The London School of Economics and Political Sciences

Essays on intergenerational mobility and occupational choice

Maria Ventura

A thesis submitted to the Department of Economics for the degree of Doctor of Philosophy

London, August 2024

Declarations

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of approximately 33,000 words.

Statement of co-authored work

I confirm that Chapter 3 was jointly co-authored with Robert Blackburn, and Stephen Machin, and I contributed 35% of this work.

Statement of inclusion of previous work

I can confirm that chapter 3 was the result of previous studies I undertook at the Centre for Economic Performance (CEP).

Abstract

This thesis analyzes the determinants and consequences of different occupational choices, and how they relate to intergenerational dynamics and the decision to engage in self-employed work.

The first chapter provides descriptive evidence on the intergenerational transmission of occupations. Using a rich set of administrative data from the Netherlands, I find that children are twice as likely to enter an occupation when it is the same as their parents'. This rate substantially increases for those above the top quartile of the parental income distribution. I observe wide heterogeneity across fields and establish the existence of a positive correlation between rates of transmission and average income in the occupation. I also uncover a gender matching pattern, with sons (daughters) following fathers (mothers) more often, even when the parental field is moderately gendered.

The second chapter explores private returns from occupational persistence. It goes beyond a correlational measure and connects intergenerational occupational transmission to income mobility. To this purpose, I exploit a unique policy experiment in the context of the Netherlands to estimate children's private gains when choosing the same profession as their parents'. Instrumental variable estimates show that "dynastic" doctors experience a 23% income boost relative to individuals who did not follow their parents.

The third chapter continues the study of occupational choice in a different context, i.e. self-employment in the United Kingdom. Using the UK Labour Force Survey, as well as newly collected survey data, I document recent trends in self-employment and describe the deleterious impact that the Covid-19 crisis had on this previously

growing group of workers. In this part, I uncover some of the broad heterogeneity characterising the pool of self-employed and show that it matters both for studying its recent trends and its reaction to the pandemic shock. Thanks to this novel data, I also find that four out of ten individuals currently in self-employment state a preference towards moving to a position of regular employment, if they could secure the same income level. Yet, they remain in self-employment due to lack of skills or adequate opportunities.

Acknowledgements

I owe the deepest gratitude to my supervisors and advisors, who have supported me throughout this journey and gently nudged me to do more and better. Camille Landais has been the first one to believe in my projects: I know my research agenda, which now excites me tremendously, would not be the same without his guidance. Johannes Spinnewijn provided the encouragement I needed to move forward and pushed me to always keep this excitement alive. Stephen Machin offered invaluable help with my research even before he became my supervisor and inspired in me the passion for social mobility and inequalities. I also want to thank Oriana Bandiera, who has been a mentor since before my PhD and a role model throughout, and Michele Piccione, who, despite not being one of my supervisors, has generously gifted me with hours of his time and attention over the last six years.

I am grateful for the feedback, comments, and emotional support from the brilliant researchers and professors who have accompanied my journey at the LSE and the CEP, as well as to my sisters and brothers in arms in the PhD: Alan Manning, Andres Felipe Fajardo Ramirez, Anton Heil, Bernardo Mottironi, Hugo Reichardt, Edoardo Leonardi, John Van Reenen, Guglielmo Ventura, Guy Michaels, Gustave Kenedi, Matteo Sandi, Shadi Farahzadi, Tiernan Evans and Tom Monk. I also enjoyed working with and learning from Robert Blackburn and Silvia Barbareschi.

My good friends Daniel, Giorgio, Inês, Iria, Felix, Mathilde, Mary, Nicolas, and Sophocles made London feel like home, and brightened up these years. I am also deeply thankful to Alice, Claudia, Enrico, Giulia, and Laura for regularly reminding me that I am more than my job and for helping me find my place when I return to Italy.

My parents, Chiara and Luigi, have instilled in me their love for learning, and

always supported me in countless ways throughout my educational journey, showing me that I could achieve anything with commitment and discipline. For this, and for the love they show me everyday, I am eternally grateful. My brother, Enrico, is a true example of resilience, and, somewhat involuntarily, has inspired me to become stronger: thank you for being there for me, I love you. I also wish to thank my grandparents Carolina, Lucia, Mario and Pietro, my cousins Francesco and Giorgia and my aunts and uncles Elisabetta, Sabina and Sergio, for their constant encouragement along the way.

Last, but very much not least, I am incredibly grateful for Nico, who brought sunshine into my life and unconditionally believed in me when I did not. It is thanks to you that I made it here.

It has been a wild ride, but worth the way.

Contents

List of Figures

List of Tables

Chapter 1

Intergenerational Transmission of Occupations: Trends and Gender Patterns

"All women become like their mothers. That is their tragedy.
No man does. That's his."
– Oscar Wilde, The Importance of Being Earnest

1.1 Introduction

The last decades have seen a renewed and growing interest in economic intergenerational dynamics and their implications for inequalities. However, an abundance of studies have focused on a few transmittable traits. Famously, and justifiably, income has often populated the debate on socioeconomic mobility across generations, sometimes accompanied by other factors such as education and wealth. On these fronts, evidence from Europe has repeatedly painted an heterogeneous picture, with the UK being among the least mobile countries, comparable to the US, and Scandinavian countries appearing about half as "sticky" (Jantti et al., 2006; Major and Machin, 2020). At the same time, studies analysing transmission of occupation from parents to children have remained relatively scarce and scattered across social sciences.

This work aims at enriching the existing body of evidence on intergenerational dynamics by providing novel and systematic evidence on intergenerational occupational persistence. In particular, the focus on the Netherlands will paint a picture of a context ideally comparable to several western countries. For the purpose of this project, my proxy for occupation is the field of the highest achieved level of education. This measure includes 82 fields, and it has the advantage of being more stable over an individual's life cycle than standard occupational titles.

The chapter provides novel stylized acts on occupational persistence and focuses in particular on gender differences within this type of mobility. The latter is justified by the potential role this channel may have in pushing offsprings in or out of gender segregated occupations, which in turn can perpetuate, or fight, social imbalances such as the gender wage gap. To this purpose, I first document the incidence of intergenerational occupational transmission and find that occupational persistence is higher than in a counterfactual scenario with no transmission mechanisms and only "lucky" matches. I show that children are, on average, twice as likely to choose a certain occupational field when this is shared with a parent. The rate of transmission remains constant along most of the parental income distribution but increases sharply above the fourth quartile, with children from families at the very top of the distribution (99th percentile) approximately twelve times as likely to pick a given field when a parent has done the same. Persistence rates also exhibit broad heterogeneity across occupational fields.

I then consider gender differences and uncover a pattern of transmission where daughters tend to follow their mothers, while sons are prone to choose the paternal occupation. This is robust to the inclusion of several controls and survives when restricting the sample to parental occupations that are moderately gendered, i.e. where the share of workers of the opposite sex are above the median.

Although further investigation will be required to confirm the results and move towards more causal estimation, the analysis provides a clean and formalized overview of the relevance of intergenerational occupational persistence in a European country. The new findings give preliminary insights on potential mechanisms behind the phenomenon and suggest it could represent a novel and important factor exacerbating social immobility. With previous literature widely recognising the relation between intergenerational mobility and inequalities (OECD, 2018), the relevance of conducting research on this topic becomes evident.

Related literature. Intergenerational transmission of occupations has been typically studied in the field of sociology, and mostly as a measure of social mobility. In this context, the general rationale for using occupations relies on the belief that occupational structure is closely related to socioeconomic status, or prestige, and approximates the hierarchy of skills and subsequent rewards (Parkin, 1972). By extrapolating these elements from the information on occupation, it is then possible to analyse social stratification in a way that has been reputed more stable than income and consequently more representative of the lifetime earnings profile (Goldthorpe and McKnight, 2006). With the exception of a few more recent contributions (Lo Bello and Morchio, 2018; Mocetti et al., 2020), and a contemporaneous work on fields of study (Altmejd, 2024), economists have usually considered jobs in the context of intergenerational mobility in absence of (or to complement) income data. The use of occupation as stand-alone element of interest in the intergenerational transmission game and as a potential factor in the persistence of earnings, remains widely undocumented.

This project aims at bridging this gap by providing fresh evidence on trends and determinants of intergenerational persistence in occupation in a European country. As a matter of fact, for the last decades economists have notably focused on the transmission of earnings (e.g., Becker and Tomes, 1986; Bell et al., 2018; Chetty and Hendren, 2018a,b; Chetty, Hendren, et al., 2014; Manduca et al., 2020; Solon, 1992, 2002). From this work, we have learnt that elasticity of income varies across countries and across cohorts with, for instance, the United States particularly and persistently immobile and Scandinavian countries showing steadily high mobility rates. Common subjects in the study of intergenerational mobility have also encompassed education (Blanden et al., 2005; Checchi et al., 1999), house ownership (Bell et al., 2018) and wealth (e.g., Adermon et al., 2018; Benhabib et al., 2019; Clark and Cummins, 2015; Piketty, 2000). The tight relationship all these measures have with various indexes of inequality has made them particularly salient, over the years, in the challenge of levelling up the gap in opportunities. So far, only a handful of studies have focused on the role of occupation, but they have either documented persistence within broad 1-digit categories (Constant and Zimmermann, 2003; Lo Bello and Morchio, 2022), which could possibly proxy for socioeconomic status instead, or have pivoted on very specific traits of occupations (e.g., Mocetti et al., 2020, on the case of regulated occupations) without providing a complete picture of the phenomenon in their context. Nevertheless, the study of occupational choice can enclose a number of aspects, such as skills, preferences and cultural and social issues among others, that go beyond the purely monetary ones. This work is one of the first attempts at considering occupation as a per se relevant factor and lays the foundations for its study in relation to economic mobility.

The paper also speaks to the existing work on gender inequalities by showing differential transmission across genders (both on the parent and on the children side). Studies on intergenerational transmission have generally neglected gender differences, or even excluded mothers and daughters altogether. While a few papers have looked at women's preference-induced occupational segregation (Escriche, 2007) and daughters' occupational choices with respect to their fathers' (Hellerstein and Morrill, 2011), to my knowledge no study has yet provided a comprehensive gender heterogeneity analysis[1]. Uncovering differential dynamics of parental following could give insights on daughters' occupational choices and in turn shed light on women's different representation in higher paid, often male dominated, jobs (Altonji and Blank, 1999; Bertrand et al., 2019; Joy, 2006).

The paper is organized as follows. Section 2 describes the data and provides information on the Dutch education system. Section 3 presents a few stylized facts on occupational transmission. Section 4 focuses on gender differences and their robustness in different scenarios. Section 5 concludes and highlights directions for future research.

[1]The one notable exception is the contemporaneous contribution of Altmejd (2024), whose results are generally consistent with those reported in this chapter.

1.2 Data and context

This section provides information on the data used in the analysis as well as background on the Dutch education system.

1.2.1 The administrative data

The data used for the analysis include comprehensive administrative records provided by the Dutch Bureau of Statistics (CBS). These contain several datasets, linkable at the individual level using a unique person identifier.

Demographics and family connections These include information on gender, date and country of birth, nationality and marital status as well as the link with both parents. They are available for all Dutch residents from 1994.

Education Records of the highest attained education, with detailed program, field and level are utilised in this chapter. Fields are coded according to the International Standard Classification of Education for fields of education and training (ISCED-F 2013), a standard utilised internationally which includes over 80 detailed categories. In this project, the proxy for occupation is the field of the highest achieved level of education, whenever this is above high school. Table 2.9 in the appendix of Section 1.9 reports a list of the fields with the respective codes, as used in this paper. This information is systematically available for the universe of students since 1999 and on a sample basis for those having completed education before then. The coverage is very high for recent cohorts and included almost 11 million people (65% of the population) in 2015, with increases every year. Nevertheless, parental generations will naturally display lower coverage, which will in turn affect the sample to which the descriptive analysis can be restricted to. Table 2.10 in Section 1.9 presents a few correlations that describe how the individuals whose at least one parent can be matched to the education records differ from the rest of the population. The correlations are quite small in magnitudes, as highlighted by the standardizes beta coefficients.

Labour market outcomes and structure Information on individual (primary, gross and net) income is collected from tax records and available within the CBS since 2006. The measure of income used in the analysis will be primary income, i.e. gross labour income from salaried employment as well as own business. The same dataset also includes employment status and sector of activity. As several observations are available for each individual, I use the average between 30 and 40 years old for children and between 50 and 60 years old for parents.

Tax data on the universe of salaried employees from 2003, with information wages and monthly hours, as well as type of contract (part-time, temporary, etc) is also included. The information available in this dataset is also measured as an average between 30 and 40 years old.

The final sample includes individuals born between 1963 and 1991 where at least one parental link exists and for whom the field of the highest level of education is available. Additionally, I restrict it to consider individuals with income and earnings records in the relevant time frame as well as income information for at least one parent. This leaves 962,437 observations, of which 937,426 with positive income. Table 2.1 reports summary statistics for some demographic traits. The sample appears representative of the general population[2] for share of women and education levels, although individuals born in the Netherlands are overrepresented compared to foreign born. This is not surprising as the parents of those belonging to the latter group are likely not to be Dutch residents. They would therefore not appear in the administrative records and can not be linked for the purpose of intergenerational study.

1.2.2 Education in the Netherlands

The Dutch schooling system is built upon the early separation of vocational and scientific education, establishing distinct paths for students from as early as 12 years of age (Figure 1.6 in Section 1.8). After a first common track in elementary school (*Basisschool*) pupils have a choice of three possible secondary education options: pre-vocational education (*Voorbereidend middelbaar beroepsonderwijs*, or VMBO),

[2]As of 2022 official statistics.

"general" education (*Hoger algemeen voortgezet onderwijs*, or HAVO) and preparatory scientific education (*Voorbereidend wetenschappelijk onderwijs*, or VWO). Each track allows access to a different type of post-secondary education, which reflects the same broad separation in: junior college track (*Middelbaar beroepsonderwijs*, or MBO), vocational university (*Hoger beroepsonderwijs*, or HBO) and scientific education (*Wetenschappelijk onderwijs*, or WO). While the first two prepare technical (such as mechanic or nursing assistant) and professional figures (like accountants and tax advisors) respectively, WOs correspond to a traditional university track, with the possibility of pursuing undergraduate and postgraduate degrees. Additionally, some WO programs may require students to have passed high school exams in certain subjects. In 2020 more than 80% of the population between 25 and 64 years of age had achieved a form of education corresponding to MBO or above (OECD, 2021). The representativeness of this measure, together with its greater stability over the life cycle compared to job titles, makes it a particularly suitable proxy for occupational fields. Moreover, it reflects occupational aspirations as measured early in the choice process and, in the case of technical and vocational tracks, it maintains a very close relationship to actual occupation[3].

1.3 Stylized facts on occupational persistence

Previous work on the inheritance of occupations and fields of study have used different metrics to quantify its magnitude. In this section, I will consider two. One simple way is to calculate the raw share of children that have chosen the same field of at least one of their parents. In the sample this is 6.2% and it can be interpreted as the probability that an individual randomly drawn from the population has made the same occupational choice as at least one of their parents. It is worth noticing that this representation of occupational persistence, although simple to calculate and interpret on any sample, is particularly dependent on the definition of occupation used, and will produce potentially very different values depending on its granularity.

[3] Junior college tracks (MBO) include very specific programs, such as bakers, cleaners and bicycle technicians training.

I therefore compare it to the average size of an occupation at the same level of detail, i.e. the average probability of the same random individual to be in any occupation, which is 3.2%, i.e. almost half the above persistence rate.

A different way of visualising the significance of the intergenerational transmission of occupation is to think about how the probability of choosing a certain field changes when at least one parent is in it. This corresponds to running a conditional logit model (McFadden et al., 1973) in which an individual chooses occupation j from the alternatives in their choice set C_i such that

$$P(choice_i = j) = \frac{\exp(x_{ij}\beta)}{\sum_{k \in C_i} \exp(x_{ik}\beta)} \tag{1.1}$$

where x_{ij} is a dummy indicating whether a parent is in occupation j, and in this model represents the relevant attribute for the jth choice alternative, as perceived by the ith individual. The coefficient β can be interpreted as an average estimate across all fields[4]. The odds ratio of choosing a field when shared by at least one parent, as opposed to when both parents are in different fields, can the be expressed as e^{β}. Table 1.2 reports the estimation results, in odds ratios. Column (1) shows the unconditional coefficient, revealing that children are on average twice as likely to choose a field when a parent has made the same choice, compared to when no parent is in the same field. The result is robust to the inclusion of controls for gender and cohort (columns (2)), level of study attained (column (3)), as well as parental income (column (4)). I conclude that both measures imply that the transmission rate of occupational fields is higher than what would be produced by "lucky", or accidental, occasional matches.

Does the persistence rate obtained in the previous paragraph change for different economic backgrounds? Figure 1.1 illustrates the correlation between the odds ratios of choosing a given occupation when shared with a parent[5], and percentile rank of parental average income[6]. If occupation was transmitted in the same way at all

[4]Although the coefficients of a fixed effect logit model are biased, Greene (2004) shows that the bias quickly dissipates with more than three choices. As in this case the number of fields is above 80, the bias should be very close to zero.

[5]The same pattern is obtainable using shares instead of odds ratios.

[6]The coefficients are obtained by augmenting Equation 1.1 with occupation specific interactions.

points of the parental income distribution, we would expect to see a straight line. In fact, we observe a first gradual rise after the 75th percentile, as well as a very steep increase in slope from in the top decile. As an example, the 99th percentile of parental income corresponds to an odds ratio that is about 4 times those of the percentiles below the 75th[7]. The finding is striking as it reveals strong distributional differences that could have implication for social mobility, especially at the top of the income distribution. As long as every occupational choice ranks individuals similarly in terms of income, higher persistence at the top will imply higher immobility of rich families. This would be further exacerbated in case of a positive correlation between occupational transmission and own income.

I further explore the heterogeneity in the rate of transmission of occupational choice by field. Figure 1.2 illustrates, for each of them, the odds ratios of choosing a given occupation when shared with a parent[8] and highlights a few salient occupations. Interestingly, the levels of persistence in occupational choice present large variation across fields. Moreover the heterogeneity is significantly correlated to the average income in the occupations (Figure 1.3), with an increase in 10% in income associated to a 9% increase in the odds ratio of choosing the occupation when shared by a parent. This, of course, resonates with the previous evidence on persistence rates across the income distribution.

