | -131.7 | -911.2 | -3,801.4 | -159.8 | -876.2 | -9.0 |
| Q(0.10) | -7.5 | -8.0 | -33.5 | -3.6 | -3.6 | -1.3 |
| Q(0.25) | -3.1 | -4.0 | -18.5 | -1.8 | -1.1 | -0.3 |
| Median | -2.0 | -1.9 | -11.5 | -0.2 | 0.0 | 0.1 |
| Q(0.75) | -0.4 | -0.4 | -4.1 | 0.3 | 1.0 | 0.8 |
| Q(0.90) | 1.8 | 0.8 | -1.7 | 1.8 | 2.9 | 1.4 |
| IQR | 2.7 | 3.6 | 14.3 | 2.1 | 2.1 | 1.1 |
| IDR | 9.4 | 8.8 | 31.8 | 5.4 | 6.5 | 2.7 |
| Max | 8.0 | 5,479.5 | 19.5 | 5,120.8 | 318.8 | 80.2 |

**Table OA.4: Principal component analysis team quality**

This table presents the results of a principal component analysis of the standardized team quality variables.

Panel (a): Correlation team quality measures

|  | Publications | Experience | Big Data | Position | #Members |
|---|---|---|---|---|---|
| Publications |  | 0.34 | 0.10 | 0.54 | 0.30 |
| Experience |  |  | -0.18 | 0.25 | 0.12 |
| Big Data |  |  |  | 0.14 | 0.14 |
| Position |  |  |  |  | 0.16 |

Panel (b): Fraction of variance explained

|  | PC1 | PC2 | PC3 | PC4 | PC5 |
|---|---|---|---|---|---|
| Variance explained | 38.3% | 23.6% | 17.1% | 12.4% | 8.6% |

Panel (c): Loading of principal components on variables

|  | Publications | Experience | Big Data | Position | #Members |
|---|---|---|---|---|---|
| PC1 | 0.61 | 0.40 | 0.13 | 0.55 | 0.37 |
| PC2 | -0.01 | -0.55 | 0.79 | 0.05 | 0.26 |
| PC3 | -0.10 | 0.06 | -0.21 | -0.46 | 0.86 |
| PC4 | -0.20 | 0.71 | 0.56 | -0.35 | -0.12 |
| PC5 | -0.76 | 0.14 | -0.02 | 0.60 | 0.22 |

## Table OA.5: Stage-1 quantile regressions with all team quality variables

This table replicates Table 3 in main paper. The only difference is that the first principal component of the team quality variables is replaced by the quality variables themselves. $^*/^{**}$ correspond to significance at the 5/0.5% level, respectively.

| | Q(0.10) | Q(0.25) | Q(0.50) | Q(0.75) | Q(0.90) |
|---|---|---|---|---|---|
| Top publications (standardized/scaled) | -0.200$^{**}$ (0.063) | -0.213$^{**}$ (0.020) | 0.000 (0.009) | 0.056$^{**}$ (0.013) | 0.648$^{**}$ (0.037) |
| Experience in field (standardized/scaled) | -0.370$^{**}$ (0.040) | 0.031$^*$ (0.015) | 0.000 (0.007) | 0.002 (0.012) | -0.106$^{**}$ (0.028) |
| Experience with big data (standardized/scaled) | 0.074 (0.048) | -0.025 (0.015) | 0.000 (0.007) | 0.030$^*$ (0.011) | 0.138$^{**}$ (0.027) |
| Academic seniority (standardized/scaled) | 1.666$^{**}$ (0.058) | 0.171$^{**}$ (0.019) | 0.001 (0.008) | -0.026$^*$ (0.013) | -1.022$^{**}$ (0.035) |
| Team size (1 or 2 members) (standardized/scaled) | 0.819$^{**}$ (0.044) | 0.068$^{**}$ (0.015) | 0.004 (0.007) | -0.061$^{**}$ (0.011) | -0.150$^{**}$ (0.028) |
| Reproducibility score (standardized/scaled) | 0.494$^{**}$ (0.043) | 0.186$^{**}$ (0.015) | -0.001 (0.007) | -0.116$^{**}$ (0.011) | -0.521$^{**}$ (0.028) |
| Average rating (standardized/scaled) | 0.521$^{**}$ (0.041) | 0.158$^{**}$ (0.014) | -0.001 (0.007) | -0.067$^{**}$ (0.011) | -0.486$^{**}$ (0.029) |
| Dummy RT-H1 Efficiency | -31.553$^{**}$ (1.032) | -6.779$^{**}$ (0.354) | -1.107$^{**}$ (0.164) | 0.844$^{**}$ (0.271) | 9.262$^{**}$ (0.664) |
| Dummy RT-H2 RSpread | -22.463$^{**}$ (0.990) | -4.543$^{**}$ (0.356) | -0.032 (0.165) | 3.846$^{**}$ (0.270) | 20.961$^{**}$ (0.647) |
| Dummy RT-H3 Client Volume | -6.548$^{**}$ (1.007) | -3.776$^{**}$ (0.354) | -3.315$^{**}$ (0.164) | -2.361$^{**}$ (0.270) | 0.023 (0.661) |
| Dummy RT-H4 Client RSpread | -18.079$^{**}$ (0.952) | -2.875$^{**}$ (0.351) | 0.158 (0.164) | 4.365$^{**}$ (0.270) | 18.189$^{**}$ (0.645) |
| Dummy RT-H5 Client MOrders | -3.169$^{**}$ (1.007) | -0.853$^*$ (0.354) | -0.003 (0.164) | 0.286 (0.270) | 1.817$^{**}$ (0.625) |
| Dummy RT-H6 GTR | -178.868$^{**}$ (0.993) | -20.853$^{**}$ (0.351) | -0.020 (0.166) | 5.672$^{**}$ (0.265) | 60.704$^{**}$ (0.676) |
| #Observations | 984 | 984 | 984 | 984 | 984 |

**Table OA.6: Dispersion in research team beliefs**

This presents test statistics on whether the beliefs of research teams about dispersion across research teams matches realized dispersion. More precisely, the test statistic is defined as the difference between the average belief on the SD across teams, minus the realized SD, where this difference is divided by the realized SD. The *p*-values in parentheses are obtained by bootstrapping 1000 times. $^{*}/^{**}$ correspond to significance at the 5/0.5% level, respectively.

| | RT-H1 Efficiency | RT-H2 RSpread | RT-H3 Client Volume | RT-H4 Client RSpread | RT-H5 Client MOrders | RT-H6 Client GTR | All |
|---|---|---|---|---|---|---|---|
| Estimate | -99.5%$^{**}$ (0.00) | -95.4%$^{**}$ (0.00) | -9.0% (0.74) | -97.5%$^{**}$ (0.00) | -45.3% (0.63) | -83.3%$^{**}$ (0.00) | -71.7%$^{**}$ (0.00) |

## Table OA.7: Statistical significance under multiple testing

This table tests how many stage-4 estimates are statistically significant, where significance levels for the individual tests are adjusted to account for multiple testing. The number of significantly negative tests and significantly positive tests is reported in brackets. The reported family $p$-value is the probability that out of all test statistics, at least one is larger than the reported value, under the null of a multivariate normal with zero means, and a covariance matrix with squared SEs (reported by the RTs) on the diagonal and off-diagonals that are either zero (Bonferroni) or based on the multiverse analysis (Section 3.3).

