September 2012

Methodology

Green goods and services in Denmark

For information on obtaining additional copies, permission
to reprint or translate this work, and all other correspondence,
please contact:

DAMVAD A/S
Badstuestræde 20
DK-1209 Copenhagen K
Tel. +45 3315 7554
info@damvad.com
damvad.com

Contents

1 Coverage and data for the statistics

This methodology includes a description of the data for the establishment of statistics on the green economic activities in Danish enterprises. The methodology also describes the sources of data used to identify the green activities as well as their coverage rates relative to the overall data for the identification.

1.1 Definition of green products

The definition of green products is based on Eurostat's definition as described in *Handbook on data collection on Green Goods and Services (Eurostat, 2009)*.

Green products include goods and/or services that

- measure, control, restore, prevent, treat, minimise, research and retrieve knowledge about green damages to air, water and soil, as well as manage problems related to waste, noise, biodiversity and landscapes. This implies cleaner technologies, products and services that prevent or reduce pollution;

- measure, control, restore, prevent, treat, minimise, research and lead to better resource exploitation, primarily through resource optimisation in relation to technologies, products and services, which minimises the use of natural resources.

Eurostat's definition of green products was chosen because it is the standard used in statistical contexts within the EU, and because future European Union legislation will be based on Eurostat's definition of green products.

This definition does however present some challenges. Firstly, despite extensive international work, the definition of green products is not a perfectly well-defined and complete definition.

In a number of borderline cases, the decision as to whether a product is a green product or not will depend to some extent on subjective judgment. It is therefore very likely that these statistics will have to be adjusted and fine-tuned in the future.

1.2 Data for the identification of green activities

The primary data for the identification of green economic activities in Danish enterprises are the general enterprise statistics[1] which are prepared and updated by Statistics Denmark. These general enterprise statistics provide a coherent and consistent description of the Danish business community through financial, employment, and accounting information at enterprise level. The enterprise statistics include all industries and sectors in Denmark.

The general enterprise statistics cover all enterprises in Denmark which are deemed to have real commercial activity. These include enterprises that have at least 0.5 man-years and/or a turnover of between DKK 150,000 and DKK 200,000 for manufacturing, and of at least DKK 500,000 for wholesale trade.

The basis for the identification includes enterprises in the private sector except for mining and quarrying (DB07 industries A and C to N). Public institutions and authorities are not included[2]. The identification is further limited to enterprises with at least one man-year (one full-time employee),

[1] http://www.dst.dk/da/Statistik/dokumentation/Varedeklarationer/generel-firmastatistik.aspx

[2] State-owned enterprises are included in the basis for the identification, if they are registered as being a part of the private sector (e.g. DONG Energy).

which means that enterprises with no real activity are not included.

The identification of green activities is conducted on the basis of the enterprise statistics for 2009 (the most current statistics published at the time of writing). Publication of the enterprise statistics for 2010 and 2011 by Statistics Denmark has been delayed due to changes to the data collection procedures at the Danish Central Customs and Tax Administration (SKAT).

The identification includes data about a total of 114,588 enterprises. This corresponds to almost 82% of all enterprises in Denmark with at least one full-time employee, see table 1.1. The data for the identification covers almost 60% of employed persons in Denmark. This should be seen in light of the fact that a large part of those employed in Denmark are employed in the public sector. The data covers 97.8% of the turnover and 98.2% of the exports of Danish enterprises.

TABLE 1.1
Coverage of data for the identification of green enterprises in Denmark, 2009

	Denmark in total	Coverage	
	Number	Number	%
Enterprises	139,525	114,588	82.1
Full-time employees	2,175,569	1,301,201	59.8
Turnover	2,726	2,667	97.8
Exports	666	654	98.2

Source: General enterprise statistics
Note: Includes enterprises with at least one man-year. Turnover and exports are stated in DKK billion.

1.3 Sources of data to identify green enterprises

The green economic activities in Danish enterprises are identified by determining which of the 114,588 enterprises in this data have green economic activities. Enterprises with green economic activities will be referred to as green enterprises in the following, regardless of the proportionate size of their green economic activities (green share).

The identification of the green enterprises in the data is based on four independent sources of information: (i) detailed product codes; (ii) a questionnaire survey; (iii) a review of the websites of enterprises; and (iv) a statement of organic farms in Denmark. The four sources and their relation to the overall data are described in more detail below.

1.3.1 Detailed product codes

Micro-level data about the products sold by the individual enterprise is used to identify the green enterprises on the basis of detailed product codes.

Statistics Denmark's commodity statistics state production of commodities by Danish enterprises at a very high level of detail. Thus, these statistics contain information about the exact products that each enterprise manufactures in a given year.

The basis for this analysis is the following statistics, *Industriens salg af varer* (Industrial Commodity Statistics) and *Udenrigshandlen* (International/External trade), which include products defined according to the Combined Nomenclature. The Combined Nomenclature covers almost 10,000 different products. The Combined Nomenclature is a classification system of products in which the criteria for classification

are related to the characteristics of the product traded.

For each enterprise the statistical basis shows which of the 10,000 product codes the enterprise manufactures, as well as its turnover for each product. If an enterprise manufactures a green product, the enterprise will be characterised as green. For the list of green products, see the appendix.

This method based on micro-level data requires that a definition of green products has been established. *Statistisches Bundesamt Deutschland (SBD)* has been one of the driving forces behind the development work in Eurostat to define the green sector in Europe. This work has resulted in the establishment of a list of detailed green product codes. The list includes 252 green product codes stated as 9-digit GP codes, which is a German classification system.

To apply this list in a Danish context, the codes had to be converted into KN codes, which is the classification system for products used in Denmark. The KN codes contain an even higher level of detail than the German, which means that the final list ends at 403 green product codes. Due to the more detailed KN nomenclature, there may be cases in which a green GP code covers both green and non-green KN8 codes. Efforts have been made to take this into account.

In connection with the establishment of the statistics, a panel of 19 experts was set up. These experts assisted in verifying the list from SBD in relation to a Danish setting, and they assisted in ensuring that the list did not contain products that fall outside of the definition of green products. Furthermore, the panel of experts helped identify products that are green but do not appear on the SBD list. Thus, a total of 528 green product codes form the basis for the identification of green activities. The experts also identified some products that are only partially green products. These products have not been included in this statement.

It is likely the KN codes will change overtime, some will be excluded while others will be subdivided into more codes. A drop-out and intake analysis was therefore performed in order to ensure comparability of the analysis over time.

However, the statement of the sales of enterprises by product code is not available for all enterprises in Denmark. Thus, only enterprises in manufacturing and with more than ten employees are obliged to report to the Industrial Commodity Statistics. Furthermore, the International/External Trade Statistics naturally only include exporting enterprises.

Product codes can be linked to 11% of the total of 114,588 enterprises in the data. However, there is a great difference in coverage across industries. Coverage in manufacturing (DB07 industries CA to CM) is generally higher, for example covering up to 74.7% of enterprises in manufacture of chemicals, whereas the lowest coverage of 26.9% is in the manufacture of food products, beverages and tobacco, see table 1.2

The high degree of coverage in manufacturing is due to the fact that the product codes only cover physical goods and therefore do not cover services. Thus for the private service sector (DB07 industries D to S), coverage is somewhat lower. With a 22% coverage, trade however is an exception because activity in this industry is based primarily on physical products.

