**Box xx: Path dependency of imputation routines**

As outlined above, the process from “raw” trade data to published trade results encompasses several distinct steps. In general, these steps follow a logical and cogent sequence. For instance, cleaning data and weeding out errors is done before other processes such as imputation or balancing can take place. There is, however, also rationale for more ambiguity in the sequence of the various steps and not all steps follow a cogent and unequivocal logic. What is more, there is no a priori (conceptual) reason to prefer one step over the other in the sequence of the implementation, e.g. to undertake the mirroring step before the unit value imputations or vice versa.

Intuitively, one may be inclined to first mirror missing data so that the unit value calculations are being put on a broader footing. As outline in the box xx on the use of unit values, such a decision would have direct implications on the choice of the unit value, whereby more observations will eventually lead to a decision as to whether country-specific or global unit values are employed in the imputation process. The intuition is likely to be confirmed when the mirrored data for a country are at about the same level as the average unit value of a given reporter/partner. If, however, a mirrored flow is far out of the normal range, this will also affect and indeed distort the unit value used in the next step of the process and, as these unit values would be applied to all flows with missing information, many imputed data would be affected.

Without prior knowledge of the correct sequence of steps, the final results become path-dependent on the sequence of imputation steps. To gauge the magnitude of the potential effect, a number of empirical tests have been conducted. The results are compiled in table xx.

Table xx:

The results suggest that, at least in general, the differences that arise from different imputation sequences are negligibly small. There are, however, a few commodity specific exceptions. Particularly where flows are characterized by large product quality differences across reporting countries, poor product differentiation in reporting, etc., path dependency can influence the results in a significant way. Overall, however, such commodity-specific differences were not deemed important to deviate from the proposed sequence of data processing steps outlined in table xy above.