Application for licence to use statistical data

Instructions for filling in the application for licence to use statistical data can be found on the applications page.

The application for licence to use statistical data and the appendices are submitted to the Registrar's Office of Statistics Finland:
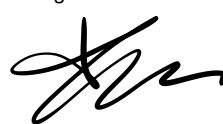- scanned by email: kirjaamo@stat.fi
- by post to: Registrar's Office, FI-00022 STATISTICS FINLAND.
- Street address: Työpajankatu 13, Helsinki

☐ New licence

☒ Extension of licence (or extension of data file), ref. *TK/206/07.03.00/2020*

☐ Addition of user, ref.

☐ Continuation of licence, ref.

1. Applicant	Project's responsibility person and organisation Pengpeng Xiao VATT and Duke University
	Email address pengpeng.c.xiao@gmail.com — Telephone +351 965427991
	Organisation's postal address VATT valtiontaloudellinen tutkimuskeskus, Arkadiankatu 7
2. Invoicing details Online invoicing address or similar, possible own reference	Invoicing customer (if other than the applicant's organisation) Ciprian Domnisoru Aalto University Department of Economics P.O. Box 21210 00076 Aalto, Finland
3. Names of persons who will be handling the data All who will handle the data file with direct or indirect identifiers in addition to the responsibility person	Name, organisation, email and telephone number Pengpeng Xiao, VATT pengpeng.c.xiao@gmail.com +351 965427991 Ciprian Domnisoru, Aalto University ciprian.domnisoru@aalto.fi +3580503205605 Veera Nippala, Aalto University email: veera.nippala@aalto.fi
4. Intended use of the data	☒ Scientific research ☐ Statistical survey
	Name of the project: Inequality in higher education and the labor market

Postal address:
00022 STATISTICS FINLAND

Street address:
Työpajankatu 13, Helsinki

Telephone:
+358 29 551 1000

www.stat.fi

Business ID: 0245491-1

Statistics Finland

| A short standard language summary (of a few sentences) for forming a general idea of the project: what are the data used for, for which purpose? | Purpose of use of the data:
This project studies gender inequality in education fields, occupations and wages. It aims to quantify the mismatches in the labor market for both men and women, and evaluates policies to mitigate the inefficiencies, reduce inequality and facilitate individual flourishing.

Purpose of new data:

We would like to add information on student performance in college (as measure by grades, credit accumulation, and enrollment patterns) available from the VIRTA modules.

Students enrolled in tertiary education face a series of fundamentally economic questions: to graduate sooner rather than later (or not at all), to work during their studies or not, and, if they do work, how much time to allocate to their studies versus paid work. Research in labor economics has shown that these decisions may have long-lasting impacts on career progression and lifetime earnings. In turn, individual decisions on time to graduation impact university and government budgets. Understanding and quantifying the labor market effects of these individual decisions is vital to our project, and can provide useful analyses for university decisionmakers.

We propose a set of analyses on 1) the determinants and 2) the labor market consequences of time to degree in Finnish academic masters' programs. These analyses will be conducted with Statistics Finland population register data (already available in our project) merged to student grade information in college. As such, we hope to achieve full coverage of student cohorts and, using administrative data, improve on insights from alumni surveys and update previous analyses of graduate outcomes in Finland.

Understanding the determinants and consequences of time to graduation is vital to informing policymaking, as policies influencing students to graduate earlier may have unwanted individual and societal effects on the transition to working life, the rate of youth unemployment, and the provision of family care. While a longer time to graduation may be seen by employers as a negative productivity signal, it allows students to search for jobs during their studies, potentially facilitating their transition to working life. The decision on how long to wait until graduation is particularly pressing during recessionary periods, or situations like the Corona pandemic. In such uncertain economic times, students anticipate poor labor market prospects, and may extend the duration of their studies as a buffer against periods of unemployment. Decisions on time to graduation are also crucially impacted by family and childcare responsibilities.

Example analysis 1: A hazard model of time-to-degree that accounts for unobserved heterogeneity and indicates how individual attributes and circumstances influence time to graduation.

