

# Supplement to “Forecaster’s Dilemma: Extreme Events and Forecast Evaluation”

**Sebastian Lerch, Thordis L. Thorarinsdottir, Francesco Ravazzolo and Tilmann Gneiting**

Table 1 in this supplement is a version of Table 1 in Section 1 of the main paper with direct links to the original media sources.

Figure 1 is a version of Figure 5 in Section 3.3 of the main paper, where simulation-based critical values have been used to adjust the level of the tests to the nominal level. Figures 2 and 3 demonstrate that the model rankings and relative score differences in Section 4.3 are consistent when the threshold in the weight functions is varied.

---

*Sebastian Lerch is PostDoc, Heidelberg Institute for Theoretical Studies (HITS), and Institute for Stochastics, Karlsruhe Institute of Technology, HITS gGmbH, Schloss-Wolfsbrunnengasse 35, 69118 Heidelberg, Germany (e-mail: [sebastian.lerch@h-its.org](mailto:sebastian.lerch@h-its.org)). Thordis L. Thorarinsdottir is Senior Research Scientist, Norwegian Computing Center, P.O. Box 114, Blindern, 0314 Oslo, Norway (e-mail: [thordis@nr.no](mailto:thordis@nr.no)). Francesco Ravazzolo is Associate Professor, Free University of Bozen/Bolzano, Universitätsplatz 1, 39100 Bozen-Bolzano, Italy (e-mail: [francescoravazzolo@gmail.com](mailto:francescoravazzolo@gmail.com)). Tilmann Gneiting is Group Leader, Heidelberg Institute for Theoretical Studies (HITS), and Professor of Computational Statistics, Institute for Stochastics, Karlsruhe Institute of Technology, HITS gGmbH, Schloss-Wolfsbrunnengasse 35, 69118 Heidelberg, Germany (e-mail: [tilmann.gneiting@h-its.org](mailto:tilmann.gneiting@h-its.org)).*

TABLE 1

*Media coverage illustrating the focus on extreme events in public discussions of the quality of forecasts. The sources were accessed April 28, 2016.*

| Year | Headline  | Source                            |
|------|---|-----------------------------------|
| 2008 | Dr. Doom  | The New York Times <sup>1</sup>   |
| 2009 | How did economists get it so wrong?   | The New York Times <sup>2</sup>   |
| 2009 | He told us so   | The Guardian <sup>3</sup>         |
| 2010 | An exclusive interview with Med Yones - The expert who predicted the financial crisis | CEO Q Magazine <sup>4</sup>       |
| 2011 | A seer on banks raises a furor on bonds   | The New York Times <sup>5</sup>   |
| 2013 | Meredith Whitney redraws 'map of prosperity'  | USA Today <sup>6</sup>            |
| 2007 | Lessons learned from Great Storm  | BBC <sup>7</sup>                  |
| 2011 | Bad data failed to predict Nashville Flood  | NBC <sup>8</sup>                  |
| 2012 | Bureau of Meteorology chief says super storm 'just blew up on the city'               | The Courier-Mail <sup>9</sup>     |
| 2013 | Weather Service faulted for Sandy storm surge warnings                                | NBC <sup>10</sup>                 |
| 2013 | Weather Service updates criteria for hurricane warnings, after Sandy criticism        | Washington Post <sup>11</sup>     |
| 2015 | National Weather Service head takes blame for forecast failures                       | NBC <sup>12</sup>                 |
| 2011 | Italian scientists on trial over L'Aquila earthquake                                  | CNN <sup>13</sup>                 |
| 2011 | Scientists worry over 'bizarre' trial on earthquake prediction                        | Scientific American <sup>14</sup> |
| 2012 | L'Aquila ruling: Should scientists stop giving advice?                                | BBC <sup>15</sup>                 |

<sup>1</sup> <http://www.nytimes.com/2008/08/17/magazine/17pessimist-t.html>

<sup>2</sup> <http://www.nytimes.com/2009/09/06/magazine/06Economic-t.html>

<sup>3</sup> <http://www.guardian.co.uk/business/2009/jan/24/nouriel-roubini-credit-crunch>

