the review of income and wealth

Review of Income and Wealth
Series 62, Number 3, September 2016
DOI: 10.1111/roiw.12172

# UNDERSTANDING NEW RESOURCE PROJECTS

BY KENNETH W. CLEMENTS,* JIAWEI SI AND THOMAS SIMPSON

*Business School, The University of Western Australia*

The surge in new resource projects has been a prominent feature of the recent strong performance of the Australian economy, with mining and energy investment accounting for almost one-half of all private investment. Although the current round of resource investment has now peaked, as swings in the resource sector tend to repeat themselves, there is an ongoing need to carefully understand the available information sources. We use a specially developed panel of matched projects from three widely followed, but under-researched, sources to analyze cost inflation, the biases, the degree of independence, and timeliness of each source. This information is of use to policy makers who have to closely monitor these developments, analysts following the resources sector, and project proponents wanting to know something about the typical cost profile of a project.

**JEL Codes**: C43, Q30, Q33

**Keywords**: cost escalation, information flows, predicting project costs, resource investment projects

## 1. INTRODUCTION

High commodity prices over the last decade or so have led to an unprecedented resources boom in Australia. This has stimulated a large-scale program of investments in new resource prospects that, in part at least, helped Australia avoid recession and perform better than most other high-income countries. As a result, analysts of the Australian economy now give considerable prominence to information on resource investment plans. As some resource projects are very large, the nature of their cost is of considerable relevance to public policy regarding infrastructure. A recent inquiry by the Australian Productivity Commission into public infrastructure has highlighted the inadequacy of presently available project data to systematically explain the source and nature of cost pressures, hampering accurate cost projections and optimal investment decisions (Productivity Commission, 2014). Internationally, there is considerable evidence of serious problems with investment in "megaprojects" as they suffer from widespread over-optimism, cost overruns, and delays (Flyvbjerg, 2009, 2014). Flyvbjerg goes so far as to describe the situation as one where "the worst projects get built."

*Note*: For comments and help, we thank Adrian Arundell, Aiden Depiazzi, Grace Gao, Peter Hartley, Liang Li, seminar participants at UWA and the Economic Measurement Group Workshop, Sydney, 2013, and the Editor and referees of this journal. We are also grateful to Steve Smith, from Deloitte Access Economics, for providing us with their "Investment Monitor" subscription database and responding to our inquiries. This research was supported in part by the ARC, BHP Billiton, and the UWA Business School. All errors are our own.

*Correspondence to: Kenneth Clements, Economics (UWA Business School), The University of Western Australia (M251), 35 Stirling Highway, Crawley WA 6009, Australia (ken.clements@uwa.edu.au).

© 2015 International Association for Research in Income and Wealth

This paper helps to deepen the understanding of project costs by identifying the strengths and weaknesses of three data sources of Australian resource projects that have tended to be underused in research in the past. The three sources are: The WA Department of State Development's *Prospect Magazine*, Deloitte Access Economics' "Investment Monitor subscription database," and the ABARES/BREE "Mining Projects" database.[1,2] Using a unique database in which projects are matched across these sources, we analyze the quality and usefulness of these data.

The structure of the paper is as follows. The next section describes the process of matching projects across the three sources. The matched data mean there are three readings on the cost of each project through time, from which preliminary estimates of the biases are obtained. Section 3 uses these rich data to construct a hedonic index of cost escalation that is not contaminated by the entry and exit of radically different projects, each with their own inflation rate. The hedonic model also leads to more refined estimates of biases in the three sources. Later sections deal with understanding the pattern of information flows regarding cost escalation and an analysis of the degree to which sources rely on each other. The estimates of the cost of a project made before its conclusion can be considered as a forecast of the final cost, and the quality of these forecasts is investigated toward the end of the paper. The paper concludes with an overall assessment of the three sources.

## 2. Matching Projects

To qualify for inclusion in our matched database, a project has to have at least one period of capital expenditure reported contemporaneously in all three sources. Projects that had triplicate matches in some periods but not others are included only for the matching periods.[3] The included projects are usually those that attract more attention from the various stakeholders, that is, the larger ones. The period covered is September 2006 to September 2012. Table 1, which summarizes the matched data, clearly shows that the mean project size (as measured by cost, termed "capex" in the table) is of the order of $A3.5b (second last row of the table). In total, 354 triplets of projects are matched and Table 1 gives the number per period.

The last three columns of Table 1 compare the cost in each of the three sources in the form of deviations from the overall mean. The second last element of the column for *Prospect* (column 10) reveals that according this source, the size of projects is lower, on average, by more than $100m than the overall mean. On the basis of a t-test, this difference is significant (see the last entry of this column). The bias is in the opposite direction for BREE projects, which are larger, on average, by about $100m (also significant). The *Investment Monitor* (*IM*) data are approximately unbiased. As the projects are matched exactly, these differences cannot be attributed to differing coverage of the three sources. Of course, in the context of

---

[1] For an earlier analysis of the *Investment Monitor* data dealing with cost escalation, lead time, and probability of success of projects, see Clements and Si (2011).

[2] Details of the ABARES/BREE source are as follows. This is the Bureau of Resources and Energy Economics' publication "Resources and Energy Major Projects." Before July 2011, this publication was the Australian Bureau of Agricultural and Resource Economics and Sciences' "Minerals and Energy Major Development Projects." For brevity, the ABARES/BREE publications will be referred to as BREE.

[3] For further details of the data, see the online Appendix.

