

## Cross-Company and Within-Company Features for Software Effort Estimation

### Analysis of features to define cross-company projects

the date of receipt and acceptance should be inserted later

Address(es) of author(s) should be given

| Dataset    | Criterion          | Subsets              | Subsets Size |
|------------|--------------------|----------------------|--------------|
| cocomo81   | project type       | cocomo81e            | 28           |
|            |                    | cocomo81o            | 24           |
|            |                    | cocomo81s            | 11           |
| nasa93     | development center | nasa93_center_1      | 12           |
|            |                    | nasa93_center_2      | 37           |
|            |                    | nasa93_center_5      | 39           |
| desharnais | language type      | desharnaisL1         | 46           |
|            |                    | desharnaisL2         | 25           |
|            |                    | desharnaisL3         | 10           |
| finnish    | application type   | finnishAppType1      | 17           |
|            |                    | finnishAppType2345   | 18           |
| kemerer    | hardware           | kemererHardware1     | 7            |
|            |                    | kemererHardware23456 | 8            |
| maxwell    | application type   | maxwellAppType1      | 10           |
|            |                    | maxwellAppType2      | 29           |
|            |                    | maxwellAppType3      | 18           |
| maxwell    | hardware           | maxwellHardware2     | 37           |
|            |                    | maxwellHardware3     | 16           |
|            |                    | maxwellHardware5     | 7            |
| maxwell    | source             | maxwellSource1       | 8            |
|            |                    | maxwellSource2       | 54           |

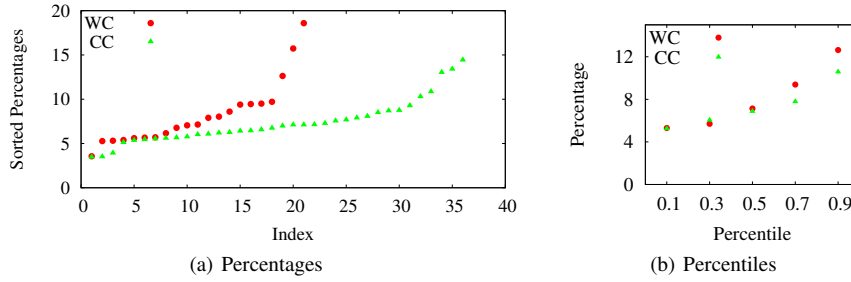
**Fig. 1** Selected datasets, their division criterion and the size of the subsets.