TABLE 1.2
Coverage of product codes

	Population in total	Coverage, %	
	Enterprise	Enterprise	Turnover
A Agriculture	12,563	1.7	10.3
CA Food products	1,345	26.9	95.2
CB Textiles	451	49.9	93.2
CC,CD Wood, paper	1,156	43.0	92.6
CE Chemicals	178	74.7	98.2
CF Pharmaceuticals	52	53.8	99.4
CG Plastic, concrete	804	62.9	96.5
CH Metals	2,230	38.3	82.9
CI Electronic components	395	70.1	98.1
CJ Electrical equipment	313	65.2	97.2
CK Machinery	1,239	65.0	97.9
CL Transport equipment	225	61.8	97.0
CM Furniture and other	1,884	32.9	87.7
DE Energy supply	926	5.0	48.3
F Construction	17,109	1.1	17.9
G Trade	28,810	22.3	67.5
H Transportation	6,810	4.6	66.2
I Hotels	7,709	0.2	2.9
JKM Knowledge-based services etc.	17,168	4.2	43.8
LN Other services	10,288	1.6	13.6
Total	114,588	11.1	62.3

Source: DAMVAD based on Statistics Denmark
Note: Turnover excluding financial and insurance.

If we measure coverage in relation to turnover, the product codes cover a considerably greater proportion of the basic data. The high coverage for turnover is important, because turnover will be used when calculating the green activities of enterprises in Denmark. Those enterprises to which product codes can be linked, account for more than 61% of the total turnover in the basic data. Within manufacturing, more or less all turnover is covered. With 82.9%, the metals industry has the lowest coverage in manufacturing measured by turnover.

1.3.2 Questionnaire survey

The purpose of the questionnaire survey is to verify and supplement the identification of green enterprises on the basis of product codes, including with an aim to improve coverage for the private service sector. Furthermore, the questionnaire survey helps provide a segmentation of the green sector.

The questionnaire survey was conducted on the basis of a representative sample of the private sector in Denmark. The respondent group was a random sample of 10% of the total enterprise population. The sample was selected so that the percentage distribution by size and industry is the same as for the total data of 114,588 enterprises forming the basis for the identification of green enterprises.

Given that the detailed product codes only cover the service sector to a lesser extent, it was decided to adjust the sample so that service enterprises were over-represented by 15%. The final sample covered 12,391 enterprises.

Almost 28% of the enterprises invited to take part in the questionnaire survey actually took part, see table 1.3. The response rates distribute relatively evenly across enterprise size and industry, so that

the questionnaire survey is representative at these two strata. For no combination of size and industry was the response rate lower than 25.4%.

TABLE 1.3|
Response rates from the questionnaire survey, %

Industry	Number of employees			
	0-19	**20-100**	**100+**	**All**
A Agriculture	26.8	44.4	100	26.9
C Manufacturing	25.9	26.6	28.6	26.1
DE Energy supply etc.	34.8	75.0	100	41.9
F Construction	25.4	30.8	40	25.8
G Trade	26.7	25.5	31	26.6
H Transportation	42.2	36.4	45.5	41.7
I Hotels	27	26.7	50	27.1
JKM Knowledge-based services etc.	29	29.9	35.3	29.1
LS Other service	26.8	43.8	35	27.9
Total	27.8	30.7	36.4	28.1

Source: DAMVAD based on Statistics Denmark

The questionnaire survey covers almost 28% of the small enterprises questioned, almost 31% of the medium-sized enterprises, and 36% of the large enterprises. Similarly, the survey covers between 25.8% and 41.9% of the enterprises questioned across industries.

From the original sample of 12,391 enterprises, 1,528 enterprises did not participate. This is due primarily to the fact that these enterprises either did not exist anymore, or that they were recorded with an invalid e-mail address. The response rates listed above were calculated on the basis of the 10,863 enterprises that received the invitations to participate in the questionnaire survey, and which therefore had the opportunity to respond.

The 1,528 enterprises that dropped out of the survey only slightly affect the representativeness of the sample. There is no systematic drop-out across industries. On the other hand, there is a small over-representation of small enterprises (0-19 employees) among the drop-outs.

The questionnaire survey was completed in the spring of 2012, whereas the basis for the identification, as mentioned above, was the enterprise statistics for 2009. The time difference between the data collection (2012) and the statistical statement (2009) means that some of the enterprises that were invited to take part in the questionnaire survey are not included in the enterprise statistics from 2009. These are mainly enterprises which, in 2009, had not yet been established or had activities below one man-year.

In *The environmental goods and services sector* (2009), Eurostat divides green products (technologies, goods and services) by function and characteristics. For the purpose of these statistics, a general breakdown based on Eurostat's classification is applied.

This breakdown means that the green products are environmental or energy-specific or adapted. The environmental or energy-specific products are characterised by having either resource management (RM) or environmental (and/or climate) protection (EP) as their primary purpose. The adapted products are characterised by resulting in less pollution or less resource consumption than similar products with the same uses (e.g. an energy-saving light bulb, the primary purpose of which is to provide lighting but which has been adapted so that it consumes less energy than the original product).

This breakdown of green products, which was made for the purpose of establishing these statistics, is however an estimate, in that it is based solely on the questionnaire survey.

1.3.3 Qualitative review of enterprises

The purpose of the qualitative review of enterprises is to serve as a supplement to the other sources in the identification of green enterprises.

The qualitative review was conducted by reviewing a selection of the websites of Danish enterprises. For each industry, the eight to twelve largest enterprises in terms of number of employees were selected. This provides for the greatest coverage for each industry.

On the basis of the information provided in their websites, it was assessed whether the enterprise in question could be characterised as green, and, if so, it was assessed how large a share of its activities could be considered to be green.

In the review of the individual enterprise, the enterprise's green share was assessed systematically on the basis of the following parameters:

- Face value - what is the general and immediate impression based on the information available?
- The enterprise's statement of business areas - how many of these are green, and how large a share is assessed to be green?
- The enterprise's use of green standards and policies.
- If possible, the enterprise's own statement of how large a share of its business is based on green products, e.g. as stated in green policies, annual reports, etc.

If identified as being green, each enterprise was given a score on the basis of the characteristics above This score shows the size of the enterprise's green share.

The review of websites covers up to 8.6% of the enterprises across industries. Manufacturing has the best coverage, while hotels and restaurants has the lowest coverage. Coverage measured on the basis of turnover goes up to 25.48%

Compared with the total enterprise population, the review of websites covers 1.4% of enterprises and around 10% of turnover.

TABLE 1.4
Coverage from websites, %

Industry	Population in total	Coverage, %	
	Enterprise	**Enterprise**	**Turnover**
A Agriculture	12,563	1.68	8.40
C Manufacturing	10,275	8.61	25.48
DE Energy supply	926	8.42	24.41
F Construction	17,109	0.09	0.53
G Trade	28,810	0.19	0.45
H Transportation	6,810	0.65	6.32
I Hotels	7,709	0.01	0.01
JKM Knowledge-based services etc.	20,098	1.37	19.04
LS Other service	10,288	0.38	7.94
Total	114,588	1.40	10.21

Source: DAMVAD based on Statistics Denmark

Like the questionnaire survey, the review of websites was conducted in the spring of 2012. This means that some of the enterprises examined in the questionnaire survey do not appear in the 2009 enterprise statistics.

1.3.4 Organic farms

Organic products are characterised by being produced with consideration for the environment and without the use of pesticides or artificial fertilisers in production. Furthermore, in organic livestock farming there are special requirements for the design of livestock sheds and stables, access to outdoor areas etc. For this reason organic farms are included as green enterprises in the statement of green activities in Denmark.