<sup>4</sup> <http://www.ceoqmagazine.com/whopredictedfinancialcrisis/index.htm>

<sup>5</sup> <http://www.nytimes.com/2011/02/08/business/economy/08whitney.html>

<sup>6</sup> <http://www.usatoday.com/story/money/business/2013/06/05/meredith-whitney-book-interview/2384905/>

<sup>7</sup> <http://news.bbc.co.uk/2/hi/science/nature/7044050.stm>

<sup>8</sup> [http://www.nbc15.com/weather/headlines/January\\_13\\_Report\\_Bad\\_Data\\_Failed\\_To\\_Predict\\_Nashville\\_Flood\\_113450314.html](http://www.nbc15.com/weather/headlines/January_13_Report_Bad_Data_Failed_To_Predict_Nashville_Flood_113450314.html)

<sup>9</sup> <http://www.couriermail.com.au/news/queensland/bureau-of-meteorology-under-fire-after-a-weekend-of-wild-weather-and-storms-in-queensland-left-many-unprepared/story-e6freoof-1226519213928>

<sup>10</sup> <http://www.nbcnewyork.com/news/local/Sandy-Report-Weather-Storm-Surge-Warnings-207545031.html>

<sup>11</sup> <http://www.washingtonpost.com/blogs/capital-weather-gang/wp/2013/04/04/weather-service-changes-criteria-for-hurricane-warnings-after-sandy-criticism/>

<sup>12</sup> <http://www.nbcnews.com/storyline/blizzard-15/national-weather-service-head-takes-blame-forecast-failures-n294701>

<sup>13</sup> <http://edition.cnn.com/2011/09/20/world/europe/italy-quake-trial/>

<sup>14</sup> <http://www.scientificamerican.com/article/trial-such-as-that-star/>

<sup>15</sup> <http://www.bbc.co.uk/news/magazine-20097554>

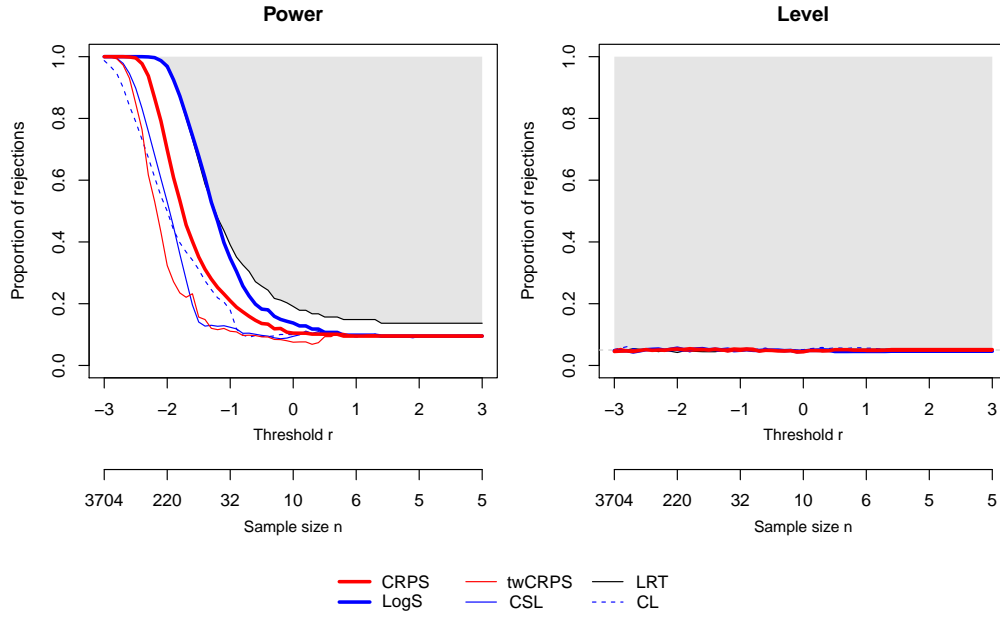


FIGURE 1. *Power (left) and level (right) of the likelihood ratio test (LRT) and one-sided Diebold-Mariano tests in the simulation setting described in Section 3.2 in the paper. In the panel for power, the shaded area above the curve for the LRT corresponds to theoretically unattainable values for a test with nominal level. In the panel for level, the dashed line indicates the nominal level. In contrast to Figure 5 in the paper, the critical values of the Diebold-Mariano tests are here determined based on Monte Carlo simulations instead of using the asymptotic values.*