Overall, the descriptive facts illustrated in this section hint at a potentially relevant role for occupational persistence in the construction of intergenerational income dynamics and social mobility. They have highlighted its tight connection with top incomes, which is also reflected in the types of occupations in which transmission is prevalent. Chapter 2 will further explore the implications for children's labour market outcomes and social mobility.

Confidence intervals are omitted for readability, and reported instead in 1.7 in Section 1.8

[7]It may be objected that higher persistence could be driven by a smaller number of fields chosen by individuals at the top of the income distribution. Although this differential concentration of fields is observed in the data, it does not seem to drive the relationship above. As a check, Figure 1.8 in the appendix in Section 1.8 reproduces the same graph, this time controlling for how common the parental occupation is at the considered income percentile. While magnitudes change, the pattern remains virtually unchanged.

[8]Statistically insignificant estimates are set equal to 1 for readability. Confidence intervals are omitted and reported instead in Figure 1.9 in the appendix in Section 1.8.

1.4 Gender patterns

I now turn to exploring whether and to what extent occupational persistence rates differ by gender. The question is relevant as it may speak to the social utility of the transmission of occupation with regards to occupational segregation and gender gaps.

1.4.1 Rates and trends

For this part, I proceed similarly to Section 1.3 and start by displaying in Figure 1.4 the raw shares of children, by gender, who are in the same occupational field as each parent. The "Random" column represents the average size of the parental occupations and serves as a benchmark for what we would expect as the accidental matches in occupations between parents and children. Panels A and D show that using this simple measures results in a same-sex following pattern, with sons both more prone to choose the same occupation as their fathers than that of their mothers, and more likely to follow their fathers than daughters are. Symmetrically, daughters tend to go in the maternal field more than in the fathers', and are substantially more likely than their male counterparts to follow their mothers. Additionally, while same-sex matches appear consistently higher than the random comparison, this is not the case for cross-gender pairs. In fact, the cases of daughters being in the same occupation as their fathers look virtually the same we would expect because of lucky matches. Interestingly, the mother-son pairs are even below that, which would hint at a resistance from sons to follow their mothers in the same field. Note that these shares are unconditional, as they do not control for any other individual or parental characteristic.

In order to formalise these relationships, Table 1.3 reproduces the conditional logit model proposed by Equation 1.1, this time differentiating by children's and parents' gender. The results broadly confirm those from the raw shares, and are robust to the inclusion of both individual and parental controls. Sons are always more likely to follow fathers than mothers, for which the odds ratio is, consistently with Figure 1.4, below one. They also choose fathers' occupations twice as often as daughters. Girls, on the other hand, choose to follow mothers more than fathers and are three times as likely to go into the maternal field than boys.

We may expect the last decades to have brought, together with developments in women participation in the labour market, relevant changes in the transmission rates by gender. Figure 1.5 displays persistence shares by cohort, for each parent-child pair. The picture reveals little to no change for paternal following, but growth in occupational persistence with respect to mothers. This has affected both sons and daughters, possibly because the higher rate of female labour market participation seen in recent years may have opened up opportunities for these matches. Unsurprisingly, the same-sex matching pattern holds at all times, although recent cohorts of daughters are now also visibly more likely to follow their mothers than their fathers.

1.4.2 Gender-dominated fields

The findings above open up the question of how much of the observed pattern is actually driven by emulation mechanisms of children towards parents, for instance in the form of parental pressures or role model effects. In fact, the same results may be explained by gender segregation into fields, with daughters following mothers more than fathers only when the occupation is generally deemed typical, for a girl, and similarly for sons.

Table 1.4 tries to address this issue by reproducing the model in Table 1.3 for specific splits of the sample. In particular, columns (1) and (2) restrict the population of interest to the child-father pairs in which the paternal field exhibits a share of women above the median or the 75th percentile, respectively. Parallelly, columns (3) and (4) do the same for the share of men in mothers' occupations. Two main facts emerge from this picture. First, all odds ratios for same-sex child-parent pairs remain above one. This implies that even when the parental occupation is dominated by workers of the opposite gender, transmission rates across these matches keep being over and above what we would expect in a random counterfactual driven by the distribution of occupations. Secondly, the gender pattern observed for the full sample survives when considering parental fields that are mildly gender-dominated, i.e. with shares of the opposite sex above the median. In extreme scenarios, namely with shares of men (women) in the maternal (paternal) field above the 75th percentile, gender segregation prevails over transmission mechanisms, making odds ratios for either child virtually the same.

Overall, the above suggests some relevance of gender, and particular same-sex matches, in occupational following across generations. This is socially relevant, as it may imply that having a larger cohort of women in STEM in a given generation, may trigger a virtuous circle through their transmission to daughters. This could in turn lead to greater gender balance across occupational fields. Future research should deepen our understanding of this aspect and explore potential mechanisms and channels leading to it.

1.5 Conclusions

This chapter provides a descriptive analysis of the intergenerational transmission of occupations, highlighting the persistence and patterns of occupational choices across generations. The evidence presented suggests that occupational inheritance is a significant phenomenon, with parent-child occupational matches occurring at rates higher than what would be expected by random chance. This finding underscores the importance of family background in shaping the career paths of the next generation.

A particularly striking aspect of this study is the pronounced increase in occupational transmission at the top of the parental income distribution. This suggests that the advantages associated with higher-income occupations may be more effectively passed from parents to children, possibly due to better access to resources, networks, and education, which facilitate the continuation of these high-income careers within families. An important caveat to this is that the proxy for occupation used in this chapter is the field of the highest level of education achieved, wherever applicable, after high school. This maintains the population of interest fairly representative, due to the structure of the Dutch education system. Nevertheless, it also inevitably excludes individuals who stopped their educational journey at the secondary level or earlier, who may have added relevant information to the relationship between income and occupational persistence.

Moreover, the analysis reveals heterogeneity in occupational transmission across different fields, with a correlation between the profitability of occupations and the likelihood of intergenerational persistence. This confirms that economic incentives and the potential for financial success may play a role in the decision to follow in a parent's professional footsteps.

The study also uncovers gender patterns in occupational transmission, with sons more likely to follow their fathers' career paths and daughters more often entering their mothers' fields. This is potentially policy relevant as it may suggest that incentivizing women to approach fields that are typically male-dominated but highly profitable, such as STEM, could also shape the occupational composition of the next generation. This gendered transmission may be influenced both by direct inheritance mechanisms and by broader patterns of occupational gender segregation, where daughters are more prone to enter fields dominated by women and vice versa. This aspect is partially explored in this chapter by looking at the persistence behavior in gender-dominated occupations. While the same pattern survives when considering moderately gendered parental fields, it fades with very high levels of opposite gender workers populating the occupation.

Overall, the findings of this chapter contribute to the understanding of how occupations are transmitted across generations, with important implications for social mobility, economic inequality, and the perpetuation of occupational hierarchies. These insights are particularly relevant for policymakers and researchers interested in addressing barriers to equal opportunity and reducing the persistence of economic inequality across generations.

References

Adermon, Adrian, Mikael Lindahl, and Daniel Waldenström (2018). "Intergenerational wealth mobility and the role of inheritance: Evidence from multiple generations". In: *The Economic Journal* 128.612, F482–F513.

Altmejd, Adam (2024). "Inheritance of fields of study". In.

Altonji, Joseph G and Rebecca M Blank (1999). "Race and gender in the labor market". In: *Handbook of labor economics* 3, pp. 3143–3259.

Becker, Gary S and Nigel Tomes (1986). "Human capital and the rise and fall of families". In: *Journal of labor economics* 4.3, Part 2, S1–S39.

Bell, Brian, Jack Blundell, and Stephen Machin (2018). "The changing geography of intergenerational mobility". In.

Benhabib, Jess, Alberto Bisin, and Mi Luo (2019). "Wealth distribution and social mobility in the US: A quantitative approach". In: *American Economic Review* 109.5, pp. 1623–47.

Bertrand, Marianne et al. (2019). "Breaking the glass ceiling? The effect of board quotas on female labour market outcomes in Norway". In: *The Review of Economic Studies* 86.1, pp. 191–239.

Blanden, Joe, Paul Gregg, and Stephen Machin (2005). "Intergenerational mobility in europe and north america". In: *Report supported by the Sutton Trust, Centre for Economic Performance, London School of Economics.*

Checchi, Daniele, Andrea Ichino, and Aldo Rustichini (1999). "More equal but less mobile?: Education financing and intergenerational mobility in Italy and in the US". In: *Journal of public economics* 74.3, pp. 351–393.

Chetty, Raj and Nathaniel Hendren (2018a). "The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects". In: *The Quarterly Journal of Economics* 133.3, pp. 1107–1162.

— (2018b). "The impacts of neighborhoods on intergenerational mobility II: County-level estimates". In: *The Quarterly Journal of Economics* 133.3, pp. 1163–1228.

Chetty, Raj, Nathaniel Hendren, et al. (2014). “Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility”. In: *The American Economic Review* 104.5, pp. 141–147.

Clark, Gregory and Neil Cummins (2015). “Intergenerational wealth mobility in England, 1858–2012: surnames and social mobility”. In: *The Economic Journal* 125.582, pp. 61–85.

Constant, Amelie F and Klaus F Zimmermann (2003). “Occupational choice across generations”. In.

Escriche, Luisa (2007). “Persistence of occupational segregation: The role of the intergenerational transmission of preferences”. In: *The Economic Journal* 117.520, pp. 837–857.

Goldthorpe, John H and Abigail McKnight (2006). “The economic basis of social class”. In: *Mobility and inequality: Frontiers of research in sociology and economics*, pp. 109–136.

Greene, William (2004). “The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects”. In: *The Econometrics Journal* 7.1, pp. 98–119.

Hellerstein, Judith K and Melinda Sandler Morrill (2011). “Dads and daughters the changing impact of fathers on women’s occupational choices”. In: *Journal of Human Resources* 46.2, pp. 333–372.

Jantti, Markus et al. (2006). “American exceptionalism in a new light: a comparison of intergenerational earnings mobility in the Nordic countries, the United Kingdom and the United States”. In.

Joy, Lois (2006). “Occupational differences between recent male and female college graduates”. In: *Economics of Education Review* 25.2, pp. 221–231.

Lo Bello, Salvatore and Iacopo Morchio (2018). *Like father, like son: Occupational choice, intergenerational persistence and misallocation*. Tech. rep.

— (2022). “Like father, like son: Occupational choice, intergenerational persistence and misallocation”. In: *Quantitative Economics* 13.2, pp. 629–679.

Major, Lee Elliot and Stephen Machin (2020). *What Do We Know and What Should We Do About Social Mobility?* Sage.

Manduca, Robert et al. (2020). “Trends in Absolute Income Mobility in North America and Europe”. In.

McFadden, Daniel et al. (1973). “Conditional logit analysis of qualitative choice behavior”. In.

Mocetti, Sauro, Giacomo Roma, and Enrico Rubolino (2020). “Knocking on parents’ doors: regulation and intergenerational mobility”. In: *Journal of Human Resources*, 0219–10074R2.

OECD (2018). *A Broken Social Elevator? How to Promote Social Mobility*, p. 352. DOI: https://doi.org/https://doi.org/10.1787/9789264301085-en. URL: https://www.oecd-ilibrary.org/content/publication/9789264301085-en.

— (2021). *Education at a Glance 2021: Netherlands - Country Note.*

Parkin, Frank (1972). *Class inequality and political order: Social stratification in capitalist and communist societies*. Paladin.

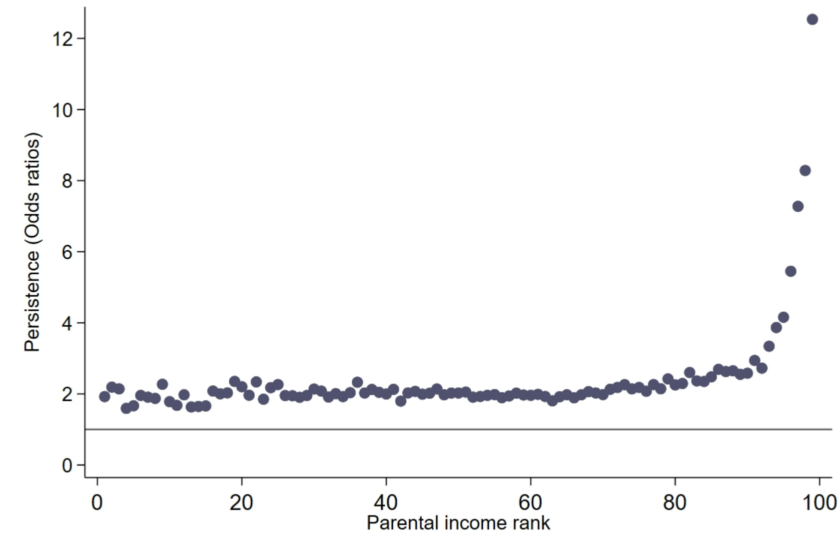
Piketty, Thomas (2000). “Theories of persistent inequality and intergenerational mobility”. In: *Handbook of income distribution* 1, pp. 429–476.

Solon, Gary (1992). “Intergenerational income mobility in the United States”. In: *The American Economic Review*, pp. 393–408.

— (2002). “Cross-country differences in intergenerational earnings mobility”. In: *Journal of Economic Perspectives* 16.3, pp. 59–66.

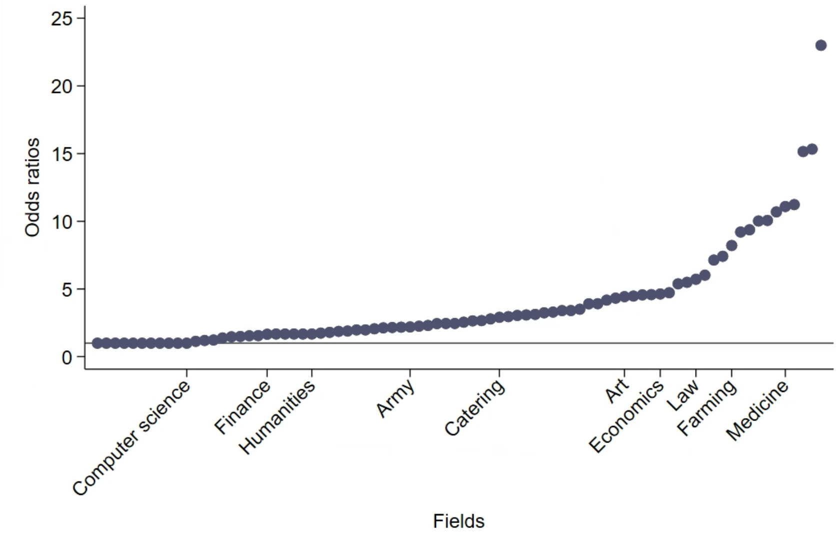
1.6 Figures

Figure 1.1: Persistence and family economic background



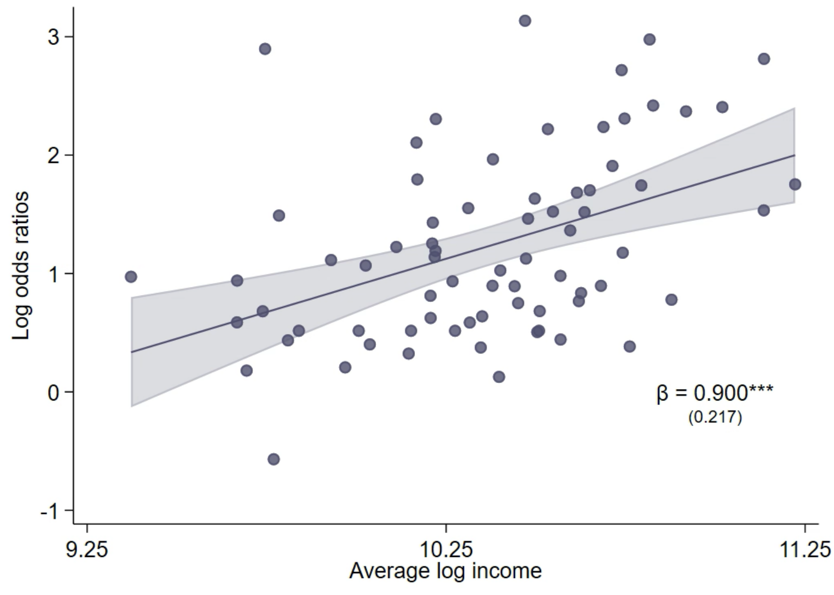
Notes: Persistence is measured as the odds ratios from the logistic model described in Equation 1.1. The specification is augmented to include interactions with parental income rank. The latter is calculated as the average between maternal and paternal income percentile rank. Parents' income is measured as described in Section 1.2. The black line corresponds to odds ratios equal to one, where an individual is as likely to choose an occupation when having a parent in it as when no parent shares the same occupation.

Figure 1.2: Persistence and fields



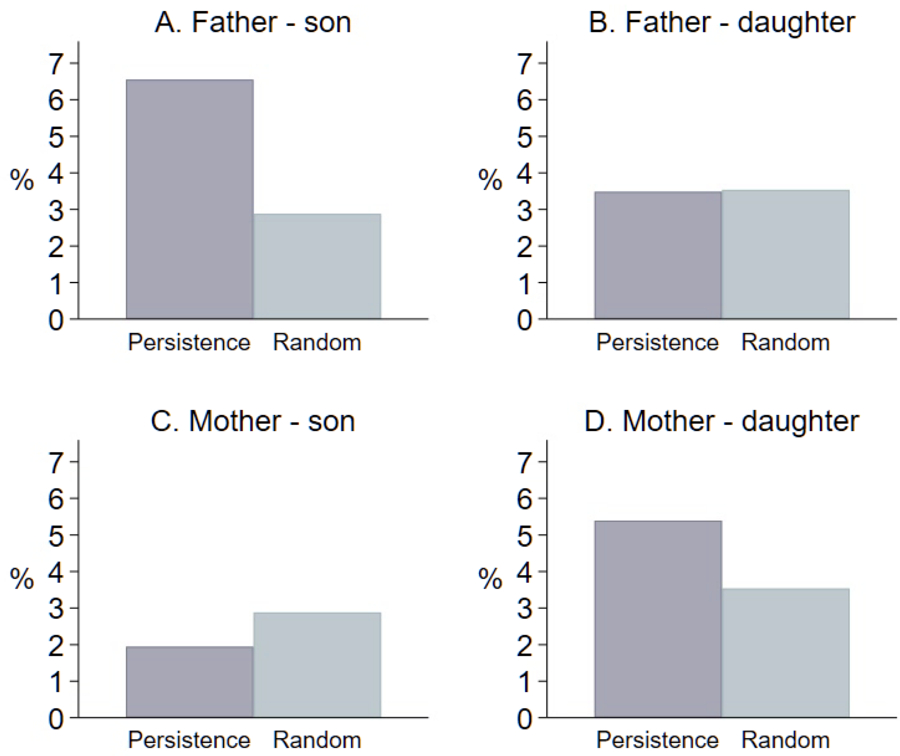
Notes: Persistence is measured as the odds ratios from the logistic model described in Equation 1.1. The specification is augmented to include interactions with ISCED-F 3 digit fields. The black line corresponds to odds ratios equal to one, where an individual is as likely to choose an occupation when having a parent in it as when no parent shares the same occupation.