| Dataset              | MAR |    |    |        |        |     | MMRE |    |    |     |     |     | MdmRE |    |    |     |     |     | Pred(30) |    |    |     |     |    |
|----------------------|-----|----|----|--------|--------|-----|------|----|----|-----|-----|-----|-------|----|----|-----|-----|-----|----------|----|----|-----|-----|----|
|                      | W   | T  | L  | WC     | CC     | WC  | W    | T  | L  | WC  | CC  | WC  | W     | T  | L  | WC  | CC  | WC  | W        | T  | L  | WC  | CC  |    |
|                      |     |    |    |        |        |     |      |    |    |     |     |     |       |    |    |     |     |     |          |    |    |     |     | WC |
| cocomo81e            | 0   | 20 | 0  | 1.0E+3 | 1.1E+3 | 2.4 | 0    | 16 | 4  | 0.9 | 0.7 | 0.9 | 4     | 16 | 0  | 0.7 | 0.9 | 0.1 | 4        | 16 | 0  | 0.1 | 0.1 |    |
| cocomo81o            | 0   | 20 | 0  | 8.2E+2 | 8.1E+2 | 2.7 | 2    | 18 | 0  | 0.8 | 0.8 | 0.9 | 2     | 18 | 0  | 0.8 | 0.9 | 0.1 | 2        | 18 | 0  | 0.1 | 0.2 |    |
| cocomo81s            | 18  | 2  | 0  | 3.6E+1 | 1.8E+2 | 1.0 | 15   | 5  | 0  | 8.6 | 0.5 | 1.7 | 13    | 5  | 2  | 0.5 | 1.7 | 0.2 | 13       | 5  | 2  | 0.2 | 0.1 |    |
| nasa93_center_1      | 0   | 20 | 0  | 1.4E+2 | 1.3E+2 | 1.2 | 0    | 20 | 0  | 2.0 | 0.8 | 0.8 | 0     | 20 | 0  | 0.8 | 0.8 | 0.6 | 0        | 20 | 0  | 0.6 | 0.5 |    |
| nasa93_center_2      | 4   | 16 | 0  | 1.8E+2 | 2.1E+2 | 1.3 | 2    | 18 | 0  | 2.8 | 0.7 | 0.8 | 2     | 18 | 0  | 0.7 | 0.8 | 0.2 | 2        | 18 | 0  | 0.2 | 0.2 |    |
| nasa93_center_5      | 0   | 20 | 0  | 6.9E+2 | 8.9E+2 | 0.9 | 0    | 12 | 8  | 0.7 | 0.6 | 0.8 | 8     | 11 | 1  | 0.6 | 0.8 | 0.2 | 8        | 11 | 1  | 0.2 | 0.2 |    |
| desharnaisL1         | 11  | 9  | 0  | 9.9E+2 | 2.0E+3 | 0.6 | 9    | 11 | 0  | 2.4 | 0.4 | 1.7 | 9     | 11 | 0  | 0.4 | 1.7 | 0.4 | 9        | 11 | 0  | 0.4 | 0.3 |    |
| desharnaisL2         | 0   | 20 | 0  | 2.8E+3 | 2.8E+3 | 0.5 | 0    | 20 | 0  | 0.6 | 0.5 | 0.5 | 0     | 20 | 0  | 0.5 | 0.5 | 0.2 | 0        | 20 | 0  | 0.2 | 0.3 |    |
| desharnaisL3         | 0   | 20 | 0  | 2.8E+3 | 3.2E+3 | 0.5 | 2    | 18 | 0  | 0.5 | 0.4 | 0.5 | 2     | 18 | 0  | 0.4 | 0.5 | 0.2 | 2        | 18 | 0  | 0.2 | 0.2 |    |