© 2015 International Association for Research in Income and Wealth

TABLE 1

Capex, Matched Projects

| Date (1) | Number of Projects (2) | Prospect Mean (3) | Prospect S.D. (4) | Investment Monitor Mean (5) | Investment Monitor S.D. (6) | BREE Mean (7) | BREE S.D. (8) | Grand Mean (9) | Difference from Grand Mean Prospect (10) | Difference from Grand Mean Investment Monitor (11) | Difference from Grand Mean BREE (12) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Sept 06 | 40 | 1,464 | 2,232 | 1,589 | 2,657 | 1,590 | 2,707 | 1,548 | 84 | 41 | 42 |
| Mar 07 | 41 | 1,549 | 2,196 | 1,719 | 2,692 | 1,712 | 2,673 | 1,660 | –111 | 59 | 52 |
| Sept 07 | 39 | 1,258 | 1,601 | 1,427 | 2,118 | 1,499 | 2,238 | 1,394 | –137 | 32 | 104 |
| Mar 08 | 34 | 1,237 | 1,948 | 1,291 | 1,937 | 1,346 | 2,077 | 1,291 | –55 | 0 | 55 |
| Sept 08 | 31 | 1,488 | 2,457 | 1,418 | 2,097 | 1,654 | 2,527 | 1,520 | –32 | –102 | 134 |
| Mar 09 | 28 | 1,733 | 2,397 | 1,967 | 2,434 | 2,077 | 2,625 | 1,926 | –193 | 42 | 151 |
| Sept 09 | 28 | 3,610 | 8,112 | 3,740 | 8,076 | 3,751 | 8,114 | 3,700 | –90 | 39 | 51 |
| Mar 10 | 24 | 4,200 | 8,656 | 4,979 | 10,071 | 5,126 | 9,222 | 4,768 | –569 | 211 | 357 |
| Sept 10 | 21 | 5,560 | 10,024 | 5,061 | 9,292 | 5,661 | 9,757 | 5,428 | 133 | –366 | 234 |
| Mar 11 | 20 | 5,130 | 10,189 | 4,860 | 9,341 | 5,126 | 9,788 | 5,039 | 91 | –179 | 87 |
| Sept 11 | 18 | 5,471 | 10,695 | 5,894 | 11,334 | 5,918 | 11,320 | 5,761 | –290 | 133 | 157 |
| Mar 12 | 15 | 5,100 | 10,588 | 5,550 | 10,527 | 5,444 | 10,539 | 5,365 | –265 | 185 | 79 |
| Sept 12 | 15 | 7,209 | 12,231 | 7,437 | 12,210 | 7,166 | 12,166 | 7,271 | –62 | 167 | –105 |
| Average | 27 | 3,462 | 6,410 | 3,610 | 6,522 | 3,698 | 6,596 | 3,590 | –128 | 20 | 108 |
| t-value |  |  |  |  |  |  |  |  | 2.86 | 0.42 | 3.17 |

*Notes*: Except for the first two columns and the last row, all entries are in $Am. The correlations between the sources are as follows: Prospect-IM = 0.977, Prospect-BREE = 0.983, and IM-BREE = 0.986.

© 2015 International Association for Research in Income and Wealth

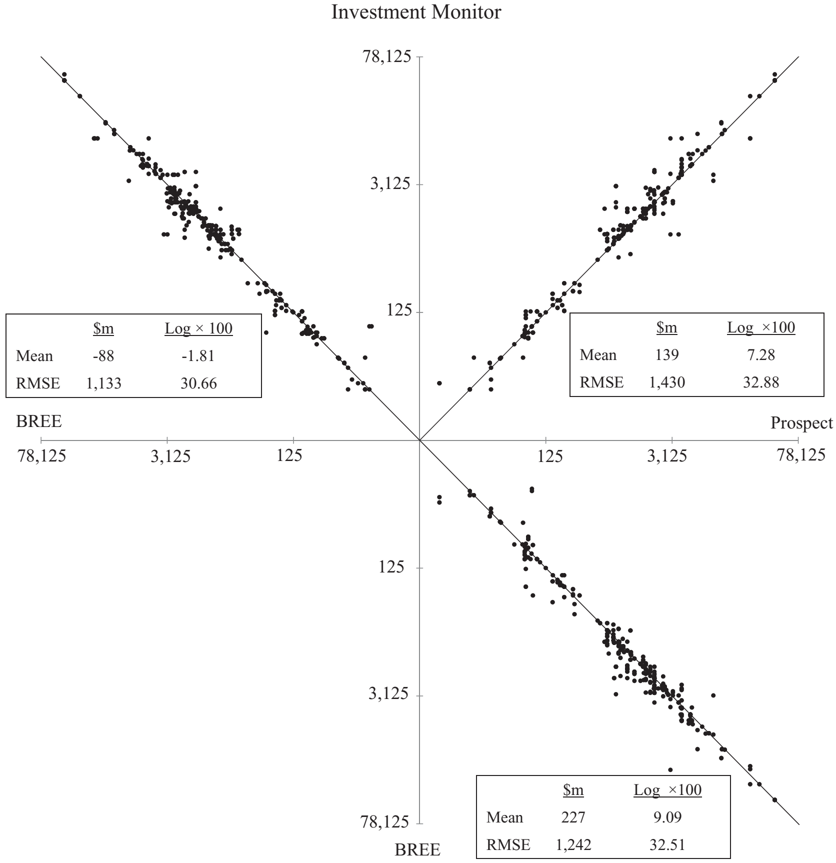

Figure 1. Three-Way Comparison of Capex, Matched Projects ($ million)

*Notes*: The rays from the origin are 45-degree lines, along which capex from pairs of sources coincide. The boxes contain the error statistics. The mean is the average difference between capex according to the source on the vertical axis minus that for the horizontal. Because the number of observations in each period is not the same in Table 1, the mean errors in this figure are not completely consistent with those of that table, but the differences are small. The RMSE is the root-mean-squared error.

a project costing $3.5b, a $100m error is less than 3 percent, so the economic significance is modest.

The second last row of Table 1 also shows that the dispersion, as measured by the standard deviation, of projects in *IM* and BREE are very similar, while that of *Prospect* is somewhat lower. The correlations among the three sources are high (at least 0.98, as indicated in the notes to the table), which is to be expected as the same projects are involved and, most likely, each source looks at its two neighbors. More will be said about this later. Figure 1 provides a visual comparison of the data by giving the three pairwise contrasts. The clustering of observations around the 45-degree lines illustrates the substantial agreement across sources. But still for the BREE/*Prospect* contrast in the south-east quadrant, there is a noticeable tendency for the points to lie below the 45-degree line, reflecting the underpricing

© 2015 International Association for Research in Income and Wealth

in *Prospect* and overpricing in BREE. The high root mean square errors (RMSEs) of each quadrant, which are of the order of 30 percent, show a reasonably large degree of cross-source variance. In summary, although project values are highly correlated, there is still a significant difference between *Prospect* and BREE.