### Panel (a): Multiple tests (Bonferroni)

|        | Reject at 0.5%? | $p$-value of family test | Mean (SD) correlation test statistics | Effective number of tests |
|--------|-----------------|--------------------------|---------------------------------------|---------------------------|
| RT-H1  | Yes (31, 4)     | < 0.0001                 | 0.00 (0.00)                           | 164                       |
| RT-H2  | Yes (38, 3)     | < 0.0001                 | 0.00 (0.00)                           | 164                       |
| RT-H3  | Yes (123, 2)    | < 0.0001                 | 0.00 (0.00)                           | 164                       |
| RT-H4  | Yes (15, 8)     | < 0.0001                 | 0.00 (0.00)                           | 164                       |
| RT-H5  | Yes (13, 9)     | < 0.0001                 | 0.00 (0.00)                           | 164                       |
| RT-H6  | Yes (3, 3)      | < 0.0001                 | 0.00 (0.00)                           | 164                       |

### Panel (b): Multiple tests (based on multiverse analysis)

|        | Reject at 0.5%? | $p$-value of family test | Mean (SD) correlation test statistics | Effective number of tests |
|--------|-----------------|--------------------------|---------------------------------------|---------------------------|
| RT-H1  | Yes (32, 5)     | < 0.0001                 | 0.01 (0.24)                           | 73                        |
| RT-H2  | Yes (47, 3)     | < 0.0001                 | 0.03 (0.28)                           | 41                        |
| RT-H3  | Yes (125, 2)    | < 0.0001                 | 0.14 (0.53)                           | 13                        |
| RT-H4  | Yes (15, 8)     | < 0.0001                 | 0.03 (0.28)                           | 91                        |
| RT-H5  | Yes (16, 11)    | < 0.0001                 | 0.10 (0.46)                           | 21                        |
| RT-H6  | Yes (3, 3)      | < 0.0001                 | 0.01 (0.25)                           | 51                        |

**Figure OA.1: Distribution of the median/mean estimate**

This figure plots the empirical probability density functions (PDFs) of the median estimate and the mean estimate across 164 researchers. They are plotted based on 1000 draws from a multivariate normal with means equal to the realized #fincap medians, and a covariance matrix with squared SEs (reported by the RTs) on the diagonal and off-diagonals that are based on the multiverse analysis (Section 3.3). This distribution is also used to generate panel (b) of Table 2.
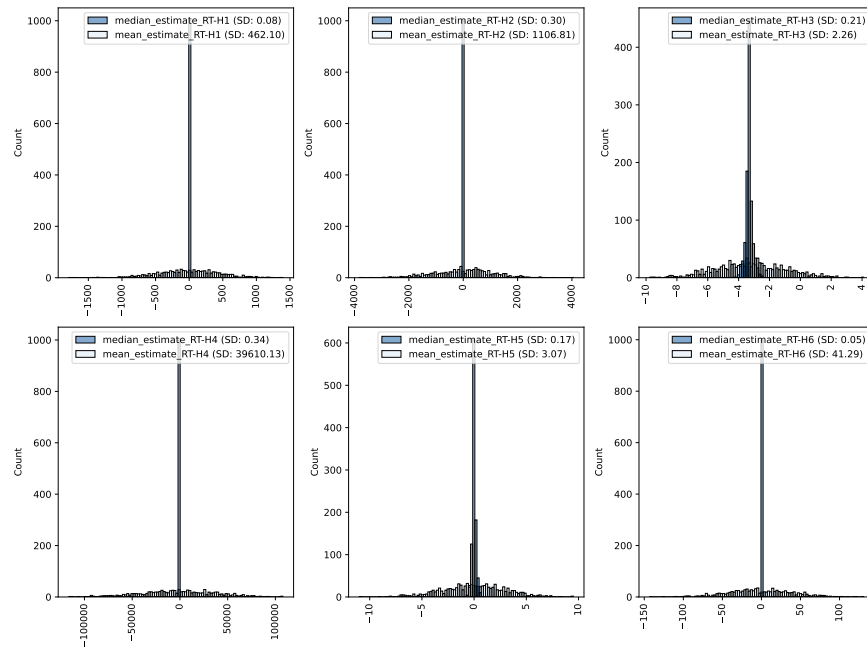
## Figure OA.2: Countries of #fincap community

This plot illustrates how the #fincap community is dispersed around the globe. The top plot depicts how the members of the research teams are dispersed across countries. The bottom plot does the same for the peer evaluators.

Residence of research-team members

Residence of peer evaluators
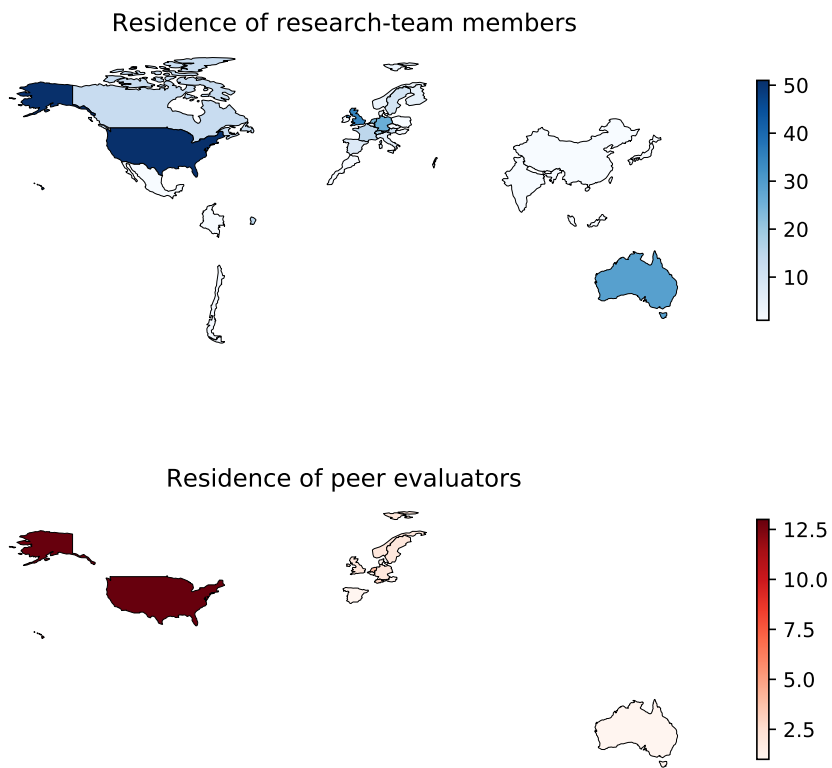
**Figure OA.3: Dispersion estimates across all feedback stages**

This plot illustrates the dispersion in RT estimates across all feedback stages. The boxes depict the first and third quartile. The horizontal line in the box corresponds to the median. The whiskers depict the 2.5% and 97.5% quantile.
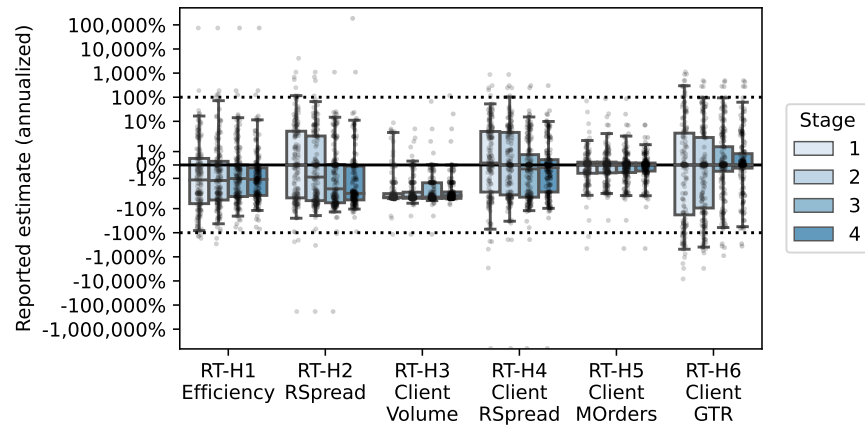
**Figure OA.4: Multiverse fit**