Figure 1.3: Persistence and income by field



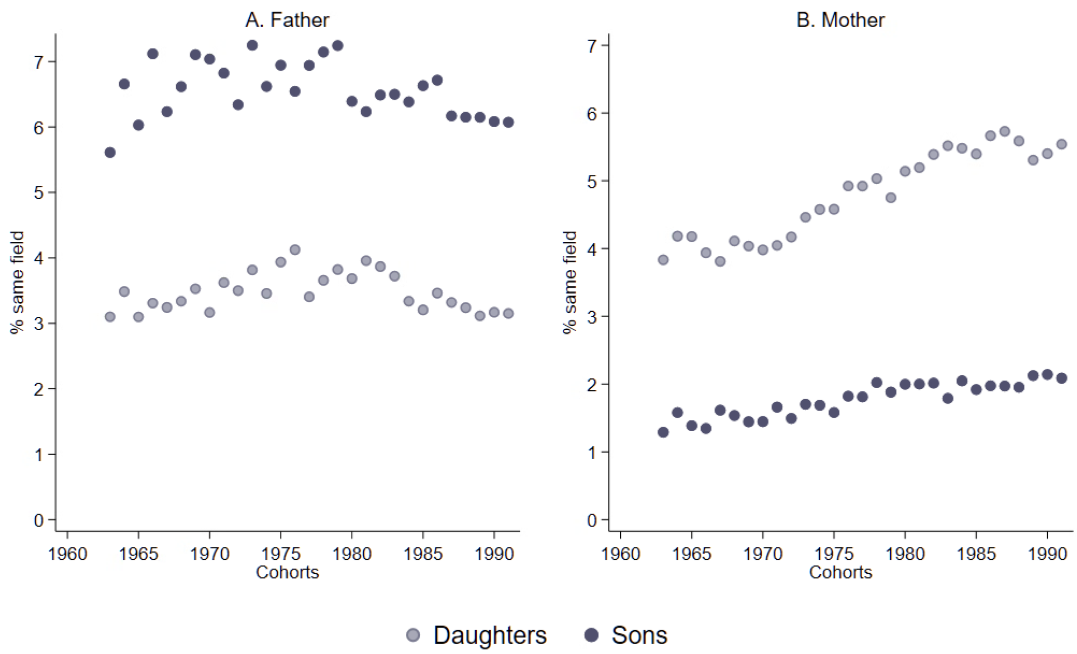
Notes: Persistence is measured as the odds ratios from the logistic model described in Equation 1.1. The specification is augmented to include interactions with ISCED-F 3 digit fields. Average income is calculated for each occupation as the mean primary income for workers in the field.

Figure 1.4: Persistence rates, by gender



Notes: Persistence is measured as share of sons/daughters in the same field as the indicated parent. Random corresponds to the average size of a field of study at the same level of granularity.

Figure 1.5: Persistence rates, by gender and across cohorts



Notes: Persistence is measured as share of sons/daughters in the same field as the indicated parent. Cohorts refer to the year of birth of children.

1.7 Tables

Table 1.1: Summary statistics - All fields

	Mean	SD	N
Female	0.510	0.500	962,437
Dutch	0.947	0.224	962,437
Year of birth	1983	5.963	962,437
Obtained higher education (VO)	0.200	0.400	962,437
Obtained higher professional education (HBO)	0.347	0.476	962,437
Father obtained higher education (VO)	0.070	0.255	962,437
Father obtained higher professional education (HBO)	0.150	0.257	962,437
Mother obtained higher education (VO)	0.029	0.167	962,437
Mother obtained higher professional education (HBO)	0.126	0.331	962,437
Parent is a doctor	0.010	0.102	962,437
Sample with positive income	0.974	0.159	937,426

Notes: The sample includes individuals born between 1963 and 1991 for which a parental link exists and for whom the field of the highest level of education, own income and parental income is available.

Table 1.2: Magnitude of persistence

	(1)	(2)	(3)	(4)
	Chooses occupation j			
Parent is in occupation j	2.051***	2.051***	2.052***	2.055***
	(0.0095)	(0.0095)	(0.0095)	(0.0095)
Controls for demographics		X	X	X
Study level FE			X	X
Controls for parental income				X
Observations	962,510	962,510	962,510	962,510

Notes: * $p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Robust standard error in parentheses. Coefficients are odds ratios. The model is a logit with multiple options, run as a binary one on a stacked sample, as illustrated in Equation 1.1. Column (2) adds controls for gender and cohort fixed effects. Column (3) adds study level fixed effects, which include three categories: WO, HBO and MBO. Column (4) adds average income for both parents, calculated as described in Section 1.2.

Table 1.3: Persistence, by parent and gender

	(1)	(2)	(3)	(4)
	Chooses occupation j			
Father is in occupation j	1.577***	1.579***		
	(0.0150)	(0.0150)		
Father is in occupation j x Son	2.087***	2.087***		
	(0.0252)	(0.0252)		
Mother is in occupation j			0.876***	0.878***
			(0.0110)	(0.0110)
Mother is in occupation j x Daughter			3.075***	3.074***
			(0.0449)	(0.0449)
Controls		X		X
Observations	695,947	695,947	734,018	734,018

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard error in parentheses. Coefficients are odds ratios. The model is a logit with multiple options, run as a binary one on a stacked sample, as illustrated in Equation 1.1. Column (2) and (4) add controls for gender, cohort fixed effects, study level fixed effects (WO, HBO and MBO) and average income for both parents (calculated as described in Section 1.2).

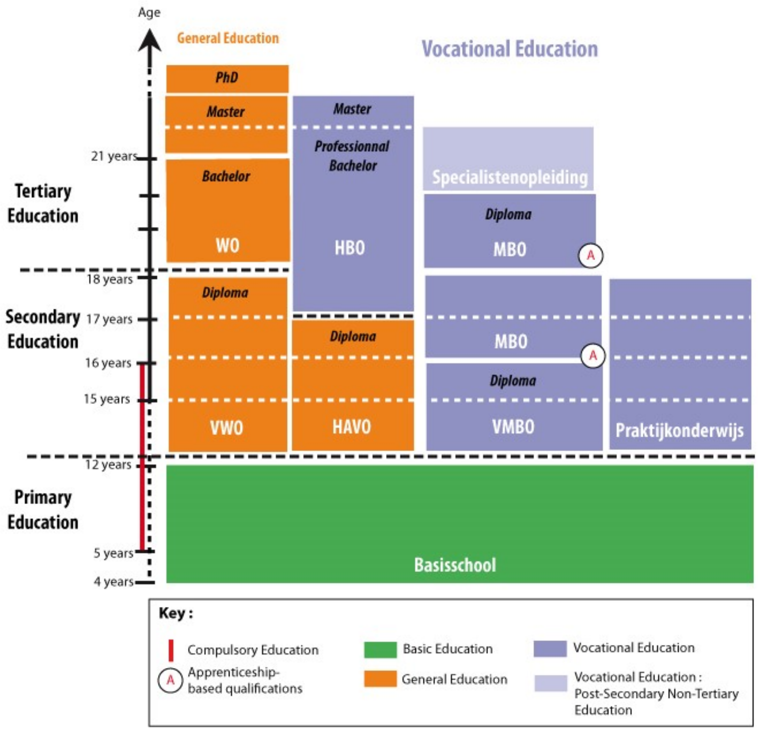
Table 1.4: Persistence and gender segregated fields

	(1)	(2)	(3)	(4)
	Chooses occupation j			
	Women above median	Women above 75th pct	Men above median	Men above 75th pct
Father is in occupation j	1.797***	2.719***		
	(0.0279)	(0.0588)		
Father is in occupation j x Son	1.408***	0.998***		
	(0.0276)	(0.0281)		
Mother is in occupation j			1.613***	1.853***
			(0.301)	(0.0417)
Mother is in occupation j x Daughter			1.270***	0.874***
			(0.0309)	(0.261)
Controls		X		X
Observations	201,736	101,251	193,883	95,954

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard error in parentheses. Coefficients are odds ratios. The model is a logit with multiple options, run as a binary one on a stacked sample, as illustrated in Equation 1.1. Column (2) and (4) add controls for gender, cohort fixed effects, study level fixed effects (WO, HBO and MBO) and average income for both parents (calculated as described in Section 1.2).

1.8 Appendix Figures

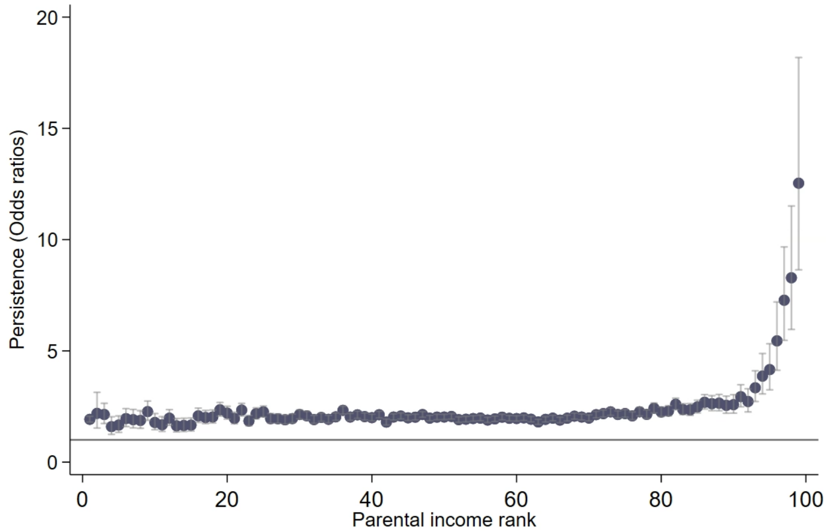
Figure 1.6: Dutch education system



Source: European Observatory of Learnings-Service in Higher Education.

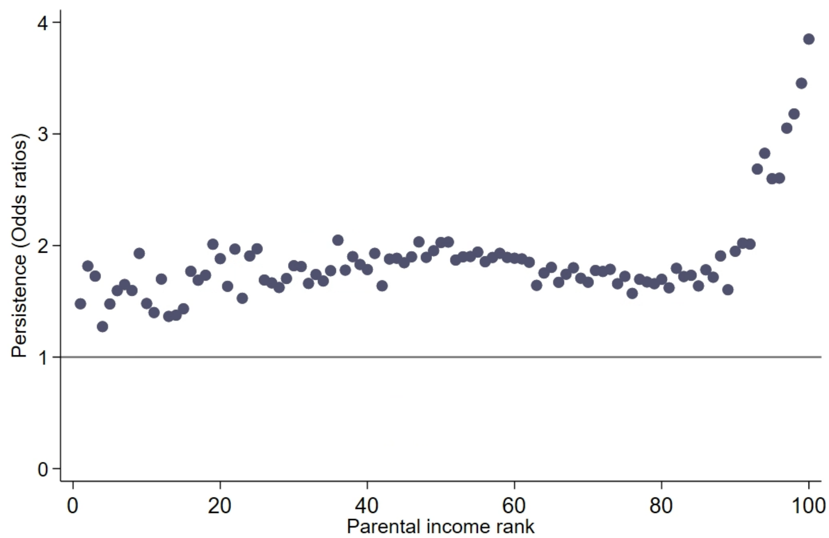
Notes: The figure graphically summarizes the structure of the Dutch education system, with the age at which the pupils enter each stage. After primary education, children can enter one of the following blocks and proceed accordingly. Changes of blocks are possible, but require obtaining additional prior education.

Figure 1.7: Persistence and family economic background, with confidence intervals



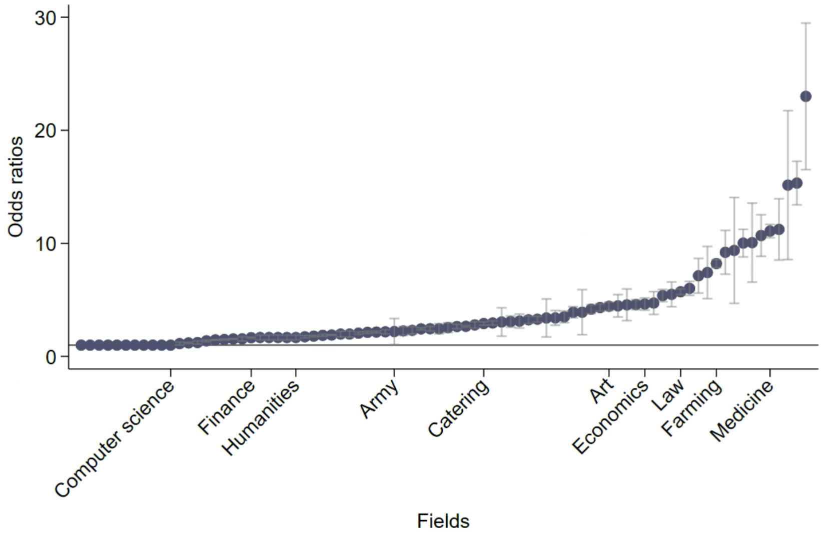
Notes: Persistence is measured as the odds ratios from the logistic model described in Equation 1.1. The specification is augmented to include interactions with parental income rank. The latter is calculated as the average between maternal and paternal income percentile rank. Parents' income is measured as described in Section 1.2. The black line corresponds to odds ratios equal to one, where an individual is as likely to choose an occupation when having a parent in it as when no parent shares the same occupation. Spikes show 95% confidence intervals.

Figure 1.8: Persistence and family economic background (Robustness)



Notes: Persistence is measured as the odds ratios from the logistic model described in Equation 1.1. The specification is augmented to include interactions with parental income rank. The latter is calculated as the average between maternal and paternal income percentile rank. Parents' income is measured as described in Section 1.2. The black line corresponds to odds ratios equal to one, where an individual is as likely to choose an occupation when having a parent in it as when no parent shares the same occupation. The model controls for the size (as share) of parental occupations within the considered income rank.

Figure 1.9: Persistence and occupation, with confidence intervals



Notes: Persistence is measured as the odds ratios from the logistic model described in Equation 1.1. The specification is augmented to include interactions with ISCED-F 3 digit fields. Statistically insignifcant estimates have been set to be equal to 1 for readability. The black line corresponds to odds ratios equal to one, where an individual is as likely to choose an occupation when having a parent in it as when no parent shares the same occupation. Spikes show 95% confidence intervals.

1.9 Appendix Tables

Table 1.5: Field coding

Code	Field	Code	Field
10	Generic and basic education	713	Electrical and energy technology
30	Personal skills	714	Electronics and industrial automation
100	Educational science	715	Mechanical engineering and metal working
200	Art	716	Vehicle, ship and aircraft engineering
211	Audiovisual engineering and media production	719	Other technology and technical services
212	Fashion, interior and industrial design	720	Food processing
215	Music and theatre	722	Woodworking, paper, plastic processing and ceramics
220	Humanities	723	Textile, clothing, footwear and leather manifacture
221	Theology and philosophy	724	Mineral extraction
222	History and archeology	731	Architecture and urban planning
231	Foreign languages	732	Construction and civil engineering
232	Native language, literature and linguistics	810	Crop and livestock production
300	Journalism, behaviour and society	812	Horticulture
311	Economics and econometrics	818	Interdisciplinary agriculture courses
312	Political and social sciences	820	Forestry and fishing
313	Psychology	841	Veterinary medicine and care
314	Sociology and cultural sciences	910	Healthcare
322	Library	911	Dentistry
410	Business administration	912	Medicine
412	Financial services	913	Nursing and obstetrics
413	Management business and personnel science	914	Medical diagnostics and technology
414	Marketing and public relations	915	Therapy and rehabilitation
415	Secretarial and administrative support	916	Pharmacy
416	Wholesale and retail	919	Other healthcare
417	Working skills	920	Wellbeing
420	Law	921	Care for disabled adults, elderly and family
541	Mathematics	922	Youth pedagogical work and childcare
511	Biology	923	Social work and career choice work
512	Biochemistry	1010	Personal services
521	Environment	1011	Home economics, facility services and cleaning
531	Chemistry	1012	Beauty and haircare
532	Earth sciences	1013	Catering
533	Physics	1014	Sport
542	Statistics	1015	Tourism and leisure
600	Computer sciences	1020	Hygiene and working condition
611	Computer use	1021	Public cleaning, water management and distribution
612	Database and network design and management	1022	Work safety and ergonomics
613	Software development and system analysis	1030	Public safety
619	Other computer science	1031	Armed forces
710	Technology and technical services	1032	Public order and safety
711	Chemical engineering and process engineering	1040	Transport and logistics
712	Environmental protection and technology		

Notes: Codes follow the ISCED-F 2013 system and courses are categorised by CBS. Numbering may slightly differ as a result of this assignment process. Moreover, a few similar fields have been subsequently aggregated during the data cleaning to avoid categories with too few observations.