The Danish AgriFish Agency (formerly the Danish Plant Directorate) is the agency responsible for supervision of organic farms in Denmark. The Danish AgriFish Agency authorises farms to carry out organic agricultural and livestock production. The farm's products are only organic if the farm has obtained authorisation. If an enterprise has obtained authorisation, its entire turnover will be included as green activities in the statement. This is considered a reasonable assumption because 93.3% of the land area of authorised organic farms is organic, see the Danish Plant Directorate: *Statistics on organic farms 2009 - Authorizations And Production.*

In 2009 there were 2,689 authorised organic farms in Denmark. Of these, 744 are included in the data forming the basis for the identification of green enterprises in Denmark. The drop-out is due to the fact that many of the organic farms are small and therefore do not meet the activity requirement for inclusion in the enterprise population. On the other hand, since the Danish AgriFish Agency has to authorise all farms, we know that the population is complete in numbers.

1.4 Total coverage of sources used to identify green enterprises

As mentioned above, the information about the green economic activities of enterprises was collected on the basis of four sources:

- Product codes
- Questionnaires
- Websites
- Organic farms

All in all the four sources contain information about a total of 15,617 unique enterprises. In other words, for each of the 15,617 enterprises observed, data exists as to whether the enterprise is green, and, if so, how large a share of its turnover can be attributed to green activities.

The 15,617 unique enterprises that are included in the four sources cover 13.6% of the population analysed, see table 1.5. Measured in terms of turnover, the observed enterprises cover 65.7% of the total turnover of the 114,588 enterprises in the population.

At the same time, the obligation to report to Statistics Denmark is relatively limited for small enterprises. In other words, Statistics Denmark's records can only contribute some data for the identification of green enterprises, e.g. through product codes.

For all of the industries under manufacturing, the coverage of the four sources is high. Measured in terms of number of enterprises, a minimum of 27.5% of the enterprises in the food products industry are covered, and a maximum 80.9% of the enterprises in the pharmaceuticals industry. Measured in terms of turnover, coverage is especially high for manufacturing. Thus, a

minimum of 83.9% of enterprises are covered in the metals industry and as many as 99.5% of enterprises are covered in the pharmaceuticals industry.

TABLE 1.5
Total coverage, %

	Population in total	Coverage, %	
	Enterprise	**Enterprise**	**Turnover**
A Agriculture	12,563	9.8	20.3
CA Food products	1,345	27.5	95.3
CB Textiles	451	52.8	93.8
CC,CD Wood, paper	1,159	44.2	92.8
CE Chemicals	178	80.9	99.3
CF Pharmaceuticals	52	55.8	99.5
CG Plastic, concrete	804	65.2	96.7
CH Metals	2,230	40.4	83.9
CI Electronic components	395	71.4	98.3
CJ Electrical equipment	313	66.1	97.3
CK Machinery	1,239	67.4	98.6
CL Transport equipment	225	65.3	97.2
CM Furniture and other	1,884	35.1	88.4
DE Energy supply	926	13.4	62.5
F Construction	17,109	2.9	20.7
G Trade	28,810	23.2	68.1
H Transportation	6,810	7.7	70.9
I Hotels	7,709	1.6	4.7
JKM Knowledge-based services etc.	20,098	6.0	54.1
LN Other services	10,288	3.5	23.1
Total	114,588	13.6	65.7

Source: DAMVAD based on Statistics Denmark
Note: By far the majority of enterprises in Denmark are small or medium-size enterprises. As many as 95,612 out of the total of 114,588 enterprises in the population have less than ten employees.

Within service, the coverage is lower but still relatively high for several industries. For trade (wholesale and retail trade), the four sources cover 23.2% of enterprises and just below 70% of turnover. However, there are also service industries for which coverage is low. These include hotels, construction and knowledge-based services etc.

All in all, the data coverage is very high compared with other work to identify green activities in Denmark. Where coverage is low, the statistics will have greater uncertainty. However, the statistical method used to impute the missing data was selected so that it exploits the information available in the extensive data across industries. This statistical method will therefore contribute to reducing the uncertainty for industries with low coverage, see chapter 2.

Overall, this means that coverage by the four sources of small and medium-sized enterprises is far smaller than coverage of large enterprises, see figure 1.1. On the other hand, the observed enterprises cover a large proportion of the turnover, which minimises uncertainty in the final data used for the statistics.

FIGURE 1.1
Coverage in terms of number of employees

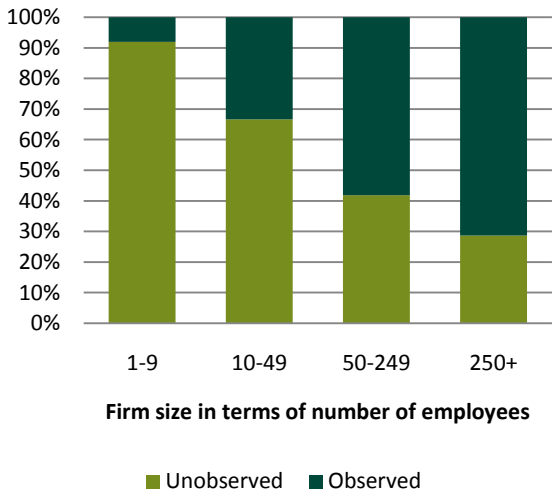


TABLE 1.6
Overlap across sources

PC	Q	WS	OF	Enterprises	Turnover
				Number	**DKK bn.**
0	0	0	0	98,971	915
0	0	1	0	665	62
0	1	0	0	1,459	20
0	1	1	0	29	5
1	0	0	0	11,523	1,428
1	0	1	0	868	195
1	1	0	0	292	28
1	1	1	0	37	11
0	0	0	1	744	3
All				114,588	2,666

Source: DAMVAD based on Statistics Denmark
Note: PC = product codes; Q = questionnaires; WS = websites; OF = organic farms; 0 = no observed data; 1 = observed data.

1.4.1 Overlap across sources

Collection of information from the four sources was performed separately. As a consequence, there will be some degree of overlap across sources.

The product codes make up the most extensive source of information, contributing information about 11,523 enterprises. Secondly, the questionnaire survey contributes information about 1,459 enterprises. The review of websites and the organic farms contribute information about 665 and 774 enterprises, respectively, see table 1.6.

For 1,354 enterprises, information about whether they are green or not is available from more than one source, and for 98,971 enterprises no source is available.

2 Identification of green activities

This chapter describes how the green enterprises were identified, along with the method for calculating the green economic activities of enterprises where no observed data was available.

2.1 Observed green enterprises

A total of 15,617 unique enterprises are included in the four sources of data used to identify the green enterprises. The sources of data contain information for each enterprise as to whether the enterprise has green economic activities or not. An enterprise is defined as "green" if at least one of the four sources identifies it as such, i.e.:

- An enterprise included in the data source with the detailed product codes is defined as being green if it manufactures a green product.
- An enterprise included in the questionnaire survey is defined as being green if it has stated so in the questionnaire form.
- An enterprise included in the review of websites is defined as being green, if the review of information about the enterprise identified it as being green.
- Finally, an enterprise which is an authorised organic farm is defined as being green.

Of the 15,617 enterprises observed in the four data sources, 35.7% were identified as green, corresponding to 5,580 enterprises, see table 2.1.

TABLE 2.1
Observed green enterprises

	Observed in total	Green enterprises	
	Number	Number	%
Agriculture	1,226	823	67.1
Manufacturing	4,850	2,038	42.0
Service	9,541	2,719	28.5
Total	15,617	5,580	35.7

Source: DAMVAD based on Statistics Denmark

2.1.1 Share of green activities at enterprise level

The size of the share of the enterprise's green activities is stated for each enterprise. This is calculated as the share of the enterprise's turnover that can be attributed to green activities.