## 3. Hedonic Costing

Next, we use the matched data to estimate a hedonic model of the form

$$\log v_{it}^{s} = \gamma^{s} + \lambda_{i} + \theta_{t} + \varepsilon_{it}^{s}, \quad s = 1, 2, 3; i = 1, \ldots, N_{t}; t = 1, \ldots, 13.$$

Here, $v_{it}^{s}$ denotes capex from source $s$ ($s = 1, 2, 3$, representing *Prospect*, *IM*, and BREE, respectively) in period $t$ ($t = 1, \ldots, 13$) for project $i$ ($i = 1, \ldots, N_t$). This capex depends on source effects, $\gamma^s$, that allow for the biases in the three publications; projects effects, $\lambda_i$, to control for projects that differ in nature and scale; time effects, $\theta_t$, to capture cost escalation (time is measured in half-yearly intervals, from September 2006 to September 2012); and random factors as measured by the disturbance term $\varepsilon_{it}^{s}$. The project effects allow for projects with different idiosyncrasies entering and dropping out of the system. The estimates of the time effects provide an index of cost escalation that measures "pure" inflation that in no way reflects extraneous influences. Similarly, estimated source biases are insulated from compositional issues.

The estimates of the hedonic model are given in Table 2. Across all projects, the annual rates of cost escalation in column 2 are highly significant, but there are some noticeable year-to-year fluctuations. Over the whole period, project cost escalation averages about 13 percent p.a., which is much larger than CPI inflation over the same period.[4] According to the estimated source effects (column 2), the costs of projects in *Prospect* are understated by about 5 percent, those in BREE overstated by 4 percent (both estimates are significant), while for *IM* the bias is positive but insignificantly different from zero. These results broadly agree with those of Figure 1.

We also split the 74 projects into three equal sized groups (25, 25, 24) based on their average starting cost. The year-to-year cost escalation for small projects (column 3) is largely insignificant but on average costs increase by 8 percent p.a., which is significant. The majority of cost escalations for large and mega projects (columns 4 and 5) are statistically significant with the average change around 13 and 16 percent p.a., respectively. *Prospect* also continues to understate costs and BREE continues to overstate for large and mega projects. The source effects are the most extreme for large projects, where *Prospect* understates costs by 9.3 percent and *IM* and BREE overstate by 5.7 and 3.6 percent, respectively. For small projects, none of the source effects is significant.

Lastly, we repeat the analysis for the projects based on the industry (LNG, Iron Ore, or Others) they belong to in columns 6 to 8 of Table 2. Apart from the year-to-year fluctuations, LNG and Iron Ore projects have an average cost

---

[4] The average logarithmic difference (·100) of the CPI from September 2006 to September 2012 is 2.8 percent p.a. (Source: ABS Cat No. 64010.0).

© 2015 International Association for Research in Income and Wealth

TABLE 2

Hedonic Costing of Projects, Matched Projects

$$\log v_{it}^{s} = \gamma^{s} + \lambda_{i} + \theta_{t} + \varepsilon_{it}^{s}$$

| | | Projects Distinguished by Size and Industry | | | | | |
|---|---|---|---|---|---|---|---|
| | | Size | | | Industry | | |
| Variable (1) | All (2) | Small (3) | Large (4) | Mega (5) | LNG (6) | Iron Ore (7) | Others (8) |
| *Cost escalation* $(\theta_t - \theta_{t-2}) \cdot 100$ | | | | | | | |
| Sept 06 | – | – | – | – | – | – | – |
| Sept 07 | 10.30 (3.43) | 3.72 (6.05) | 5.37 (5.33) | 23.30 (6.28) | 21.92 (6.65) | 11.33 (7.16) | 5.14 (4.45) |
| Sept 08 | 11.13 (3.58) | 3.40 (5.80) | 14.67 (5.40) | 22.01 (7.78) | 31.78 (8.86) | 8.12 (6.37) | 10.58 (4.72) |
| Sept 09 | 14.69 (3.86) | 5.22 (8.23) | 14.77 (5.55) | 14.36 (7.36) | 11.00 (8.23) | 12.00 (6.24) | 13.15 (6.06) |
| Sept 10 | 14.50 (4.32) | 11.09 (11.14) | 19.27 (6.56) | 8.72 (6.49) | 10.57 (7.49) | 9.54 (6.94) | 17.12 (7.89) |
| Sept 11 | 15.75 (4.80) | 9.42 (11.16) | 16.81 (7.16) | 16.41 (7.68) | 10.16 (7.43) | 33.04 (8.00) | −0.89 (8.89) |
| Sept 12 | 11.34 (5.23) | 17.38 (16.41) | 7.45 (7.95) | 12.56 (7.56) | 9.17 (8.43) | 17.30 (8.16) | 3.18 (10.26) |
| Average | 12.95 (0.88) | 8.37 (2.52) | 13.05 (1.32) | 16.23 (1.44) | 15.77 (1.60) | 15.22 (1.52) | 8.05 (1.53) |
| *Source effects* $\gamma^{s} \cdot 100$ | | | | | | | |
| Prospect | −5.46 (1.06) | −2.01 (2.04) | −9.28 (1.60) | −2.81 (1.84) | −5.53 (2.03) | −8.96 (1.81) | −1.93 (1.57) |
| IM | 1.83 (1.06) | −0.60 (2.04) | 5.70 (1.60) | −1.76 (1.84) | 0.01 (2.03) | 2.70 (1.81) | 1.93 (1.57) |
| BREE | 3.63 (1.06) | 2.61 (2.04) | 3.58 (1.60) | 4.56 (1.84) | 5.53 (2.03) | 6.26 (1.81) | 0.00 (1.57) |
| *Project effects included* | | | | | | | |
| SEE | 0.244 | 0.237 | 0.245 | 0.234 | 0.217 | 0.262 | 0.228 |
| $R^2$ | 0.999 | 0.997 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 |
| No. of observations | 1,062 | 270 | 468 | 324 | 225 | 417 | 420 |

*Notes*: Standard errors are in parentheses. As the data are semi-annual, cost escalation measured by $\theta_t - \theta_{t-2}$ is in terms of a change over the year. Only September years are shown. For project size, the sample is split into three equal sized categories (small, large, and mega) based on the average of their earliest matched cost across the three sources. The cost ranges are: Small, less than $470 million; Large, greater than Small but less than $1,740 million; and Mega, greater than Large and up to $13.5 billion. For details of size and industry classification, as well as a listing of the data, see the online Appendix.

© 2015 International Association for Research in Income and Wealth

escalation of about 15 percent p.a., while costs rise by 8 percent for those in Other industries, all of which are significant. The source bias of *IM* continues to be bracketed by *Prospect* and BREE for both the LNG and Iron Ore industries, whilst the source biases are insignificant for Other industries. The degree of underestimation in *Prospect* is more substantial for Iron Ore projects compared to LNG (−9 percent vs. −5.5 percent); likewise, BREE overstates costs slightly more for Iron Ore projects (6.3 percent) as compared to LNG projects (5.5 percent).