Table 1.6: Matches with parental education

	Matched parental education data		
	(1)	(2)	(3)
Born after 1985	0.1983***	0.1968***	0.1956 ***
	(0.0006)	(0.0006)	(0.0006)
	[0.196]	[0.194]	[0.193]
Male	-0.0035**	-0.0035**	-0.0002
	(0.0006)	(0.0006)	(0.0007)
	[-0.0035]	[-0.0035]	[-0.0002]
Dutch nationality	-0.1015***	-0.1035***	-0.1031***
	(0.0013)	(0.0013)	(0.0013)
	[-0.0453]	[-0.0462]	[-0.0461]
Parental income		0.0000	0.0000
		(0.0000)	(0.0000)
		[0.0087]	[0.0094]
Fields FE:			
Generic programmes and qualification			0.0022
			(0.0071)
			[0.0002]
Education			0.0056**
			(0.0012)
			[0.0031]
Arts and Humanities			0.0200***
			(0.0015)
			[0.0088]
Social Sciences, Journalism and Information			0.0043***
			(0.0.0013)
			[0.0021]
Natural Sciences, Mathematics and Statistics			-0.0196***
			(0.0022)
			[-0.0056]
Information and Communication Technologies			0.0257***
			(0.0018)
			[0.0093]
Engineering, Manifacturing and Construction			-0.0022**
			(0.0010)
			[-0.0016]
Agriculture, Forestry, Fisheries and Veterinary			-0.00258***
			(0.0020)
			[-0.0081]
Health and Welfare			0.0157***
			(0.0009)
			[0.0123]
Services			0.0298***
			(0.0010)
			[0.0203]
Constant	0.5900***	0.5871***	0.5772***
	(0.0013)	(0.0013)	(0.0014)
Observations	2,605,439	2,605,439	2,605,439
R-Squared	0.040	0.040	0.041

Notes: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$. Standard errors in parentheses. Beta coefficients in squared brackets. The dependent variable is a dummy that equals 1 if at least one parent has a matched education level. Fields are coded according to the 1-digit ISCED-F 2013 system.

Chapter 2

Following in the Family Footsteps: Returns from Occupational Persistence

2.1 Introduction

Children often follow in their parents' footsteps, sometimes by choosing the same occupations. From medieval artisans and craftsmen (Clark, 2015) to modern political dynasties (Dal Bó et al., 2009), occupational persistence has characterized both ancient and modern societies (Sorokin, 1927). Even though researchers have long been aware of and often documented the intergenerational transmission of occupations, only a handful of studies have focused on its economic consequences. While we may expect it to affect individuals' returns from their professions and, accordingly, inequalities overall, the direction of these effects is unclear ex ante. Intergenerational persistence in occupation may, in principle, boost children's performance through transfers of skills, information, and professional networks, as well as impair it through a worse job match. In particular, the former types of mechanisms can be thought of as a treatment effect of occupational transmission, while the latter refers to the type of selection associated with it.

Indeed, individuals typically choose occupations based on ability or other unobservable characteristics before actually experiencing gains from such choices. Consequently, one's performance on the labor market may be attributed to both the occupation they are in and to additional idiosyncratic characteristics, such as pre-existing skills or motivation. The same selection issue exists in the choice of occupational persistence. Individuals choosing the same profession as their parents may do so with the expectation of particular benefits from this intergenerational match, such as parental support in the job search. It follows that, in the absence of a way to identify selection bias, any attempt to study the effects of occupational persistence will be unable to disentangle the two. The issue is relevant because measuring the impact of occupational transmission on individuals is a necessary ingredient, together with its prevalence in different population segments, to quantify its repercussions on inequalities.

This paper investigates the effects of intergenerational occupational persistence on the labor market outcomes of dynastic individuals, i.e., those with a parent in the same occupation, and how this is related to social mobility. I do so by using a rich set of administrative data from the Netherlands and a unique policy experiment that provides exogenous variation in occupational persistence for a field of high value for society, namely medicine. I organize the analysis into two main parts. In the first part of the paper, I draw on some of the descriptive facts discussed in Chapter 1 to document intergenerational occupational transmission. For the purpose of this exercise, my proxy for occupation is the field of the highest achieved level of education. This measure includes 82 fields, and it has the advantage of being more stable over an individual's life cycle than standard occupational titles. In particular for the purpose of this work, it is relevant to remember that the rate of transmission increases sharply above the fourth quartile, with children from families at the very top of the distribution (99th percentile) approximately twelve times as likely to pick a given field when a parent has done the same. This fact introduces the relevance of measuring returns from occupational persistence, as the existence of gains or losses from reproducing parental fields together with the differential prevalence by economic background will determine whether this increases or hampers social mobility.

Therefore, I start by considering the OLS estimated returns associated with the transmission of occupations, that amount, on average, to an increase of 2.8% in income. The results also indicate wide heterogeneity across fields but do not vary significantly with parental income. Because individuals select into occupations and, potentially, into choosing the same ones as their parents', these naive returns will encompass both the causal effects of having a parent in the same occupation and the selection bias associated with this choice. I then compare estimates of intergenerational correlation in income for the whole population and the sample excluding children who have followed their parents in the same occupation. The difference in estimates reveals that occupational persistence accounts for 21% of the rank-rank correlation observed in the top decile of the parental income distribution.

The second part of the paper aims to decompose the estimated returns into the two elements highlighted above. In order to do so, I focus on the medical profession. This is a particular relevant case for several reasons. First, medical doctors are part of a sector that is highly valued by every society and in which policy makers may be especially interested given the high share of GDP spent in Healthcare[1]. Second, they are more likely than average to follow in their parents occupational footsteps. This is true in my sample, where doctors' children are more than ten times as likely to choose Medicine than the children of other professionals, and it has been confirmed by Polyakova et al. (2020) and Altmejd (2023) in Sweden and by Friedman and Laurison (2020) in the United Kingdom. Finally, previous literature has highlighted the the importance of socioeconomic diversity in Medicine for patients' outcomes (Alsan et al., 2019; Greenwood et al., 2018), which could instead be hindered by occupational persistence.

I exploit a policy experiment in which lottery admission to medical school provides random assignment into occupational persistence. In particular, I use data on applicants and their outcomes in the period 1988-1999. In this context, while all candidates by definition select into Medicine, only those with a parent in the medical profession are also selecting into occupational persistence. Focusing on the latter group and comparing individuals marginally admitted to medical school (and thus, allowed to be persistent) to those randomly rejected, I obtain the causal effect of

[1]Health expenditure as a share of GDP was 8.8% in the OECD38 in 2019 (OECD, 2021b).

becoming a doctor for children of doctors. Clearly, this estimate alone is not enough to isolate the effect of occupational transmission, as it may only capture returns from medical school, rather than from persistence. In particular, this will happen if the returns from medical school are the same for dynastic and non dynastic doctors. Following this intuition, I can contrast the obtained estimate with the same measure for the children of non-doctors and retrieve the difference in treatment effects between the two groups. This can be interpreted as the change in returns from the medical profession due to persistence.

I first focus on labour income. I find that occupational persistence increases income of doctors by 23%. This corresponds to 54% higher returns from the medical profession, compared to the baseline estimate for non dynastic doctors. Using these estimates to calculate a counterfactual income distribution where returns from occupational persistence are canceled shows that these gains account for 11% of the total intergenerational correlation in income observed in the top 10% of family income distribution. This corresponds to about half of the increase in stickiness attributable to occupational persistence.

The comparison of the naive and causal returns estimated in the paper suggests the existence of considerable negative selection into occupational persistence, that biases the OLS estimates towards zero. Using the lottery context, I am able to formalise this idea and provide a measure of the selection bias. The estimation confirms that this is large and negative, such that not only it can on aggregate counteract the effect of persistence, but it also makes the observed OLS returns from occupational transmission for Medicine negative. This indicates that individuals choosing to follow their parents in the the medical profession would perform on average worse when they cannot become doctors than those who do not select to follow their doctor parents in the first place.

I then explore additional margins and find that, although aggregate gains are driven by labour income from own business, doctors in salaried employment also benefit from occupational persistence, in the form of higher hourly earnings and less hours worked. Moreover, the difference in treatment effects also differ by high school GPA and gender, with medium GPA students and women gaining more from occupational persistence in medicine. Considering outcomes in different time frames also reveals that, while gains take some years to accrue, they appear stable in the long run. Finally, I provide some suggestive evidence on the mechanisms that may be at work in improving outcomes of dynastic doctors. This points against a role of genetics while it hints towards benefits that individuals may obtain through working for a family practice or for the same employer as their parents, which highlights the importance of connections and professional network. The descriptive analysis also excludes that different outside options faced by children of doctors and non doctors drive the observed gains, as the former move to similar, and if anything better, alternative fields when they cannot enroll in medicine. Regarding my results on negative selection, I document how children of doctors are more likely to specialise towards the same profession from early stages of education, which may then lead them to perform poorly when in different occupations. Given the higher likelihood of persistent medicine students to drop out from their studies, I also argue that external forces pushing these students into medicine, even in lack of the adequate skills or motivation, may produce a mismatch between desired and realised occupation, making them generally less fit for the field they choose to undertake.

While focused on the medical profession, the findings highlight how, for one of the occupations most represented at the top of the income distribution, occupational persistence may hamper equality of opportunities by giving some children an advantage that is uncorrelated to their own abilities. If generalised to more occupational fields, the results of this paper would bring no good news for social mobility. Occupational persistence's large effect on income, paired with its prevalence among high income families, identifies this type of transmission as one more channel through which children of rich parents remain rich. The argument strengthens with the negative selection. As a matter of fact, individuals benefiting from medical dynasties are the ones that, by otherwise performing worse than their peers in other fields, could have contributed to a decrease in income inequality.

Related literature. This paper contributes to several lines of research. First, it integrates intergenerational occupational persistence into the literature on social mobility and inequality of opportunities. Mobility in income is taken as the main measure and driver by several papers in economics, such as Becker and Tomes (1979), Björklund and Jäntti (1997), Solon (1992, 2002), Blanden et al. (2005), Chetty et al. (2014) and Olivetti and Paserman (2015). While occupations have been used in this literature (Bell, Blundell, et al., 2023; Long and Ferrie, 2013), they usually proxy for individuals' socioeconomic status, especially when data on income is not available, rather than per se as a potential channel for inequalities. Given the interaction of treatment and selection effects in the choice of occupation and occupational persistence, descriptively controlling for it (Haeck and Laliberté, 2023) cannot be enough to identify its impact on social mobility. By focusing on one high-paying occupation, I am able to isolate the contribution of persistence to the incomes of some of the wealthiest individuals in the labour market. Paired with the higher incidence of occupational transmission at the top of the income distribution, this approach, although partial in its coverage of the population, allows to extract more information than when working with average conditional correlations.

More specifically, this work deepens our understanding of intergenerational persistence in occupations and fields of study as an economic relevant phenomenon. Although these mechanisms have been documented before both in sociology (Buscha and Sturgis, 2018; Erikson et al., 1979; Goldthorpe and Erikson, 1992; Sorokin, 1927) and economics (Aina and Nicoletti, 2018; Altmejd, 2023; Dribe and Helgertz, 2016; Long and Ferrie, 2018; Mocetti et al., 2022; Sinha, 2016), I contribute to this descriptive literature with evidence of an additional margin of heterogeneity in transmission rates, namely by economic background. Furthermore, and despite the recent attention this topic has also received in the policy sphere (OECD 2022; SMC 2022), we still lack convincing evidence on the effects occupational persistence may produce. In fact, the studies who have moved in this direction mostly present descriptive estimates of the returns associated to it (Altmejd, 2023; Lo Bello and Morchio, 2022). Two exceptions using quasi-experimental variation from an Italian policy change are Basso et al. (2021) and Raitano and Vona (2021). However, their works focus on how the estimated (OLS) returns change when entry in an occupation is made more competitive, rather than causally estimating them in the first place. Moreover, their analysis is limited to the case of lawyers, and their measure of "connections" in the profession reflects a larger set of individuals inside and outside the parental sphere[2]. Here, I focus on a different profession, of high value for society, and embed the estimation in the broader context described by the stylized facts in the first part of the paper. I use experimental variation to causally identify the returns from medical school for both dynastic and non-dynastic doctors, and compare the two to determine the additional gains from occupational transmission. The lottery setting allows me to convincingly deal with the issue of selection into the field of Medicine and into occupational persistence, as well as to measure the selection bias associated to the latter.

Finally, my research speaks to the strand of literature studying differential payoffs from the choice of occupations and field of study (Altonji, Arcidiacono, et al., 2016; Altonji, Blom, et al., 2012; Kirkeboen et al., 2016). In particular, I use the same experimental context as Ketel et al. (2016), which found substantial income premium for doctors, to explore an additional factor affecting returns from specific occupation, i.e. parental transmission. In the case considered in this paper, the payoff from the medical profession substantially increases for dynastic doctors, suggesting that occupational persistence, with its different incidence across fields, appears to be a relevant factor in the determination of income differentials.

The rest of the paper is organized as follows. Section 2.2 describes the institutional context of Dutch education, medical school and profession. Section 2.3 provides details on the administrative records and lottery applicants data used in the analysis. Section 2.4 presents the descriptive results on occupational persistence. Section 2.5 illustrates the empirical challenge in estimating returns from occupational following and describes the proposed identification strategy. Section 2.6 displays the findings both for the heterogeneous treatment effects from persistence and selection bias. Section 2.7 provides a series of extensions to the main results. Section 2.8 proposes

[2] "Connections" in the profession are measured as lawyers with the same last name in Basso et al. (2021) and parents or close relatives working as lawyers in Raitano and Vona (2021).

and tests some potential mechanisms. Section 2.9 concludes and highlights directions for future research.

2.2 Institutional context

Occupations often change over individuals' lifecycle, and studying occupational choice can require researchers to measure it at a particular point in time which may not necessarily be representative of the main field in which they have worked. The Dutch context represents a particular suitable laboratory for this type of questions. The first part of the paper will proxy for occupational choice using the field of the highest achieved education, when this is above the high school level. Even though it may differ from realized occupation, I consider this a good approximation as it is more stable than occupational titles across an individual lifetime, and it reflects occupational aspirations as measured early in the choice process. Moreover, the strong emphasis placed on vocational and professional education and the many options offered in this sector[3] by the Dutch system produces, in many cases, a tight connection between education and field of occupational choice. The rest of this section introduces the context of Dutch medical schools and their admission system, and gives a few information on the work and conditions of medical doctors in the country.

2.2.1 Medical school and lottery admission

Medicine programs are taught in research universities and require students to have graduated from a VWO school with profiles in "Nature and health" or "Nature and Technology". In practice, this corresponds to have taken exams at VWO level in mathematics, physics, chemistry and biology. There are currently eight medical schools in the Netherlands, each linked to an academic hospital.

Contrary to most university and vocational courses, Medicine is a *numerus fixus* program, i.e. a limited number of places are available to be filled every year. The rule

[3]Junior college tracks (MBO) include very specific programs, such as bakers, cleaners and bicycle technicians training.

was introduced for the first time in 1972 to face the growing number of applicants. Starting in 1976 the system in place was a lottery weighted by seven high school GPA groups, such that applicants with higher GPAs would be more likely to be admitted. In 1999 a threshold for unconditional acceptance for individuals with GPA above 8 was introduced. From 2000, the lottery system was gradually combined with a decentralised selection mechanism, in which individual universities could set specific criteria, other than high school GPA, to grant admission, until the lottery was completely dismissed in 2017. For this reason, the sample used for this paper will only consider applicants until 1999, when the lottery was the only admission channel. Figure 2.1 shows the average empirical probability of being admitted by GPA category, between 1988 and 1999, the period considered for the analysis, while Figure 2.12 in the appendix Section 2.12 displays the shares of applicants in each GPA group, by whether they have a parent who is also a doctor. The distributions are similar, with children of doctors being slightly more likely to appear in the highest GPA categories. Figures 2.13 and 2.14 in the same section present the probability of applying multiple times and the number of applications, conditional on re-applying, respectively, both for children of doctors and non doctors. Interestingly, while children of doctors are not more likely to re-apply after a rejection, when they do they submit a greater number of applications.

The number of available places is established every year by the government. This was fixed at 1,458 until 1993 and then gradually expanded over time. The share of applicants that was admitted has been on average 47% between 1988 and 1999. Applicants rejected at a given round can successively re-apply, but the number of applications to Medicine is limited to three from 1999.

Students apply through a centralised system, rather to schools directly, and can express up to three preferences. Each preferences is a combination of program and university (e.g., Medicine at the University of Amsterdam) and only one *numerus fixus* program can be indicated at each application round.

The standard duration of the foundational medicine program is six years, with three years of theoretical studies followed by an equal period of practical training.

Statutory fees are currently around 2,000£ per year[4]. After the first medical degree graduates can proceed to a specialty training of variable length. They can also choose to become a GP or another type of non specialized professional following one of the so called "public health" tracks, such as occupational medicine, youth care and environmental medicine. These options still require additional qualifications, usually up to three years. Alternatively, they can remain *basisarts* ("junior doctors"), i.e. professional figures with no further training who often work as consultants. In practice, due to the limited number of specialty spots, recent years have seen many young medicine graduates either acquiring clinical experience as basisarts or pursuing a PhD before moving to further training.

When unable to attend medical school in the Netherlands, it is common for students to enroll for this program abroad. Medicine graduate can then obtain recognition of their foreign qualification and register as doctors in the Netherlands. Table 2.8 in the appendix Section 2.13 shows that dynastic doctors are approximately 40% less likely than other doctors to have obtained a degree abroad. Somewhat surprisingly, this is true even when unconditionally comparing the two groups.

2.2.2 Medical doctors, the labour market and healthcare

In 2015 the Netherlands had on average 3.7 doctors per 1,000 inhabitants, slightly below the EU average of 3.9 (OECD et al., 2021). General practitioners are usually self-employed and manage their own practice, either individually or in collective organizations with other GPs. In a small amount of cases they may be employed by a practice. Specialists can work as salaried employee or self-employed, and the two work statuses co-exist even in hospitals. Self-employed specialists often work through partnerships of doctors ("maatschappen"), which have agreements with hospitals regarding the services they offer and their fees. In 2021, approximately 43% of medical specialists were working as self-employed (Statistics Netherlands, 2023). While doctors on hospitals' payroll are paid a flat salary, self-employed in maatschappen are paid a fee depending on the service offered, and as previously

[4]Higher "institutional fees" apply instead to students that do not meet the requirements for statutory fees, i.e. that are not EEA citizens or that have already obtained a Dutch degree.

negotiated with the hospital. An analysis of six European countries (Belgium, Denmark, England, France, Germany and the Netherlands; SEO (2012)), finds that Dutch self-employed doctors earned, on average, 1.8 times the average gross income of a (primarily) salaried doctor in 2009, making them the highest paid self-employed physicians among these countries. Outside hospitals, doctors can also work as regular employees for private companies or freelance contractors.