Example analysis 2: A regression model of the effects of time to graduation on labor market outcomes such as earnings and unemployment status. In such a model, we plan to illustrate the sensitivity of the coefficient on time to graduation to the gradual addition of controls. This is equivalent to a decomposition that could answer the question: "What fraction of the observed effect of delayed time to graduation on labor market outcomes is explained by a) family responsibilities b) student ability and academic performance c) student employment during studies d) other observable individual and institutional factors and e) unobserved factors". While such a model remains descriptive, the stability of the effect of time to graduation on labor market outcomes after the inclusion of many observable controls could signal that potential biases are small. We are also exploring using policy quasi-experiments that influenced time to graduation to identify the causal effect of time to graduation on labor market outcomes.

These analyses could shed light on the labor market effects of time to graduation and deepen our understanding of the potential impacts of policies seeking to reduce time to graduation.

Such analyses would be vital building blocks for our research project but also provide timely analyses to universities on the labor market outcomes of their graduates, particularly on the effect of time to graduation, employment during studies, and academic performance on later labor market outcomes. |
| **5. Data requested from Statistics Finland**

Where necessary, data can be recorded in a separate appendix. | A. Microsimulation
SISU microsimulation model and related basic data files   ☐

Other data files connected to microsimulation: |

Statistics Finland

	B. Ready-made data files / service data files: Applying for all VIRTA files (LYHENNE_Opiskelijat_vvvv, LYHENNE_Lukukausi_ilmoit_vvvv, LYHENNE_Opiskeluoikeudet_vvvv, LYHENNE_Opiskeluoikjakso_vvvv, LYHENNE_Opintosuoritus_xxxx) for the following institutions: HY 01901 Helsinki University of AA 01903 Åbo Akademi OY 01904 University of Oulu TAU 10122 University of Tampere JY 01906 University of Jyväskylä SHH 01910 Hanken School of Economics University of Vaasa 01913 Vaasa University LUT 01914 Lappeenranta University of Technology, Lahti, LUT LAY 01918 University of Lapland AYO 10076 Aalto University ISY University 10088 Eastern Finland TY 10089 University of Turku TAIY 10103 University of the Arts Helsinki (TAY 01905 University of Tampere (old)) (TUT 01915 Tampere University of Technology (old)) If the request concerns the total data file of individual-level ready-made data files (e.g. FOLK modules), give reasons for the need of the total data file:
	C. Register variables (or topic) / other data (e.g. aggregated data file):
	Data are requested for the year/years: all years available
6. Other authorities' data	Other authorities' register data/data files to be attached to Statistics Finland's data.
7. Estimated duration of use of the data file The licence can be granted at most for five years.	5 years
8. Publicity of the application documents (see instructions)	The application for licence should be compiled so that it does not include confidential information. Where necessary, confidential information should be presented in a separate appendix. ☐ The applicant considers that the research plan appended in the application for licence is confidential. Reasons:
9. Method of using the data file	The licence to use the data file is applied for: ☒ FIONA remote access system ☐ for a data file to be released to your own organisation (see instructions)
	The data file will be used ☒ in Finland ☒ outside Finland Record here the countries and organisations where the data file will be used: Portugal, University of Porto

Statistics Finland

10. Appendices	**Must always be submitted:** Pledges of secrecy (granting of a licence to use data is conditional to obtaining a pledge of secrecy from the applicant as well as from all persons who will be handling the data): ☐ appended ☒ submitted earlier Summary of the research plan/the plan of the statistical survey, max. 3 pages (no plan needed on microsimulation): ☐ appended ☒ submitted earlier ─── **Must be submitted where necessary** (see instructions): Data protection description (register description) ☐ appended ☒ submitted earlier Licences of other authorities: ☐ appended ☒ submitted earlier Free-form organisation-level account of the use of data to be released abroad: ☐ appended ☒ submitted earlier Model for an information/consent form given to survey respondents: ☐ Appended ☒ submitted earlier

11. Date and signatures	Place Porto, Portugal	Time Nov. 22, 2021
	Applicant's signature Printed name: Pengpeng Xiao	Signature of the representative of the applicant's organisation Printed name: Essi Eerola