These hedonic results provide some insight into the pressures faced by the resources sector during the recent boom: if, on average, the cost of a project rises by, say, 15 percent p.a., after six years, the cumulative escalation is almost 150 percent. Cost escalations are prevalent for all types of projects on average, but are substantially greater for the larger ones, which tend to be in the LNG and Iron Ore industries. This is consistent with evidence presented by Flyvbjerg (2014) that initial project costs are substantially understated. Regarding source biases, project costs in *Prospect* are understated, BREE overstates them, and *IM* is usually bracketed between the other two.

## 4. Modeling Information Flows

The three sources of capex data refer to the same projects, but in many instances report different values. Over time, it might be expected that the values converge through a Darwinian process of "good information driving out bad." Suppose, for example, that source 2 initially has more accurate data on a certain project than source 1. Then, the "updating" process could be direct in the form of source 1 using previously published data by source 2 with a lag, which we can write as $2 \rightarrow 1$. This situation would also occur when source 2 responds rapidly to new information on the project, while 1 responds only slowly. Although there is no overt copying of one source by another, as it is observationally equivalent, the process can still be described as $2 \rightarrow 1$. The process could also be indirect involving a third source of the form $2 \rightarrow 3 \rightarrow 1$, a sequence that might extend over a longer period. For other projects, the reverse situation may apply with source 1 being more accurate than 2, so when all projects are considered together, there would be a two-way flow of information. In this section, we use a VAR model to measure this type of information exchange. This approach considers flows in all directions and sheds some light on which sources tend to excel in publishing new information.

Let $g_{it}^{s} = \log\left(v_{it}^{s} / v_{i,t-1}^{s}\right)$ be the revision, or growth rate, in the projected capex $v_{it}^{s}$ for project $i$ from period $t-1$ to $t$ according to source $s$. The $3 \cdot 1$ vector of growth rates for project $i$, $[g_{it}^{1}, g_{it}^{2}, g_{it}^{3}]'$, is taken to be a first-order vector autoregressive process, the s^th^ member of which is

$$g_{it}^{s} = \alpha_{i}^{s} + \sum_{r=1}^{3} \beta_{i}^{sr} g_{i,t-1}^{r} + \varepsilon_{it}^{s}, \tag{1}$$

where $\alpha_{i}^{s}$ and $\beta_{i}^{sr}$ are coefficients and $\varepsilon_{it}^{s}$ is a disturbance term. The intercept $\alpha_{i}^{s}$ measures the role of other sources of cost escalation that occur independent of the past; these can be called "autonomous" cost increases. The own-coefficient $\beta_{i}^{ss}$

© 2015 International Association for Research in Income and Wealth

refers to the degree to which current inflation depends on its own past history. The size of the cross-coefficient $\beta_i^{sr}$, for $s \neq r$, measures the direct flow of information from source $r$ to $s$ over 1 period. Equation (1) for $s = 1, 2, 3$ is the VAR model for project $i$.

As there is insufficient time-series data to estimate model (1) for each of the 74 distinct projects, we pool the data by taking the coefficients to be the same over projects to estimate

$$g_{it}^{s} = \alpha^{s} + \sum_{r=1}^{3} \beta^{sr} g_{i,t-1}^{r} + \varepsilon_{it}^{s}.$$

for $s = 1, 2, 3$. Panel A of Table 3 uses that matched data to estimate this model.[5] Looking at the first row that refers to *Prospect*, the estimate of the intercept is 0.050, which means that autonomous inflation according to this source is about 5 percent per half year (and significant). Next, the estimate of the own autoregressive coefficient, $\beta^{11}$, is –0.043. The negative sign means that higher inflation in the last period tends to be followed by lower inflation in this period, other things remaining unchanged. Thus, rather than inflation inertia, there is some degree of mean reversion in the level of capex. However, this coefficient is relatively small and not significantly different from zero. The estimated cross-lag coefficient $\beta^{13}$ of 0.115 implies that about 12 percent of past growth in costs in BREE passes through into current growth in costs reported in *Prospect*. This estimate is significant and considerably larger than that for *IM* → *Prospect* $\left(\hat{\beta}^{12}\right)$. Thus, there is a more substantial flow of information from BREE to *Prospect*, than from *IM*. The estimates of the coefficients of the two other equations have a similar interpretation.

The F-statistics in column 8 of the table test the hypothesis that all the three lagged source coefficients are jointly zero. The null is rejected in the case of both *IM* and BREE, but not for *Prospect*. The last column tests if in each case, the two alternate sources play no role. BREE has a lower F-value than *IM*, suggesting the possibility that BREE is informed less by the other sources than is *IM*, and may rely more on its own research to revise its data. Further results below would also seem to point to this conclusion. As discussed in the online Appendix, there are slight asynchronies in the dates of the publication of the three sources that may in part explain the result of Table 3 that BREE appears to be more influenced by *IM* than vice versa (that is, $\hat{\beta}^{32} > \hat{\beta}^{23}$).

The insignificant F-value for *Prospect* in the last column of panel A of Table 3 means that other sources play no role in contributing to this publication. Taken in isolation, the interpretation of this result is ambiguous. It could be that *Prospect* is the “market leader” in disseminating new information and does not need to absorb information from the other sources. Alternatively, it could be that they do not go to the trouble of “learning” from the other sources. From the second and third rows of columns 3–5 of Table 3, both *IM* and BREE appear to take on less information from *Prospect* than from the other sources $\left(\hat{\beta}_{21} < \hat{\beta}_{23}, \hat{\beta}_{31} < \hat{\beta}_{32}\right)$; this

---

[5] See Clements *et al.* (2014) for further details.

