2011

Summary of G. Van Harten, 'The Use of Quantitative Methods to Examine Possible Bias in Investment Arbitration' and 'Reply' [to Franck, Garbin, and Perkins] in the Yearbook on International Investment Law & Policy (2011)

Gus Van Harten
Osgoode Hall Law School of York University, gvanharten@osgoode.yorku.ca

Summary of G. Van Harten, "The Use of Quantitative Methods to Examine Possible Bias in Investment Arbitration" and "Reply" [to Franck, Garbin, and Perkins] in the *Yearbook on International Investment Law & Policy* (2011).

This is a summary of two articles that I produced in an exchange with Susan D. Franck, Calvin P. Garbin, and Jenna M. Perkins in the *Yearbook on International Investment Law & Policy*. The first article, "The Use of Quantitative Methods…", discussed the use of empirical methods to examine possible bias in investment arbitration. It supported empirical methods but cautioned against over-statements of findings. To elaborate on the latter point, two studies – Franck (Harvard Int'l L.J., 2009) and McArthur and Ormachea (Rev of Litigation, 2009) – were critiqued.

My first article was accepted for publication in the *Yearbook*, subject to a process of reply and counter-reply with Susan Franck. My second article, the "Reply", responded to a detailed rebuttal by Franck, Garbin, and Perkins of my first article. In the Reply, I made some further points about Franck (2009), summarized below.

- **Faulty reasoning behind the conclusions in Franck (2009)**

  In my first article, Franck (2009) was criticized for making the erroneous claim that a lack of reliable evidence of a connection between development status and outcome entailed that there was no such connection and that investment treaty arbitration had procedural integrity, was functioning fairly, etc. In my Reply, I elaborated on why this reasoning, repeated by Franck, Garbin, and Perkins, was erroneous.

  First, it assumed that the competing claims about the presence or absence of bias were contradictory. In fact, in Franck (2009), the researcher was able to test only whether there was reliable evidence of bias or reliable evidence of a lack of bias. These were contrary (not contradictory) statements because they cannot both be true but can both be false (i.e. one can lack reliable evidence of both claims).

  Second, following from the first error, the conclusions in Franck (2009) were based merely on the author's allocation of the burden of proof. The author opted to put the burden on those claiming an imbalance in the system rather on those making the counterclaim that there is no imbalance. This was a prerequisite to the author drawing conclusions about the integrity and fairness of the system, having found no reliable evidence of a link between development status and outcomes. Had the burden been allocated the other way round, then on the author's reasoning this would have been taken to mean that the system was not functioning fairly, etc.

  Third, there are many factors that can affect the integrity and fairness of the system and that were not tested in Franck (2009). For this additional reason, it was inappropriate for the author to suggest that the results in Franck (2009) entailed that the system was functioning fairly, etc.

- ***Non-disclosure of important information in the reporting of coding outcomes***

  In Franck (2009), the author did not communicate to readers that most of the cases she classified as being against developed countries were in fact cases against Mexico, the Czech Republic, or the Slovak Republic. It was important to disclose this because the expectations that Franck (2009) sought to test were based on concerns laid out in other sources that treated

Mexico and former East Bloc countries as among the countries expected to suffer from an imbalance in the system.

Further, it became apparent—based on a review of the codebook for the dataset used in Franck (2009) — that the author was aware, as far back as 2006, of limitations of possible measures of development status (and of arbitrator nationality) and opted not to code either of these variables. Yet both variables are integral to the analysis in Franck (2009), and readers were never informed about when and why the decision was taken by the author to code these variables in spite of earlier reservations.

- **Problems with a further analysis by Franck, Garbin, and Perkins**

  In their rebuttal, Franck, Garbin, and Perkins provided a new analysis of Franck's original dataset. This analysis had various problems, many tracking flaws in Franck (2009). First, in the new analysis, the number of cases classified as being against developed countries was reduced from 18 to 7. As in Franck (2009), the minimal data led to an unacceptably high risk of error. The appropriate finding in both the new analysis and Franck (2009) was simply that there was insufficient data to draw any conclusions about the presence or absence of bias, not that the system was functioning fairly, etc.

  Second, in the new analysis, six of the seven cases classified as being against developed countries involved claims, under NAFTA, by Canadian investors against the U.S. or vice versa. The one remaining case against a developed country was against Spain under a bilateral investment treaty. Thus, the data on developed countries related almost exclusively to the experience of two countries under one treaty. This re-emphasized that it was inappropriate for the author in Franck (2009) to draw conclusions about the integrity and fairness of the system.

  Third, in the new analyses and in Franck (2009), readers were not informed properly of how several data substitutions by the researchers – described as winsorizing (i.e. truncation) of the data – affected the results. Only after obtaining and reviewing the author's original dataset, and reviewing Franck, Garbin, and Perkins rebuttal, did I learn that data substitutions had reduced the mean amount awarded against developing countries in the data from $12.7 million to $1.4 million, without changing the mean amount awarded against developed countries ($630,000). Readers should have been informed from the outset of this major implication of the data substitutions.

Overall, the results in Franck (2009) did not provide reliable evidence of the presence or absence of actual bias in investment arbitration. Ironically, this reinforces the important role of institutional safeguards of judicial independence to counter reasonable perceptions of bias in adjudication. In the absence of such safeguards in investment treaty arbitration, reasonable perceptions of inappropriate bias may continue to arise.

The Reply was subject to a further rebuttal by Franck, Garbin, and Perkins. Readers are encouraged to examine the full exchange in the *Yearbook of International Investment Law & Policy* (2011).