Workplace and Employee Survey - Imputation

There are three types of no response in WES: unit non-response, item non-response and wave non-response. Unit non-response occurs if it is not possible to obtain the survey information for all variables of a selected unit (workplace or employee) due to a refusal or the impossibility to make a contact. Item non-response occurs if we are able to obtain only partial information from a selected unit. This could be due to a refusal or the impossibility to respond to some questions or inconsistencies in the data collected. Finally, wave non-response occurs when we have at least partial information at a previous wave for a selected unit but no information at the current wave. In the current non-response treatment strategy, a weight adjustment for the respondents is computed to deal with unit non-response while item and wave non-response are treated using different imputation methods. Cross-sectional versions of these methods are used for units appearing at the current wave for the first time. Otherwise, if historical data are available, longitudinal versions are used.

In the case of item non-response, some processing and editing is done before proceeding to imputation in order to remove inconsistencies in the data collected. Editing is based on a set of rules that must or should likely be satisfied. This process leads to either creating additional missing values or imputing by deduction the values that should have been reported. This type of imputation is used when a single missing field can be deduced uniquely from the given information. For example, if one component of a sum is missing and the remaining components including the sum are present, then the missing component can be determined uniquely.

Once this process is completed, the remaining missing values are imputed using one of the four methods described below. To avoid producing inconsistencies in the imputed data, most interrelated fields are imputed as a block. Since there are a number of questions falling into this category, a post-imputation system has been developed to preserve all inter-field relationships.

There are four main imputation methods being used both for the employer and employee portions of WES: carry-over, distributional, weighted hot deck and nearest-neighbour. Carry-over imputation is used when historical data is available. It consists simply of transferring the value from the previous wave to the current wave. For continuous variables, the value may be adjusted by a trend from an auxiliary variable. Obviously, there is no cross-sectional version of carry-over imputation.

Distributional imputation is used for questions where the respondent is asked to provide a total and its breakdown into multiple categories when either two or

more of the categories are missing. The distribution of the categories is computed at a macro level and applied at the micro level. To illustrate this approach, let us assume that the respondent gave us total employment but was unable to provide a breakdown by occupational group. We would apply the distribution of the occupational groups computed at the industry/size level to the total employment figure to impute the missing fields. This method can only be applied cross-sectionally.

For weighted hot deck, a missing field is imputed using the response of a randomly-selected donor within an imputation class; either the value of the donor for the missing variable is imputed directly or the ratio from the donor between the missing variable and an auxiliary variable. In the latter, the ratio is then multiplied by the auxiliary value from the recipient. The method is longitudinal if the auxiliary variable or the imputation classes are determined using previous wave information, otherwise it is cross-sectional. The donor is selected randomly with a probability of selection equal to the ratio of its sampling weight over the sum of the sampling weights of all units in the corresponding imputation class. The weighted hot deck approach was adopted for the following two main reasons: i) the method is easy to implement and ii) it leads to approximately unbiased point estimates provided that all units within each imputation class can be assumed to have the same propensity to respond (Rao, 1996).

Finally, nearest-neighbour imputation is used to preserve relationships between certain variables. It is a donor imputation method like weighted hot-deck. This method replaces the missing values of a given recipient by the corresponding values from the donor which is the closest to the recipient with respect to a few matching variables. Similarly to weighted hot-deck, a ratio from the donor can be imputed to the recipient, which is then multiplied by an auxiliary value from the recipient. Again, the method is longitudinal if the auxiliary variable, the imputation classes or the matching variables are determined using previous wave information, otherwise it is cross-sectional. It is very similar to weighted hot-deck imputation. It differs only in the way donors are selected. With weighted hot-deck imputation, donors are randomly selected while they are deterministically selected according to some matching variables with nearest-neighbour imputation.