This figure plots the quality of the fit between the paths in the multiverse analysis and the paths that RTs followed when conducting their analyis. More specifically, for the multiverse analysis, each RT was asked to select among a set of pre-defined alternatives for a number of key forks on the analyis path (see Table 5). After selecting an alternative, they were they were asked to rate how close the selected alternative was to what they actually did on a scale from 1 ("Far from what we did.") to 5 ("Very close to what we did."). The graph plots the overall average across all RTs and all forks.
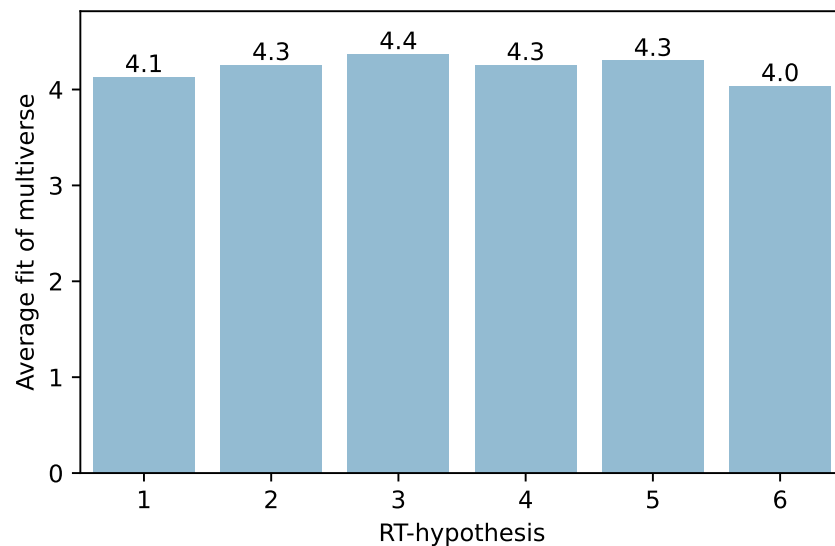
# Figure OA.5: Distribution decisions at the forks

This figure plots the distribution of decisions for all forks in the multiverse. It depicts the information that is also available in Table 5.

**Figure OA.6: Details on the data sample**

# about the data

The data pertain to 17 years (2002-2018) of trading of [EuroStoxx 50](#) futures, which are among the world's most actively traded index derivatives. They give investors exposure to "Europe," or, more precisely, to a basket of euro-area blue-chip equities. All trading is done through an electronic limit-order book (see, e.g., [Parlour and Seppi, 2008](#)). Please find more background information on the futures in this [factsheet](#).

The data consist of 720 million trade records and will be made available in monthly gzipped semicolon separated text files ("csv"). Each zipped monthly file is no larger than 50 MB. The data is clean in the sense that for all files the format is identical. Please find below the first ten lines of the December 2018 file as an example.

```
DATETIME; EXPIRATION; BUY_SELL_ID; TRADE_SIZE; MATCH_PRICE; AGGRESSOR_FLAG;ACCOUNT_ROLE; EXEC_TYPE_ID
2018-12-03 08:00:06.400; 201812; S; 2; 3229; N; A; F
2018-12-03 08:00:06.410; 201812; S; 1; 3229; N; A; F
2018-12-03 08:00:06.410; 201812; S; 1; 3229; N; A; F
2018-12-03 08:00:06.410; 201812; B; 4; 3229; Y; A; F
2018-12-03 08:00:06.540; 201812; S; 1; 3229; N; A; F
2018-12-03 08:00:06.550; 201812; B; 2; 3229; Y; A; F
2018-12-03 08:00:06.550; 201812; S; 1; 3229; N; A; F
2018-12-03 08:00:06.630; 201812; B; 1; 3229; Y; A; F
2018-12-03 08:00:06.630; 201812; S; 1; 3229; N; A; F
```

The variables are defined as follows (the characterizations are short and therefore imprecise, please refer to any standard textbook on futures to get a detailed description of what futures are and how they are traded):

- **DATETIME:** Time stamp of the trade denoted as YYYY-MM-DD hh:mm:ss.sss where ss.sss denotes seconds up to a third decimal (i.e., the precision is tens of milliseconds as the last digit is always zero).
- **EXPIRATION:** The expiration date of the futures contract being traded. All data pertain to Eurex trading in EuroStoxx 50 (SX5E) futures contracts. Expiration months are: March, June, September, and December. Contracts expire on the third Friday of the expiration month. The notation of expiration is YYYYMM (where MM is in [03, 06, 09, 12]).
- **BUY_SELL:** This indicator shows if the trade record is for a buyer "B" (who goes long the index) or for a seller "S" (who goes short the index).
- **TRADE_SIZE:** This is the size of the trade expressed in number of contracts. The contract value per index point is EUR 10 (e.g., per contract traded, the long side is entitled to receive 10 euro from the short side of the trade each time the index increases by one point).
- **MATCH_PRICE:** The price at which the trade between buyer and the seller is concluded (i.e., the long and the short side of the trade, respectively).
- **AGGRESSOR_FLAG:** If the trade record pertains to a market order (or marketable limit order) that is executed against a standing limit order, this flag takes the value "Y". If the record pertains to a limit order, resting in the book before being matched with an incoming market order, or to an order in an auction (e.g., the opening and closing auction), then this flag takes the value "N". This flag became available as of November 2009.
- **ACCOUNT_ROLE:** This variable is either:
  - A: Agency trade (i.e., a trade an exchange member does for a client).
  - M: Market-maker principal trade (i.e., a trade an exchange member does for his own account in his role as market maker).
  - P: Non-market-maker principal trade (i.e., a trade an exchange member does for his own account).
  - P.S.: The distinction between M and P is not an economically meaningful one for the purpose of this project.
- **EXEC_TYPE_ID:** This variable is:
  - F if the full order was executed in the trade.
  - P if the order was only partially executed in the trade.
  - N if not assigned.

**Figure OA.7: Sign-up sheet research team**

# sign-up as a research team

Please fill in the following form if you wish to participate as a research team (RT). You can either participate **alone** or **in a team of two** (in case you want to join the project as a team of two researchers, the form below only needs to be submitted once). Upon submission of the sign-up form, a link to the entry survey will be forwarded to the e-mail address(es) you enter below. Please make sure that the entry survey is filled out by each team member within three days. Looking forward to collaborating with you in this project!

---

**team member #1:**

✎  full name: *

> first name, last name

✉  e-mail address: *

> e-mail address the entry survey will be sent to

---

**team member #2 (optional):**

✎  full name:

> first name, last name

✉  e-mail address:

> e-mail address the entry survey will be sent to

---

**informed consent:** *

I understand that the project requires research expertise and experience in empirical finance (with a particular focus on market liquidity) and the analysis of large datasets. Also, at least one of the team members has to hold a PhD in finance or economics. I consent to elaborate on why our team fulfills the requirements for participating in the study in the form field below.

☐  I agree

---

**fulfillment of RT requirements:** *

Please provide a brief explanation why you (and your potential team member) are sufficiently skilled to participate as a research team in #fincap. In particular, briefly sketch your background in empirical finance research, research methodology, etc. Please note that two to three sentences are sufficient.

> Our team fulfills the participation requirements because …

**Figure OA.8: Non-disclosure agreement research teams**

# non-disclosure agreement for research team members

Whereas,

Data Supplier intends to provide records of executions of the EuroStoxx® future by means of a temporary electronic access via a shared drive to Data Recipient ("Data").

