

Industry Codes in the FirmStat Registers

Anders Humlum*

September 5, 2021

1 Introduction

This notes assigns time-consistent industry codes to firms in the FIRM and FIGT registers from 1992 to 2018, according to the [DB07](#) industry classification.

The measurement challenge is that Statistics Denmark (DST) has revised its industry classification throughout the years (DSE77 before 1992, DB93 from 1993 to 2002, DB03 from 2003 to 2006, and DB07 after 2007), and that no official m:1 keys exist between these revisions.

Our starting point is the [m:m keys](#) between the different classifications: [DSE77 to DB93](#), [DB93 to DB03](#) with [title changes](#), and [DB03 to DB07](#). DST has [estimated](#) the DB07 code of each firm from 2000 to 2006. Similarly, DB93 codes have been estimated for 1992. For these years, we use the estimated DB07 codes directly. Furthermore, these estimates allow us to infer a m:1 crosswalk by linking each DB93 code to its most frequent DB07 code.

Section [2](#) describes how we construct the m:1 crosswalk from DB93 to DB07. The crosswalk is available for download [here](#). Using this crosswalk, Section [3](#) details a procedure for assigning DB07 codes to all firm-year observations in the FirmStat registers. Section [4](#) shows that, once this procedure has been applied, the number of firms by industries and industry switching rates are stable over the underlying change in industry nomenclatures.

*E-mail: humlum@uchicago.edu. Please contact me if you have suggestions for improving the crosswalk or the industry assignment procedure.

2 Crosswalk From DB93 to DB07

In this section, we construct a $m:1$ crosswalk from DB93 to DB07. Our starting point is the DB07 codes that DST has estimated for firms in 2000, 2001, and 2002. For each DB93, we calculate the share of observations and total employment that are linked to each DB07 code. We use these shares, together with whether the detailed codes are consistent at the aggregate level, to create the crosswalk.¹ We assign each DB93 code to a unique DB07 code using the following procedure

1. If a single DB07 has the highest share of observations and employment, and has consistent aggregates, assign that code automatically.
2. If the DB07 codes with the highest observations and employment differ, but both have consistent aggregates, pick manually between the two codes.
3. If one of the DB07 codes with highest observations or employment have inconsistent aggregates, and another linked DB07 code with consistent aggregates exists, choose manually among these codes.
4. For DB93 codes not active in 2000-2002, pick manually among the DB07 codes in the official crosswalk, taking into account whether the codes have consistent aggregates.

Table 1 shows the number of DB93 codes we assign in the different steps.

Table 1: Linking DB93 Codes to DB07 Codes

Match Method	Codes	Percent
1. Unique top candidate with consistent aggregates	646	77.6
2. Both top candidates have consistent aggregates, but top differ	14	1.6
3. Choose among top candidates or those with consistent aggregates	131	15.7
4. Manual (DB93 code not active 2000-2)	41	4.9

The crosswalk is available for download [here](#).

¹To check for aggregate consistency, we select for each “53 grouping” of DB93, the “36 grouping” of DB07 that has the highest share of observations or employment. This check for aggregate consistency also controls for the fact that, for a few firms, Statistics Denmark had external information that allowed them to assign a firm to a DB07 code outside the official crosswalk.

3 Assigning DB07 Codes to Firms

We assign DB07 codes to firms in the FIRM/FIGT registers using the following procedure

1. From 2000, use the DST estimates.
2. If firm’s DB93 code is the same as in 2000, use the DB07 code from 2000.
3. If firm’s DB93 code is different from 2000 but the same in 2001 or 2002, use the most frequently estimated DB07 in those years.
4. If firm’s DB93 code is not used in 2000-2, use crosswalk from Section 2.
5. If firm has missing industry code, use (next, then last) non-missing code for that firm.

Table 2 shows the number of observations 1992-2006 that are assigned to a DB07 code in each step.