© 2015 International Association for Research in Income and Wealth

TABLE 3

Estimates of VAR Model of Information Flows, Matched Projects, Semi-Annual, 2006–12

$$g_{it}^{s} = \alpha^{s} + \sum_{r=1}^{3} \beta^{sr} g_{i,t-1}^{r} + \text{dummies} + \varepsilon_{it}^{s}$$

| | | Independent Variables, Lagged Values, $g_{i,t-1}^{r}$ | | | | | F-tests, $H_0 : \beta^{sr} = 0$ for | |
|---|---|---|---|---|---|---|---|---|
| Dependent Variable $g_{it}^{s}$ (1) | Intercept (2) | Prospect (3) | IM (4) | BREE (5) | $R^2$ (6) | SEE (7) | $r = 1, 2, 3$ (8) | $r = 1, 2, 3; r \neq s$ (9) |
| | | | **A. Intercepts included** | | | | | |
| Prospect | 0.050 (0.014) | –0.043 (0.057) | 0.017 (0.074) | 0.115 (0.059) | 0.37 | 0.198 | 1.43 | 2.07 |
| IM | 0.038 (0.012) | 0.089 (0.050) | –0.072 (0.065) | 0.101 (0.052) | 0.25 | 0.174 | 2.83 | 4.10 |
| BREE | 0.041 (0.017) | 0.056 (0.068) | 0.210 (0.088) | –0.190 (0.070) | 0.12 | 0.235 | 3.97 | 3.43 |
| | | | **B. Intercepts suppressed** | | | | | |
| Prospect | – | –0.009 (0.058) | 0.070 (0.075) | 0.145 (0.060) | 0.34 | 0.204 | 2.74 | 3.94 |
| IM | – | 0.115 (0.051) | –0.032 (0.065) | 0.123 (0.052) | 0.22 | 0.178 | 4.46 | 6.59 |
| BREE | – | 0.084 (0.068) | 0.253 (0.087) | –0.166 (0.070) | 0.10 | 0.237 | 4.54 | 5.60 |

*Notes*: The dummies deal with transition from the last observation on one project to the first observation of the next. For details, see Clements *et al*. (2014). Standard errors in parentheses. The model of panel A has an AIC value of –0.703; panel B AIC = –0.635. These are insignificantly different ($p = 0.97$).

© 2015 International Association for Research in Income and Wealth

would seem to point to *Prospect* not being the market leader. Coupled with *Prospect*'s systematic bias noted above, the indications are that *Prospect* does not seem to lead in information dissemination, but is somewhat divorced from the system as a whole. This conclusion is reinforced by the extremes of the cross-effects coefficients: in absolute terms, the smallest is for *IM* → *Prospect* $\left(\hat{\beta}^{12} = 0.017\right)$, while the largest is *IM* → BREE $\left(\hat{\beta}^{32} = 0.210\right)$. That is, *IM* plays a minor role in the revisions to data published in *Prospect*, but a major one regarding BREE, pointing to the apparent "insulation" of *Prospect*.[6]

## 5. Bilateral and Multilateral Information Balances

The interactions among pairs of sources in the VAR model are bidirectional. For example, from panel A of Table 3, information from *IM* in the past is associated with revisions to BREE $\left(\hat{\beta}^{32} = 0.210\right)$, while there is also a reciprocal flow from BREE to *IM* $\left(\hat{\beta}^{23} = 0.101\right)$. The difference between these two gross flows is the net flow of $\hat{\beta}^{32} - \hat{\beta}^{23} = 0.210 - 0.101 = 0.109$, which can be interpreted as saying that if costs in both sources grow by the same rate in the previous period, BREE will receive about 10 percent more information from *IM* than it gives in return. The flow of information is measured by that part of revision to costs in one source that can be attributed to past growth in another source, all other factors remaining unchanged. If the $3 \cdot 3$ matrix of estimated coefficients of the lagged terms, $\hat{\boldsymbol{\beta}} = \left[\hat{\beta}^{sr}\right]$, is symmetric, then the reciprocal trade flows are exactly equal, net flows are zero, and no source is a net sender to or receiver of information from the others. Accordingly, the degree of asymmetry of $\hat{\boldsymbol{\beta}}$ provides a measure of the bilateral information flows. It is convenient to formulate asymmetry with the skew symmetric matrix $\boldsymbol{\Gamma} = \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}'$. The elements of the upper triangle of this matrix, $\gamma^{sr} = \hat{\beta}^{sr} - \hat{\beta}^{rs}$, $s < r$, give the signs of the net flows from source $r$ to $s$, $s < r = 1, 2, 3$. The elements in the lower triangle are the net flows from $r$ to $s$, $\gamma^{sr} = \hat{\beta}^{sr} - \hat{\beta}^{rs}$, $s > r$, which are the negative of those in the upper triangle, so $\gamma^{sr} = -\gamma^{rs}$. In words, if s receives information from $r$ ($\gamma^{sr} > 0$), then obviously $r$ sends it to $s$ ($\gamma^{rs} = -\gamma^{sr} < 0$). As a source can neither receive or send a net flow to itself, $\gamma^{ss} = \hat{\beta}^{ss} - \hat{\beta}^{ss} = 0$. The benchmark case is when $\boldsymbol{\Gamma}$ contains all zero elements, as then the original coefficient matrix $\hat{\boldsymbol{\beta}}$ is symmetric and the bilateral trades are balanced.

Panel C of Table 4 gives the $\boldsymbol{\Gamma}$ matrix associated with the estimates of panel A of Table 3 (the other two panels contain intermediate steps). The lower triangle contains the three independent measures of net flows; as these are all non-zero, bilateral trade is unbalanced. For the pairs *IM*/*Prospect* and BREE/*IM*, the net flows are positive, so more information is received by the former source than sent back in return. The reverse is true for BREE/*Prospect*. But as these measures have relatively large standard errors, not too much reliance can be placed on these results.

---

[6] Panel B of Table 3 shows the results when the intercepts are omitted from the VAR model. This has the effect of: (i) decreasing (in absolute value) the own-lag coefficients, so now there is not as much mean reversion; and (ii) increasing most of the cross coefficients. But as these changes are not huge and as the general pattern remains more or less the same, in what follows we use the estimates of the model for the case in which the intercepts are included.

© 2015 International Association for Research in Income and Wealth

TABLE 4

Relative Information Flows

| Source (1) | Prospect (2) | IM (3) | BREE (4) | Row Sum (5) |
|---|---|---|---|---|
| | | **A. Coefficient matrix** $\left[\hat{\beta}^{sr}\cdot 100\right]$ | | |
| Prospect | −4.3 (5.7) | 1.7 (7.4) | 11.5 (5.9) | |
| IM | 8.9 (5.0) | −7.2 (6.5) | 10.1 (7.0) | |
| BREE | 5.6 (6.8) | 21.0 (8.8) | −19.0 (7.0) | |
| | | **B. Transpose** $\left[\hat{\beta}^{rs}\cdot 100\right]$ | | |
| Prospect | −4.3 (5.7) | 8.9 (5.0) | 5.6 (6.8) | |
| IM | 1.7 (7.4) | −7.2 (6.5) | 21.0 (8.8) | |
| BREE | 11.5 (5.9) | 10.1 (7.0) | −19.0 (7.0) | |
| | | **C. Net information flows** $\boldsymbol{\Gamma} = (\hat{\beta}^{sr} - \hat{\beta}^{rs})\cdot 100$ | | |
| Prospect | 0 | −7.1 (9.0) | 6.0 (9.0) | −1.1 (12.2) |
| IM | 7.1 (9.0) | 0 | −10.9 (10.2) | −3.8 (13.3) |
| BREE | −6.0 (9.0) | 10.9 (10.2) | 0 | 4.9 (13.1) |
| Total | 1.1 (12.2) | 3.8 (13.3) | −4.9 (13.1) | 0.0 |

*Notes*: Panel A is from Table 3. In panel C the elements of the matrix refer to the bilateral information balances. A positive element indicates that the row source receives more information from the column source than it sends in return; vice versa for a negative element. The row sums refer to the multilateral balances. A positive row sum indicates the source receives more information from the others than it sends in return; vice versa for a negative row sum. Standard errors in parentheses.