The Data comprises the following semicolon separated fields for executions in the simple instruments of the EuroStoxx® Future between 1st January 2002 and 31st December 2018:

| FieldName | Description | Format | Example |
|---|---|---|---|
| DATETIME | Date and Time of the execution | String: yyyy-mm-dd hh:mm:ss.000 | 2018-10-15 08:00:04.840 |
| EXPIRATION | Expiration months of the contract | Integer: yyyymm | 201809 |
| BUY_SELL_ID | Buy/Sell Indicator | Char: 1 = Buy, 2 = Sell | 1 |
| TRADE_SIZE | Number of contracts executed | Integer | 2 |
| MATCH_PRICE | Price of execution in points | Integer | 3176 |
| AGGRESSOR_FLAG | Flag indicating if an order was aggressive or passive | Char: Y = aggressive, N = not aggressive | Y |
| ACCOUNT_ROLE | Account Role | Char: M = MarketMaker, P = Proprietary, A = Agency | M |
| EXEC_TYPE_ID | Full/partial execution indicator | Char: F: Full execution, P: Partial Execution | F |

Therefore,

Data Recipient and Data Supplier agree as follows:

1. Data Recipient shall only use the Data for empirical analyses of the Data and/or for evaluations of other data recipients' analyses within a project where multiple decentral teams report their results and/or evaluations to the core research team (represented by Data Recipient) ("Project"). Data Recipient shall only use the Data for academic purposes and not make any commercial use of the Data in whatsoever form. The Data Recipient shall indicate the origin of the Data (i.e. Deutsche Börse AG, Mergenthalerallee 61, 65760 Eschorn) in any publication resulting from or in connection with the Project by a proof of source and shall provide Data Supplier with a voucher copy by e-mail (pdf) or mail.

2. Data Recipient shall not disclose or reveal any Data to any third party without the prior written consent of the Data Supplier.

3. Data Recipient shall use all reasonable efforts to keep the Data in confidence and to safeguard the Data. In so doing, the Data Recipient shall take at least the same precautions which it would take to safeguard its own similarly valued proprietary and confidential data, but shall in no event take less than commercially reasonable precautions.

4. Data Recipient acknowledges that Data Supplier does not make any representation, warranty or undertaking, express or implied, as to the accuracy, reliability, completeness or reasonableness of the Data.

5. Except as set forth in Section 1, this Agreement does not transfer to the Data Recipient or any other person any ownership, license, intellectual property rights, or other rights including, but not limited to, patent and copyright rights, in or to the Data. All Data shall remain the exclusive property of the Data Supplier or the third party which owns it.

6. The Data Recipient shall be allowed to access and use the Data for research purposes within the scope of the Project for a period of 18 months. **After the period of 18 months the Data Recipient shall promptly permanently destroy, erase or delete all Data in its possession or control**, in particular, but not limited to, from any computer, word processor, mobile telecommunication device or similar device into which it was stored or programmed, and provide to the Data Supplier written confirmation of such destruction, erasure or deletion.

7. No provision of this Agreement creates a partnership between the Parties or makes a Party the agent of the other Party for any purpose.

8. This Agreement, including this Section 8, may only be supplemented, amended, varied or modified in writing. In case any provision in or obligation under this Agreement shall be held invalid, illegal or unenforceable, the validity, legality and enforceability of the remaining provisions or obligations of this Agreement shall not be affected. The invalid, illegal or

unenforceable provision shall be replaced by a valid, legal and enforceable provision which reflects as close as possible the original intention of the Parties. The foregoing shall apply mutatis mutandis in case of any gaps of this Agreement.

9. This Agreement, and any non-contractual obligations arising out of or in connection with it, shall be governed by, construed and enforced in accordance with the laws of the Netherlands without reference to or inclusion of the principles of choice of law or conflicts of law of that jurisdiction and the Parties hereby submit to the jurisdiction of the courts of Amsterdam.

**Figure OA.9: Non-disclosure agreement peer evaluators**

# non-disclosure agreement for peer evaluators

Whereas,

Project Coordinator intends to pass on short scientific papers (**"Short Papers"**) to Peer Evaluator as part of a project where multiple decentral teams report their results and/or evaluations to the core research team (represented by Project Coordinator) (**"Project"**).

The Short Papers are written by various research teams (**"Research Teams"**), reporting their results on a predefined set of research questions as part of the Project.

Therefore,

Peer Evaluator and Project Coordinator agree as follows:

1. Peer Evaluator shall only use the Short Papers for the purpose of evaluating them within the scope of the Project.

2. Peer Evaluator shall not disclose or reveal any content of or any information related to the Short Papers to any third party. This applies particularly, but not limited to, the names of the members of the Research Teams.

3. Peer Evaluator shall use all reasonable efforts to keep the Short Papers in confidence and to safeguard the Short Papers. In so doing, the Peer Evaluator shall take at least the same precautions which it would take to safeguard its own similarly valued proprietary and confidential data, but shall in no event take less than commercially reasonable precautions.

4. The Peer Evaluator shall be allowed to access and use the Short Papers for evaluation purposes within the scope of the Project until 31 December, 2021. **After 31 December, 2021, the Peer Evaluator shall promptly permanently destroy, erase or delete all Short Papers in its possession or control**, in particular, but not limited to, from any computer, word processor, mobile telecommunication device or similar device into which it was stored or programmed.

5. No provision of this Agreement creates a partnership between the Parties or makes a Party the agent of the other Party for any purpose.

6. This Agreement, including this Section 6, may only be supplemented, amended, varied or modified in writing. In case any provision in or obligation under this Agreement shall be held invalid, illegal or unenforceable, the validity, legality and enforceability of the remaining provisions or obligations of this Agreement shall not be affected. The invalid, illegal or unenforceable provision shall be replaced by a valid, legal and enforceable provision which reflects as close as possible the original intention of the Parties. The foregoing shall apply mutatis mutandis in case of any gaps of this Agreement.

7. This Agreement, and any non-contractual obligations arising out of or in connection with it, shall be governed by, construed and enforced in accordance with the laws of the Netherlands without reference to or inclusion of the principles of choice of law or conflicts of law of that jurisdiction and the Parties hereby submit to the jurisdiction of the courts of Amsterdam.

**Figure OA.10: Registration as a team member**

# registration as a team member

To register as a research team member, please fill out the form below. Fields indicated with an asterisk are required. For your contact information, please make sure to provide the details in the way they should be added to the final paper.

✎ first name: *

first name

✎ middle name(s):

middle name(s)

✎ last name: *

last name

✉ e-mail address: *

e-mail address

🌐 webpage:

webpage url (incl. 'http://' or 'https://' prefix)

🪪 your ORCID:

if available (see https://orcid.org/)

🏛 full affiliation: *

department, organization, country

📅 years since PhD: *

please enter 0 if you do not have obtained your PhD yet

⚤ gender: *

please select...

👥 country of residence: *

country you currently work in

🎓 highest degree: *

please select...

*Other:*

highest degree: if other, please specify

field of education: *

please select...

*Other:*

field of education: if other, please specify

current position: *

please select...

*Other:*

current position: if other, please specify

What is your number of Google Scholar citations? (If you don't have a Google Scholar profile, please provide an approximate number of citations.) *

number of citations

Have you published (or had a paper accepted for publication) in at least one of the following journals: Journal of Finance, Journal of Financial Economics, Review of Financial Studies, American Economic Review, Econometrica, Journal of Political Economy, Review of Economic Studies, or Quarterly Journal of Economics? *

please select...

What is the largest dataset you have worked with so far (in terms of observations)? *

please select...

Do you consider the data to be analyzed in this project (652,000,000 observations) "big data"? *

please select...

expertise in empirical finance: *

please select...

expertise in market liquidity: *

please select...

any other comment:

anything you would like to tell us?

**Figure OA.11: Instruction sheet for research teams**

# instruction sheet for research teams

This three-page instruction sheet clarifies what is expected of you as a research team in the #fincap project. It first provides some context for the hypotheses you are expected to test, then presents the assignment, and finally lists the hypotheses you are asked to test with only the Deutsche Börse data that is made available to you by the #fincap team. These data contain trade information on the EuroStoxx 50 futures.