The Netherlands have a semi-public healthcare system. Most hospitals and health insurance providers are private, but the sector is heavily subsidised. Insurers are non-profit organizations and any profit should, by law, concur to lower the premiums. There is a mandate for universal healthcare, with compulsory basic insurance. Citizens can choose the provider and have the option to purchase additional coverage from the same.

2.3 Data

The main data used for the analysis includes comprehensive administrative records provided by the Dutch Bureau of Statistics (CBS). These encompass several datasets, which can be linked at the individual level using a unique person identifier. To these, I link data on medical school applicants. This section presents short list of the datasets used and of how the analysis samples are constructed.

2.3.1 Administrative records and lottery data

Demographics and family connections These include information on gender, date and country of birth, nationality and marital status as well as the link with both parents. They are available for all Dutch residents from 1994.

Education Records of the highest attained education, with detailed program, field and level are utilized for the descriptive facts in the first part of the paper. Fields are coded according to the International Standard Classification of Education for fields of education and training (ISCED-F 2013), a standard utilised internationally which includes over 80 detailed categories. Table 2.9 in the appendix of Section 2.13 reports a list of the fields with the respective codes, as used in this paper. This information is systematically available for the universe of students since 1999 and on a sample basis for those having completed education before then. The coverage is very high for recent cohorts and included almost 11 million people (65% of the population) in 2015, with increases every year. Nevertheless, parental generations will naturally display lower coverage, which will in turn affect the sample to which the descriptive analysis can be restricted to. Table 2.10 in the appendix of Section 2.13 presents a few correlations that describe how the individuals whose at least one parent can be matched to the education records differ from the rest of the population. The correlations are quite small in magnitudes, as highlighted by the standardizes beta coefficients.

The governmental agency for the execution of education policy, known as DUO (*Dienst Uitvoering Onderwijs*) conducts admission lotteries and registers enrollments of students in higher education. I observe applications to medical school, as well as lottery outcomes, and high school GPA category, for all students who applied in 1987-2013.

Labour market outcomes and structure Information on individual (primary, gross and net) income is collected from tax records and available within the CBS since 2006. The measure of income used in the analysis will be primary income, i.e. gross labour income from salaried employment as well as own business. The same dataset also includes employment status and sector of activity. As several observations are available for each individual, I use the average between 30 and 40 years old for children and between 50 and 60 years old for parents.

Tax data on the universe of salaried employees from 2003, with information wages and monthly hours, as well as type of contract (part-time, temporary, etc) is also included. The information available in this dataset is also measured as an average between 30 and 40 years old.

I also have access to information on employers and firms owners, which are linkable to employee records. I use this data to explore the role of family connections in the workplace.

Registry of health professions The BIG register (*Beroepen in de Individuele*

Gezondheidszorg, "Professions in the individual healthcare") lists all officially recorded medical professionals, including doctors, with date of first registration and medical specialty. Registrations are mandatory in order to practice as a health professional. The registry was established in 1995 but also contains data for all doctors who had qualified before then.

2.3.2 Final samples construction

The final sample for the descriptive analysis includes individuals born between 1963 and 1991 for which a parental link exists and for whom the field of the highest level of education is available. Moreover, I further restrict it to individuals with income and earnings records in the considered time frame as well as income information for at least one parent. This leaves me with 962,437 observations. Of these, 937,426 display positive income. Table 2.1 reports summary statistics for a few demographic variables. The sample appears in line with the general population[5] for share of women and education levels, but individuals born in the Netherlands are overrepresented compared to foreign born. This is not surprising as the parents of those belonging to the latter group are likely not to be Dutch residents. They would therefore not appear in the administrative records and can not be linked for the purpose of intergenerational study.

For the second part of the paper, I construct an instrumental variable by only considering first time of applications, as applicants who did not make in the first round are allowed to reapply. For the same reason, all candidates in 1987 are excluded as it is not possible to observe previous applications[6]. In the interest of robustness, and following Ketel et al. (2016), I only include candidates that are no older than 20 years old at the time of application. As decentralised selection was gradually introduced from 2000, I also limit the analysis to applicants up to 1999. Given that in the first periods of the data the first two categories were admitted with certainty (see Figure 2.11), I exclude from the sample individuals with GPA

[5]As of 2022 official statistics.

[6]As repeated applications usually happen in consecutive years, we can assume individuals appearing in 1988 and afterwards are first time applicants if they are not observed to apply in 1987.

above 8.5 until 1992, and those with GPA between 8 and 8.5 until 1990. The same two categories are also omitted for year 1999 as they were, by law, automatically granted admission from that year. Table 2.2 reports a few summary statistics for the full sample of applications between 1988 and 1999. The applicants data can be linked to the administrative records described in Section 2.3.1, so that the final sample is constructed using all the applicants in the group above for which a match with education enrollment, labour market outcomes and parental links was found. Applicants' outcomes on the labour market are measured as an average between 15 and 20 years from the first application. Finally, as the analysis will require to control for a number of parental characteristics, I need to consider the applicants for which these are available. Once I restrict it to include matches with parental income, the final sample counts 10,419 observations.

2.4 Descriptive evidence

Chapter I in this thesis has established that: i) children tend to follow their parents at a rate that is over and above what we would expect from accidental matches and ii) the persistence rate increases exponentially for the top decile of the parental income distribution. The last aspect is particularly relevant when discussing intergenerational mobility as it uncovers strong distributional differences that may translate into stickier incomes at the top of the distribution. To the extent to which every occupational choice places individuals in a particular segment of the income ladder, higher persistence at the top may naturally convert in a larger tendency of rich parents to have rich children. This would be further exacerbated in case of a positive correlation between occupational transmission and own income, which motivates the interest in the existence of returns from persistence and its correlation with intergenerational income mobility.

2.4.1 Returns from occupational persistence

The analysis below presents a first attempt to uncover these potential gains. To this purpose, the model employed takes the form of the following loglinear regression

$$log(Y_i) = \alpha + \gamma p_i + \mathbf{X}'_i\delta + \epsilon \tag{2.1}$$

with p_i the dummy taking value of 1 when the individual has chosen the same field as at least one parent and $\mathbf{X}_i$ representing a vector of individual controls. The latter always includes occupation fixed effects, so that the changes in income can be interpreted as within occupation returns. Figure 2.2 displays the results of the specification with the iterative addition of control variables. Appendix Table 2.12 in Section 2.13 reports the corresponding estimates for clarity. Returns are positive and significant, and the full set of controls accounts for two thirds of the observed within occupation returns. The most conservative coefficient (Column (4)) indicates a 2.8% average increase in primary income linked to occupational persistence.

Given the wide heterogeneity in persistence rates along the income distribution uncovered in Chapter I, I investigate whether this is reflected by variation in returns in the same dimension. Figure 2.3 addresses this question by showing the relationship between estimated returns and parental income rank. No clear pattern is discernible in the plot, with most coefficients gravitating around the average return rate. This suggests that observed gains may be orthogonal to parental placement within the occupation. Consequently, as long as income rank may be interpreted as indicative of skills, it could imply that the role of field specific skills transmission is relatively small.

In contrast to the homogeneity of returns along the income distribution, Figure 2.4 displays large variation by field. Interestingly, changes in income associate to persistence in occupational choice are not strongly correlated to their persistence rate. Panel (a) in Figure 2.5 confirms it: a 1 percentage point increase in the descriptive returns from occupational persistence is associated with only 0.7% increase in odds ratios. Higher income in an occupation is also weakly correlated to the estimated changes in income, with a 10% increase in average income corresponding to a change in returns of 1.5 percentage points (Panel (b)).

Overall, the aggregate correlational gains associated with occupational persistence appear fairly limited in size. Moreover, no clear pattern emerges with respect to what drives differences in them. Section 2.5 will shortly describe the technical and conceptual reasons why a correlational analysis cannot uncover the real effect of occupational persistence and propose an empirical approach to disentangle it from selection bias.

2.4.2 Intergenerational income mobility and occupational persistence

It may be reasonable to think that, to the extent to which occupations are tied to positions in the income distribution, their family transmission may play a role in determining intergenerational income correlations. Whether and how this happens will depend both on the incidence, i.e. its prevalence across different population segments, and the returns of occupational persistence. Chapter 1 and Section 2.4.1 have shown these two elements separately in their correlational form. This section considers the contribution of occupational persistence to intergenerational income mobility.

Table 2.3 displays overall rank-rank correlations for the sample of non persistent individuals only, as well as the full one. The difference, although significant, shows that 0.6% of the total correlation is attributable to occupational persistence. The modesty of this contribution is easily explained by the overall incidence of occupational persistence in the population. As a matter of fact, and despite the relevance played by parental occupation in own occupational choice, the share of persistent individuals, at the considered level of granularity, is possibly too small to make a difference on average. Consequently, we would expect occupational persistence to play a more visible role where its size is greater, i.e. at the top of the income distribution. Table 2.4 makes this point by reporting the rank-rank correlation estimates for the top decile of the parental income distribution. Two main facts emerge from it.

First, occupational dynasties now account for 21% of overall correlation in income. Secondly, without persistent individuals in the sample, the correlation coefficient is only 1.5 percentage points above the one presented in Table 2.3, possibly singling out occupational persistence as a crucial factor in decreasing mobility at the top.

2.5 Identification strategy

To uncover the potential benefits of occupational inheritance, an ideal, though impractical, experiment involving the random assignment of occupations to offspring across all parental vocations would allow to obtain the average impact of occupational persistence. In practice, the choice of field and, as a consequence, of whether to be persistent in occupation, is undoubtedly endogenous. This section briefly explains the double selection issue affecting occupational persistence, the proposed identification strategy and the assumptions necessary to implement it.

2.5.1 Empirical challenge

When deciding to follow in parental footsteps, individuals are in practice making an occupational choice in the first place. As a matter of fact, different occupations, or fields of study, can cause significant variation in one's payoff on the labour market (Altonji, Arcidiacono, et al., 2016; Altonji, Blom, et al., 2012; Kirkeboen et al., 2016), regardless of whether there is also a match with the career choices of parents. This characteristic of occupational persistence, naturally brings with it a double selection issue. To exemplify it, let us take the example an average individual in the medical profession. Individuals select into becoming doctors based on unobservable characteristics, such as ability or intrinsic motivation. This self-selection will also be a factor affecting observed outcomes for doctors, and will make simple differences in outcomes a biased estimate for the treatment effect of that occupation (Angrist and Pischke, 2009). Now, let us consider a doctor who also has a doctor parent. Their choice will not only driven by the same type of unobservable elements as the average doctor, but may also also incorporate factors deriving from the occupational match, such as inherited preferences or expected benefits in the field. Consequently, observed outcomes of dynastic doctors will also reflect this additional margin.

This paper focuses on isolating variations in outcomes in presence of occupational transmission, over and above these standard effects. Disentangling the effects of occupational persistence requires observing who selects into following their parents, as well as how they differ from those selecting into the occupation alone. Moreover, exogenous variation establishing who gets to become persistent in occupation is necessary to estimate the heterogeneity in treatment effects for dynastic individuals.

2.5.2 Empirical solution

The lottery design helps us to deal with the two steps of selection. For all applicants, who by definition select into the medical profession, the random assignment allows to estimate the causal effect of becoming a doctor, conditional on selection. We can do this both for children of doctors and children of non-doctor as, to the extent to which the independence assumption is satisfied (see Section 2.5.3), the exogeneity of assignment will hold within subgroups. The analysis will use an instrument variable (IV) approach, where enrollment in medical school is instrumented by the outcome of the first lottery in which the applicant participates. In particular, the following equations are estimated simultaneously in the second stage on the sample of applicants:

$$y_i = \begin{cases} \alpha_{\mathrm{P}} + \beta_{\mathrm{P}} D_i + \pi_{\mathrm{P}} \mathbf{X_i} + \varepsilon_i & \text{if} \quad P_i = 1 \\ \alpha_{\mathrm{NP}} + \beta_{\mathrm{NP}} D_i + \pi_{\mathrm{NP}} \mathbf{X_i} + \nu_i & \text{if} \quad P_i = 0 \end{cases} \tag{2.2}$$

where D_i equals 1 if an individual enrolls in medical school, P_i is a dummy for at least one of the parents being a medical doctor and $\mathbf{X_i}$ is a vector of individual controls, including the GPA class of the applicant. The latter is necessary as the lottery is weighted on GPA categories, i.e. the assignment will be truly random within these. In the two specifications β_j, for $j = \{P, NP\}$, identifies the average treatment effect on the treated (ATT) of becoming a doctor on outcome y_i for each group, respectively, or

$$\beta_{\mathrm{P}} = \mathbb{E}[Y_{1i}|S_i = 1, P_i = 1] - \mathbb{E}[Y_{0i}|S_i = 1, P_i = 1] = [ATT|P_i = 1]$$
$$\beta_{\mathrm{NP}} = \mathbb{E}[Y_{1i}|S_i = 1, P_i = 0] - \mathbb{E}[Y_{0i}|S_i = 1, P_i = 0] = [ATT|P_i = 0]$$

where S_i denotes whether the individual has applied to medical school.

For the subgroup of children of doctors, these estimates correspond to the ATT of becoming a doctor when having a parent in the same profession. However, they cannot alone identify how occupational persistence affects doctors' outcomes. In fact, in the case in which ATTs are homogeneous for children of doctors and non doctors, the β_P coefficient would carry no additional information with respect to β_{NP}. In other words, the meaningful estimate is the difference between the ATTs for the two groups. If $\beta_P - \beta_{NP} > 0$ dynastic doctors experience higher returns from their profession than non dynastic ones. This double difference design, allows to net out all the common characteristics of those entering medical school (regardless of the parental background) and of having parents in the medical profession (notwithstanding the kids' occupation) and isolate the treatment of interest, i.e. the combination of the two.

In the remainder of the paper, I will refer to the difference in the two coefficients as the effect of persistence. Conservatively speaking, it measures the difference in returns from medical school between those who have a parent who is a doctor and those who do not. It could be argued that while the lottery assignment is, conditional on applying and within each class of GPA, exogenous by design, parental occupation is not. The concern would then be that the additional difference in outcomes observed for dynastic doctors is, in fact, due to some alternative unobservable margin in which they may differ from non dynastic individuals, that is also correlated to having a parent who is a doctor. In the measure in which these characteristics affect outcomes across all occupations, this will be picked up by the constant term α_{P}, and will not affect the additional returns specific to medical schools β_{P}. The latter, will reflect these potential confounding factors only when their impact on outcomes changes when in the medical profession. An instance of this may be genetics. As an example, suppose there exist a characteristic that is more common among doctors, is genetically transmitted and affects the performance of doctors only, such as having a "steady hand". I argue that this type of elements are also relevant in how we think about occupational persistence, and should be included in the effects of occupational persistence. In other words, factors working in this way may represent a channel through which having a parent in the same occupation affects children's outcomes, and its impact should therefore be taken into account when considering the overall returns from occupational persistence.

The same lottery design allows to retrieve measures for the two types of selection that would normally threaten OLS estimates, i.e. into the medical profession and into occupational persistence. Using the relationship obtained by the literature on potential outcomes (Cunningham, 2021), we can measure the selection bias as[7]

$$SB = ADO - ATT$$

where SB represents the selection bias and ADO is the average difference in outcomes obtained by naively comparing dynastic and non dynastic individuals

$$ADO = \mathbb{E}[Y_{i1}|P_i = 1] - \mathbb{E}[Y_{i0}|P_i = 0]$$

Running a OLS model on the specifications in Equation 2.2, using the augmented sample that also includes individuals that did no apply to medical school, i.e. did not select into the medical profession, allows to obtain the ADO estimates

$$\beta_{\text{P, ADO}} = \mathbb{E}[Y_{1i}|P_i = 1] - \mathbb{E}[Y_{0i}|P_i = 1]$$
$$\beta_{\text{NP, ADO}} = \mathbb{E}[Y_{1i}|P_i = 0] - \mathbb{E}[Y_{0i}|P_i = 0]$$

In order for this additional sample group to be comparable to the medical school applicants, I consider all students starting a WO level program in the same years as the lottery candidates. As the type of previous education required to apply to the lottery is the same as the other WO courses, this restriction should ensure that I am

[7]The decomposition is extensively worked out in the appendix in Section 2.14.

comparing applicants to individuals that did not apply to Medicine but could have done so.

In parallel with the estimation of the treatment effects, I am then able to obtain two measures of the selection into medical school for the groups of dynastic and non dynastic doctors

$$\hat{SB}_P = \hat{\beta}_{\text{P, ADO}} - \hat{\beta}_{\text{P}} \tag{2.3}$$

$$\hat{SB}_{NP} = \hat{\beta}_{\text{NP, ADO}} - \hat{\beta}_{\text{NP}} \tag{2.4}$$

In a similar fashion, we can interpret the difference between the two selection effects $\hat{SB}_P - \hat{SB}_{NP}$ as the selection bias into occupational persistence. This is the additional selection bias for children of doctors, over and above the "standard" selection bias into medical school.

As the estimates obtained with this model are local average treatment effects (LATE), the appendix in Section 2.14 shortly discusses how they would differ from an ATT, as well as how to interpret the selection bias.

2.5.3 Identification assumptions

For the LATE to be properly identified, the four assumptions of relevance, monotonicity, independence and exclusion need to be satisfied. Table 2.15 shows the results from the first stage. Following Ketel et al. (2016) I start by considering three potential independent variables for the second stage, i.e. enrolling in medical school, completing it and officially registering as a doctor. All of these show strong correlation with the instrument, making it relevant. For the rest of the paper, the variable "Enrolled in medical school" will be the one used in the second stage analysis. This allows to examine educational outcomes (i.e. prior to completing the program or registering as a doctor) as well as the labour market ones. I consider the monotonicity requirement to be fulfilled based on two facts. First, the share of students who won the lottery but did not enroll, is very small (4.87%), which points against a systematic role of the lottery pushing individuals out of medical school. Secondly, most of the other education programs do not have a *numerus fixus*, and only one application to any *numerus fixus* program was allowed each year. When applying for any other degree applicants would be granted admission unconditionally. Therefore, applying to the lottery discloses, by revealed preferences, that medicine is the first choice of the candidate, and any defiance of the lottery outcome should be motivated by reasons that are independent of it. The independence assumption is also satisfied, as shown by the balancing checks in Table 2.13. Note that the checks are made on the full sample of applicants, as this is the level at which the randomization was performed. Finally, I argue that the exclusion restriction is satisfied as the lottery as no value per se, but only in the measure in which it allows access to medical school. Thus, any additional consequence following the lottery assignment derives, in fact, by its implication on the probability of becoming a doctor.