Table 2: Assigning DB07 Codes to Firms 1992-2006

Source of DB07 code	Observation Count	Count Share (%)	Employment Share (%)
1. DST estimate (raw data)	3,603,775	58.0	61.5
2. Firm value in 2000	1,200,405	19.3	27.5
3. Firm-db93 mode	369,658	5.9	0.9
4. Crosswalk	1,027,964	16.5	9.8
5. Next/last non-missing	4,065	0.0	0.0
Missing	975	0.0	0.0

4 Evaluating the Industry Codes

Figure 1 shows the number of active firms by one-digit industry codes.² The first vertical dashed line marks the change in data source from the FIGT to FIRM register in 1999. While FIGT only included private VAT liable firms, FIRM includes all industries.³ Figure 1 focuses on the industries that are covered throughout the years. Figures 3 and 4 show industry switching rates (across six-digit codes) for firms whose industries are covered throughout the years.

²Active firms employ at least 0.5 FTE workers. The FIRM register applies this restriction.

³Section 4 of this documentation discusses the differences in sample coverage between the FIRM and FIGT registers.

Figure 1: Active Firms (in Thousands) by One-Digit DB07 Industries

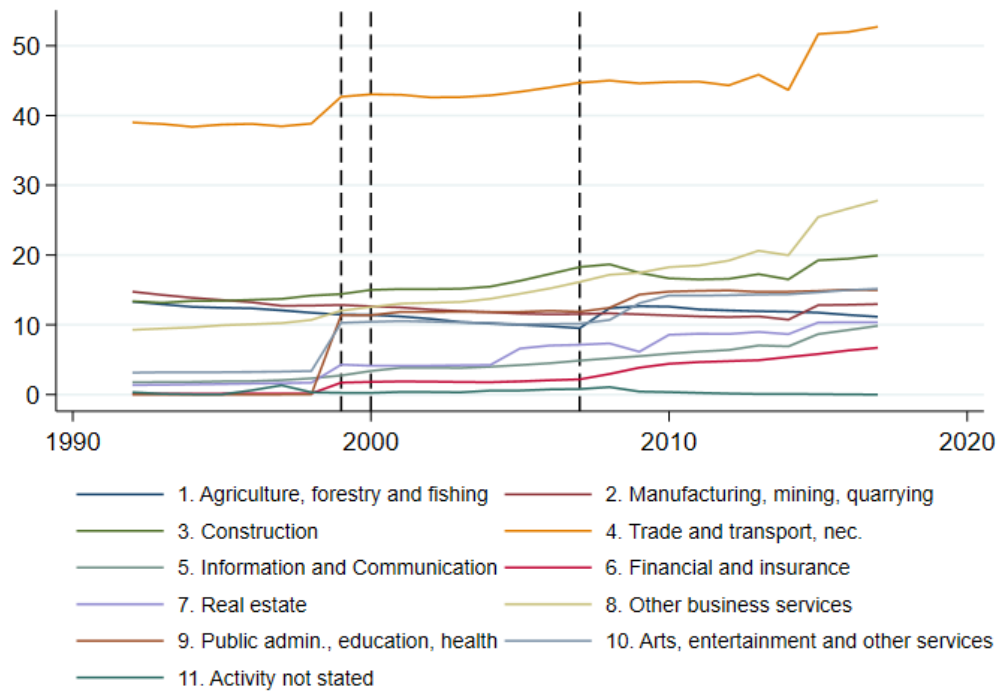


Figure 2: Active Firms (in Thousands) by One-Digit DB07 Industries (Full Data Coverage)