The calculation of this green share at enterprise level is especially important in terms of large enterprises, because these often have activities within several business areas, not all them green. Large enterprises account for a significant share of the Danish economy and the "non-green" activities therefore have to be deducted in order to arrive at a true picture of green economic activities in Denmark.

All four sources contain information about the share of green activities at enterprise level. The green share was calculated as follows:

- For the enterprises in the detailed-product-codes data source, the green share was calculated as sales of green product codes divided by sales of all product codes at enterprise level.
- In the questionnaire survey, the enterprises themselves provided information about the ratio of green activities to total turnover.

- For enterprises in the review of websites, the green share was assessed on the basis of information about business areas and financial statements.

- For authorised organic farms, the entire turnover was defined as green.

If an enterprise was observed to have more than one green weight, the following hierarchy was used to determine the source of weight: The statement of organic farms was applied first, then the questionnaire survey, then the product codes, and, finally, the review of websites. Thus the hierarchy reflects which source is assumed to provide the truest picture.

2.2 Computation of green activities

The four sources of information cover 15,617 enterprises, corresponding to 13.6% of the total enterprise population. We therefore had to estimate the green enterprises among the remaining 98,971 enterprises, including their individual green weights. This is primarily enterprises in agriculture and in the private service sector, see table 2.2.

TABLE 2.2
Breakdown of observed and unobserved enterprises by sector, 2009

	Observed	Unobserved	Total
		Number	
Agriculture	1,226	11,337	12,563
Manufacturing	4,850	5,425	10,275
Service	9,541	82,209	91,750
Total	15,617	98,971	114,588

Source: DAMVAD based on Statistics Denmark

2.2.1 Imputation strategy and choice of models

We used a two-step imputation strategy, partly to identify the unobserved enterprises that are green, and partly to determine the green share of activities for each green enterprise.

- Step 1: The green status (green or not) was determined for the 98,971 enterprises that were not observed in the four information sources.

- Step 2: The green weight was determined for the unobserved enterprises that were estimated to be green under step 1.

Because the number of green enterprises is an important result in itself, and because the green weights should be estimated as accurately as possible, a two-step approach was utilised. Then, the two most suitable econometric models were selected for steps 1 and 2, respectively.

The green status and the green weight are explained on the basis of a number of explanatory variables. These include:

- enterprise size (measured as number of man-years)
- turnover per man-year
- imports and exports

- educational composition
- industry grouping (the standard 127-grouping in the DB07 classification)
- geographical region

The correlation between green activity and the explanatory variables was determined on the basis of the enterprises included in the green information sources. When this correlation was determined, the results were transferred to the observed enterprises for which there is no information on the green weight.

An alternative approach would have been to determine the green status and weight in one single step. Such an approach, however, is likely to result in an overestimation of the number of green enterprises and an underestimation of green weights among the actual green enterprises.

The reason for this is that econometric models (which seek to describe average trends, in our case green weights) have difficulties coping with several observations concentrated in a single point, in our case the value zero for "no green activity". Econometric models will therefore seek to even out the concentration by distributing these enterprises to positive green weights. In this way the number of green enterprises will be overestimated. To maintain the average trend, the model will then underestimate the green weights for the actual green enterprises.

Note that organic farms were *not* included in the econometric models, because the Danish AgriFish Agency's statement of organic farms has complete coverage. This leaves a total of 14,873 (15,617 observed enterprises less 744 organic farms) unique non-organic enterprises which have been included in the econometric models.

The population of enterprise in these statistics is based on the year 2009. However, there are enterprises in the period 2005 to 2008, and in 2010, which did not exist in 2009. For example, this could be enterprises that went bankrupt before 2009 or enterprises that were not established until in 2010. Information on these enterprises was imputed from the four data sources and, otherwise, according to the same principle as described in section 2.2. Thus, a new population was not estimated for each year.

The green share of enterprises in the population was therefore calculated for 2009. The same share is used for all years. For enterprises that did not exist in 2009, but existed in the period 2005 to 2008 and in 2010, the green share was imputed.

Step 1: Imputation of green status

The purpose of this step is to identify the green status of the 98,971 enterprises for which there is no observed data. The approach can be split into two phases.

In the first phase, the characteristics of significance for whether an enterprise is green are identified on the basis of knowledge about the 14,873 observed enterprises.

In the second phase, this information is used to determine which of the unobserved enterprises are green. The second phase is further split into two distinct operations.

- Phase 2A: Computation of the probability of being green for unobserved enterprises. The probability depends on the parameters that were estimated in the first phase.

- Phase 2B: Assessment of whether an enterprise is green or not. This assessment

was conducted on the basis of the probability computed in phase 2A.

Phase 1: Firstly, information about the characteristics linked to green and non-green enterprises, respectively, has to be obtained. This can only be done on the basis of the 14,873 enterprises for which information about green status has been provided, of which 4,836 are green[3].

Technically, a binomial logistic model[4] was used to estimate parameters for the characteristics of being green. After the model's parameters had been estimated, it was possible to compute the probability of being green (based on enterprise characteristics) for those enterprises for which information about green status has not been provided. This is done in phase 2. Please note that the model's parameters are not *identified;* they are *estimated.* This is expressed by the model's parameters being stated within a confidence interval.

Phase 2: The next step was to determine which of the 98,971 enterprises, for which there is no observed data, are green. This was done using multiple imputation[5]. The multiple imputation involved a number of repeated assessments of the green status of each enterprise. An assessment is denoted as a single simulation run. The outcome of the individual enterprise assessments may vary from simulation run to simulation run. This analysis used 100 simulation runs. In other words, there were 100 separate assessments of the green status of each enterprise. In each simulation run, during multiple imputation, two separate operations were performed. First, the enterprise's probability of being green was estimated (phase 2A). Then an assessment (green or not green) was made on the basis of the estimated probability (phase 2B).

Phase 2A: The operation in this step used the parameter estimations from the binomial logistics model (estimated in phase 1) to compute the probability of being green for each of the 98,971 enterprises for which there are no observations about green status. During each simulation run, new parameters were used which were drawn randomly within the confidence interval that the logistics model (during phase 1) estimated for a given parameter. This was done in order to take account of the precision with which the binomial model's parameters were initially determined (during phase 1). New parameters were drawn in each simulation run in the realisation that the model's parameters are not *identified* but *estimated* with a given element of uncertainty.

As a result of phase 2A, for each simulation run, the unobserved enterprises were allocated a probability of being green. The same enterprise will have different probabilities of being green from simulation to simulation. This is because new parameters were drawn for each simulation in order take account of the uncertainty with which the logistics model in phase 1 was estimated.

Phase 2B: For each simulation run it was determined which of the 98,971 unobserved enterprises are assessed to be green and which are non-green. For all enterprises in each simulation run, the following took place: Assume that an enterprise has a 75% probability of being

3 5,580 observed green enterprises less 744 organic farms

4 The binomial logit model was selected because it describes an event (in our case green activity) with two possible outcomes (green activity or no green activity). Alternatively a binomial probit model could have been used. The two models are based on different underlying distribution assumptions, but otherwise they have the same properties. An important property of the two models is that the estimated probabilities are restricted to the logical interval between 0 and 1. For example, this is not the case in the linear probability model.

5 The MI module under STATA 12 was used

green. The enterprise is allocated its green status at random. There is a 75% probability the enterprise will be characterised as green and a 25% probability it will be characterised as non-green. After all of the 100 simulation runs, the enterprise will have been assessed to be green approx. 75 times, and to be non-green approx. 25 times.

