ARDx Capital Stock User Guide

1. Overview

The ARDx capital stock ('capstock') is an imputed dataset intended to generate approximate reporting unit (RU) level capital stocks for use in productivity analysis. It was developed from the ARD2 capital stock (sometimes called the Martin capital stock after its original author, Ralf Martin of Imperial/LSE)[1].

The data is created by a simple Perpetual Inventory Model (PIM):

$$k_t = (1 - \delta)k_{t-1} + I_t$$

where k_t is the capital stock in period t, I_t is investment in period t and δ is the depreciation rate. Investment data is taken from the ARDx files. Depreciation is a parameter in the model. The model creates capital stock data at RU level for each of the asset types specific in the model; in the default model these are currently 'buildings' (b), 'vehicles' (v) and 'other'.

There are two fundamental difficulties with creating such a model in the ARDx. The first is that the initial capital stick k_0 is unknown. The second is that most RUs do not have investment data for most years, and some have none at all. Hence the capital stock values are almost all imputed; the ARDx capstock file should be considered a *synthetic* dataset.

The original Martin and Harris capstock files were created to improve productivity analysis. These were subsequently adopted by other researchers and, in the case of the Martin stock, a decision was taken to create a 'standard' capital stock by the VML team to provide consistency amongst researchers; researchers would be free to create their own (code was available) but in practice none did.

The ARDx capital stock is created and maintained by the VML team at ONS. However, the code is freely available, and researchers can choose to customise a version to reflect their own preference for depreciation rates, real versus nominal prices, tolerance for missing values and so on.

This guide explains how the user can create his or her own version. The next section discusses the structure and use of the data. Section 3 illustrates how the missing data problems are addressed; and the final section discusses how to create different capital stocks, and the options which have been set up for the user to change the program easily. For a more complete understanding of the code, we suggest you look at the ARDx Capstock Data Managers guide, which goes into the technical detail.

2. Using ARDx capstock

ARDx capstock is a single file covering all years of the ARDx, organised at the RU level. The key variables (all lower-case) are

ruref	RU reference (str11)
year	Year
capstock_[asset-type]	Capstock for 'asset-type' for that RU-year combination

[1] Richard Harris at Durham developed and maintains an alternative capital stock at Local Unit (LU) level. This is referred to as the Harris capital stock

At present, the only asset-types coded are ‘b’, ‘v’ and ‘other’. The dataset also contains the year-on year investment values used to create the capital stock; again these are specific to asset types.

Ruref-year combinations are unique within the dataset, so the capstock data can be merged on ruref alone for single-year analysis or both variables for multi-year analysis. Ruref is an 11-character string, consistent with the main ARDx files.

All calculations are carried out by industry, at the ‘letter’ level. This is the smallest industry grouping, slightly smaller than (and encompassing) 2-digit SIC code. See the SIC codebook or Data Managers Guide for details.

The code can, as it stands, produce negative capital stocks. The code carries out several corrective actions but at present those still left after ‘correction’ are left in the dataset; users may decide to set these to zero or missing. The exact amount of negative capstocks can be found in the log files creating the cap stock.

Important: specificity of the dataset

The capstock dataset looks over the existence of the RU in the ARDx universe, values for years out of scope are interpolated, and the investment data uses information form the whole period to calculate investment in years where this is not observed. Therefore it is likely that, as years are added to the dataset, values for capstock in previous years will now be re-estimated.

It is therefore important that researchers wishing to have reproducible results take careful note of the name of (or simply take a copy of) the capstock version they have previously used. The capstock file name generally incorporates all the information on parameters set.

Using ARDx capstock: a warning

Confusion sometimes arises amongst researchers about the status of the capstock file, with it often being treated as respondent data and occasionally even used as the dependent variable in models. As noted above, the ARDx capstock is a **synthetic** dataset. **We therefore would like to emphasise that researchers using the ARDx capstock are cautious in their interpretation of results** as the inclusion of a constructed variable in a model may affect some of the observed properties in other variables which were used in the capstock construction (such as employment). In particular, direct interpretation of coefficients should be avoided, and findings should be treated as indicative irrespective of the size of the standard errors.