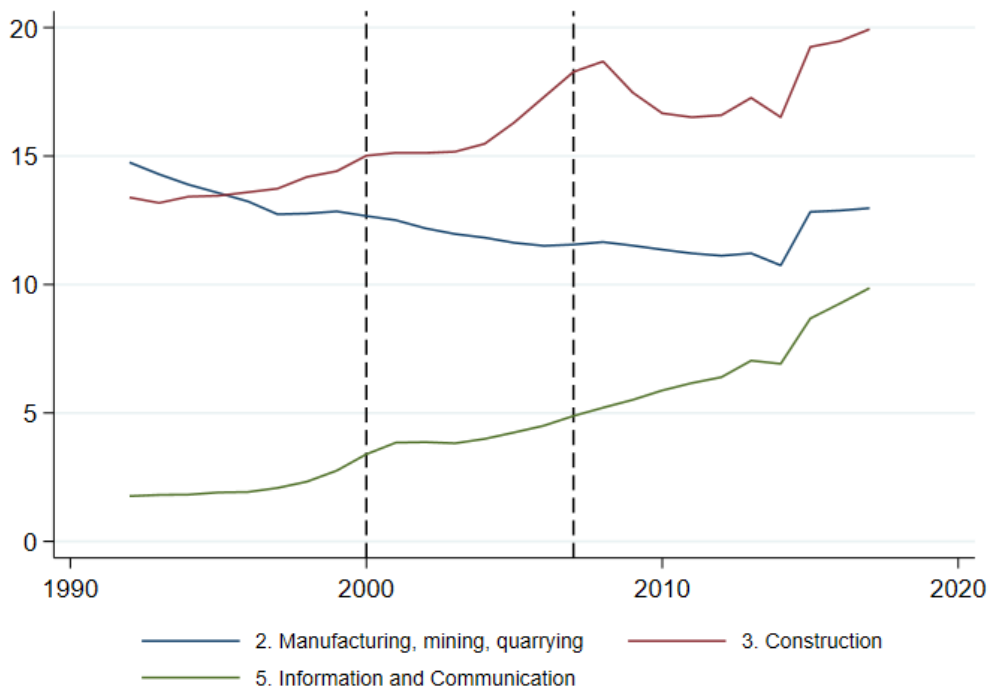


Figure 3: Industry Switching Rates (Full Data Coverage)

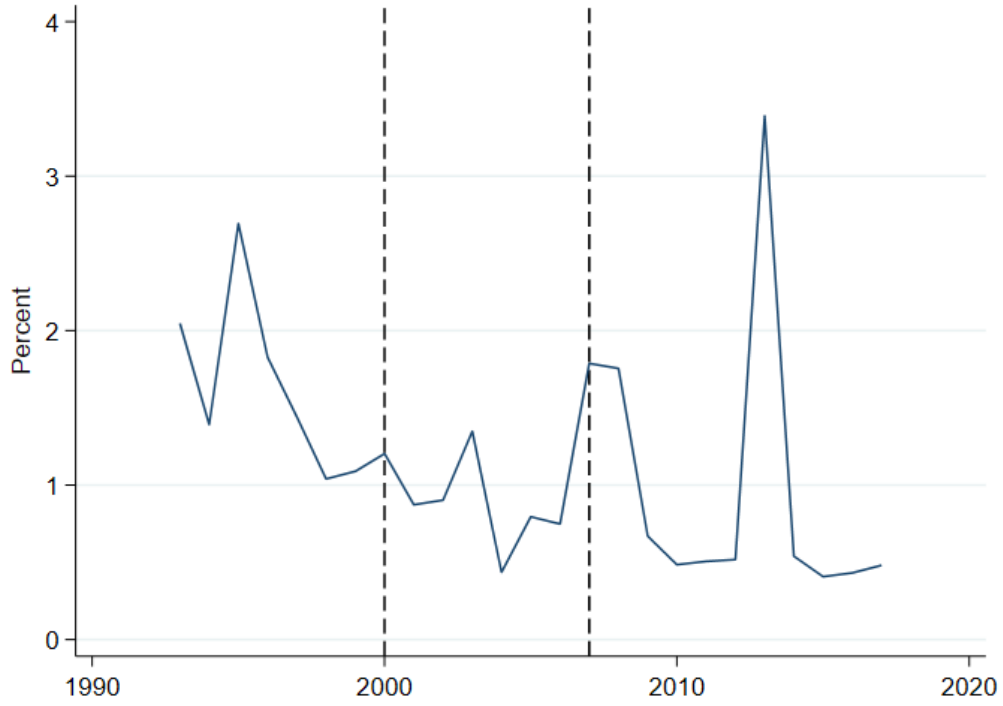


Figure 4: Industry Switching Rates by One-Digit DB07 Industries (Full Data Coverage)