In any given simulation run, an enterprise with a high estimated probability can therefore be assessed to be non-green, whereas an enterprise with lower estimated probability can be assessed to be green. However, seen across a large number of simulation runs, there will be a positive correlation between the enterprise's estimated probabilities and the number of times it is assessed to be green. This approach was chosen because it addresses the fact that some of the observed enterprises that have a low estimated probability of being green, are actually green (and vice versa), c.f. the information sources. Since this applies for the observed enterprises, it should also apply for the unobserved enterprises. Therefore, the imputation of green status was performed through multiple imputation rather than merely basing the assessment on a simple threshold value (e.g. by consistently choosing those enterprises that have an estimated probability above 50%)[6].

If the logistic model's parameters are determined at high accuracy, largely the same enterprises will be assessed to be green across the 100 simulation runs. This will further mean that the imputed green enterprises can largely be assumed to have been chosen correctly.

In the final report on the green sector, only the results from a single simulation run are used (the median run), see next section.

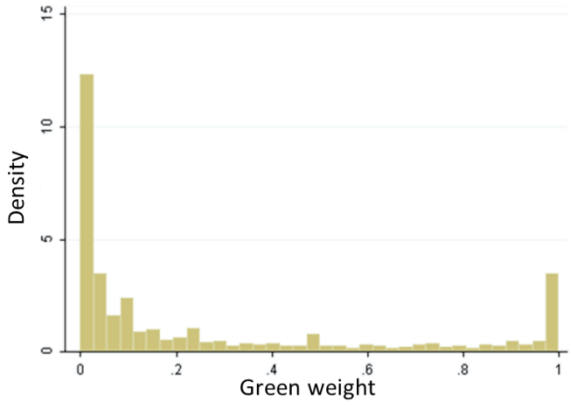
Step 2: Imputation of green weights

When the green status of the enterprises had been assessed for each of the 100 simulations, the green weight had to be calculated for those enterprises that were expected to be green.
The choice of specific econometric model was made on the basis of the distribution of green weights among the observed green enterprises. This distribution is illustrated in figure 2.1.

As the figure clearly shows, the distribution of green weights is far from a normal distribution. This initially excludes using an OLS model to describe the green weight. Instead a Poisson model was chosen as the econometric model. Using a Poisson model is advantageous when the empirical distribution "peaks" early on the left side of the distribution and then declines towards the right.

FIGURE 2.1

The distribution of the green weight of observed enterprises



Source: DAMVAD based on Statistics Denmark

In particular, the empirical distribution concentration early on the left side, combined with

6 The advantages of multiple imputation have been described in Rubin, D. (1987) *"Multiple Imputation in Nonresponse Surveys"*, J. Wiley & Sons, New York.

an approximately declining trend towards the right side, motivated the choice of a Poisson model.

This choice of a Poisson model is however not without challenges. Firstly, the empirical distribution is not asymptotically declining but has a natural upper limit at 100% green weight. This means that, for certain enterprises, a Poisson model will compute a green weight greater than 100%.

Secondly, the empirical distribution does not decline steadily, which is seen in particular in the pronounced concentration on the right close to 100%. This concentration on the right side is especially challenging because it will amplify the probability that the computed green weight for some enterprises will be greater than 100%.

The deviations between the empirical distribution and the Poisson model's theoretical distribution assumption will result in a model with a relatively lower explanatory power. Despite the mentioned weaknesses of the Poisson model, this model was chosen because it is the best candidate among obvious alternatives. For the enterprises for which the model computes a green weight greater than 100%, the green weight was set manually to 100%.

In contrast to step 1, the econometric model under step 2 was not performed during *multiple imputation*. The Poisson model was therefore performed without the addition of a stochastic component[7]. This means that the green weight will be identical in each of the 100 simulation runs in which the enterprise is assessed to be green. For simulation runs in which the enterprise is assessed to be non-green, the green weight was set to zero.

[7] During step 1, the random draw made up the stochastic component.

Note that although the weights of the enterprises (conditioned by their green status) do not vary across the 100 simulation runs, the total green turnover will vary, because the composition of expected green enterprises may vary from simulation to simulation.

On the basis of the 100 computations, a single simulation run was selected for use in the final identification of the green sector. The total green turnover was calculated for each of the 100 simulations. The 100 simulations were then sorted from the lowest to the highest total turnover. The simulation run placed in the middle (the median) was selected for the final identification of the green sector.

Alternative model specification

An alternative way to model the green profile of the enterprises, is by using an alternative measure for their green activity.

If the green weight of the enterprises is converted into the natural logarithm value of the enterprise's green turnover per full-time employee, an approximate normal distribution is achieved. In this way the green activity of the enterprises can be described using an OLS model. Analyses have suggested that this approach can produce a model which on the face of it has greater precision than the Poisson model.

However, this alternative approach has the same weakness as the Poisson model: the green turnover may be computed as greater than the enterprise's total turnover. Measured in money terms, this problem initially appears to be considerably smaller for the alternative OLS approach than for the Poisson model. However, as is evident in the following, monetary calculations of the two models cannot readily be compared.

What decisively led to rejection of the OLS approach was the logarithm transformation required to achieve the desired normal distribution. This approach calculates the expected logarithm value of the enterprises' green turnover per full-time employee. Then the expected logarithm value is converted into a monetary amount. However, this monetary amount will (on average) be underestimated, since the logarithmic transformation is not a linear transformation (a transformation with a constant scale factor). In the logarithm transformation the large values are scaled down by an exponentially larger factor than smaller values. Enterprises with a large green turnover will therefore be weighted relatively less, and the total green turnover will be underestimated.

An example is given below to illustrate this. The average is calculated for three different distributions (symmetrical, left-skewed and right-skewed). The average is to symbolise the OLS estimator which models the conditioned average. The three distributions are given as:

- symmetrical distribution: (1;13;19;21;22;22;23;23;23;24;24;25;27;33;45)
- left-skewed distribution: (1;33;39;40;42;42;43;43;44;44;44;44;45;45;45)
- right-skewed distribution: (1;1;1;2;2;2;2;3;3;4;4;6;9;15;45)

The table below shows the average for the three distributions in column 2. In column 3, the average has been computed for the logarithm-transformed distributions. This is to symbolise the value of the green turnover from the logarithm-transformed OLS model. These values are much lower than the average in column 2, because they have not yet been transformed back to the original scale. This re-transformation is performed by an exponential transformation of the values from column 3. These are provided in column 4. As can be seen (in column 5) the re-transformed averages are systematically lower than the average of the original values. This suggests that the average is underestimated in the re-transformation. This problem is especially apparent for the right-skewed distribution, in which the average is underestimated by 49%.

Distribution	Average of original values	Average of logarithm-transformed values	Re-transformed logarithm average	Deviation from the "pure" average:
Symmetrical	23	2,954	19.2	-16.6%
Left-skewed	39.6	3,493	32.90	-16.9%
Right-skewed	6.6	1,216	3.4	-49%

The distribution for green turnover per employee among green enterprises has shown to be strongly right-skewed (the average is 4.9 times greater than the median). For this reason, it is expected the alternative OLS approach will calculate a significantly underestimated total imputed green turnover. Computations have shown that the total imputed green turnover from the alternative OLS approach is almost 40% less than the green turnover imputed using the Poisson model (DKK 50 and 82 billion, respectively).

2.3 Specification of econometric models

Having decided on the choice of econometric models, the next task was to determine the data to be included for estimation in the model as well as the explanatory variables to be used.