| finnishAppType1      | 0   | 20 | 0  | 3.2E+3 | 3.8E+3 | 1.1 | 0    | 20 | 0  | 1.0 | 0.5 | 0.6 | 0     | 20 | 0  | 0.5 | 0.6 | 0.3 | 0        | 20 | 0  | 0.3 | 0.2 |    |
| finnishAppType2345   | 0   | 20 | 0  | 7.1E+3 | 5.4E+3 | 2.2 | 0    | 17 | 3  | 0.9 | 0.8 | 0.7 | 0     | 17 | 3  | 0.8 | 0.7 | 0.1 | 0        | 17 | 3  | 0.1 | 0.2 |    |
| kemererHardware1     | 0   | 0  | 20 | 1.4E+2 | 5.5E+1 | 1.3 | 0    | 0  | 20 | 0.3 | 1.1 | 0.3 | 0     | 0  | 20 | 1.1 | 0.3 | 0.4 | 0        | 0  | 20 | 0.4 | 0.5 |    |
| kemererHardware23456 | 0   | 20 | 0  | 2.0E+2 | 2.0E+2 | 0.7 | 0    | 20 | 0  | 0.7 | 0.6 | 0.5 | 0     | 20 | 0  | 0.6 | 0.5 | 0.1 | 0        | 20 | 0  | 0.1 | 0.1 |    |
| maxwellAppType1      | 6   | 14 | 0  | 1.4E+3 | 3.2E+3 | 0.8 | 1    | 19 | 0  | 1.9 | 0.4 | 0.7 | 1     | 19 | 0  | 0.4 | 0.7 | 0.3 | 1        | 19 | 0  | 0.3 | 0.3 |    |
| maxwellAppType2      | 0   | 18 | 2  | 6.6E+3 | 5.4E+3 | 1.2 | 0    | 19 | 1  | 0.9 | 0.5 | 0.4 | 0     | 19 | 1  | 0.5 | 0.4 | 0.3 | 0        | 19 | 1  | 0.3 | 0.3 |    |
| maxwellAppType3      | 0   | 20 | 0  | 5.6E+3 | 6.6E+3 | 1.0 | 1    | 19 | 0  | 1.0 | 0.5 | 0.6 | 1     | 19 | 0  | 0.5 | 0.6 | 0.2 | 1        | 19 | 0  | 0.2 | 0.2 |    |
| maxwellHardware2     | 0   | 20 | 0  | 5.6E+3 | 5.3E+3 | 0.8 | 0    | 20 | 0  | 1.0 | 0.5 | 0.5 | 0     | 20 | 0  | 0.5 | 0.5 | 0.2 | 0        | 20 | 0  | 0.2 | 0.2 |    |
| maxwellHardware3     | 0   | 20 | 0  | 5.3E+3 | 5.9E+3 | 0.9 | 0    | 20 | 0  | 0.7 | 0.4 | 0.5 | 0     | 20 | 0  | 0.4 | 0.5 | 0.3 | 0        | 20 | 0  | 0.3 | 0.4 |    |
| maxwellHardware5     | 0   | 20 | 0  | 3.6E+3 | 3.6E+3 | 3.7 | 0    | 20 | 0  | 2.8 | 0.7 | 0.8 | 0     | 20 | 0  | 0.7 | 0.8 | 0.1 | 0        | 20 | 0  | 0.1 | 0.1 |    |
| maxwellSource1       | 6   | 14 | 0  | 1.5E+3 | 3.3E+3 | 0.3 | 1    | 19 | 0  | 0.4 | 0.1 | 0.4 | 1     | 19 | 0  | 0.1 | 0.4 | 0.8 | 1        | 19 | 0  | 0.8 | 0.8 |    |
| maxwellSource2       | 0   | 20 | 0  | 6.0E+3 | 6.0E+3 | 1.2 | 0    | 20 | 0  | 1.9 | 0.6 | 0.7 | 0     | 20 | 0  | 0.6 | 0.7 | 0.2 | 0        | 20 | 0  | 0.2 | 0.2 |    |