2.6 Results

2.6.1 Returns from occupational persistence

I proceed to estimate the returns from occupational persistence using the strategy described in Section 2.5.2. I focus the analysis on personal primary income, which includes earnings from salaried employment as well as income from own business. Table 2.5 reports the point estimates for the IV regressions of both groups, namely children of doctors and non doctors, as well as their difference $\hat{\beta_P} - \hat{\beta_{NP}}$. I run four models, enlarging each time the set of controls.

The results highlight positive and significant treatment effects from medical school, both for dynastic and non dynastic doctors. However, these are 54% higher for children of doctors. The difference is significant and translate into a 23% income increase with respect to applicants with a different occupation than their parents[8].

[8]Table 2.16 reports the results for the same model estimated via ordinary least squares, which are biased downward with respect to the IV ones. This can be justified in light of how LATE works in the Dutch context of the medical school. While the vast majority of accepted candidates enrolls in medical school (see Section 2.5.3), applicants rejected by the lottery, are allowed to re-apply for a number of times. As illustrated in Section 2.5.3, always-takers represent the main difference between the individuals that are actually treated and the ones that are assigned to treatment by the instrument. The downward bias of the OLS estimator would then imply that always-takers display on average worse OLS outcomes than individuals treated under the lottery. Under the

The result is striking as it implies effects of persistence that are about ten times the OLS average estimates in Section 2.4.1. Moreover, as discussed in section 2.5.3, this is likely to be an underestimation of the total ATT due to the local nature of my estimates.

Counterfactual income mobility

Section 2.4.2 has shown the contribution of occupational persistence to the intergenerational income correlation in the top decile of the parental income distribution. Reproducing parental occupations can mechanically increase intergenerational income correlations in dynastic individuals. Yet, a counterfactual calculation based on the gains experienced through occupational persistence measured in Section 2.6.1 allows to roughly estimate how much of the additional income correlation is attributable to additional on-the-job returns.

For the purpose of this exercise, I assume that the same returns experienced by dynastic doctors can be extended to all children who have followed their parents in the same occupations. The income increase corresponds to 19.98% of the average for persistent doctors. Reducing the income of all dynastic individuals by the same amount would then be equivalent to shutting down gains from persistence. I do so and then calculate counterfactual percentile ranks from this artificial income measure.

Table 2.7 compares the rank-rank income correlations for the top decile of the parental income distribution in the actual data and in the constructed counterfactual with no returns from occupational persistence. The difference in coefficients, displayed in Column (3), shows that shutting down within profession gains from occupational persistence reduces intergenerational correlation by 4 percentage points, i.e. 11%. In other words, these returns account for about half of the additional stickiness in income produced by dynastic families at the top of the distribution (8 percentage points, as shown in Section 2.4.2).

assumption explained in Section 2.14 of always-takers having higher treatment effects, this would also imply more negative selection.

2.6.2 Selection

This section draws on the theory of potential outcomes outlined in Section 2.5.2 and focuses on providing a measure of the selection bias into occupational persistence. In particular, I use the relationships in Equations 2.3 and 2.4 to calculate the bias identified by the lottery strategy. This will require to obtain $\hat{\beta}_{P,ADO}$ and $\hat{\beta}_{NP,ADO}$, i.e. the coefficients from the sheer average difference in outcomes. The estimation can be performed using the same model in Equation 2.2 on a sample that also includes individuals who chose not to apply to the lottery, as opposed to the applicant-only population used for the IV. It is then critical that the analysis sample is constructed so to comprise students that could have potentially applied, such that the two groups are comparable. In this context, I have included all students that, in the same years as the lottery sample, have enrolled in a higher education program (WO). This is justified by the fact that, as explained in Section 2.14, candidates must have completed the most advanced type of high school in order to access a research university, making the two groups reasonably homogeneous in terms of potential academic choices and study level. In the same fashion as the treatment effect estimation strategy, the difference in coefficients for dynastic and non dynastic doctors will be interpreted as the additional gains associated to medical school for those who have and do not have a parent who is a doctor. Once the mean difference in outcomes is calculated, the selection bias is obtained as the difference between $\hat{\beta}_{j,ADO}$ and $\hat{\beta}_{j,ATT}$ for $j \in P, NP$. Note that combining linearly the two estimates requires their full variance-covariance matrix. To this purpose, I run the IV and OLS models as a unique regression with stacked samples, so that the two equations are estimated simultaneously and the coefficients can be compared.

Again, I focus on labour income. Table 2.6 reports, for both dynastic and non-dynastic doctors, the coefficients associated with the enrollment in medical school for both models, as well as the difference in the two identifying the selection bias. The IV results correspond to the ones reported in Table 2.5. The last column, presents the difference in estimated selection biases, i.e. the additional bias observed for medical school applicants whose parent is also a doctor. This is negative, and large enough to completely counteract the treatment effect estimated. Consistently, the coefficient for the average difference in outcomes of non dynastic doctors is not statistically different from that for dynastic individuals.

In practice, a negative selection bias tells us that those students who selected into persistence (applied to the medical school lottery and have a parent who is a doctor) but are exogenously forced out of it exhibit worse outcomes than those who did not. Again, it can be thought as the difference between selection into medical school for children whose parent is a doctor and children who do not have a parent who is a doctor. The additional difference is necessary, as restricting the calculation to only children of doctors may capture a general selection bias which is not related to parental status. On the other hand, the full sample paired with the interaction term allows to obtain a "net" selection bias, which takes into account what selection into medical school would be in absence of persistence. Note that, as discussed in Section 2.5.3, this is likely to be an underestimation of the actual selection bias, as it is only calculated on compliers.

The negative selection bias explains why descriptive returns of persistence substantially underestimate the gains experienced by dynastic individuals, as estimated in Section 2.4.1. The co-existence of a positive treatment effect and negative selection can be interpreted in two different ways. Applicants' abilities may be positively correlated across fields (e.g., someone who performs relatively well as a medical doctor would also perform relatively well as a biologist). If this was the case, we could hypothesize that children choosing to be persistent in occupation have, on average, lower ex-ante abilities than those who do not, and yet gain when in the same field as their parent because of the additional available support provided by their connection. On the other hand, should abilities in different jobs display negative or no correlation, we could still interpret the results as dynastic children having developed high and strongly specialised skills in the parental occupation, which make them better than their peers when occupationally persistent but ill-suited for alternative professions.

2.7 Extensions and robustness

The analysis above has focused on describing the relevance of intergenerational occupational persistence with respect to total labour income. This section presents results on additional margins as well as some relevant heterogeneity.

2.7.1 Alternative labour market outcomes

The measure of income utilized above comprises both salary earnings and labour income from own business. As previous literature has often highlighted the relevance of family firms and entrepreneurship in intergenerational transmission dynamics, we may be interested in exploring whether the heterogeneous returns found Section 2.6.1 are solely driven by business income, or if employee doctors also enjoy some benefits. In order to do this, I run the same model described in Section 2.5.2, using this time as dependant variables annual and hourly salary earnings, as well as hours worked in employment contracts. Table 2.17 reports the results. Column (1) suggests that aggregate results on differential income returns are, as a matter of fact, driven by labour income from own business, as no difference in treatment effects is observed in annual earnings. However, when decomposing the latter into hours worked and hourly earnings rates, a different story emerges. In particular, these regressions tell us that although dynastic and non dynastic doctors earn, on aggregate, the same salaries, children of doctors still experience gains from occupational persistence. As a matter of fact, they work about 13% less hours per month than non persistent individuals, and yet earn 18% higher hourly wages. This finding suggests that different mechanisms may be at play among salaried and self-employed doctors when their parents are also in the medical profession, and call for further investigation of the channels involved.

2.7.2 Heterogeneity by GPA category

The treatment effects from the medical profession estimated in Section 2.6.1 represent averages across GPA groups, and thus across students with different ex-ante ability.

While controlling for the relevant groups is essential for identification, as assignment to medical school is only random within those, we may also be interested on whether children with different abilities gain differentially from occupational persistence. Figure 2.6 displays the difference in coefficients $\hat{\beta_P} - \hat{\beta_{NP}}$ from the model in which the treatment variable is interacted with dummies for the GPA categories. Table 2.18 reports the estimates for clarity. The differential payoffs linked to occupational persistence vary non-linearly with GPA. In particular, while smaller for low GPA, the difference in treatment effects peaks for the 7-7.5 category, and then becomes large and negative for very high GPAs. The pattern is consistent with a model in which parental investment is a complement of ex-ante ability at first, but a substitute for higher abilities. On one hand, this would imply that a basic level of previous skills needs to be developed in order to fully exploit the additional resources provided by the transmission of occupation. On the other hand, the negative returns from persistence for high ability students suggest that the extra support does not benefit this category of doctors but, if anything, it locks them into worse outcomes. While we would think that the transmission of occupation specific knowledge and skills may be a complement to previous developed talent throughout the GPA distribution, professional network may be an instance of parental contribution acting in such non linear fashion.

2.7.3 Gender differences

The descriptive evidence in Chapter I has highlighted gender differentials in the rates of transmission of occupations. I therefore ask the question of whether these differences also translate into a gap in returns. Table 2.19 attempts to answer by reporting coefficients from the same model described in Section 2.5.2, with the inclusion of interaction terms with gender, and shows substantially higher gains for women. In particular, it is worth noticing that average returns from medical school are virtually equivalent in absolute terms but considerably higher for women relatively to the their average income. Moreover, female doctors appear to benefit more from having a parent in the medical profession, with gains from persistence that are more than twice those of male doctors, although the latter are imprecisely estimated. The findings are striking as, depending on the mechanisms at play, they may point at occupational transmission as a way of bridging the income gender gap.

Table 2.20 also reproduces the same specification using having father and mother in the medical profession separately. The estimates suggest that having a father in the medical profession may be the real driver of the gains experienced by dynastic doctors. However, the model using mothers is underpowered, due to the very limited number of mother doctors present in the sample. This is not surprising as most of the applicants in 1988-1999 have parents born in the 1940s.

2.7.4 Robustness across different time frames

The outcomes used in the main analysis to evaluate gains from persistence have been measured as averages between 15 and 20 years from the first application. Given the structure of medical studies and careers, we can consider these as medium term estimates. Nevertheless, we may also be interested on whether the same premium is observed in outcomes measured at different times of the doctors' career. To answer this question, Figure 2.7 shows the difference $\hat{\beta}_P - \hat{\beta}_{NP}$ obtained through the same empirical design of Section 2.5.2, but using as dependent variable average labour income in all 6-year windows between 10 and 30 years from application[9]. The results highlight a time pattern in which benefits appear 15-20 years from application and slightly increase in time, plateauing after 20-25 years. The absence of gains in the first time frames may suggest, on one hand, that gains from persistence only manifest at a more mature career stage. On the other hand, the results may be partly driven by the different patterns of re-applications, which may in turn affect the timings at which doctors finish their studies and move on to more rewarding phases. In particular, as children of doctors apply on average more times after a rejection, they may still not be on the labour market 10 to 15 years after their first application. This would drive their income down and wipe out the gains we observe at later stages.

[9] Not that different time frames may catch different individuals in the sample.

2.7.5 Placebo tests

The descriptive statistics of Table 2.14 showed how persistent and non persistent applicants displayed some differences in terms of relevant observable characteristics. To support the idea that the heterogenous treatment effects in section 2.6.1 reflect the contribution of occupational persistence, rather than of other traits correlated with having a parent who is a doctor, I perform a series of placebo tests. In these, the specification proposed by Equation 2.2 is used to explore heterogeneity according to different margins than having a parent who is a doctor and assess whether the same differences in returns across groups is observed in these cases. In particular, the considered characteristics are gender, parental income and high school GPA, i.e. those exhibiting differences across the groups of children of doctors and non doctors. Table 2.21 reports the differences in the coefficients estimating returns from the medical professions for those who fall into the different categories of the considered heterogeneity variable. Reassuringly, none of these sample splits can reproduce the additional returns of dynastic doctors, confirming that intrinsic differences between children of doctors and non doctors do not drive the observed gains.

2.8 Mechanisms

The positive income gains found in Section 2.6.1 may be due to several different mechanisms. Here, I hypothesize a few channels through which the positive returns and negative selection found above may realize, and I use the available data to assess the likelihood of each. Note that the results below are suggestive, rather than conclusive, and as different factors are likely to concur to the final result, evidence in favour of one mechanism need not be in contradiction with the existence of others.

2.8.1 Returns

Outside options

Differential returns from fields and occupations need to be interpreted in light of the the outside options faced by those who select into them, which may not be equivalent even among those choosing the same profession (Kirkeboen et al., 2016). If children of doctors applying to medical school face different best alternatives than children of non doctors, this may be an important driver of the observed differences in gains. As a matter of fact, one could hypothesize that, when losing the lottery, those who initially selected into occupational persistence may remain tied to the broader field, i.e. health professions in the case of doctors, and move towards on average lower paying but related occupations, such as nursing or physiotherapy. On the other hand, children of non doctors may feel more free to change field and go into other more rewarding occupations, such as law or finance. Such scenario would justify observing larger returns from medical school for dynastic doctors.

While the data does not allow to observe the second choice of all applicants, Figure 2.8 compares the distributions of best alternatives for children of doctors and non doctors who lose the lottery at the time of their first application[10]. Overall, the two do not appear very dissimilar and do not point in general at children of doctors being in lower paid occupations when not being able to enroll in medical school. On the contrary, they are more likely to move to Dentistry and Law, both among the highest paid professions. On the other hand, children of non doctors choose more often nursing, physiotherapy or natural sciences, which are usually not as well financially rewarded.

Family practices and connections

The role of nepotism, or the favouritism enjoyed by a child in the same profession as their parent, has been sometimes brought up in the literature to justify the potential advantage experienced by dynastic individuals (Basso et al., 2021; Raitano and Vona,

[10]Figure 2.15 reproduces, for clarity, the same plot excluding those who eventually enroll in medical school.

2021). This can manifest as a support in the job search, both through professional network and own resources, such as an existing business[11].

I test this hypothesis by using data on two margins. In particular, exploiting the full sample of doctors for which I can observe parental connections I check whether dynastic individuals are more likely to work for a firm owned by their parents or to have the same employer as their parents. The estimates reported in Table 2.22 show that indeed doctors with a parent in the medical profession are 2.7 and 2.9 times more likely than average to work for their parents or for their same employer, respectively. While these coefficients will, at least partly, be mechanically driven by the occupational match between parents and children, they are made particularly relevant by the fact that dynastic doctors exhibit the highest difference in income with respect to the non dynastic ones exactly when they are connected to their parents in the workplace (Figure 2.9).

Genetics

Does a "gene of the doctor" exist? If that was the case, the positive gains associated with occupational persistence may be explained by a higher intrinsic ability of children of doctors in tasks related to Medicine. In other words, this could be the case of the "steady hand" trait hypothesized in Section 2.5.2. To rule out this option, I build on Bell, Chetty, et al. (2019)'s research design. Their identification relies on the idea that genetic differences in the ability of inventors are unlikely to lead to differences in propensities to innovate across narrowly defined technology classes. In the medical context, I can exploit information on the medical specialty chosen by dynastic doctors to perform a similar exercise.

Figure 2.10 displays the increase in the probability of choosing a given medical specialty when the individual has a parent in the same one[12]. If children of doctors chose to follow their parents and performed better in the profession because of differences in field specific inherited abilities, we should not expect to observe any systematic pattern in the choice of specialisation. In fact, the coefficients are all

[11]The results in Lo Bello and Morchio (2022) also support this idea, as they find that individuals following their parents in the same occupations find a job faster.

[12]For this exercise I consider all specialties in which I observe at least 500 doctors.

positive and significant, thus suggesting against this hypothesis.

2.8.2 Selection

Early specialisation

Section 2.6.2 has found negative selection bias into occupational persistence. In practice, this implies that the difference in baseline outcomes for those who selected into medical school and those who did not is lower for children of doctors than for children of non doctors. One possible explanation for this finding is that individuals selecting into following their parents may have, since early age, worked towards specialising in the parental field, for instance by developing preferences or skills making them fit for the job, which would not be rewarded in the same way in different professions. In the case of medical doctors, this could imply, among other things, undertaking the highest tier of high school (VWO) and choosing the track and profile required for eligibility to medical school applications[13].

I explore this hypothesis by observing the choices of two cohort of students that I am able to follow from the end of primary school, in years 2006 and 2007, to the choice of university program. Table 2.23 confirms that it is indeed the case that children of doctors are, in the most conservative specification, 81% more likely to choose a VWO school, even when controlling for parental level of education. Consistently, Table 2.24 also show that, conditional on enrolling in a VWO school, they also face 41% higher probability of choosing a study track that makes them eligible to apply to medical school. When believing talents and abilities across fields to be mostly uncorrelated, the estimates above suggest a potential role for an excess investment in the medical field, at the cost of other skills that would instead increase the rewards in the best alternative options.

Occupational mismatch

One alternative factor that may justify negative selection into persistence is a mismatch in desired and realised occupation. This may due, for instance, to parental

[13] See Section 2.2.1 for further details on application requirements.

pressures or to the idea, developed by the child through the years, that the career path traced by their parents would be the natural one to follow. If that was the case, we may still observe gains from occupational persistence attributable to parental support and investment, but we may also catch signs of struggle when own ability and motivation play a stronger role than external resources, e.g. during college studies.

Even though no information on assessment and performance is available in the data, I can observe completion dates of medical school applicants and, consequently, dropouts. I use this information to run the same model that explored income returns in Section 2.6.1, with the probability of dropping out as dependent variable. The estimates in Table 2.25 show that, although imprecisely estimated, dynastic doctors are more likely to drop out of medical school than non dynastic ones. This suggests that students choosing to follow their parents into Medicine studies may, on average, be less suited for this path than children of non doctors applying to medical school.

2.9 Conclusions

Intergenerational transmission of occupation is a known and academically documented phenomenon, whose implications have not been convincingly and causally explored so far. Given its prevalence, understanding what the associated gains are contributes to our knowledge on social mobility and its drivers. This paper has presented new evidence on the returns of transmission of occupations across generations. Using rich administrative data, I disentangle treatment effects from and selection into the medical profession for children of doctors.