With regard to the choice of data to be included, there is the possibility of estimating the models for all industries in a single model. Naturally, there will be differences in enterprises' probabilities of being green, depending on the industry they belong to. There will also be differences in green weights across industries. These differences can be addressed partially by introducing dummy variables for the relevant industries. In this context, it is important to note that this approach only takes account of differences in level across industries.

However, it is very likely that the importance of enterprise size, internationalisation and educational composition for the green profile of enterprises varies depending on which industry the enterprise belongs to. The above approach does not take account of this. In order to check for these industry differences, it was therefore necessary to estimate the models separately for relevant groups of industries.

Of course, the potential number of industry-specific models that can be set up is limited by the data available. On the one hand, it would be advantageous to divide the models by industry at the most detailed level of industry and thus to be able to estimate the model and its parameters with the greatest possible accuracy. On the other hand, however, each time the models are split up, the number of observations in the individual model will fall, and this in itself will mean less statistical certainty. It is therefore important to find an appropriate level at which to estimate the models.

For both the binomial logit model in step 1 (green or not) and the Poisson model in step 2 (green weight) the models were estimated divided into nine specific and separate industrial groupings. The models were broken down by the standard 19-groupings in the Danish adoption of the NACE Rev. 2 classification. However, due to the nature of the data material, some industries had to be combined.

Energy supply (electricity, gas, steam and air conditioning supply) and water supply (water supply, sewerage and waste management) were combined in a single model. So were the three industrial groupings hotels and restaurants (accommodation and food service activities), transportation, and information and communication. Finally, the financial and insurance industry was combined with the grouping real estate activities. These groupings were made on the basis of combining similar industries where necessary.

Another relevant approach would be a breakdown by the three sources in which the green activities were observed (product codes, questionnaire, websites). This may be relevant because there are systematic differences among the enterprises included in the individual sources. However, this approach was not chosen in our case, because if the breakdown by source is combined with the breakdown by industrial groupings, the data becomes too thin. The breakdown by industry is preferred, because it also partially catches differences in source representation, see table 3.1.
The econometric models were specified with regard to a great number of explanatory variables. These include:

- Enterprise size (measured as number of full-time employees)
- Turnover per full-time employees
- Import and export intensities
- Percentage of share, skilled employees and employees with short-, medium- and long-cycle higher educations.

These variables are entered in their level and squared forms[8]. Furthermore, the models were also specified on the basis of a number of dummy variables which indicate whether the enterprise imports, exports, has at least one employee that is skilled, has a short-, medium- or long-cycle higher education or a PhD qualification. Dummy variables were also used for the different fields of education, as well as for industry and geographical region.

[8] The squared variables were entered in order to check for marginally increasing/decreasing or U-shaped trends in the explanatory variables' depiction of green activities.

Each of the nine industry-specific models was specified separately and includes only the explanatory variables that are of significance for the individual industry. Initially, all variables which are insignificant at a 10 per cent level were removed. However, with some exceptions. For example, if the level variable or the squared variable was only significant if the other variable (level or squared, respectively) was also included. The explanatory variable which can be assumed to be different from zero at a 20 per cent level was also accepted under these conditions.

The specification and the parameter estimates for the nine binomial models in step 1 (green or not) are given in appendix 2, while the specification and parameter estimates for the nine Poisson models in step 2 (green weight) are presented in appendix 3.

Overall, the binomial logit models have an explanatory power at around 20%. The explanatory power is strongest for the model for energy supply and water supply (42%) and weakest for construction (10%).

The explanatory power of the Poisson models is in general stronger than the explanatory power of the logistic models. The explanatory power is as strong as 99% for the model for financial and insurance and real estate activities.

This should however be seen as an indication that the model for these industries is based on very limited data. This model also differs from the other models in that turnover, imports and exports are not included as explanatory variables. This is due to the fact that this information is not collected for enterprises in the financial and insurance industry and therefore does not appear in the general enterprise statistics.

As described above, the computed green weights used in the identification of the green sector were selected from the simulation run that constitutes the median with regard to the total green turnover. This assessment was performed for each of the nine industrial groupings. The final result is a total imputed green turnover of DKK 82.4 billion. This turnover was generated by a total of 17,187 enterprises. In combination with the 5,580 green enterprises that were observed directly in the data, this provides a total turnover of DKK 255.8 billion for the year 2009.

A statement of the green turnover broken down by the individual source shows that most of the turnover is identified through product codes (77%). The review of websites is the second-most important source (15%), followed by the questionnaire survey (7%) and the list of organic farms (1%). This distribution is shown in table 2.3.

The table shows that product codes account for a relatively higher share of the turnover observed directly in the data. On the other hand, the review of websites and the questionnaire survey account for a share which is approx. three-times greater in the imputed turnover column than in the observed turnover column. It must be emphasised that, for the imputed turnover, the actual distribution cannot be inferred accurately because the three data sources were combined into one in the model computations. For the imputed turnover, the distribution was estimated on the basis of how the *observed* turnover is spread over each of the nine industrial groupings.

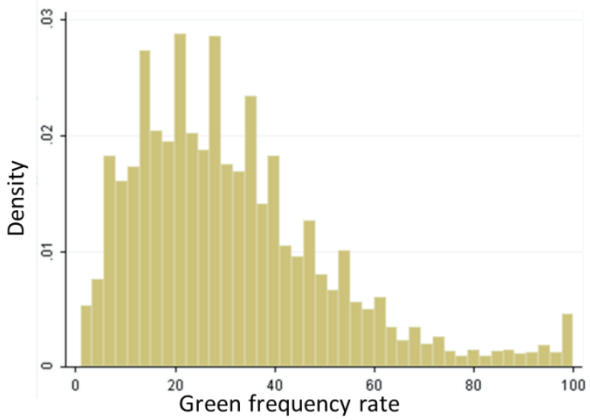
2.4 Assessment of the imputation strategy

The ideal imputation model will have a high consistency in the enterprises the model assesses to be green across the simulation runs. This is achieved in a model with a strong explanatory power. If the explanatory power of the model were low, there would be very little consistency across simulations.

A majority of the 17,187 (imputed) green enterprises included in the final statistics are assessed to be green in 5-40% of the 100 simulations. Enterprises which are assessed to be green in 65-95% of the 100 simulations, on the other hand, are relatively under-represented. This is illustrated in figure 2.2.

FIGURE 2.2

Imputed green enterprises, frequency of being green



Source: DAMVAD based on Statistics Denmark

Enterprises with green frequency rates below 50% contribute to the total imputed green turnover, see figure 2.2. This figure illustrates the accumulated green turnover across the green frequency rates of enterprises. As it appears from the figure, the curve for the accumulated green turnover rises quite considerably in the frequency interval between 10 and 50. This means that enterprises that are assessed to be green between 10 and 50 times out of the total of 100 simulation runs

TABLE 2.3

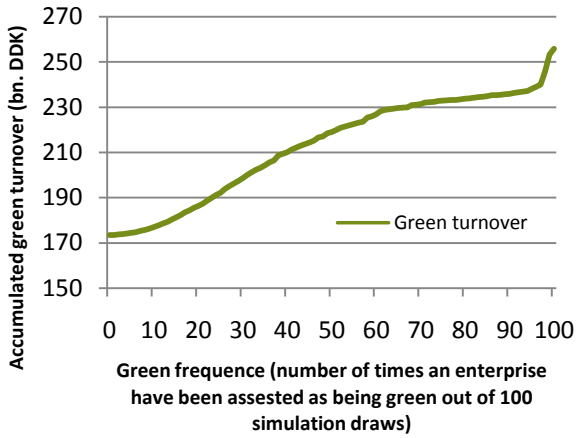
Distribution of green turnover by source of information

Source	Observed		Imputed		Total	
	Green turnover (DKK bn.)	**Distribution**	**Green turnover (DKK bn.)**	**Distribution**	**Green turnover (DKK bn.)**	**Distribution**
Product codes	148.1	85%	48.1	58%	196.2	77%
Questionnaires	6.8	4%	11.7	14%	18.5	7%
Websites	15.4	9%	22.6	27%	38	15%
Organic farms	3.1	2%	---	---	3.1	1%
Total	173.4	100%	82.4	100%	255.8	100%

Source: DAMVAD based on Statistics Denmark

contribute considerably to the total imputed green turnover. The curve starts at an accumulated green turnover of DKK 173.4 billion, which is the amount contributed by the 5,580 green enterprises that are observed directly in the data.