Suppose each source revises upwards in the previous period their cost data for project i; if the revisions are equiproportional, then $g^{1}_{i,t-1} = g^{2}_{i,t-1} = g^{3}_{i,t-1} = g^{*}_{i,t-1}$. Other things remaining unchanged, the current-period revision of source s is then $g^{*}_{i,t-1}\cdot\sum_{r=1}^{3}\hat{\beta}^{sr}$. Thus, the sum of the coefficients in the row for source s is proportional to its response to a "uniform" message from the three sources. The corresponding information supplied by s to the others as the reciprocal flow is $g^{*}_{i,t-1}\cdot\sum_{r=1}^{3}\hat{\beta}^{rs}$, implying that the net effect is $g^{*}_{i,t-1}\cdot\left(\sum_{r=1}^{3}\hat{\beta}^{sr} - \sum_{r=1}^{3}\hat{\beta}^{rs}\right) = g^{*}_{i,t-1}\cdot\sum_{r=1}^{3}\gamma^{sr}$. When $\sum_{r=1}^{3}\gamma^{sr} > 0$, source s receives more information from the others than it provides, so can be described as a net importer, and vice versa. In other words, the sign of $\gamma^{sr}$ denotes the bilateral information trade balance, while that of $\sum_{r=1}^{3}\gamma^{sr}$ denotes the multilateral balance. By construction, $\sum_{s=1}^{3}\sum_{r=1}^{3}\gamma^{sr} = 0$, so world trade is balanced. The multilateral balances are contained in the last column of panel C of Table 4 and as can be seen, *Prospect* is a small exporter of information, *IM* is a larger exporter, and BREE is an importer. In this sense, *IM* would seem to be the largest contributor to the flow of new information. But due to the high standard errors, again caution should be exercised with this specific result. Further analysis of the speed of information flows using impulse response functions and a vector error correction model suggests that *IM* and BREE respond faster to new information than *Prospect*. For details, see the online Appendix.[7]

---

[7] Another metric of the quality of information is its timeliness, as measured by the frequency and nature of cost revisions. We find that BREE is updated substantially more frequently than the other two sources; however, *IM* adds more unique information in its updates. For details, see the online Appendix.

© 2015 International Association for Research in Income and Wealth

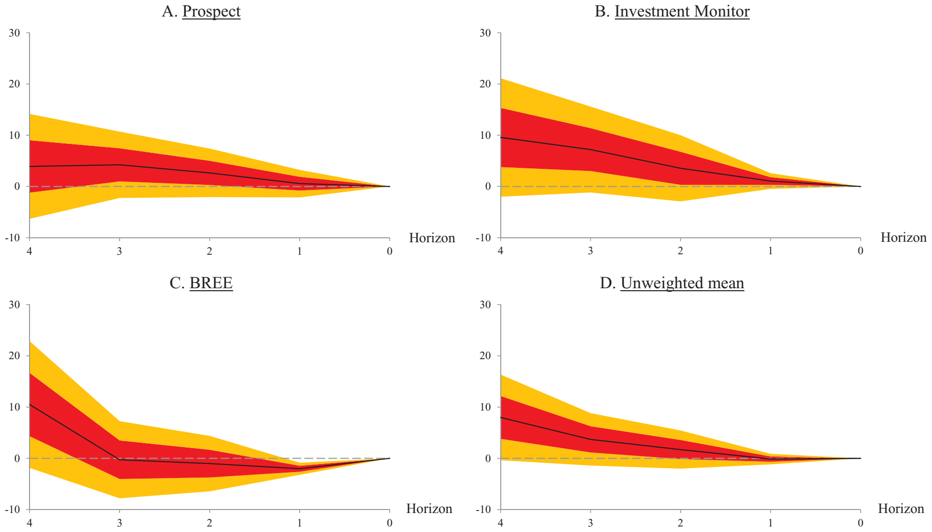

Figure 2. Capex Forecast Errors (logarithmic ratios · 100)

*Notes*: The solid back line is the average forecast error; a positive value implies actual exceeds forecast. The dark and light shaded areas below and above the solid black line represent the one- and two-standard error bands, respectively. The horizon is measured in terms of six-monthly intervals.

## 6. Predicting Future Costs

This section examines the ability of the estimated capital expenditure in each source to predict the subsequent actual cost of projects.

As before, for source $s$, the estimated capex of project $i$ at time $t$ is $v_{it}^s$. If construction of the project is completed at time $T_i$, this estimated cost is to be compared with the final cost denoted by $v_{i,T_i}^s$; this final, or actual, cost is observed $T_i - t = h$ periods in the future from $t$. The estimated cost can also be formulated in terms of the forecast horizon $h$ as $v_{i,T_i-h}^s$. If there are $N_h$ projects having horizon $h$, then the logarithmic mean forecast error at h and the corresponding standard deviation are

$$B_h^s = \frac{1}{N_h}\sum_{i=1}^{N_h}\left(\log v_{i,T_i}^s - \log v_{i,T_i-h}^s\right), \quad SD_h^s = \sqrt{\frac{1}{N_h}\sum_{i=1}^{N_h}\left(\log v_{i,T_i}^s - \log v_{i,T_i-h}^s - B_h\right)^2}.$$

These measures, for each source, are used as the basis for the fan charts of panels A, B, and C of Figure 2.[8] For each source, the mean errors are mostly positive, indicating a bias to underestimate costs.[9] The bias, however, declines with the horizon from about 5–10 percent for a two-year horizon to –2 to 1 percent for six months out. The error bands also shrink noticeably with the horizon; for $h = 4$, for

---

[8] For details of the data used in this section, see Clements *et al.* (2014).