## Context

Electronic order matching systems (automated exchanges) and electronic order generation systems (algorithms) have changed financial markets over time. Investors used to trade through broker-dealers by paying the dealers' quoted ask prices when buying, and accepting their bid prices when selling. The wedge between dealer bid and ask prices, the bid-ask spread, was a useful measure of trading cost, and often still is.

Now, investors more commonly trade in electronic limit-order markets (as is the case for EuroStoxx 50 futures). They still trade at bid and ask prices. They do so by submitting so-called market orders and marketable limit orders. However, investors now also can quote bid and ask prices themselves by submitting (non-marketable) standing limit orders. Increasingly, investors now also use agency algorithms to automate their trades. Concurrently, exchanges have been continuously upgrading their systems to better serve their clients. Has market quality improved, in particular when taking the viewpoint of non-exchange members: (end-user) clients?

## Assignment

You are expected to write an academic paper that is maximum five pages long. To make that feasible you can skip many parts of a typical academic paper. You only need to do the following for all hypotheses listed below:

1. Propose a statistical measure, briefly motivate it, and present the formula to calculate it.
2. For this measure, estimate the average per-year change in percentage terms, based on the full sample (or at least the longest possible period because some series are not available yet at the beginning of the sample). Test it against the null of no change.
3. Report this estimate along with its standard error in four decimals (e.g., "measure Z declined by 1.251% with a standard error 0.241%").
4. Briefly discuss your result.

For example, an appropriate outcome statement for testing hypothesis X which states that Y has not changed is:

"We propose measure Z to test hypothesis X because [...]. It is calculated as Z = f(DATA). Implementing it leads to the following result: We reject the null of no change. We find that Y declined as our measure Z declined by 1.251% on average per year where the standard error of this change is 0.421% and the resulting t-statistic is 2.971. This result shows [...]"

We emphasize that you are asked to report your results in a brief manner. *If the paper is longer than five pages we will not consider the paper and we will have to exclude you as co-authors from the project.*

## Hypotheses

1. Assuming that informationally-efficient prices follow a random walk, did market efficiency change over time?

   **Null hypothesis 1:** Market efficiency has not changed over time.

2. Did the (realized) bid-ask spread paid on market orders change over time? The realized spread could be thought of as the gross-profit component of the spread as earned by the limit-order submitter.

   **Null hypothesis 2:** The realized spread on market orders has not changed over time.

*The remaining hypotheses focus on client trades only (i.e., trades implemented by exchange members on behalf of their clients).*

3. Did the share of client volume in total volume change over time?

   **Null hypothesis 3:** Client share volume as a fraction of total volume has not changed over time.

4. On their market orders and marketable limit orders, did the realized bid-ask spread that clients paid, change over time?

   **Null hypothesis 4:** Client realized spreads have not changed over time.

5. Realized spread is a standard cost measure for market orders, but to what extent do investors continue to use market and marketable limit orders (as opposed to non-marketable limit orders)?

**Null hypothesis 5:** The fraction of client trades executed via market orders and marketable limit orders has not changed over time.

6. A measure that does not rely on the classic limit- or market-order distinction is *gross trading revenue* (GTR). Investor GTR for a particular trading day can be computed by assuming a zero position at the start of the day and evaluating an end-of-day position at an appropriate reference price. Relative investor GTR can then be defined as this GTR divided by the investor's total (euro) volume for that trading day. This relative GTR is, in a sense, a realized spread. It reveals what various groups of market participants pay in aggregate for (or earn on) their trading. It transcends market structure as it can be meaningfully computed for any type of trading in any type of market (be it trading through limit-orders only, through market-orders only, through a mix of both, or in a completely different market structure).

**Null hypothesis 6:** Relative gross trading revenue (GTR) for clients has not changed over time.

**Figure OA.12: Form to submit stage 1 results**

# submit results for stage 1

Please read the following instructions for submitting your results carefully.

- Please enter your team ID **before** uploading your short paper and/or your analysis scripts. You will find your team ID in the e-mail notifications sent by the project collaborators. Each research team is supposed to submit their results **only once**. Please coordinate with your team member to avoid double submissions.

- Please upload your short paper in .PDF format. Make sure that the file does not exceed the page limit of **five pages**. Note that if your paper exceeds the page limit, we will not consider it and we will have to exclude you as co-authors from the project. **Click on the "upload file" button on the right to submit the PDF.**

- To ensure anonymity throughout the project, please make sure that **no identifying information** is contained in any of the submitted files. Please note that we will lift your anonymity only for the peer evaluation process (see peer evaluation for details). That is, neither the short paper nor the analysis scripts should allow drawing any inferences on the identity of the team members.

- Please bundle all analysis and data processing scripts in a single .ZIP, .RAR, or .7ZIP file. **Please upload only the script files, not the data.** Please also include a "readme" file (in .TXT format) with precise instructions on how to reproduce your estimates from the raw data (e.g., which files to put in what folder, the order in which to run the scripts, etc.). From experience, the best guidance we can give you is to literally follow your own instructions to see if the workflow you describe yields the desired results. **Click on the "upload file" button on the right to submit the file.**

- For each of the six null hypotheses, please provide the average per-year change in percentage terms (effect size) and the corresponding standard error. Please report all estimates with **three digits after the comma** (e.g., effect size = 1.234%, standard error = 0.567%).

---

🔽 team ID: *

| ABCD |
| --- |

| team ID not yet provided |
| --- |

🔼 short paper: *

| browse | | upload file |
| --- | --- | --- |

🔼 analysis scripts: *

| browse | | upload file |
| --- | --- | --- |

---

**upload status:**

*Please note that your results can be submitted only if all three bullet point below are indicated with a check mark. To do so, please provide your correct team ID first, select the two files AND press the "upload file" buttons on the right (for both files separately).*

- ⊗ *team ID not yet provided*
- ⊗ *short paper not yet uploaded*
- ⊗ *analysis scripts not yet uploaded*

---

**null hypothesis 1 (H1):** Market efficiency has not changed over time.

⤢ **H1:** effect size: *

| | % |
| --- | --- |

⎍ **H1:** standard error: *

| | % |
| --- | --- |

**null hypothesis 2 (H2):** The realized spread on market orders has not changed over time.

⤢ **H2:** effect size: *

| | % |
|---|---|

⎍ **H2:** standard error: *

| | % |
|---|---|

**null hypothesis 3 (H3):** Client share volume as a fraction of total volume has not changed over time.

⤢ **H3:** effect size: *

| | % |
|---|---|

⎍ **H3:** standard error: *

| | % |
|---|---|

**null hypothesis 4 (H4):** Client realized spreads have not changed over time.

⤢ **H4:** effect size: *

| | % |
|---|---|

⎍ **H4:** standard error: *

| | % |
|---|---|

**null hypothesis 5 (H5):** The fraction of client trades executed via market orders and marketable limit orders has not changed over time.