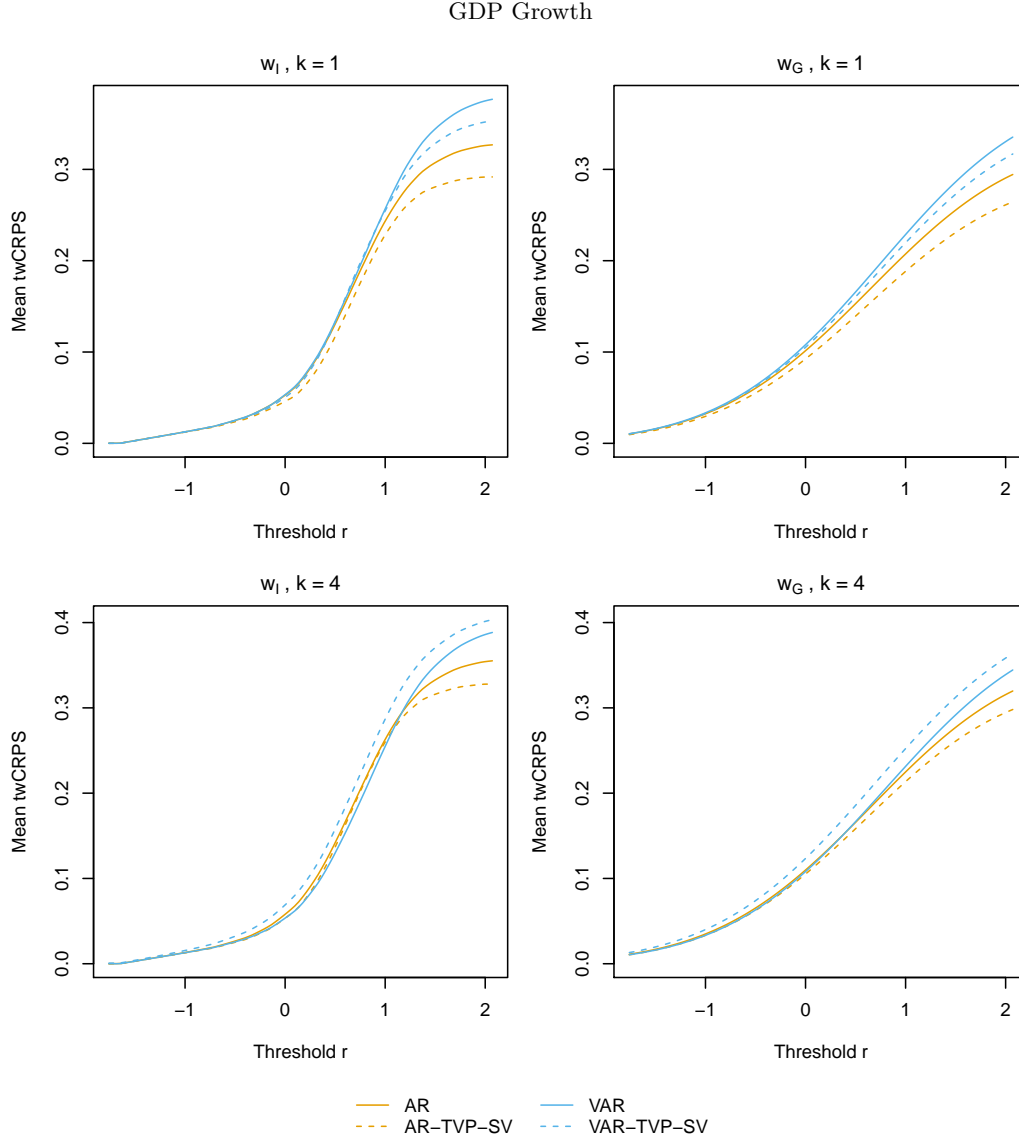


FIGURE 2. Mean threshold-weighted CRPS for probabilistic forecasts of GDP growth in the U.S. for the first quarter of 1985 to the second quarter of 2011 at prediction horizons of  $k = 1$  and  $k = 4$  quarters, respectively, shown as function of the threshold  $r$  in the weight functions  $w_I(z) = \mathbb{1}\{z \leq r\}$  and  $w_G(z) = 1 - \Phi(z | r, 1)$ .

