The tone from above: The effect of communicating a supportive regulatory strategy on reporting quality

Sanne R. van Duin – Vrije Universiteit Amsterdam
Henri C. Dekker – Vrije Universiteit Amsterdam
Jacco L. Wielhouwer – Vrije Universiteit Amsterdam
Juan P. Mendoza – Vrije Universiteit Amsterdam

Compliance with Data Policy for the Journal of Accounting Research

January 2018

1. **A description of which author(s) handled the data and conducted the analyses.**

Sanne van Duin and Jacco Wielhouwer signed a compliance agreement contract with the AFM (Authority for the Financial Markets in The Netherlands) which grants them access to anonymous raw data. Sanne prepared these raw data to generate the dataset used for the empirical analysis using the statistical software package STATA (see #2 for more details). The analysis was conducted by Sanne van Duin and by Juan Mendoza. Although Juan Mendoza had no access to the raw data, he could – without access – write and run code, and view output. Regarding the pretests, all authors were involved in analyzing the data.

2. **A detailed description of how the raw data were obtained or generated, including data sources, the specific date(s) on which data were downloaded or obtained, and the instrument used to generate the data (e.g., for surveys or experiments). We recommend that more than one author is able to vouch for the stated source of the raw data.**

**DATA:** All data were provided by the AFM. On December 9, 2016 we were granted access to the raw data of the 2016 Market Monitor (MM). Based on these, we designed the random stratification procedure, as detailed in Section 3f, and finalized the proposal for the registered report process. The AFM shared the first 2017 MM data with us in the beginning of May, 2017. The last file with data was delivered to us on June 28, 2017. Both Sanne van Duin and Jacco Wielhouwer had access to these data on an AFM laptop via a secured environment.

Sanne van Duin, Jacco Wielhouwer, and Henri Dekker have taken part in on-site visits at the AFM offices. All authors have been involved in discussions with the AFM, and vouch for the stated source of the raw data.

**INSTRUMENT:** Translated 2017 MM questions are included in the proposal and Appendix A. Translated letters used for manipulation are included in Appendix B. The original Dutch letters and the complete MM (in Dutch) are available upon request.

Section 3 of the article describes in detail which items are used for each construct and how these are measured. Access to the answers to the instrument would allow to construct all variables.
3. If the data are obtained from an organization on a proprietary basis, the authors should privately provide the editors with contact information for a representative of the organization who can confirm data were obtained by the authors. The editors would not make this information publicly available. The authors should also provide information to the editors about the data sharing agreement with the organization (e.g., non-disclosure agreements, any restrictions imposed by the organization on the authors, such as restrictions to publish certain results).

The authors included a formal letter of intent, signed by the AFM, in the submission of their proposal. The contact information of the person responsible for collecting and sharing the data with the authors has been provided to the editors. Throughout the registered report process, the authors have informed the editors that the data are collected on a proprietary basis and cannot be shared. There were no restrictions preventing us from following the accepted proposal, other than reporting the minimum and maximum values of certain variables. We therefore deviated from the proposal and reported quartiles instead.

4. A complete description of the steps necessary to collect and process the data used in the final analyses reported in the paper. For experimental and survey papers, we require information about the instructions and instruments used to generate the data, subject eligibility and/or selection, as well as any exclusion criteria. The full set of instructions and instruments can be provided in the online appendix.

All details as indicated in this point are in the paper (random stratification of the manipulated letters, constructs, and variables) or in the appendices to the article (translated instrument and manipulated letters). Exclusion criteria were all decided upon before running the tests, and are described in both the proposal and the paper.

To be provided upon acceptance of the paper and prior to publication:

5. The computer programs or code used to convert the raw data into the final dataset used in the analysis plus a brief description that enables other researchers to use this program. The purpose of this requirement is to facilitate replication and to help other researchers understand in detail how the raw data were processed, the final sample was formed, variables were defined, outliers were treated, etc. This code or programming is in most circumstances not proprietary. However, we recognize that some parts of the code or data generation process may be proprietary, including from the authors’ perspective. Therefore, instead of the code or program, researchers can provide a detailed step-by-step description of the code or the relevant parts of the code such that it enables other researchers to arrive at the same final dataset used in the analysis. In such cases, the authors should inform the editors upon initial submission, so that the editors can consider an exemption from the code sharing requirement.