[9] The tendency to underestimate project costs has been noted by others (see, e.g., Flyvbjerg, 2009, 2014).

© 2015 International Association for Research in Income and Wealth

example, the two-standard-error band is about ±10 percent, while for $h = 1$, it is smaller by a factor of almost 10. As to a first approximation these patterns apply to all three sources, they provide little basis for choosing between them. Panel D of Figure 2 will be discussed subsequently.

Rather than taking each source by itself as a predictor, we now consider a composite forecast made up of all three together. We start with a regression of actual on estimated capex for source s and horizon $h$:

$$(2) \qquad \log v_{i,T_i}^{s} = \alpha_0^{s} + \alpha_1^{s} \log v_{i,T_i-h}^{s} + \varepsilon_i^{s}, \quad i = 1, \ldots, N_h,$$

where $\varepsilon_i^{s}$ is a random disturbance. The forecasts are said to be unbiased if the intercept $\alpha_0^{s} = 0$ and efficient if the slope coefficient $\alpha_1^{s} = 1$ (Mincer and Zarnowitz, 1969). Averaging both sides of this equation over sources gives

$$(3) \qquad \log v_{i,T_i} = \beta_0 + \sum_{s=1}^{3} \beta_1^{s} \log v_{i,T_i-h}^{s} + \varepsilon_i, \quad i = 1, \ldots, N_h,$$

where $\log v_{i,T_i} = (1/3)\sum_{s=1}^{3} \log v_{i,T_i}^{s}$ is averaged actual cost, $\beta_0 = (1/3)\sum_{s=1}^{3} \alpha_0^{s}$ is the averaged intercept, $\beta_1^{s} = \alpha_1^{s}/3$ is one-third of the slope coefficient in equation (2), $s = 1, 2, 3$, and $\varepsilon_i = (1/3)\sum_{s=1}^{3} \varepsilon_i^{s}$ is the averaged disturbance.[10] The term $\sum_{s=1}^{3} \beta_1^{s} \log v_{i,T_i-h}^{s}$ on the right of equation (3) can regarded as a composite forecast; unbiasedness and efficiency of this composite requires $\beta_0 = 0$, $\sum_{s=1}^{3} \beta_1^{s} = 1$.

Table 5 gives the estimates of equation (3) for four horizons; several features should be noted. First, the estimates of the intercepts for all four horizons are positive and three are insignificant. This indicates that the forecasts when combined in this manner are approximately unbiased and agrees with Figure 2, where the error bands mostly span the zero line. Second, in all but one out of the 12 cases, the estimated slope coefficients are positive, so each source usually makes a positive contribution to the composite forecast. Third, as column 6 shows that the sums of the slope coefficients are insignificantly different from unity, it can be concluded that the composite forecast is also efficient. Finally, from columns 10 and 11, there is no strong evidence against the hypothesis of equal slope coefficients, so that the three sources can be equally weighted to form the composite.

Based on these results, we set the intercept in equation (3) to zero and the slope coefficients to 1/3. Thus, the composite becomes the unweighted mean of the three sources:

$$(4) \quad \log v_{i,T_i} = \log v_{i,T_i-h} + \xi_i, \quad \text{with} \quad \log v_{i,T_i-h} = \sum_{s=1}^{3} (1/3) \log v_{i,T_i-h}^{s}, \quad i = 1, \ldots, N_h,$$

---

[10] In most cases, "actual" capex differs by source, probably because sources update their data at different speeds. Taking the average reduces the random components of the "actuals"; and, of course, when actual is the same in each source (which occurs for some projects), the average is the common value. It is worthwhile noting that there does not seem to be any particular tendency for poor prediction to be associated with projects with diverse actuals.

© 2015 International Association for Research in Income and Wealth

TABLE 5

Combining Forecasts of Capex Costs

$$\log v_{i,T_i} = \beta_0 + \sum_{s=1}^{3} \beta_1^s \log v_{i,T_i-h}^s + \varepsilon_i, \quad i = 1, \ldots, N_h,$$

| Forecast Horizon (6-month periods) (1) | Intercept $\beta_0$ (2) | $\beta_1^s$, Coefficients of Forecast from<br>Prospect (3) | IM (4) | BREE (5) | Sum $\sum_{s=1}^{3} \beta_1^s$ (6) | SEE · 100 (7) | $R^2$ (8) | Number of Projects (9) | Equal Coefficients, $\beta_1^s = \beta_1$ $s = 1, 2, 3$<br>Probability (10) | Restricted Estimate $\beta$ (11) |
|---|---|---|---|---|---|---|---|---|---|---|
| 4 | 0.657 (0.535) | 0.494 (0.443) | 0.815 (0.783) | −0.385 (0.792) | 0.924 (0.072) | 17.15 | 0.959 | 14 | 0.681 | 0.314 (0.019) |
| 3 | 0.116 (0.225) | 0.113 (0.191) | 0.286 (0.191) | 0.587 (0.263) | 0.986 (0.030) | 12.08 | 0.986 | 20 | 0.524 | 0.326 (0.010) |
| 2 | 0.178 (0.087) | 0.287 (0.115) | 0.263 (0.119) | 0.425 (0.056) | 0.975 (0.013) | 9.35 | 0.997 | 26 | 0.246 | 0.328 (0.004) |
| 1 | 0.037 (0.024) | 0.337 (0.027) | 0.317 (0.026) | 0.339 (0.015) | 0.994 (0.004) | 2.76 | 1.000 | 28 | 0.812 | 0.331 (0.001) |

*Notes*: Standard errors are in parentheses. SEE is the standard error of estimate. Column 10 contains p-values for F-statistics of $H_0 : \beta_1^s = \beta_1, s = 1, 2, 3$.

© 2015 International Association for Research in Income and Wealth

where $\xi_i$ is the forecast error. Panel D of Figure 2 contains the corresponding fan chart and it can be seen that the averaging procedure decreases the width of the error bands noticeably—by at least 40 percent in seven of the twelve cases. In other words, averaging leads to a considerable increase in forecast precision. Columns 11–13 of Table 6 confirm that model (4) performs reasonably satisfactorily: For a two-year horizon, the mean error is about 8 percent and the RMSE is 18 percent, while for six months these fall to near zero and slightly less than 3 percent, respectively.[11]

## 7. Summary and Conclusions

The resources sector (mining and energy) has been a prominent contributor to Australia's recent strong economic performance. As estimates of future investment in resource projects are carefully monitored as an indicator of the likely future course of the Australian economy, it is surprising that there is little research assessing the quality of this information. In this paper, we examined carefully three such sources: The Western Australian Department of State Development's *Prospect Magazine*, Deloitte Access Economics' "Investment Monitor" subscription database, and the ABARES/BREE "Mining Projects" database (which, for simplicity, we refer to as just "BREE").