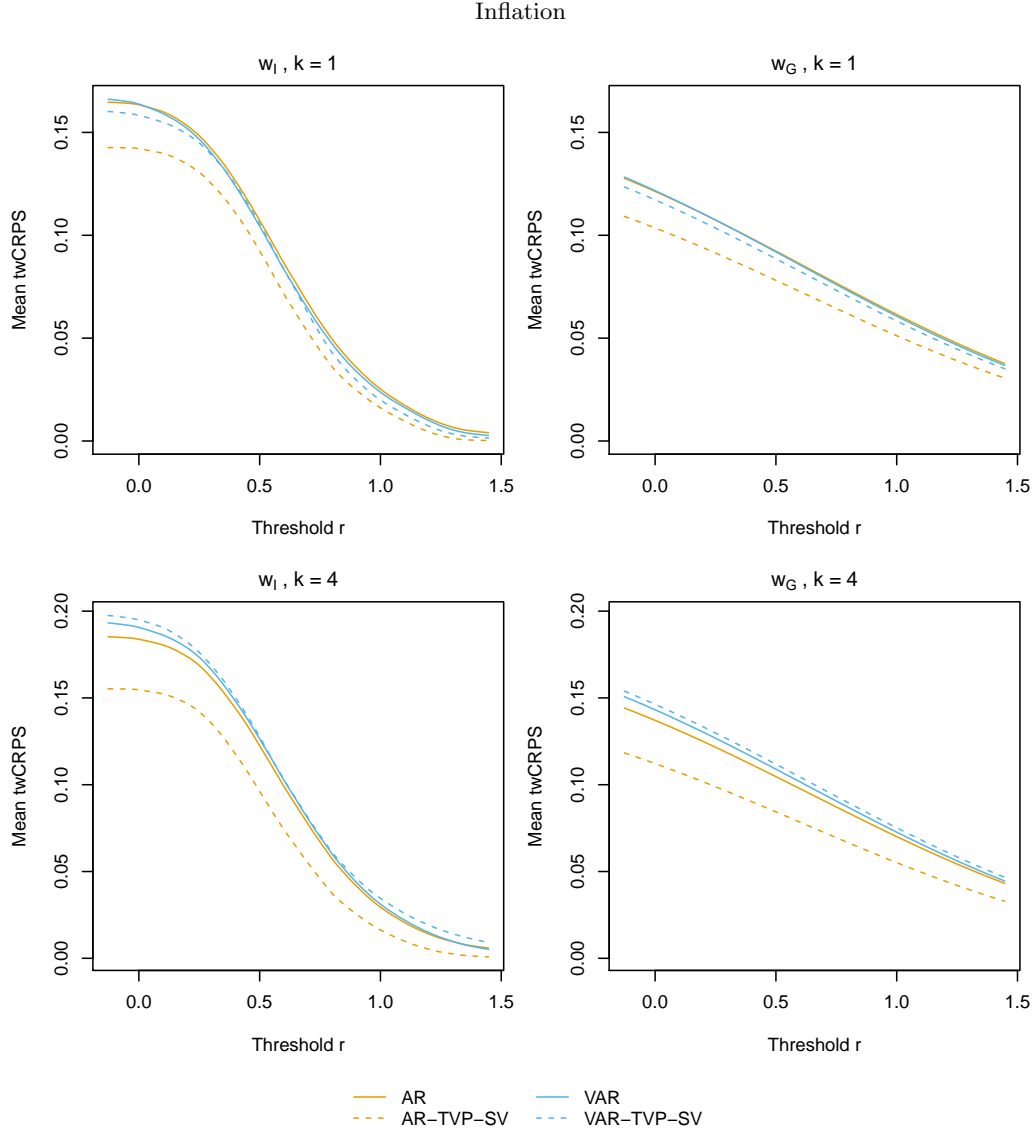


FIGURE 3. Mean threshold-weighted CRPS for probabilistic forecasts of inflation in the U.S. for the first quarter of 1985 to the second quarter of 2011 at prediction horizons of  $k = 1$  and  $k = 4$  quarters, respectively, shown as function of the threshold  $r$  in the weight functions  $w_I(z) = \mathbb{1}\{z \geq r\}$  and  $w_G(z) = \Phi(z | r, 1)$ .