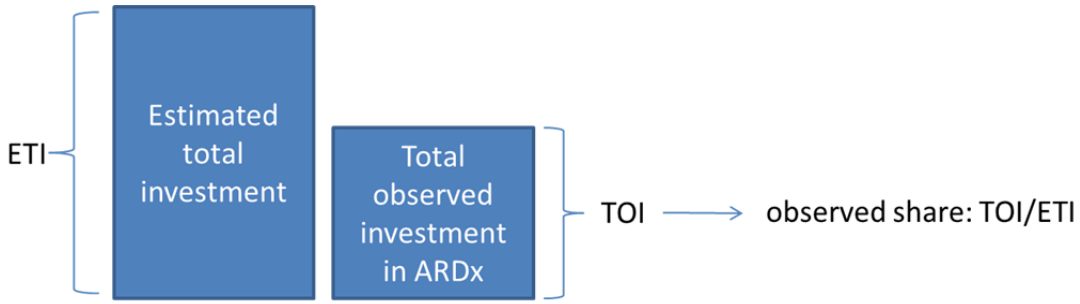
3. How the capital stock is created

As noted above, there are two fundamental problems: initial capital stocks and missing investment data. We take each of these in turn.

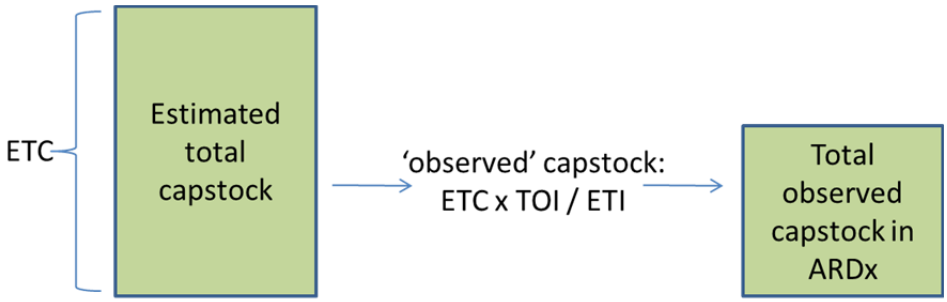
3.1 Initial values

The first time an RU is observed, it is given an initial capital stock. This is calculated as a proportion of the total capital stock in that year. However, this cannot be a simple proportion because the ARDx does not have responses from all firms. Moreover, proportions which do not take account of the firm size and industry may under-or over-allocate initial capital stocks.

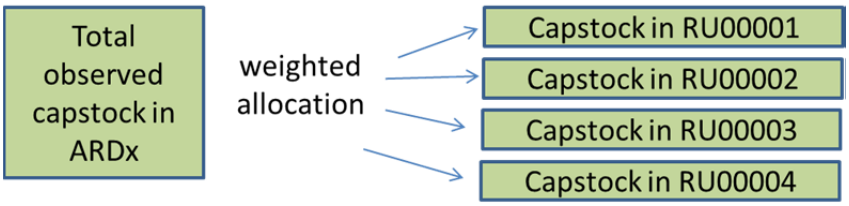
The solution has three stages. First, a number for total investment in the year is established (in the default program run by the VML, this is taken from National Accounts) for the industry. This is compared with observed investment in the ARDx to get a measure of how much of the actual investment is observed in the ARDx:



This ratio is then applied to a consistent measure of capital stocks (in the case of the default VML code, capital services estimates from Nationla Accounts, not the capital stock measures). This should then give an estimate of how much capital stocks there should be in the respondent data:



Finally, this 'observed total' is then allocated to individual RUs using the options set by the user:



The default VML setting is to use IDBR turnover as the weight.

Aft er the initial capstock has been set for a firm on its first appearance in the data, all subsequent years are created by the PIM, once investment has been determined.

3.2 Calculating missing investment data

For all but the largest companies, most of the data on investment is missing in any particular year. Missing investment data is imputed using investment-per-employee in the observed years to create a baseline, allocated by employment. The reason for the choice of employment as the weight is because employment and turnover are the only variables observed in all RUs for all years.