The proprietary raw data were provided in a MS Excel file. The table below provides details on how each of the variables was constructed. The AFM registered for each firm whether it received a high or low support letter or a letter similar to the one sent in the previous year. We removed all observations with the letter that was similar to the previous year. This resulted in the sample on which all the following steps are applied.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communicated supportiveness of regulatory strategy</td>
<td>In Excel, we generated an indicator variable with the value 1 if the firm received a high support letter, and 0 otherwise.</td>
</tr>
</tbody>
</table>
| Reporting quality                            | After importing the scores on comprehensiveness, precision, and voluntary reporting into STATA, we used the following commands to obtain standardized scores on each dimension.  
  egen z1 = std (precision)  
  egen z2 = std (comprehensiveness)  
  egen z3 = std (voluntaryreporting)  
  We then used the following command to construct a measure for reporting quality:  
  generate RQ_std = (z1 + z2 + z3)                                                                                                                                 |
| Comprehensiveness of reporting                | In Excel, we created indicator variables for all questions in the Market Monitor to indicate whether the firm completed the question (1) or not (0). The comprehensiveness score is equal to the sum of these indicator variables. We excluded item 1.18 as this item is used to determine whether a firm voluntarily provided information (or not). |
| Precision of reporting                       | To measure precision of reporting, we used items 2.11, 4.11, 5.1, 5.2, 5.5, 5.10, 5.11, 6.1 and 6.2, and followed the following steps:  
  1: Number of digits at the end of a number. To determine if the last digit was equal to zero, we used the function RIGHT(number) in Excel. If this digit was not equal to zero, the total number of zeros at the end of the number was zero. If this digit was zero, we added the last two digits by using the function RIGHT(number;2). If this sum was not equal to zero, the total number of zeros at the end of the number was 1. If the sum was zero, we summed the last three digits of the number using RIGHT(number;3) and followed the steps above. We repeated this procedure until we obtained a score for all firms.  
  2: Total number of digits. We separated all digits to obtain the total number of digits. We used function LEFT(number) to extract the first digit, MID(number; 2;1) to extract the second digit, MID(number; 3;1) to extract the third digit, until the ninth digit. We calculated the total number of digits by 9 (maximum number of digits) minus the number of cells that do not contain digits.  
  3: Precision. We then calculated: 1 – (number of zero digits/total digits) to obtain the firm’s item score on precision, and calculated the mean score over the applicable items. For example, if the firm completed four items, the score would be calculated as sum(precision items)/4. The scores were then incorporated into the STATA dataset. |
<table>
<thead>
<tr>
<th>Variable</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voluntary reporting</td>
<td>In Excel, we generated an variable indicating whether (1) or not (0) voluntary narrative (written) information was provided in item 1.18 (if item 1.18 = ““;0;1).</td>
</tr>
<tr>
<td>Firm size</td>
<td>In STATA, we calculated the natural logarithm of the number of full-time employees (item 3.2) using the command: gen ln_fte = ln(fte + 1). We mean-centered the variable using the commands: qui sum ln_fte gen mc_lnfte = ln_fte – r(mean).</td>
</tr>
<tr>
<td>Time horizon</td>
<td>In the raw data file, responses to items 3.5 and 5.18 were coded as “increase”, “stable”, or “decrease”. In Excel, we recoded these as 1, 2 and 3, using the formula: if (item 3.5 = “increase”; 3; if (item 3.5 = “stable”; 2; if (item 3.5 = “decrease”; 1; ““)). We then used the average function to calculate the average score of these two items.</td>
</tr>
<tr>
<td></td>
<td>In STATA, we used the following code to determine if firms had a long or short time horizon: generate short_TH = 0 replace short_TH = 1 if TH &lt;= 1.5 generate long_TH = 0 replace long_TH = 1 if TH &gt;= 2.5</td>
</tr>
<tr>
<td>Number of product types</td>
<td>The raw data provides Yes/No responses per product type, which we transformed into indicator variables and used these to calculate the total sum.</td>
</tr>
<tr>
<td>Type of products</td>
<td>We use three dummy variables that indicate if the firm sells mortgages, damage insurances, or income insurances (item 1.4)</td>
</tr>
<tr>
<td>Experience</td>
<td>In item 1.2, firms had to indicate why they have been inactive in the previous year. Responses 1, 2, and 4 (see appendix A) indicate that firms do not have experience with the Market Monitor. We constructed an indicator variable taking the value 1 if firms ticked one of these responses, and 0 otherwise.</td>
</tr>
<tr>
<td>Member branch organizations</td>
<td>In Excel, we generated a variable taking the value 1 if firms responded “Yes” to the question whether they are affiliated with branch organizations, and 0 if not.</td>
</tr>
<tr>
<td>Franchise</td>
<td>In Excel, we generated an indicator variable with the value 1 if firms responded “Yes” to the question whether they are franchise holders or providers (responses 1 and 2 in item 1.11), and 0 if not.</td>
</tr>
<tr>
<td>Batch</td>
<td>The AFM registered if firms received the instruction letter in batch 1, 2 or 3. We used these data to generate two dummy variables. Batch 1 takes the value 1 if the firm was in batch 1 (if(batch = 1; 1; 0), and 0 otherwise. Batch 2 takes the value 1 if the firm was in batch 2 (if(batch = 2; 1; 0), and 0 otherwise.</td>
</tr>
<tr>
<td>Variable</td>