⤢ **H5:** effect size: *

| | % |
|---|---|

⎍ **H5:** standard error: *

| | % |
|---|---|

**null hypothesis 6 (H6):** Relative gross trading revenue (GTR) for clients has not changed over time.

⤢ **H6:** effect size: *

| | % |
|---|---|

⎍ **H6:** standard error: *

| | % |
|---|---|

**Figure OA.13: Form to submit stage 2 results**

# submit results for stage 2

#### 〰 team ID: *

ABCD

team ID not yet provided

#### ⬆ short paper: *

browse | | upload file

#### ⬆ analysis scripts: *

browse | | upload file

**upload status:**

*Please note that your results can be submitted only if all three bullet points below are indicated with a check mark. To do so, please provide your correct team ID first, select the two files AND press the "upload file" buttons on the right (for both files separately).*

- ⊗ *team ID not yet provided*
- ⊗ *short paper not yet uploaded*
- ⊗ *analysis scripts not yet uploaded*

**null hypothesis 1 (H1):** Market efficiency has not changed over time.

#### ⤢ H1: effect size: *

| | % |

#### � H1: standard error: *

%

**H1:** Direction of the effect: *

--- ▾

---

**null hypothesis 2 (H2):** The realized spread on market orders has not changed over time.

**H2:** effect size: *

| | % |

**H2:** standard error: *

| | % |

**H2:** Direction of the effect: *

--- ▾

---

**null hypothesis 3 (H3):** Client share volume as a fraction of total volume has not changed over time.

**H3:** effect size: *

| | % |

**H3:** standard error: *

| | % |

**H3:** Direction of the effect: *

--- ▾

---

**null hypothesis 4 (H4):** Client realized spreads have not changed over time.

**H4:** effect size: *

| | % |

**H4:** standard error: *

| | % |

**H4:** Direction of the effect: *

--- ▾

---

**null hypothesis 5 (H5):** The fraction of client trades executed via market orders and marketable limit orders has not changed over time.

**H5:** effect size: *

| | % |

**H5:** standard error: *

|  | % |

**H5:** Direction of the effect: *

| --- | ⌄ |

---

**null hypothesis 6 (H6):** Relative gross trading revenue (GTR) for clients has not changed over time.

**H6:** effect size: *

|  | % |

**H6:** standard error: *

|  | % |

**H6:** Direction of the effect: *

| --- | ⌄ |

---

**Comments and clarifications on the revision:**

**Figure OA.14: Form to submit stage 3 results**

# submit results for stage 3

Please read the following instructions for submitting your results carefully.

- Please enter your team ID **before** uploading your short paper and/or your analysis scripts. You will find your team ID in the e-mail notifications sent by the project collaborators. Each research team is supposed to submit their results **only once**. Please coordinate with your team member to avoid double submissions.

- Please upload your revised short paper in **.PDF format**. Make sure that the file does not exceed the page limit of **five pages**. Click on the "upload file" button on the right to submit the PDF.

- To ensure anonymity throughout the project, please make sure that **no identifying information** is contained in any of the submitted files. That is, neither the short paper nor the analysis scripts should allow drawing any inferences on the identity of the team members.

- Please bundle all analysis and data processing scripts in a single .ZIP, .RAR, or .7ZIP file. **Please upload only the script files, not the data.** Please also include a "readme" file (in .TXT format) with precise instructions on how to reproduce your estimates from the raw data (e.g., which files to put in what folder, the order in which to run the scripts, etc.). From experience, the best guidance we can give you is to literally follow your own instructions to see if the workflow you describe yields the desired results. **Click on the "upload file" button on the right to submit the file.**

- For each of the six null hypotheses, please provide the **average per-year change in percentage terms (the annualized effect size)** and the corresponding **standard error**. Please report all estimates with **at least three digits after the decimal point**. In cases where this would yield 0.000 or 0.00x (e.g. 0.003), please report your result with at least two more digits after the first non-zero digit (e.g., effect size = 0.000123%, standard error = 0.00345%). To aggregate the data, it is important for us to correctly align the signs of the reported effect sizes. We thus ask you to also **indicate the direction of your results** with respect to the corresponding standardized statements in the form.

- Please be aware that you can leave comments and clarifications on the revision in the open text field at the bottom of the form.

---

### ∿ team ID: *

ABCD

team ID not yet provided

### ⬆ short paper: *

browse | | upload file

### ⬆ analysis scripts: *

browse | | upload file

---

**upload status:**

*Please note that your results can be submitted only if all three bullet points below are indicated with a check mark. To do so, please provide your correct team ID first, select the two files AND press the "upload file" buttons on the right (for both files separately).*

- ⊗ *team ID not yet provided*
- ⊗ *short paper not yet uploaded*
- ⊗ *analysis scripts not yet uploaded*

---

**null hypothesis 1 (H1):** Market efficiency has not changed over time.

### ⬈ H1: effect size: *

| | % |

**H1:** standard error: *

| | % |
|---|---|

**H1:** Direction of the effect: *

| --- | ⌄ |
|---|---|

**null hypothesis 2 (H2):** The realized spread on market orders has not changed over time.

**H2:** effect size: *

| | % |
|---|---|

**H2:** standard error: *

| | % |
|---|---|

**H2:** Direction of the effect: *

| --- | ⌄ |
|---|---|

**null hypothesis 3 (H3):** Client share volume as a fraction of total volume has not changed over time.

**H3:** effect size: *

| | % |
|---|---|

**H3:** standard error: *

| | % |
|---|---|

**H3:** Direction of the effect: *

| --- | ⌄ |
|---|---|

**null hypothesis 4 (H4):** Client realized spreads have not changed over time.

**H4:** effect size: *

| | % |
|---|---|

**H4:** standard error: *

| | % |
|---|---|

**H4:** Direction of the effect: *

| --- | ⌄ |
|---|---|

**null hypothesis 5 (H5):** The fraction of client trades executed via market orders and marketable limit orders has not changed over time.

**H5:** effect size: *

| | % |
|---|---|

**H5:** standard error: *

|  | % |

**H5:** Direction of the effect: *

| --- | ⌄ |

---

**null hypothesis 6 (H6):** Relative gross trading revenue (GTR) for clients has not changed over time.

**H6:** effect size: *

|  | % |

**H6:** standard error: *

|  | % |

**H6:** Direction of the effect: *

| --- | ⌄ |

---

**Comments and clarifications on the revision:**

| |

**Figure OA.15: Belief elicitation sheet**

# belief elicitation

Please read the following instructions for submitting your beliefs carefully.

- Please enter your team ID. You will find your team ID in the e-mail notifications sent by the project collaborators. Each research team is supposed to submit their results **only once**. Please coordinate with your team member to avoid double submissions.

- Task:

  Imagine that, at the end of the project, you receive the short papers of all teams who completed all stages in the project (more than 100 research teams registered for #fincap and were given access to the data on January 11). **What do you predict the dispersion in results to be across these teams in Stage 1?**

  More specifically, for each hypothesis, we ask you to predict...

  1. the **standard deviation of reported effect size estimates** across teams in Stage 1 (i.e., imagine that you collect the effect size estimates $y_i$ of all teams in a single data series; then, what is the standard deviation of this series?)

  2. the **standard deviation of t-statistics** across teams in Stage 1 (the t-statistic is defined as the reported effect size estimate divided by the reported standard error).

  Recall that estimates of the **effect sizes were in terms of per-year percentage changes**. Thus, your prediction of the standard deviation across teams should be in the same units. Likewise, the prediction for the standard deviation in t-statistics should be provided in the equivalent units (i.e., in t units).

  In total you will make 12 predictions (the above two questions on each of the 6 hypotheses).

- Monetary rewards:

  **One out of five (20%) RTs** that complete all stages in the project will be randomly drawn for a monetary reward. If your RT is drawn for a monetary reward, **one out of the 12 prediction questions** will be randomly drawn to determine the monetary reward. If your prediction about the standard deviation in results across the teams is within +/- 50% of the observed value you will receive a **monetary reward of $300** and otherwise you will receive a reward of $0. If, for example, the observed standard deviation is X in the prediction question randomly drawn for payment, you will be paid a $300 reward if your predicted standard deviation is between 0.5X and 1.5X and you will get paid zero otherwise.

  The observed standard deviation will be computed based on the Stage 1 results of those RTs who completed all stages in the project. The monetary reward will be paid out as an Amazon gift card (if you are two persons in the RT, the reward will be split equally between you).