The example below shows how this might work for a fictional RU, observed for six out of the four years 2003-2008. Observed investment is unchanged but missing investment is calculated from the per-employment ratio in the non-missing years:

Year	Observed?	IDBR employment	Actual investment £000	Imputed investment £000
2003	yes	43	£127	£127
2004	no	45	--	£84
2005	yes	61	£89	£89
2006	yes	54	£78	£78
2007	no	58	--	£109
2008	yes	60	£114	£114
Average per-employee investment				**1.87**

Firms which are never selected to provide financial information obviously cannot contribute anything to the determination of capital stocks and so are excluded from the dataset. This is less of a limitation as it sounds, as it means these firms have very little of analytical value apart from IDBR data.

When firms have some missing values, a judgment is needed as to how much imputation is allowable. One of the user options allows for a 'tolerance limit' to be set. Depending on the value, the capstock can contain any firm which ever supplies financial information, only those who only have complete records, and anything in between. The default VML setting allows for a large amount of imputation.

4. Creating the capital stock

The capstock code is controlled by global and local macros set at the start of the code. These are intended to make the code as parameterised as possible; that is, a variety of different outcomes can be generated by changing parameter values, rather than having to edit the Stata code. Users can edit these to customise their dataset.

The capstock code runs in two sections. First, there is a need to create the 'input' capstock. This contains the universe of observations, plus aggregate data, and imputed investment data. The user should not need to change this unless the number of years to be considered is different, or if a different set of aggregates is to be used.

This section of the code can take several hours to run, depending on the business of the system. Hence customisation of this is limited to the minimum so that only the major changes listed above necessitate a rerunning of this section of the code. The decisions in this section of the code (for example, employment weighting for creating imputed investment) should be uncontroversial.

The second section contains the main user options. This section runs much faster (half an hour or so, depending on the number of observations and the verbosity of output).

The user options in this section are very subjective: what is the appropriate depreciation rate? How should initial capital stock be weighted? What is the right tolerance level for imputed investment? Accordingly, the code has been parameterised so that users can play around with these options. These are described in the table below (for more detail, see the Data Managers Guide).

Option	Values	Purpose
verbose	0 (=no), 1 (=yes)	Generates large amount of analytical output for assessing
tolerance	Value	Tolerance value – depends on following
tolerance_type	'proportion'	If 'number', 'tolerance' is the minimum number of

	'number'	obs needed to be included in the dataset If 'proportion', 'tolerance' is the minimum proportion of non-missing observations
drop_neg_capex	'all' 'neg_only' 'none'	Ignore firms which have negative capex? 'all' – drop firms which have overall negative investment 'neg_only' – drop years which are negative [NB likely to lead to significant over-estimation] 'none' – take no action
depreciation_[]	Value	Deprecation rates for asset type []
current_price	0 (=no), 1 (=yes)	Whether the data should be in current or constant price terms NB this should have been set up in the aggregate datasets
ARDx_vars	ARDx variable names and synonyms	Variables to be used from the ARDx for weighting the initial capital stocks – **changing this requires running the whole program, including creation of the 'input' capstock.**
key	Variable name	Variable to be used for allocating initial capital stocks
capstock_methods	'simple' 'key_share' 'emp_share'	Methods to be used for creating initial capstocks – any or all can be specified. See the Data Manager Guide for details; in practice, seems to make little difference
capstocks_to_ create	'simple' 'key_share' 'emp_share'	Capstocks to be created. The program will create capstock variables of each type [b,v,other] for any and all of the capstock methods requested
backfill_method	'missing_capex' 'rebase_init_stock'	Methods for dealing with negative capstocks by backfilling. Each method makes different assumptions about the cause of the negative capstock/. See the Data Manager Guide for details
max_backfill_ attempts	Value	Number of times the backfill method will re-iterate before giving up and leaving negative values in place

5. Appendix: Stats

Probalb should have sme stats in here – perhaps no of rurefs/obs with capstock values?