FIGURE 2.3
Accumulated green turnover by green frequency rate



Source: DAMVAD based on Statistics Denmark

The above reveals some degree of inconsistency in the enterprises that are assessed to be green across the individual simulations. It is therefore relevant to examine the extent to which the characterisation of the green enterprises changes across simulations. This is illustrated in table 2.4.

In general, the table shows that the characterisation of green enterprises does not change significantly across the individual simulations. This is best illustrated by the narrow confidence interval for the individual indicators.
With a 95% certainty, the number of green enterprises is expected to lie within the interval 17,000 to 17,400. Similarly, the green turnover can be narrowed down to between DKK 80.8 to 84.1 billion, which corresponds to 4% of the average level across simulations.

The variation across simulations is also very low for enterprise size, turnover per full-time employee, export intensity and educational composition. The width of the confidence interval constitutes between 1-3% of the average level.

Although the enterprises assessed to be green are not the same from simulation to simulation, they are still comparable. This is an indication that the econometric model checks for this as required.

However, it has not been possible to include productivity (added value per full-time employee) and capital intensity (capital per full-time employee) in the econometric models. This is because these indicators have not been stated for a number of enterprises. The table also shows that variation in the productivity measure is relatively low. Presumably, this is because turnover per full-time employee serves as a proxy for the productivity measure. The variation in capital intensity, on the other hand, is relatively greater (12% relative to the average).

This suggests that detailed industrial groupings are insufficient as approximations of capital input. An analysis of the capital holding of imputed green enterprises should therefore be performed with greater caution.

Research development and innovation are not included in the econometric models. The statistics for research, development and innovation from 2009 covered only 4,500 enterprises. This corresponds to just below 4% of the enterprises included in the data used to map the green sector. It was therefore not possible to include research, development and innovation in the econometric models.

TABLE 2.4

Characteristics of imputed green enterprises across 100 simulations

	Min.	Max.	Average	Median	95% confidence interval	
					Lower limit	Upper limit
Number of green enterprises	15,143	20,267	17,191	17,164	16,983	17,399
Green turnover (DKK bn.)	54.9	103.5	82.4	82.5	80.8	84.1
Turnover (DKK bn.)	196.5	308.6	258.3	258.7	254.2	262.4
Full-time employees	7.8	10.3	9.2	9.1	9.1	9.3
Turnover per full-time employees (DKK mill.)	1.95	2.24	2.09	2.09	2.08	2.11
Export intensity	2.10%	3.27%	2.73%	2.72%	2.69%	2.76%
Share skilled employees	40.38%	45.17%	42.57%	42.58%	42.36%	42.79%
Share of employees with further education	22.60%	27.44%	24.67%	24.58%	24.48%	24.85%
Capital input per full-time employees (DKK mill.)	0.825	3.387	1.371	1.245	1.286	1.455
Value added per man-year (DKK mill.)	576,068	649,461	599,022	595,421	596.303	601,742

Source: DAMVAD based on Statistics Denmark

3 Appendix: Product codes

Overview of KN8 product codes (2009 nomenclature)										
25041000	39172210	40021930	46019905	59080000	72042190	73090059	84129080	84742010	85061015	87089291
25173000	39172290	40021990	46019910	59100000	72042900	73090090	84135040	84742090	85061019	87089299
25210000	39172310	40022000	46019990	59111000	72043000	73101000	84135080	84743910	85061091	90251920
25222000	39172390	40023100	46021100	59112000	72044110	73218900	84136039	84743990	85061095	90251980
26201100	39172912	40023900	46021200	59114000	72044191	73219000	84136069	84778091	85061099	90258040
26203000	39172915	40024900	46021910	59119010	72044199	73229000	84136070	84778099	85063030	90258080
26211000	39172919	40025900	46021991	59119090	72044910	73251050	84136080	84798200	85063090	90261021
27101999	39172990	40026000	46021999	63090000	72044990	73251092	84137021	84818099	85064010	90261029
27109900	39173210	40027000	46029000	68061000	73030010	73251099	84137029	84821010	85064030	90261081
28011000	39173231	40028000	48193000	68062010	73030090	73269070	84137030	84821090	85065010	90261089
28151100	39173235	40029910	54050000	68062090	73043180	74040010	84137035	84822000	85065030	90262020
28151200	39173239	40029990	54071000	68069000	73043910	74040091	84137045	84823000	85065090	90262040
28161000	39173251	40030000	54072011	68071010	73043952	74040099	84137051	84824000	85066030	90262080
28164000	39173291	40040000	54072019	68080000	73043958	75030010	84137059	84825000	85066090	90268020
28182000	39173299	40052000	54072090	68101110	73043992	75030090	84137065	84828000	85068005	90268080
28183000	39173912	40059100	54073000	68101190	73043993	76012099	84137075	84829110	85068011	90271010
28230000	39173915	40059900	54081000	68101910	73043998	76020011	84137081	84829190	85068015	90271090
28272000	39173919	40081100	54082100	68101931	73044100	76020019	84137089	84829900	85068090	90272000
28273100	39173990	40082190	54082210	68101939	73044910	76020090	84138100	84833032	85069000	90273000
28273200	39174000	40091100	54082290	68101990	73044993	76110000	84138200	84833038	85073020	90278011
28273500	39199090	40092100	54082300	68109110	73044995	76129091	84162020	84833080	85073089	90279080
28273910	39201023	40101100	54082400	68109190	73044999	78020000	84178030	84834030	85074000	90321020
28273920	39201024	40101200	54083100	68151010	73045112	79020000	84178050	84834051	85078020	90321089
28273930	39201026	40101900	54083200	68151090	73045118	80020000	84178070	84834059	85078030	90322000
28273985	39201027	40112090	54083300	69029000	73045189	81083000	84186100	84834090	85078080	90328100
28363000	39201028	40122000	54083400	69041000	73045910	81122200	84191100	84836020	85079020	90328900
29051100	39201040	40129020	56021011	69091200	73045932	84021910	84191900	84836080	85079090	94060011
29171200	39201081	40161000	56021019	69091900	73045938	84021990	84193910	84839020	85142080	94060020
38012090	39201089	44119410	56021031	70010010	73045992	84022000	84193990	84839081	85143099	
38021000	39202090	44119490	56021038	70080020	73045993	84031010	84194000	84839089	85144000	
38089410	39203000	44181010	56021090	70080081	73045999	84031090	84195000	85013100	85369020	
38089420	39211100	44181050	56022100	70080089	73049000	84051000	84196000	85016120	85389011	
38089490	39211310	44181090	56022900	70099100	73063041	84061000	84198910	85016180	85389019	
38151100	39211390	45041091	56029000	70099200	73063049	84068110	84211970	85016300	85389091	
38151200	39211900	45041099	56031110	70191100	73063072	84068190	84212100	85021380	85414090	
38151910	39219011	46012110	56031190	70191200	73063077	84068211	84213920	85023100	85437010	
38151990	39219019	46012190	56031210	70191990	73063080	84068219	84213940	85030010	85437030	
38159010	39219090	46012210	56031290	70193100	73064020	84068290	84213960	85030091	85437060	
38159090	39232100	46012290	56031310	70193200	73064080	84069010	84213980	85030099	85437090	
39031100	39232910	46012910	56031390	70193900	73065080	84069090	84213990	85043300	85481010	
39151000	39232990	46012990	56031410	70194000	73066110	84101100	84219900	85044040	85481021	
39152000	39239090	46019205	56031490	70195100	73066199	84101200	84289091	85044055	85481091	
39153000	39251000	46019210	56039110	70195200	73066910	84101300	84289095	85044081	86090010	
39159011	39252000	46019290	56039190	70195900	73066990	84109010	84411040	85044084	86090090	
39159080	39253000	46019305	56039210	70199010	73071110	84109090	84411080	85044088	87052000	
39159090	39259010	46019310	56039290	70199030	73071190	84118100	84623910	85059010	87059010	
39162010	39269050	46019390	56039310	70199091	73071910	84118220	84623991	85059030	87059030	
39162090	39269092	46019405	56039390	70199099	73082000	84118260	84623999	85059050	87059090	
39172110	40021910	46019410	56039410	72041000	73090030	84118280	84659600	85059090	87089220	
39172190	40021920	46019490	56039490	72042110	73090051	84119900	84741000	85061011	87089235	