---

🖊 team ID: *

> ABCD

> team ID not yet provided

⊗ *team ID not yet provided*

---

**Null hypothesis 1 (H1):** Market efficiency has not changed over time.

⤢ **H1:** Standard deviation of effect size estimates: *

> | %

🔀 **H1:** Standard deviation of t-statistics: *

> | "t"

**Null hypothesis 2 (H2):** The realized spread on market orders has not changed over time.

**H2:** Standard deviation of effect size estimates: *

| | % |
|---|---|

**H2:** Standard deviation of t-statistics: *

| | "t" |
|---|---|

**Null hypothesis 3 (H3):** Client share volume as a fraction of total volume has not changed over time.

**H3:** Standard deviation of effect size estimates: *

| | % |
|---|---|

**H3:** Standard deviation of t-statistics: *

| | "t" |
|---|---|

**Null hypothesis 4 (H4):** Client realized spreads have not changed over time.

**H4:** Standard deviation of effect size estimates: *

| | % |
|---|---|

**H4:** Standard deviation of t-statistics: *

| | "t" |
|---|---|

**Null hypothesis 5 (H5):** The fraction of client trades executed via market orders and marketable limit orders has not changed over time.

**H5:** Standard deviation of effect size estimates: *

| | % |
|---|---|

**H5:** Standard deviation of t-statistics: *

| | "t" |
|---|---|

**Null hypothesis 6 (H6):** Relative gross trading revenue (GTR) for clients has not changed over time.

**H6:** Standard deviation of effect size estimates: *

| | % |
|---|---|

**H6:** Standard deviation of t-statistics: *

| | "t" |
|---|---|

**Figure OA.16: Instruction sheet peer evaluators**

# instruction sheet for peer evaluators

For each hypothesis and for the short paper, please rate the quality of the analysis, briefly clarify and provide suggestions for possible improvement (a paragraph is sufficient). For your feedback please follow the guidelines of the Journal of Finance: "(P)lease focus on what you see as central weaknesses in terms of (…) the chosen econometric strategy (…). (Y)ou should provide clear and concise reasons why you see your proposed revision as material (…)."

Please read the following information carefully:

- the instruction sheet that was given to the research teams (RT), click [here], and
- the FAQs that the RTs were alerted to, click [here].

Importantly, RTs could use any frequency of the data for their analyses (e.g., daily data, monthly data, or annual data), but they were asked to report effect sizes in per-year/annualized terms (to make them comparable across teams).

**The following list contains detailed instructions:**

- At the end of this e-mail, please find the list of links to the 10 short papers we kindly ask you to evaluate. The papers are named after the team IDs (e.g., `ABCD.pdf´).
- As outlined in the project description, you as a peer evaluator will remain anonymous to the research teams, but not the other way around. For each paper, there are two anonymous evaluators. The list below also contains, for each of your papers, the team ID, the name(s) of the team members, and your blinded evaluator ID (1 or 2).
- For your evaluations, please use the following file as a template: Evaluation Template.
- For each RT, please open a new version of the provided template.
- On the first page, enter the team ID and your evaluator ID for the paper as outlined in the attached list.
- Please use the designated cells to provide your ratings and your suggestions.
- Save the file as "TeamID_EvaluatorID.xlsx" according to the attached list (e.g., `ABCD_1.xlsx´).
- Please double-check that you have entered all six hypothesis ratings as well as the overall rating for each paper.
- Please send all 10 files to info@fincap.academy until before Friday, May 7, 2021, 23:59 GMT-12.

In case you have any questions, please do not reach out to the research teams or other potential peer evaluators. We are happy to answer any questions you might have on info@fincap.academy.

Thank you very much for your time and for your contribution to the #fincap project.

**Figure OA.17: The form that RTs completed for multiverse**

# #fincap RT decisions overview

This form is used to take stock of the key decisions that research teams (RTs) made in the FIRST STAGE of the #fincap project.

* Required

1. Please provide the four-letter code of your research team. *

_____

2. Who is/are the member(s) of the research team?  Please report first and last name. *

_____

| Stage-1 decisions for all hypotheses | The following five questions refer to decisions that affect all hypotheses.  Please tick six boxes, one for each hypothesis.  Please tick the alternative that you picked for your Stage-1 result (i.e., the result that you filled in on the reporting sheet). |
|---|---|

3. QUESTION 1: Did you remove parts of the sample to get rid of idiosyncrasies associated with the open and close of trading?  PLEASE TICK SIX BOXES, ONE FOR EACH HYPOTHESIS! *

*Mark only one oval per row.*

| | No. | Yes, we removed the first and the last 30 minutes of trading. |
|---|---|---|
| H1 | ◯ | ◯ |
| H2 | ◯ | ◯ |
| H3 | ◯ | ◯ |
| H4 | ◯ | ◯ |
| H5 | ◯ | ◯ |
| H6 | ◯ | ◯ |

4. How close were the alternatives that you picked in the previous question (Q1), to what you actually did? *

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Far from what we did. | ◯ | ◯ | ◯ | ◯ | ◯ | Very close to what we did. |

5. Comment (optional)

_____
_____
_____
_____

6. QUESTION 2: Did you exclude periods from the sample?  (Please tick six boxes, one for each hypothesis!) *

*Mark only one oval per row.*

| | No. | Yes, settlement weeks were excluded. |
|---|---|---|
| H1 | ◯ | ◯ |
| H2 | ◯ | ◯ |
| H3 | ◯ | ◯ |
| H4 | ◯ | ◯ |
| H5 | ◯ | ◯ |
| H6 | ◯ | ◯ |

7. How close were the alternatives that you picked in the previous question (Q2), to what you actually did? *

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Far from what we did. | ◯ | ◯ | ◯ | ◯ | ◯ | Very close to what we did. |

8. Comment (optional)

_____
_____
_____
_____

9. QUESTION 3: Did you deal with potential outliers in your final sample (i.e., the sample you used for your final analysis in Stage 1)?  (Please tick six boxes, one for each hypothesis!) *

*Mark only one oval per row.*

| | No. | Yes, we winsorized at 2.5 and 97.5 percentile. | Yes, we trimmed at 2.5 and 97.5 percentile (trimming=removing). |
|---|---|---|---|
| H1 | ◯ | ◯ | ◯ |
| H2 | ◯ | ◯ | ◯ |
| H3 | ◯ | ◯ | ◯ |
| H4 | ◯ | ◯ | ◯ |
| H5 | ◯ | ◯ | ◯ |
| H6 | ◯ | ◯ | ◯ |

10. How close were the alternatives that you picked in the previous question (Q3), to what you actually did? *

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Far from what we did. | ◯ | ◯ | ◯ | ◯ | ◯ | Very close to what we did. |

11. Comment (optional)

_____
_____
_____
_____

12. QUESTION 4: What is the frequency of your final sample (i.e., at what frequency did you conduct your final analysis in Stage 1)?  (Please tick six boxes, one for each hypothesis!) *

*Mark only one oval per row.*

| | Daily frequency. | Weekly frequency. | Monthly frequency. | Annual frequency. |
|---|---|---|---|---|
| H1 | ◯ | ◯ | ◯ | ◯ |
| H2 | ◯ | ◯ | ◯ | ◯ |
| H3 | ◯ | ◯ | ◯ | ◯ |
| H4 | ◯ | ◯ | ◯ | ◯ |
| H5 | ◯ | ◯ | ◯ | ◯ |
| H6 | ◯ | ◯ | ◯ | ◯ |

13. How close were the alternatives that you picked in the previous question (Q4), *
to what you actually did?

*Mark only one oval.*

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Far from what we did. | ◯ | ◯ | ◯ | ◯ | ◯ | Very close to what we did. |

14. Comment (optional)

_____
_____
_____
_____
_____

15. QUESTION 5: What is the statistical model/method that you used to estimate *
change? (Please tick six boxes, one for each hypothesis!)