Transmission rates increase at the top of the parental income distribution. The differential incidence highlights the importance of investigating the effects of occupational persistence on the income of offsprings as together they could strongly contribute towards income inequality. OLS regressions uncover positive gains associated with occupational transmission. These are largely heterogeneous across occupations but quite stable along the income distribution.

The descriptively estimated gains may, in principle, be driven by both the effects of occupational transmission and the specific selection of individuals following their parents in the same field. In order to disentangle the two and obtain causal estimates for the returns from occupational inheritance, I focus on the medical profession and exploit the exogenous variation provided by lottery assignment to medical school. I find that children of doctors who enter Medicine earn, on average, 23% than their individuals whose occupation is different from their parents'. This corresponds to 54% higher returns from the medical profession relative to their non dynastic counterparts. The positive treatment effects are associated with a large negative selection bias into intergenerational occupational persistence, which biases OLS estimates of returns towards zero. Together, the two aspects point at field inheritance in Medicine as increasing economic inequality, as it helps children of high income parents remaining rich, even when they have ex ante lower abilities than non persistent children. Suggestive evidence presented in the paper point at a potential role for family businesses and professional connections in boosting outcomes of dynastic individuals.

My findings contribute to the growing literature on intergenerational transmission of occupations and fields of study by providing the first account of how this affects private returns and inequalities. More broadly, they introduce additional evidence on a channel through which social mobility could be fostered, or hampered.

References

Aina, Carmen and Cheti Nicoletti (2018). "The intergenerational transmission of liberal professions". In: *Labour Economics* 51, pp. 108–120.

Alsan, Marcella, Owen Garrick, and Grant Graziani (2019). "Does diversity matter for health? Experimental evidence from Oakland". In: *American Economic Review* 109.12, pp. 4071–4111.

Altmejd, Adam (2023). *Inheritance of fields of study*. Tech. rep. IFAU-Institute for Evaluation of Labour Market and Education Policy.

Altonji, Joseph G, Peter Arcidiacono, and Arnaud Maurel (2016). "The analysis of field choice in college and graduate school: Determinants and wage effects". In: *Handbook of the Economics of Education*. Vol. 5. Elsevier, pp. 305–396.

Altonji, Joseph G, Erica Blom, and Costas Meghir (2012). "Heterogeneity in human capital investments: High school curriculum, college major, and careers". In: *Annu. Rev. Econ.* 4.1, pp. 185–223.

Angrist, Joshua D and Jörn-Steffen Pischke (2009). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Basso, Gaetano et al. (2021). "Quality and selection in regulated professions". In.

Becker, Gary S and Nigel Tomes (1979). "An equilibrium theory of the distribution of income and intergenerational mobility". In: *Journal of political Economy* 87.6, pp. 1153–1189.

Bell, Alex, Raj Chetty, et al. (2019). "Who becomes an inventor in America? The importance of exposure to innovation". In: *The Quarterly Journal of Economics* 134.2, pp. 647–713.

Bell, Brian, Jack Blundell, and Stephen Machin (2023). "Where is the land of hope and glory? The geography of intergenerational mobility in England and Wales". In: *The Scandinavian Journal of Economics* 125.1, pp. 73–106.

Björklund, Anders and Markus Jäntti (1997). "Intergenerational income mobility in Sweden compared to the United States". In: *The American Economic Review* 87.5, pp. 1009–1018.

Blanden, Joe, Paul Gregg, and Stephen Machin (2005). "Intergenerational mobility in europe and north america". In: *Report supported by the Sutton Trust, Centre for Economic Performance, London School of Economics.*

Buscha, Franz and Patrick Sturgis (2018). "Declining social mobility? E vidence from five linked censuses in E ngland and W ales 1971–2011". In: *The British Journal of Sociology* 69.1, pp. 154–182.

Chetty, Raj et al. (2014). "Where is the land of opportunity? The geography of intergenerational mobility in the United States". In: *The Quarterly Journal of Economics* 129.4, pp. 1553–1623.

Clark, Gregory (2015). *The Son Also Rises: Surnames and the History of Social Mobility: Surnames and the History of Social Mobility.* Princeton University Press.

Cunningham, Scott (2021). *Causal inference: The mixtape.* Yale university press.

Dal Bó, Ernesto, Pedro Dal Bó, and Jason Snyder (2009). "Political dynasties". In: *The Review of Economic Studies* 76.1, pp. 115–142.

Dribe, Martin and Jonas Helgertz (2016). "The lasting impact of grandfathers: class, occupational status, and earnings over three generations in Sweden 1815–2011". In: *The Journal of Economic History* 76.4, pp. 969–1000.

Erikson, Robert, John H Goldthorpe, and Lucienne Portocarero (1979). "Intergenerational class mobility in three Western European societies: England, France and Sweden". In: *The British Journal of Sociology* 30.4, pp. 415–441.

Friedman, Sam and Daniel Laurison (2020). *The class ceiling: Why it pays to be privileged.*

Goldthorpe, John and Robert Erikson (1992). *The constant flux: a study of class mobility in industrial societies.*

Greenwood, Brad N, Seth Carnahan, and Laura Huang (2018). "Patient–physician gender concordance and increased mortality among female heart attack patients". In: *Proceedings of the National Academy of Sciences* 115.34, pp. 8569–8574.

Haeck, Catherine and Jean-William Laliberté (2023). "Careers and Intergenerational Income Mobility". In.

Ketel, Nadine et al. (2016). “The returns to medical school: Evidence from admission lotteries”. In: *American Economic Journal: Applied Economics* 8.2, pp. 225–54.

Kirkeboen, Lars J, Edwin Leuven, and Magne Mogstad (2016). “Field of study, earnings, and self-selection”. In: *The Quarterly Journal of Economics* 131.3, pp. 1057–1111.

Lo Bello, Salvatore and Iacopo Morchio (2022). “Like father, like son: Occupational choice, intergenerational persistence and misallocation”. In: *Quantitative Economics* 13.2, pp. 629–679.

Long, Jason and Joseph Ferrie (2013). “Intergenerational occupational mobility in Great Britain and the United States since 1850”. In: *American Economic Review* 103.4, pp. 1109–37.

— (2018). “Grandfathers matter (ed): occupational mobility across three generations in the US and Britain, 1850–1911”. In: *The Economic Journal* 128.612, F422–F445.

Mocetti, Sauro, Giacomo Roma, and Enrico Rubolino (2022). “Knocking on Parents’ Doors Regulation and Intergenerational Mobility”. In: *Journal of Human Resources* 57.2, pp. 525–554.

OECD (2021a). *Education at a Glance 2021: Netherlands - Country Note.*

— (2021b). *Health at a Glance 2021*, p. 274. DOI: https://doi.org/https://doi.org/10.1787/ae3016b9-en. URL: https://www.oecd-ilibrary.org/content/publication/ae3016b9-en.

OECD, European Observatory on Health Systems, and Policies (2021). *Netherlands: Country Health Profile 2021*, p. 24. DOI: https://doi.org/https://doi.org/10.1787/fd18ea00-en. URL: https://www.oecd-ilibrary.org/content/publication/fd18ea00-en.

Olivetti, Claudia and M Daniele Paserman (2015). “In the name of the son (and the daughter): Intergenerational mobility in the United States, 1850–1940”. In: *American Economic Review* 105.8, pp. 2695–2724.

Polyakova, Maria et al. (2020). “Does medicine run in the family—Evidence from three generations of physicians in Sweden: Retrospective observational study”. In: *bmj* 371.

Raitano, Michele and Francesco Vona (2021). “Nepotism vs. specific skills: the effect of professional liberalization on returns to parental background of Italian lawyers”. In: *Journal of Economic Behavior & Organization* 184, pp. 489–505.

Roy, Andrew Donald (1951). “Some thoughts on the distribution of earnings”. In: *Oxford economic papers* 3.2, pp. 135–146.

SEO (2012). “Remuneration of medical specialists. An international comparison.” In: *SEO Economic Research, Amsterdam*.

Sinha, Rishabh (2016). “Intergenerational occupational mobility and labor productivity”. In: *Available at SSRN 2952882*.

Social Mobility Commission (2021). *State of the Nation 2022: A fresh approach to social mobility*.

Solon, Gary (1992). “Intergenerational income mobility in the United States”. In: *The American Economic Review*, pp. 393–408.

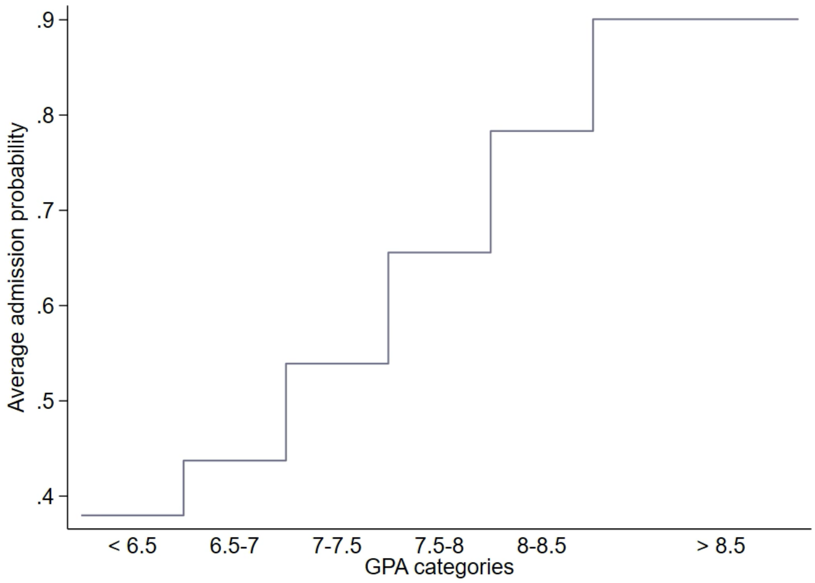
— (2002). “Cross-country differences in intergenerational earnings mobility”. In: *Journal of Economic Perspectives* 16.3, pp. 59–66.

Sorokin, Pitrim A. (1927). *Social mobility*. New York, Harper and Brothers.

Statistics Netherlands (2023). *Medisch geschoolden; arbeidspositie, werkkring, specialisme, vanaf 1999*. URL: https://www.cbs.nl/nl-nl/cijfers/detail/84760NED.

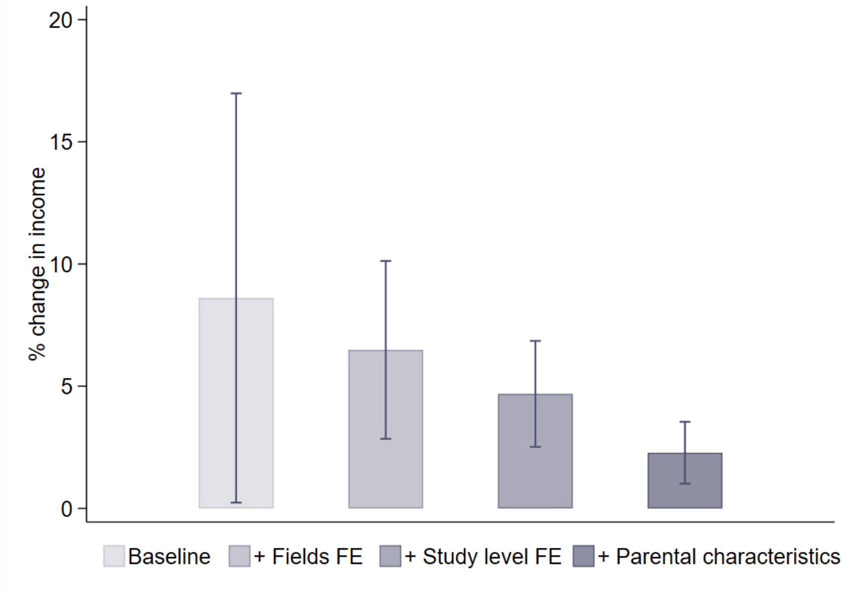
2.10 Figures

Figure 2.1: Average probabilities of winning the lottery



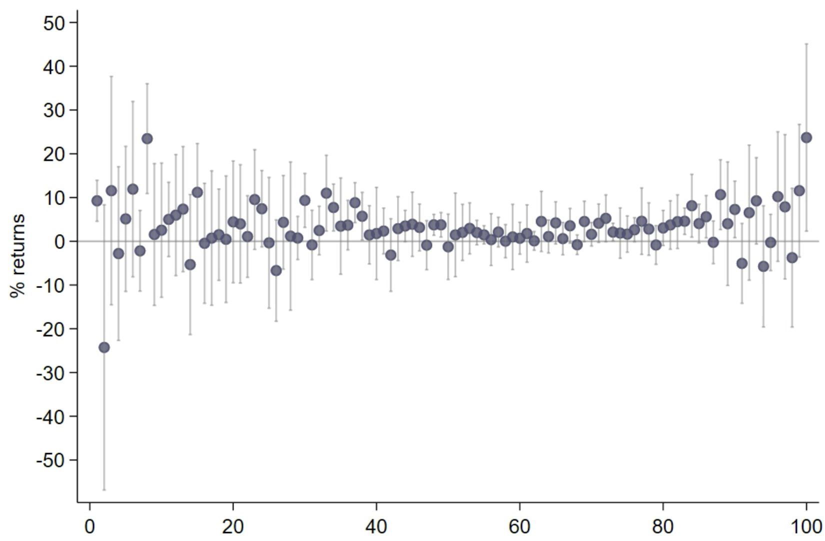
Notes: The plot displays the empirical probability of a successful application to medical school via the lottery, averaged for the period 1987-1999. The GPA groups are the ones used to weight the lottery. Figure 2.11 in the appendix of Section 2.12 reports the whole time series for admission probabilities by GPA group.

Figure 2.2: Returns associated with occupational persistence



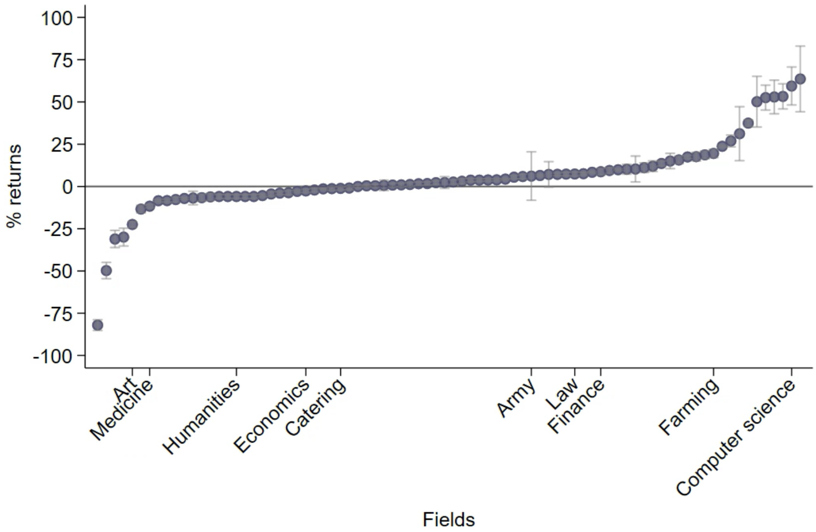
Notes: Returns are measured using the OLS specification of Equation 2.1. The "Baseline" coefficient only controls for 3 digit ISCED-F fixed effects. The demographics included in the second model are gender, Dutch nationality and year of birth. The Education level FE refer to the highest education attained. Parental controls include parental occupation and income, separately for father and mother.

Figure 2.3: Returns heterogeneity by income



Notes: Returns are measured using the OLS specification of Equation 2.1, augmented with interactions with parental income ranks. The regression controls for gender, Dutch nationality, cohort fixed effects, parental income rank, education level and ISCED-F detailed field. The latter is calculated as the average between maternal and paternal income percentile rank. Parents' income is measured as described in Section 2.3.1. The spikes show the 95% confidence intervals.

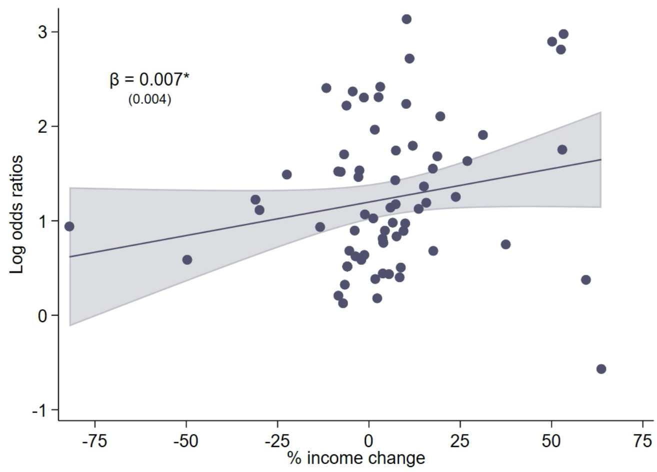
Figure 2.4: Returns heterogeneity by field



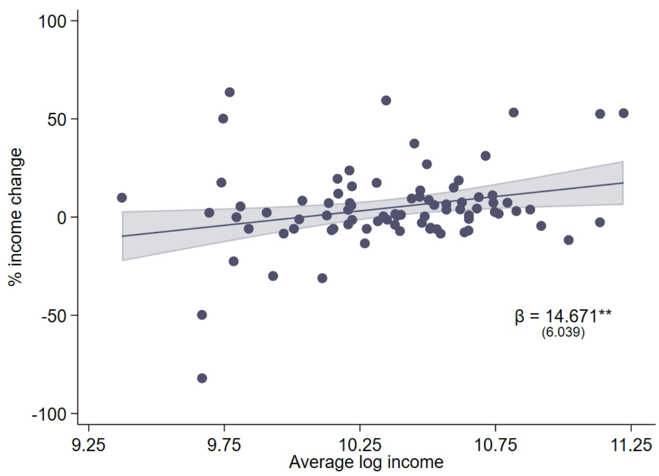
Notes: Returns are measured using the OLS specification of Equation 2.1, augmented with interactions with fields fixed effects. The regression controls for gender, Dutch nationality, cohort fixed effects, parental income rank, education level and ISCED-F detailed field. The spikes show the 95% confidence intervals.