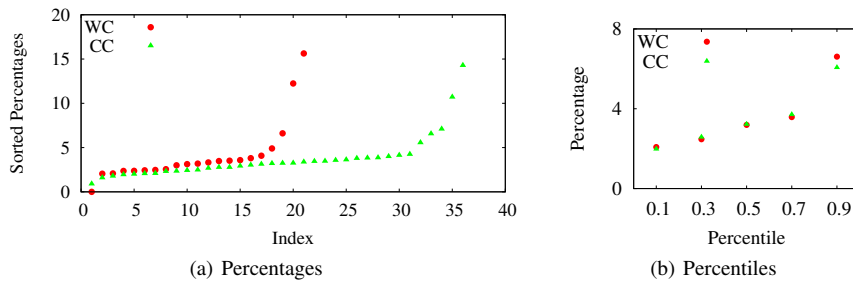
**Fig. 2** Comparison of performance between cross-company and within-company datasets w.r.t. 4 different performance measures (MAR, MMRE, MdmRE, Pred(30)). *W, T, L* statistics are for within-company and the gray colored rows are those in which within-company data wins more than half the time. Under the columns of *WC, CC*, the actual performance values associated with within and cross-company datasets -respectively- are given.

| Test Set                     | Prediction Zone | From S1     | From S2     | From S3     |
|------------------------------|-----------------|-------------|-------------|-------------|
| S1: cocomo81e (28)           | 3.7             | 1.0 (3.6%)  | 1.1 (4.8%)  | 1.6 (14.4%) |
| S2: cocomo81o (24)           | 4.3             | 1.8 (6.6%)  | 1.3 (5.6%)  | 1.1 (10.4%) |
| S3: cocomo81s (11)           | 4.1             | 1.4 (5.1%)  | 1.7 (7.0%)  | 1.0 (9.4%)  |
| S1: nasa93_center_1 (12)     | 5.6             | 1.0 (8.1%)  | 2.9 (7.9%)  | 1.7 (4.3%)  |
| S2: nasa93_center_2 (37)     | 10.0            | 1.6 (13.0%) | 4.6 (12.4%) | 3.8 (9.8%)  |
| S3: nasa93_center_5 (39)     | 5.1             | 0.8 (6.7%)  | 2.2 (6.0%)  | 2.1 (5.4%)  |
| S1: desharnaisL1 (46)        | 5.0             | 2.5 (5.5%)  | 1.7 (7.0%)  | 0.8 (7.9%)  |
| S2: desharnaisL2 (25)        | 4.8             | 2.6 (5.6%)  | 1.5 (6.1%)  | 0.7 (6.7%)  |
| S3: desharnaisL3 (10)        | 3.5             | 1.9 (4.1%)  | 1.3 (5.0%)  | 0.4 (4.0%)  |
| S1: finnishAppType1 (17)     | 3.1             | 1.6 (9.1%)  | 1.6 (8.8%)  |             |
| S2: finnishAppType2345 (18)  | 3.0             | 1.4 (8.2%)  | 1.6 (8.8%)  |             |
| S1: kemererHardware1 (7)     | 1.5             | 0.6 (8.8%)  | 0.9 (10.7%) |             |
| S2: kemererHardware23456 (8) | 1.4             | 0.5 (7.3%)  | 0.8 (10.6%) |             |
| S1: maxwellAppType1 (10)     | 3.5             | 0.7 (7.1%)  | 1.7 (5.9%)  | 1.0 (5.8%)  |
| S2: maxwellAppType2 (29)     | 3.2             | 0.4 (3.7%)  | 1.8 (6.2%)  | 1.0 (5.5%)  |
| S3: maxwellAppType3 (18)     | 2.5             | 0.6 (6.3%)  | 0.9 (3.2%)  | 1.0 (5.6%)  |
| S1: maxwellHardware2 (37)    | 2.9             | 1.7 (4.6%)  | 0.8 (4.9%)  | 0.4 (6.0%)  |
| S2: maxwellHardware3 (16)    | 3.9             | 2.5 (6.8%)  | 1.1 (6.8%)  | 0.3 (4.3%)  |
| S3: maxwellHardware5 (7)     | 3.4             | 2.3 (6.2%)  | 0.8 (5.0%)  | 0.3 (4.5%)  |
| S1: maxwellSource1 (8)       | 3.0             | 0.1 (1.6%)  | 2.8 (5.2%)  |             |
| S2: maxwellSource2 (54)      | 3.2             | 0.4 (4.6%)  | 2.8 (5.3%)  |             |

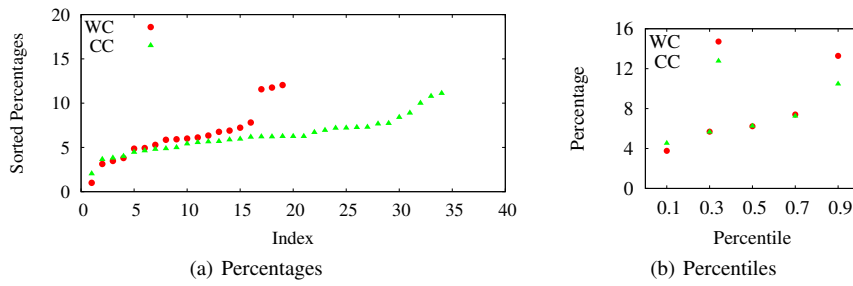
**Fig. 3** The amount of instances selected from within and cross company datasets. In parenthesis the percentage of selected instances out of the actual within company dataset is given. The diagonal entries that are highlighted with gray are the within company selection amounts and percentages.