<td>Details</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| Interaction effects            | In STATA, we generated interaction terms using the following commands:  
  generate STHxHS = Short_TH * High_support  
  generate LTHxHS = Long_TH * High_support  
  generate FTExHS = mc_lnfte * High_support                                                                                                                                                                                                                                                                                                                                                               |
| Change variables               | (Additional test A) In STATA, we generated change measures in reporting quality (and in each dimension) using the following commands:  
  generate change_RQ = RQ2016 – RQ2015  
  generate change_Comprehensiveness = Comprehensiveness2016 – Comprehensiveness2015  
  generate change_Precision = Precision2016 – Precision2015  
  generate change_Voluntaryreporting = Voluntaryreporting2016 - Voluntaryreporting2015                                                                                                                                                                                                                                                                                                                                 |
| Informational fairness         | (Additional test B) Item 8.1 asked firms if they perceived they were timely and clearly informed about the Market Monitor. Responses in the raw data file were recorded as “completely agree”, “agree”, “not agree/not disagree”, “disagree”, and “completely disagree”. In Excel, we transformed these responses into a 5-point scale (1 = completely disagree; 2 = disagree; 3 = not agree/not disagree; 4 = agree; 5 = completely agree). In STATA, we calculated the average using the command:  
  generate informationalfairness = (Clearlyinformed + Timelyinformed) / 2.                                                                                                                                                                                                                                                                                                                               |
| Timeliness                     | (Additional test D) The AFM registered the date and time when firms submitted the Market Monitor. We compared the date and time of the submission with each batch deadline. In Excel, we created a dummy variable taking the value 1 if the firm submitted the Market Monitor on time, and 0 otherwise.                                                                                                                                                                                                 | |
| Effort in responding           | (Additional test F) Registered time. The AFM registered the number of hours that each firm was online while completing the Market Monitor. In STATA, we log-transformed the hours using the following command:  
  generate lntime = ln(time + 1).  
  Self-reported time. Item 8.5 asked intermediaries to provide the number of hours used to complete the Market Monitor. In STATA, we log-transformed the hours using the following command:  
  generate lnSRtime = ln (SRtime + 1).                                                                                                                                                                                                                                                                                                                                                         |
| Extensiveness of voluntary     | reporting (Additional test G) The AFM provided data on the number of words typed in each (open-question) item. These counts can also be obtained from the text using Excel functions such as IF(LEN(TRIM(cellnumber))=0 or 0:LEN(TRIM(A2))-LEN(SUBSTITUTE(cellnumber;”;”’”))+1).                                                                                                                                                       |
| Reminder letter                | (Additional test H) We generated an indicator variable indicating whether the firm received a reminder letter (1) or not (0).                                                                                                                                                                                                                                                                                                                                                          |

Note: Item numbers refer to questions in the 2017 Market Monitor (see Appendix A).
The resulting dataset was used to conduct the analyses in STATA. We used standard functions such as ttest, tab, regress, logit, and margins to do t-tests, chi-square tests, and regression analyses. We used the function “digidis [variable], by [treatment]” to compare our frequency distributions of sales with Benford’s frequency distributions.

6. An assurance that the data and programs will be maintained by at least one author (usually the corresponding author) for at least six years, consistent with National Science Foundation guidelines.

The contract with the AFM specified that all data have to be returned to the AFM close to the end of 2018, but that the raw data and the STATA file will be stored and maintained by the AFM for at least 10 years, following the return of the data. Also, it specified that Sanne van Duin and Jacco Wielhouwer can get on-site access during this period in case this is necessary in the light of the ongoing project.

We refer in the article to two pretests. The letters (treatment), instrument (questions and assignment), and data are available upon request. These, as well as the individual answers from the respondents, will be maintained by one author for at least 10 years.