4 Appendix: Estimations

Step 1. binomia-logit(Selection)	Agriculture, forestry and fishing	Manufacturing	Energy and water supply	Construction	Wholesale and retail trade	Transportation, hotels and restaurants, information and communication	Financial and insurance, real estate activities	Knowledge-based services	Travel agent, cleaning, and other operational services
Ln(man-year)		0.547***	1.466***		0.365***	0.286***	-0.039	0.552***	0.451***
Ln(man-year)^2	0.143***						0.058		
Ln(turnover/man-year)				23.886***	2.771***	-0.496***		-0.493***	15.281**
Ln(turnover/man-year)^2	0.013***	0.016***		-0.843***	-0.088**	0.026***		0.029***	-0.497*
Export dummy	-0.455				0.257**			-0.885***	
Import dummy					0.202*				
(Export/Turnover)	-1.368**	2.990***		1.496*	3.782***			1.261***	
(Export/Turnover)^2		-2.204***			-2.811***				
(Import/Turnover)								-0.000***	
(Import/Turnover)^2								0.000***	
(Share, skilled employees)				-1.019**	-0.237*		4933	1.086*	1026
(Share, skilled employees)^2							-4915		
(Share, short-cycle)	0.983*						-7362		
(Share, short-cycle)^2									
(Share, medium-cycle)				12.147**		-2.563**	10.501***		
(Share, medium-cycle)^2				-21.663**	-1.162***	2.976**	-8.138**		
(Share, long-cycle)			-20899		-2.735***		-14.635**		
(Share, long-cycle)^2			51246	-13.308***	2.984***		17.227***		11.331**
Dummies: educational level									
D_skilled	0.612*							-0.788***	
D_short-cycle							1673		
D_medium-cycle			4.702***	-0.808*	0.288***				
D_long-cycle					0.255**		1322		

D_PhD	1.738*	0.466**			-0.471		2.448*	-0.642*	
Dummies: field of education									
D_agriculture and fisheries						0.597**	-2.757**		
D_agricultural science							1.923*	0.790**	
D_food and nutrition	-1.617**			1.126*			0.944		
D_natural science									
D_social science			-4.324**	0.733**					
D_healthcare			-5.770***		-0.188*				
D_technical	-0.815**	0.299***			0.250***		-1.382*	0.976***	
D_transport	-2.634***				0.412***				
D_humanities							-2.109**	-0.746***	
D_educational					-0.223**			-0.543*	
D_arts					-0.258*			0.554*	
Foreign-owned enterprise				-1.193*	0.203**				
Industry dummies: stdgr-127	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographical region dummies:	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	534	4840	122	492	6686	1062	200	667	270
Explanatory power (pseudo R2)	0,115	0,24	0,421	0,096	0,16	0,114	0,3	0,179	0,142
Trin2. Poisson	Agriculture, forestry and fishing	Manufacturing	Energy and water supply	Construction	Wholesale and retail trade	Transportation, hotels and restaurants, information and communication	Financial and insurance, real estate activities	Knowledge-based services	Travel agent, cleaning, and other operational services
Ln(man-year)		-0.251***			-0.355***	-0.301***	3.686***	-0.187***	-2.140***
Ln(man-year)^2	-0.056**	0.021**		-0.027***			-0.611***		0.478***
Ln(turnover/man-year)			2.170***	-12.935**	-1.795**			0.011***	-0.322
Ln(turnover/man-year)^2			-0.066***	0.453**	0.056**				
Export dummy	0.691***	-0.325***	0.794***	0.477***					
Import dummy			-0.284**		-0.256**			0.17	

(Export/Turnover)	-3.292***	-0.363***			-0.636***	-0.452*			9.085***
(Export/Turnover)^2									-9.505***
(Import/Turnover)		-0.391**	0.000***			1622		-2.144***	5.520***
(Import/Turnover)^2		0.084***				-2.166**		0.903**	
(Share, skilled employees)	0.772**	0.265*					-58.460***	-2.659**	2.852*
(Share, skilled employees)^2							52.719***	1908	-4.080***
(Share, short-cycle)				-2.051**			-61.095***	-1.891**	4.090***
(Share, short-cycle)^2	0.621**		3.563***	1295		-6.699***		2.116**	
(Share, medium-cycle)			-5.059***			6.382***	13.248***		4.794***
(Share, medium-cycle)^2			4.665**			-5.401***	-19.349***		
(Share, long-cycle)	3.986***	1.529**		2.709***	0.917	-7.318***	3.373***	-1.353***	23.155***
(Share, long-cycle)^2		-2.061**	4.527***		-1.367*	11.356***	5.277***		-51.173***
Dummies: educational level									
D_skilled							7.229***	0.765***	3.415***
D_short-cycle						0.893***	16.637***		
D_medium-cycle						-1.203***	4.747***		
D_long-cycle	-0.821***						-11.631***	0.35	
D_PhD	1.783***			-1.402*			-14.705***		4.423***
Dummies: field of education									
D_agriculture and fisheries								0.596***	-1.298***
D_agricultural science				2.212***					
D_food and nutrition	-4.243***		-0.864***						
D_natural science			-0.519***						-6.040***
D_social science			0.261						-2.820***
D_healthcare									-1.395**
D_technical					0.172***				-1.695***
D_transport	4.457***	0.200**	0.747***		0.245**				
D_humanities	-1.097***		0.657***						-1.239***
D_educational						0.517***			

D_arts	1.034***	-0.342*			-0.681***			-0.677***	7.987***
Foreign-owned enterprise			-0.674***		0.139*				-1.878***
Industry dummies: stdgr-127	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Geographical region dummies:	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	131	2028	105	183	1924	190	28	187	60
Explanatory power (pseudo R2)	0.34	0.243	0.43	0.195	0.23	0.375	0.988	0.321	0.884

OSLO . COPENHAGEN . DAMVAD.COM

Badstuestræde 20
DK-1209 Copenhagen K
Tel. +45 3315 7554

Essendrops gate 3
N-0368 Oslo, Norway
Tel. +47 970 43 859