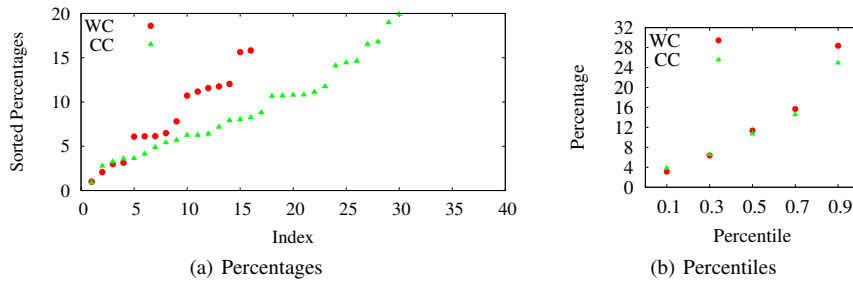
**Fig. 4** Percentages and percentiles of instances selected by TEAK from WC and CC datasets.



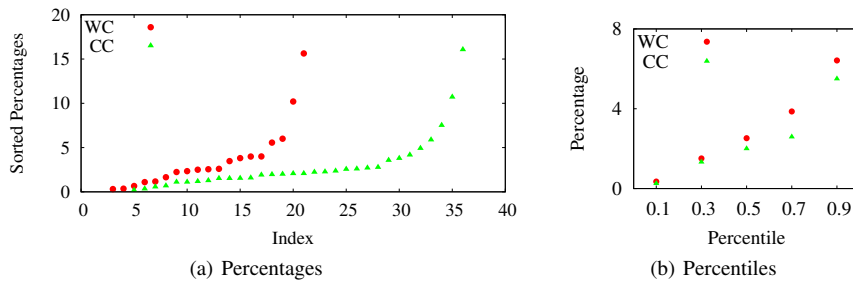
**Fig. 5** Percentages and percentiles of instances selected by ABE0 from WC and CC datasets for  $k=2$ .



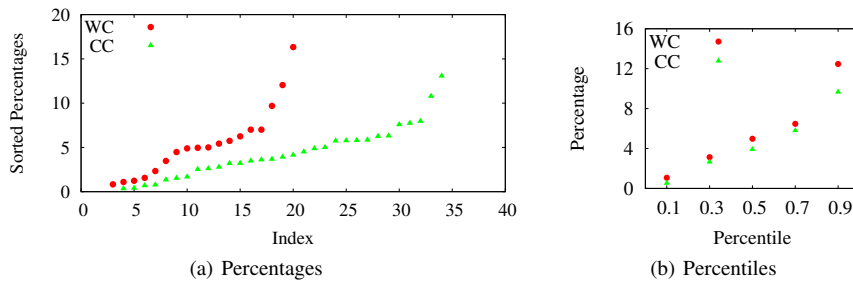
**Fig. 6** Percentages and percentiles of instances selected by ABE0 from WC and CC datasets for  $k=4$ .



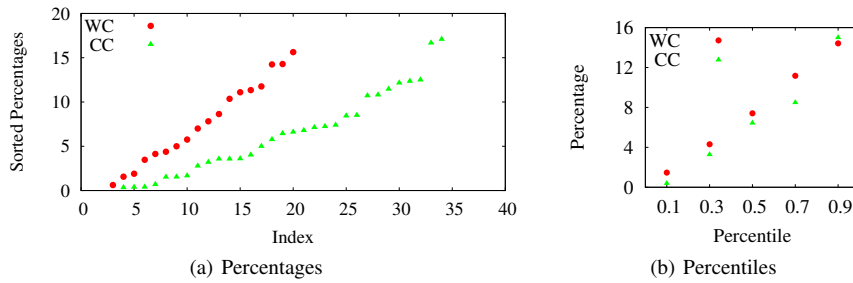
**Fig. 7** Percentages and percentiles of instances selected by ABE0 from WC and CC datasets for  $k=best$ .