The results of the paper provide guidance regarding how the sources should be assessed and ranked. Table 7 provides a convenient summary of the key results. From panel A, there are significant differences in the estimates of capital expenditure in the three sources, with those in *Prospect* the cheapest, on average, and those in BREE the most expensive. Panel B shows that the bias in *Prospect* is about −5.4 percent, while that for BREE is 3.6 percent. While modest, these are significantly different from zero. The *Investment Monitor* (*IM*) is approximately unbiased. *IM* also distinguishes itself as being the largest net exporter of information (row 8 of the table).

There are some additional important features of the three sources that should also be mentioned: BREE presents some difficulties in tracking projects over time as it does not assign a unique number to each project and also has the problem of referring to the same project by different names at different times. *Prospect* and BREE are provided free of charge by government, whereas *IM* costs $1210 for four issues (or $616 for a single issue). Another feature is timeliness and coverage: *Prospect* is published biannually and deals with major projects in the state of Washington (the location of the majority of projects); and *IM* and BREE are quarterly and report Australian projects. Finally, as discussed in Clements *et al.* (2014), *IM* attracts more media attention than BREE (and *Prospect*, which has a very low media presence), but BREE is a more recent product that is growing rapidly (in terms of citations).

Taken as a whole, the above considerations mean that *IM* is most likely the preferred source. This conclusion is reinforced by the finding that the updates to

---

[11] Table 6 also contains the corresponding error statistics for each of the three sources.

© 2015 International Association for Research in Income and Wealth

## TABLE 6

### The Accuracy of Capex Forecasts

| | Forecast Errors (Logarithmic Ratios · 100) | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Prospect | | | IM | | | BREE | | | Unweighted Mean | | | |
| Forecast Horizon (6-month periods) (1) | Mean (2) | S.D. (3) | RMSE (4) | Mean (5) | S.D. (6) | RMSE (7) | Mean (8) | S.D. (9) | RMSE (10) | Mean (11) | S.D. (12) | RMSE (13) | Number of Projects (14) |
| 4 | 3.91 | 19.15 | 19.54 | 9.56 | 21.62 | 23.64 | 10.50 | 23.15 | 25.42 | 7.99 | 15.64 | 17.56 | 14 |
| 3 | 4.22 | 14.48 | 15.09 | 7.20 | 18.74 | 20.08 | –0.28 | 16.82 | 16.82 | 3.71 | 11.43 | 12.02 | 20 |
| 2 | 2.65 | 12.07 | 12.36 | 3.55 | 16.47 | 16.85 | –1.04 | 13.79 | 13.83 | 1.72 | 9.50 | 9.66 | 26 |
| 1 | 0.55 | 7.09 | 7.11 | 1.06 | 4.04 | 4.18 | –2.05 | 3.13 | 3.74 | –0.14 | 2.74 | 2.74 | 28 |

## TABLE 7

### Summary comparison of three sources of Capex Data

| | Data source | | | | |
|---|---|---|---|---|---|
| Criterion (1) | Prospect (2) | Investment Monitor (3) | BREE (4) | Origin (5) | Comments (6) |
| | | **A. Summary statistics** | | | |
| 1. Mean ($m) | 3,462 | 3,610 | 3,698 | Table 1 | Projects in *Prospect* significantly cheaper; BREE significantly more expensive; substantial dispersion, but sources highly correlated |
| 2. Standard deviation | 6,410 | 6,522 | 6,596 | | |
| 3. t-value (from mean) | 2.86 | 0.42 | 3.17 | | |
| | | **B. Source bias** | | | |
| 4. Source bias (%) | –5.46 | 1.83 | 3.63 | Table 2 | Constant-quality biases consistent with panel A results |
| 5. $H_0$ : Bias = 0 (t-value) | –5.15 | 1.72 | 3.33 | | |
| | | **C. Information flows** | | | |
| 6. $H_0$ : All sources in past = 0 (F-value) | 1.43 | 2.83 | 3.97 | Table 3 | *IM* and BREE consumers of information from other sources |
| 7. $H_0$ : Cross sources in past = 0 (F-value) | 2.07 | 4.10 | 3.43 | | |
| 8. Multilateral information balance (elasticity) | –0.011 | –0.038 | 0.049 | Table 4 | *IM* net information exporter, BREE net importer |

© 2015 International Association for Research in Income and Wealth

the *IM* data contribute more unique information relative to the other two sources (see the online Appendix for details). Thus, in this sense, it is true that "you get what you pay for."

## REFERENCES

Clements, K. W. and J. Si, "The Investment Project Pipeline: Cost Escalation, Lead-Time, Success, Failure and Speed," *Australian Journal of Management*, 36, 317–48, 2011.

Clements, K. W., J. Si, and T. Simpson, "Understanding New Resource Projects," Economics Discussion Paper 14.17, UWA Business School (http://www.business.uwa.edu.au/__data/assets/pdf_file/0005/2530058/14.17-Understanding-New-Resource-Projects.pdf ), 2014.

Flyvbjerg, B., "Survival of the Un-Fittest: Why the Worst Infrastructure Gets Built—and What We Can Do About It," *Oxford Review of Economic Policy*, 25, 344–67, 2009.

———, "What You Should Know About Megaprojects and Why: An Overview," *Project Management Journal*, 45(2), 6–19, 2014.

Mincer, J. A. and V. Zarnowitz, "The Evaluation of Economic Forecasts," in J. A. Mincer (ed.), *Economic Forecasts and Expectations: Analyses of Forecasting Behavior and Performance*, Columbia University Press for the NBER, New York and London, 1–46 (http://www.nber.org/chapters/c1214), 1969.

Productivity Commission, "Public Infrastructure," Inquiry Report No. 71, Canberra (http://www.pc.gov.au/projects/inquiry/infrastructure/report), 2014.

## SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

**Appendix A.1:** The Data
**Table A1.1:** Matched triplets of projects, by project size
**Appendix A.2:** The Transmission of New Information
**Table A2.1:** Impulse response functions
**Appendix A.3:** Sources' Resources
**Figure A3.1:** Revision frequency (282 triplets)
**Table A3.1:** Are capex revisions independent?

© 2015 International Association for Research in Income and Wealth