*Mark only one oval per row.*

|  | Assume trend stationarity for the log measure, and estimate a linear model with trend. | "Log difference" the measure, and estimate its mean. | Compute percentage change of the measure, and estimate its mean. |
|---|---|---|---|
| H1 | ◯ | ◯ | ◯ |
| H2 | ◯ | ◯ | ◯ |
| H3 | ◯ | ◯ | ◯ |
| H4 | ◯ | ◯ | ◯ |
| H5 | ◯ | ◯ | ◯ |
| H6 | ◯ | ◯ | ◯ |

16. How close were the alternatives that you picked in the previous question (Q5), *
to what you actually did?

*Mark only one oval.*

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Far from what we did. | ◯ | ◯ | ◯ | ◯ | ◯ | Very close to what we did. |

17. Comment (optional)

_____
_____
_____
_____
_____

Stage-1 hypothesis-specific decisions

The remaining questions pertain to decisions for a subset of hypotheses. Please tick one box per hypothesis.

18. QUESTION 6: For H1, which measure did you use for market efficiency? *

*Mark only one oval.*

◯ R-squared of an autoregressive model for returns.
◯ Variance ratio for returns.

19. How close were the alternatives that you picked in the previous question (Q6), *
to what you actually did?

*Mark only one oval.*

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Far from what we did. | ◯ | ◯ | ◯ | ◯ | ◯ | Very close to what we did. |

20. Comment (optional)

_____
_____
_____
_____
_____

21. QUESTION 7: For H1, which frequencies did you use for computing your market *
efficiency statistic? (If you used the R-squared statistic, then the highest
frequency is the frequency at which you ran the regression, and the lowest
frequency drives the number of lags, e.g., second and minute frequencies imply
you sampled in seconds and used 60 lags.)

*Mark only one oval.*

◯ Highest frequency second; lowest frequency minute.
◯ Highest frequency minute; lowest frequency 5-minutes.
◯ Highest frequency 5-minutes; lowest frequency 30-minutes.
◯ Highest frequency day; lowest frequency week.
◯ Highest frequency day; lowest frequency month.

22. How close were the alternatives that you picked in the previous question (Q7), *
to what you actually did?

*Mark only one oval.*

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Far from what we did. | ◯ | ◯ | ◯ | ◯ | ◯ | Very close to what we did. |

23. Comment (optional)

_____
_____
_____
_____
_____

24. QUESTION 8: For H2, H4, and H5, did you use the "tick test" (e.g., Lee and *
Ready, 1991) to determine the aggressor side of the trade, or did you use the
aggressor flag and, therefore, accepted that you can only use part of the
sample? (Lee and Ready, 1991, "Inferring Trade Direction from Intraday Data")

*Mark only one oval per row.*

|  | We used the tick test. | We used the aggressor flag. |
|---|---|---|
| H2 | ◯ | ◯ |
| H4 | ◯ | ◯ |
| H5 | ◯ | ◯ |

25. How close were the alternatives that you picked in the previous question (Q8), *
to what you actually did?

*Mark only one oval.*

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Far from what we did. | ◯ | ◯ | ◯ | ◯ | ◯ | Very close to what we did. |

26. Comment (optional)

_____
_____
_____
_____
_____

27. QUESTION 9: For H2 and H4, which post-trade reference price did you use to compute the realized spread? *

*Mark only one oval per row.*

| | Price 5 minutes after the trade. | Price 10 minutes after the trade. | Price 30 minutes after the trade. |
|---|---|---|---|
| **H2** | ◯ | ◯ | ◯ |
| **H4** | ◯ | ◯ | ◯ |

28. How close were the alternatives that you picked in the previous question (Q9), to what you actually did? *

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Far from what we did. | ◯ | ◯ | ◯ | ◯ | ◯ | Very close to what we did. |

29. Comment (optional)

_____
_____
_____
_____

30. QUESTION 10: For H2 and H4, how did you aggregate across trades? *

*Mark only one oval per row.*

| | Trade-size-weighted average. | Equal-weighted average. |
|---|---|---|
| **H2** | ◯ | ◯ |
| **H4** | ◯ | ◯ |

31. How close were the alternatives that you picked in the previous question (Q10), to what you actually did? *

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Far from what we did. | ◯ | ◯ | ◯ | ◯ | ◯ | Very close to what we did. |

32. Comment (optional)

_____
_____
_____
_____
_____

33. QUESTION 11: For H3, which units did you use to compute volume shares? *

*Mark only one oval.*

◯ Volume expressed in number of contracts.
◯ Volume expressed in euro.

34. How close were the alternatives that you picked in the previous question (Q11), to what you actually did? *

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Far from what we did. | ◯ | ◯ | ◯ | ◯ | ◯ | Very close to what we did. |

35. Comment (optional)

_____
_____
_____
_____
_____

36. QUESTION 12: For H6, which reference price did you pick when computing "gross trading revenue" (GTR)? *

*Mark only one oval.*

◯ Last trade price in the day.
◯ Last trade price one day later.
◯ Volume-weighted average-price (VWAP) full-day.
◯ Volume-weighted average-price (VWAP) based on last five trades in the day.

37. How close were the alternatives that you picked in the previous question (Q12), to what you actually did? *

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Far from what we did. | ◯ | ◯ | ◯ | ◯ | ◯ | Very close to what we did. |

38. Comment (optional)

_____
_____
_____
_____
_____

39. QUESTION 13: For H6, to attenuate any complications arising from extreme outliers for daily GTRs, you might have aggregated using medians instead of means. Which aggregation approach did you use? *

*Mark only one oval.*

◯ Mean.
◯ Median.

40. How close were the alternatives that you picked in the previous question (Q13), to what you actually did? *

*Mark only one oval.*

| | 1 | 2 | 3 | 4 | 5 | |
|---|---|---|---|---|---|---|
| Far from what we did. | ◯ | ◯ | ◯ | ◯ | ◯ | Very close to what we did. |

41. Comment (optional)

_____
_____
_____
_____
_____

42. QUESTION 14: For H6, how did you deal with non-negative values for daily GTRs? (The idea is that client GTRs are expected to be negative to make intermediary GTRs positive, because they sum to zero. To make a series of mostly negative numbers, strictly negative, one option is to subtract the maximum of the series, and a little more to avoid one observation being zero. Another option is to set non-negative values to missing values.) *

*Mark only one oval.*

◯ If needed, we created a non-negative time series by applying f(X) = f(X) - max[f(X)] - 0.001.

◯ If needed, we created a non-negative time series by applying f(X) = f(X) - max[f(X)] - 1.

◯ If needed, we set non-negative values to missing values.

43. How close were the alternatives that you picked in the previous question (Q14), * to what you actually did?

*Mark only one oval.*

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Far from what we did. | ◯ | ◯ | ◯ | ◯ | ◯ | Very close to what we did. |

44. Comment (optional)

_____
_____
_____
_____
_____

45. QUESTION 15: For H6, when reporting average per-year change in percentages, * did you "retain the sign" of a negative trend? (For example, did you report a change from -1 to -2 as a negative number?)

*Mark only one oval.*

◯ Yes, we did retain the sign (in the example we would have reported a negative number).

◯ No, we did not retain the sign (in the example we would have reported a positive number).

46. How close were the alternatives that you picked in the previous question (Q15), * to what you actually did?

*Mark only one oval.*

|  | 1 | 2 | 3 | 4 | 5 |  |
|---|---|---|---|---|---|---|
| Far from what we did. | ◯ | ◯ | ◯ | ◯ | ◯ | Very close to what we did. |

47. Comment (optional)

_____
_____
_____
_____
_____