**Fig. 8** Percentages and percentiles of instances selected by log+ABE0 from WC and CC datasets for  $k=2$ .



**Fig. 9** Percentages and percentiles of instances selected by log+ABE0 from WC and CC datasets for  $k=4$ .



**Fig. 10** Percentages and percentiles of instances selected by log+ABE0 from WC and CC datasets for  $k=best$ .

| Dataset              | MAR |     |        | MMRE |     |        | MdmRE |     |        | Pred(30) |     |        |
|----------------------|-----|-----|--------|------|-----|--------|-------|-----|--------|----------|-----|--------|
|                      | k=2 | k=4 | k=best | k=2  | k=4 | k=best | k=2   | k=4 | k=best | k=2      | k=4 | k=best |
| cocomo81e            | o   | o   | o      | +    | o   | o      | +     | o   | o      | -        | o   | o      |
| cocomo81o            | -   | o   | o      | +    | +   | o      | +     | -   | o      | -        | -   | o      |
| cocomo81s            | o   | o   | o      | o    | o   | o      | o     | o   | o      | o        | o   | o      |
| nasa93_center_1      | -   | o   | -      | +    | o   | -      | +     | o   | -      | -        | o   | -      |
| nasa93_center_2      | o   | o   | o      | o    | o   | o      | o     | o   | o      | o        | o   | o      |
| nasa93_center_5      | -   | o   | o      | +    | +   | o      | +     | -   | o      | -        | -   | o      |
| desharnaisL1         | -   | -   | -      | o    | o   | o      | o     | o   | o      | o        | o   | o      |
| desharnaisL2         | -   | -   | o      | +    | +   | o      | +     | -   | o      | -        | -   | o      |
| desharnaisL3         | o   | o   | o      | o    | o   | o      | o     | o   | o      | o        | o   | o      |
| finnishAppType1      | o   | o   | o      | o    | o   | o      | o     | o   | o      | o        | o   | o      |
| finnishAppType2345   | o   | o   | o      | o    | o   | o      | o     | o   | o      | o        | o   | o      |
| kemererHardware1     | o   | o   | o      | o    | o   | o      | o     | o   | o      | o        | o   | o      |
| kemererHardware23456 | o   | o   | o      | o    | o   | o      | o     | o   | o      | o        | o   | o      |
| maxwellAppType1      | o   | o   | o      | o    | o   | o      | o     | o   | o      | o        | o   | o      |
| maxwellAppType2      | o   | o   | o      | o    | o   | o      | o     | o   | o      | o        | o   | o      |
| maxwellAppType3      | o   | o   | o      | o    | o   | o      | o     | o   | o      | o        | o   | o      |
| maxwellHardware2     | o   | o   | o      | o    | o   | o      | o     | o   | o      | o        | o   | o      |
| maxwellHardware3     | o   | o   | o      | o    | o   | o      | o     | o   | o      | o        | o   | o      |
| maxwellHardware5     | o   | o   | o      | o    | o   | o      | o     | o   | o      | o        | o   | o      |
| maxwellSource1       | o   | o   | o      | o    | o   | o      | o     | o   | o      | o        | o   | o      |
| maxwellSource2       | o   | o   | o      | o    | o   | o      | o     | o   | o      | o        | o   | o      |

**Fig. 11** Performance comparison between cross-company and within-company datasets w.r.t. 4 different performance measures (MAR, MMRE, MdmRE, Pred(30)) under different “log+ABE0” methods. Note that in this setting a “log” pre-processor was applied to the datasets prior to estimation with ABE0. Each cell in this table can have three values: “+”, “-” and “o”. A “+” sign indicates that WC performance won more 10 of the 20 runs, whereas a “-” sign tells that WC lost more than 10 runs. If none of these conditions occur, i.e. WC and CC performances tie, then a “o” sign is assigned to the cell. For convenience “+” signs are highlighted (see how few they are).

| Dataset              | MAR |     |        | MMRE |     |        | MdmRE |     |        | Pred(30) |     |        |
|----------------------|-----|-----|--------|------|-----|--------|-------|-----|--------|----------|-----|--------|
|                      | k=2 | k=4 | k=best | k=2  | k=4 | k=best | k=2   | k=4 | k=best | k=2      | k=4 | k=best |
| cocomo81e            | o   | o   | o      | +    | o   | -      | +     | o   | -      | -        | o   | -      |
| cocomo81o            | -   | o   | -      | +    | +   | o      | +     | -   | o      | -        | -   | o      |
| cocomo81s            | o   | o   | o      | o    | o   | o      | o     | o   | o      | o        | o   | o      |
| nasa93_center_1      | -   | -   | -      | +    | +   | -      | +     | -   | -      | -        | -   | -      |
| nasa93_center_2      | o   | o   | o      | o    | o   | o      | o     | o   | o      | o        | o   | o      |
| nasa93_center_5      | -   | o   | o      | +    | o   | -      | +     | o   | -      | -        | o   | -      |
| desharnaisL1         | -   | -   | -      | o    | o   | o      | o     | o   | o      | o        | o   | o      |
| desharnaisL2         | -   | o   | o      | +    | +   | o      | +     | -   | o      | -        | -   | o      |
| desharnaisL3         | o   | o   | o      | o    | o   | o      | o     | o   | o      | o        | o   | o      |
| finnishAppType1      | o   | o   | o      | o    | o   | o      | o     | o   | o      | o        | o   | o      |
| finnishAppType2345   | o   | o   | o      | o    | o   | o      | o     | o   | o      | o        | o   | o      |
| kemererHardware1     | o   | o   | o      | o    | o   | o      | o     | o   | o      | o        | o   | o      |
| kemererHardware23456 | o   | o   | o      | o    | o   | o      | o     | o   | o      | o        | o   | o      |
| maxwellAppType1      | o   | o   | o      | o    | o   | o      | o     | o   | o      | o        | o   | o      |
| maxwellAppType2      | o   | o   | o      | o    | o   | o      | o     | o   | o      | o        | o   | o      |
| maxwellAppType3      | o   | o   | o      | o    | o   | o      | o     | o   | o      | o        | o   | o      |
| maxwellHardware2     | o   | o   | o      | o    | o   | o      | o     | o   | o      | o        | o   | o      |
| maxwellHardware3     | o   | o   | o      | o    | o   | o      | o     | o   | o      | o        | o   | o      |
| maxwellHardware5     | o   | o   | o      | o    | o   | o      | o     | o   | o      | o        | o   | o      |
| maxwellSource1       | o   | o   | o      | o    | o   | o      | o     | o   | o      | o        | o   | o      |
| maxwellSource2       | o   | o   | o      | o    | o   | o      | o     | o   | o      | o        | o   | o      |

**Fig. 12** Performance comparison between cross-company and within-company datasets w.r.t. 4 different performance measures (MAR, MMRE, MdmRE, Pred(30)) under different “log+ABE0” methods. Note that in this setting a “log” pre-processor was applied to the datasets prior to estimation with ABE0. Each cell in this table can have three values: “+”, “-” and “o”. A “+” sign indicates that WC performance won more 10 of the 20 runs, whereas a “-” sign tells that WC lost more than 10 runs. If none of these conditions occur, i.e. WC and CC performances tie, then a “o” sign is assigned to the cell. For convenience “+” signs are highlighted (see how few they are).