Figure 2.5: Correlations of returns from occupational choice

(a) Persistence in occupational choice and returns

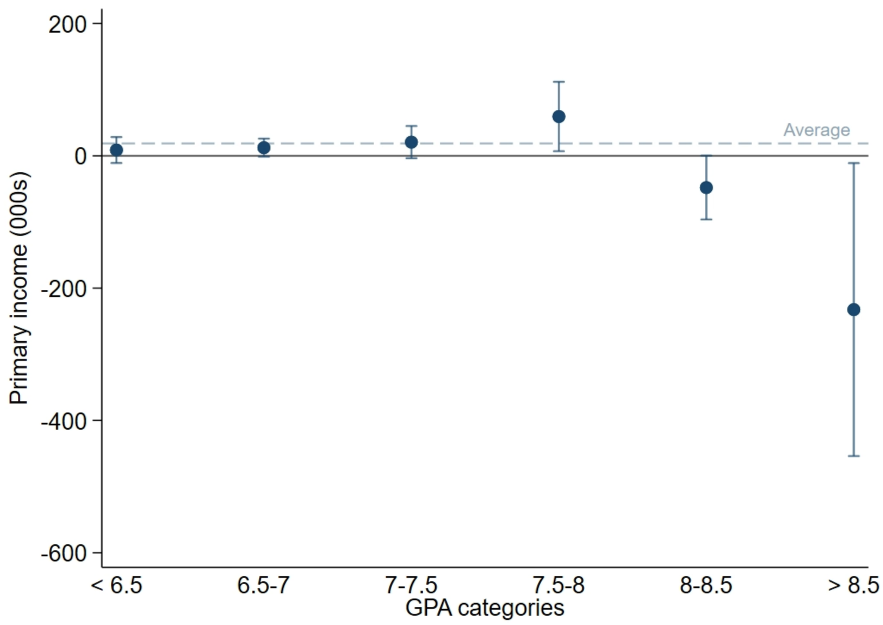


(b) Returns and average income



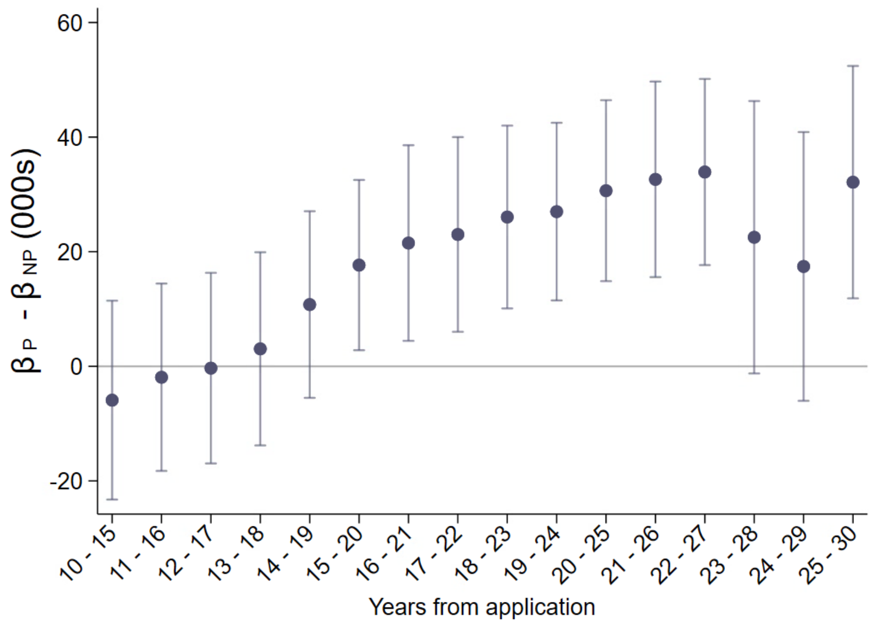
Notes: Panel (a): Persistence is measured as the odds ratios from the logistic model describing the likelihood of choosing a given field when having a parent in the same. Income changes associated with persistence in occupational choice are measured as the coefficient from the specification in Equation 2.1. Panel (b): Income changes associated with persistence in occupational choice are measured as the coefficient from the specification in Equation 2.1. Average income is calculated for each occupation as the mean primary income for workers in the field.

Figure 2.6: Heterogeneity by GPA



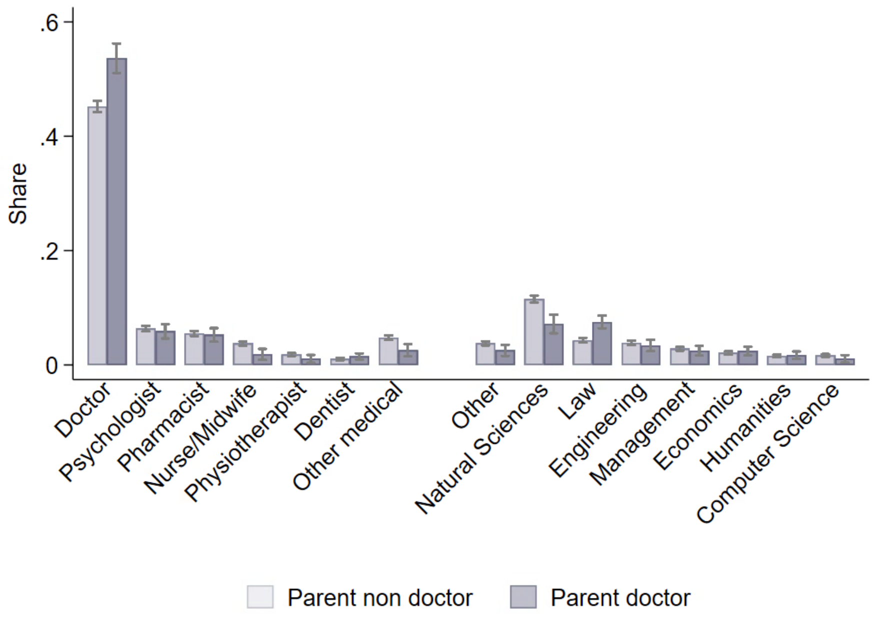
Notes: The specification controls for year of application, GPA category, years in which income is measured, gender, Dutch nationality, year of birth, parents cohort and income, separately for father and mother, which parent is a doctor, whether any other member of the family (among grandparents, uncles and aunts and older siblings) is also a doctor and 2 digit ISCED-F fields fixed effects. The spikes show 95% confidence intervals. Standard errors are clustered at the 1-digit ISCED level.

Figure 2.7: Different time frames



Notes: The graph reports the difference in causally estimated returns from medical school using different 6-year windows to calculate average income. Spikes represent 95% confidence intervals.

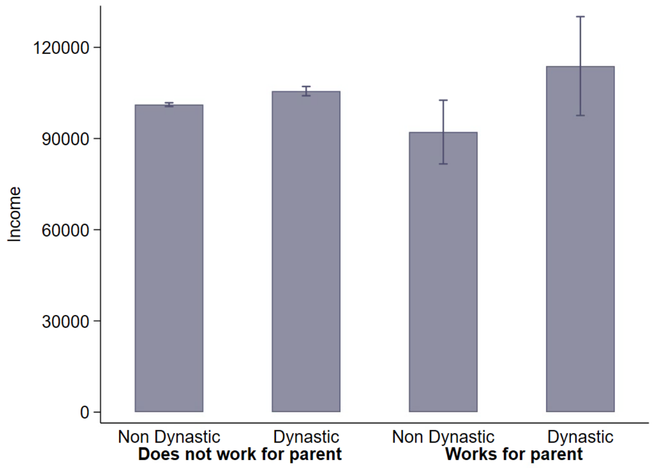
Figure 2.8: Outside options



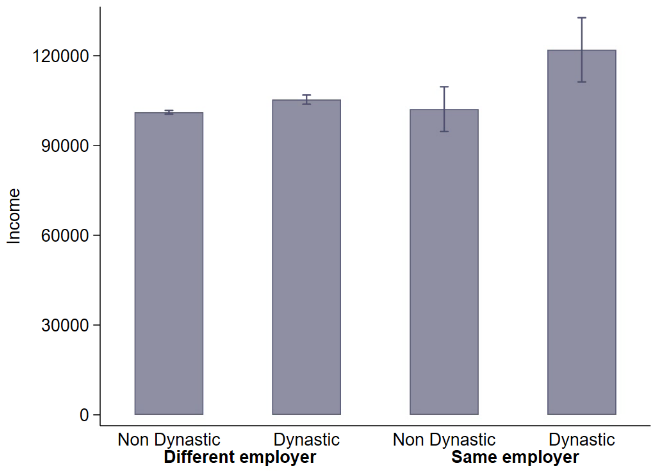
Notes: The graph reports the average shares of individuals choosing each of the considered fields when losing the lottery at the first application round.

Figure 2.9: Returns from family practices and connections

(a) Differences in income when working for parent

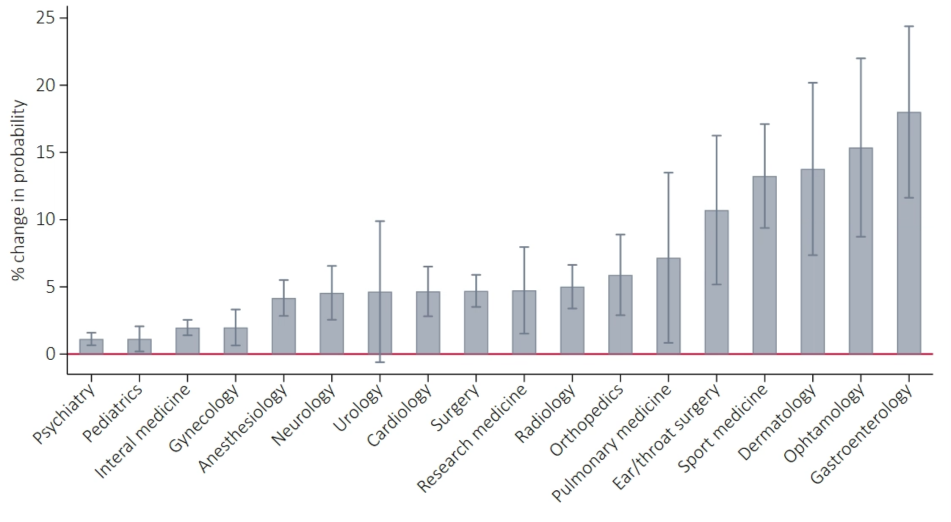


(b) Differences in income when same employer as parent



Notes: Panel (a): Bars represent the average unconditional income for dynastic doctors depending on whether they work for a firm owned by either parent. Spikes represent 95% confidence intervals. Panel (b): Bars represent average unconditional income for dynastic doctors depending on whether they work for the same employer as either parent. Spikes represent 95% confidence intervals.

Figure 2.10: Choice of medical specialty



Notes: Each coefficient is estimated using a OLS model, where the probability of entering a given medical specialty is regressed on a dummy indicating whether a parent is in that specialty. Coefficients are divided by the average size of the specialty to show percentage change in the probability of choosing it. The regression controls for gender, Dutch nationality, cohort fixed effects, year of first registration as a doctor, GPA category, application year and household income. The spikes show the 95% confidence interval.

2.11 Tables

Table 2.1: Summary statistics - All fields

	Mean	SD	N
Female	0.510	0.500	962,437
Dutch	0.947	0.224	962,437
Year of birth	1983	5.963	962,437
Obtained higher education (VO)	0.200	0.400	962,437
Obtained higher professional education (HBO)	0.347	0.476	962,437
Father obtained higher education (VO)	0.070	0.255	962,437
Father obtained higher professional education (HBO)	0.150	0.257	962,437
Mother obtained higher education (VO)	0.029	0.167	962,437
Mother obtained higher professional education (HBO)	0.126	0.331	962,437
Parent is a doctor	0.010	0.102	962,437
Sample with positive income	0.974	0.159	937,426

Notes: The sample includes individuals born between 1963 and 1991 for which a parental link exists and for whom the field of the highest level of education, own income and parental income is available.

Table 2.2: Summary statistics - Lottery sample

	Mean	SD	N
Female	0.583	0.493	40,007
Dutch	0.915	0.279	40,007
Age first app	19.43	2.58	40,007
Number applications	1.458	0.770	40,007
Won lottery	0.462	0.499	40,007
Parent is a doctor	0.123	0.329	40,007
GPA group:			
> 8.5	0.017	0.131	40,007
8 - 8.5	0.054	0.227	40,007
7.5 - 8	0.088	0.284	40,007
7 - 7.5	0.211	0.408	40,007
6.5 - 7	0.223	0.416	40,007
< 6.5	0.301	0.459	40,007

Notes: Sample includes all applications between 1988 and 1999. Individuals may appear multiple times in the sample when they re-apply after a rejection.

Table 2.3: Rank-rank correlations

	Income rank		
	(1)	(2)	(3)
	Non persistent	All	Difference
Parental income rank	0.2848***	0.2866***	0.0018*
	(0.0012)	(0.0012)	(0.0002)
R-Squared	0.200	0.201	
Observations	902,932	962,437	

Notes: * $p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Standard error from simultaneous estimation in parentheses. All specifications control for gender, cohort fixed effects, and 1-digit ISCED-F field fixed effects. Parental rank is the average of the percentile income ranks for fathers and mothers. The latter are calculated by cohort for their average income between 50 and 60 years old.

Table 2.4: Rank-rank correlations in top decile

	Income rank		
	(1)	(2)	(3)
	Non persistent	All	Difference
Parental income rank	0.3249***	0.4042***	0.0793*
	(0.1164)	(0.1063)	(0.0416)
R-Squared	0.119	0.118	
Observations	6,987	7,978	

Notes: * $p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Standard error from simultaneous estimation in parentheses. The sample includes individuals in the top income decile. All specifications control for gender, cohort fixed effects, and 1-digit ISCED-F field fixed effects. Parental rank is the average of the percentile income ranks for fathers and mothers. The latter are calculated by cohort for their average income between 50 and 60 years old.

Table 2.5: Effects of persistence on income

	(1)	(2)	(3)	(4)
	Income			
A. Parent is a doctor				
Enrolled in med school	40,729***	42,373***	41,754***	50,294***
	(5,820)	(6,389)	(6,283)	(6,315)
Mean of dep. variable	83,748	83,748	83,748	83,748
Observations	1,364	1,364	1,364	1,364
B. Parent is not a doctor				
Enrolled in med school	30,819***	29,779***	28,896***	32,630***
	(3,604)	(2,470)	(2,352)	(1,878)
Mean of dep. variable	77,936	77,936	77,936	77,936
Observations	9,055	9,055	9,055	9,055
C. Difference				
$\beta_P - \beta_{NP}$	9,910**	12,594**	11,857**	17,664**
	(4,499)	(5,696)	(5,576)	(7,579)
Mean of dep. variable for non persistent	77,634	77,634	77,634	77,634
Demographics controls		X	X	X
Family controls			X	X
Field FE				X
F-stat				30.39

Notes: * $p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Clustered standard errors at the 1-digit ISCED level in paretheses. All specifications control for year of application, GPA category, years in which income is measured. The second specification also controls for gender, Dutch nationality and year of birth. The third specification adds controls for parents' cohort and income, separately for father and mother, which parent is a doctor and whether any other member of the family (among grandparents, uncles and aunts and older siblings) is also a doctor. The last specification also control for 2 digit ISCED-F fields.

Table 2.6: Selection bias calculation

	(1) $\hat{\beta}_{ADO}$	(2) $\hat{\beta}_{ATT}$	(3) $\hat{SB}$
A. Parent is a doctor			
Enrolled in med school	16,683***	49,248***	-35,566***
	(4,040)	(5,383)	(4,350)
Mean of dep. variable		83,748	
Observations	4,801	1,364	
B. Parent is not a doctor			
Enrolled in med school	20,016***	31,925***	-11,908***
	(991)	(1,628)	(2,553)
Mean of dep. variable		77,936	
Observations	117,106	9,055	
C. Difference			
$\hat{SB}_P - \hat{SB}_{NP}$			-20,657***
			(4,715)
F-stat	196.6	169.3	

Notes: * $p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Clustered standard errors at the 1-digit ISCED level in paretheses. All specifications control for gender, Dutch nationality, cohort fixed effects, year of application, GPA category, gender of parent who is a doctor, parental cohort, parental income, average age at which income is measured, average year in which income is measured, ISCED-F 2-digit field and whether another family member is a doctor (among older siblings, grandparents, uncles and aunts). $\hat{SB}_j$ is calculated using the relationship described Equations 2.3 and 2.4.

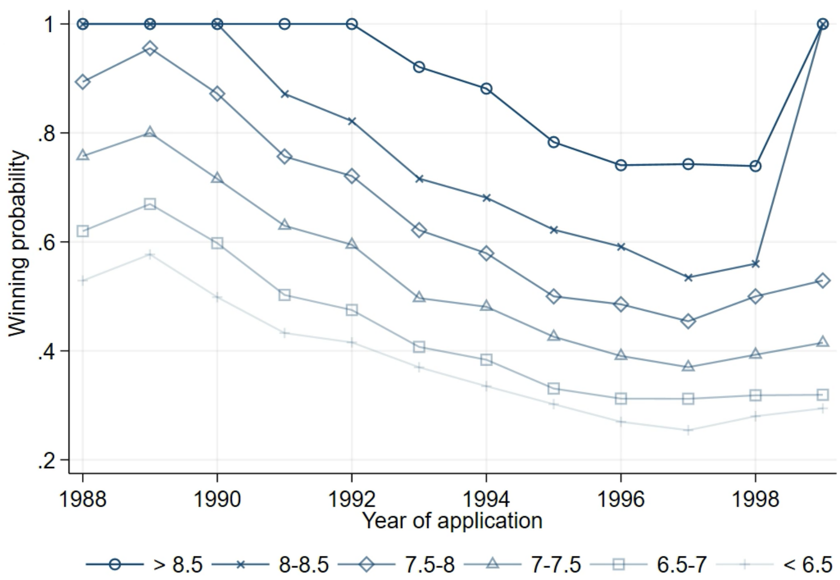
Table 2.7: Counterfactual rank-rank correlations in top decile

	Income rank		
	(1)	(2)	(3)
	Counterfactual	Real	Difference
Parental income rank	0.3612***	0.4042***	-0.0429**
	(0.1091)	(0.1063)	(0.0167)
R-Squared	0.118	0.118	
Observations	7,978	7,978	

Notes: * $p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Standard error from simultaneous estimation in parentheses. The sample includes individuals in the top income decile. All specifications control for gender, cohort fixed effects, and 1-digit ISCED-F field fixed effects. Counterfactual income ranks are calculated using income decreased by 19.98% for individuals in the same occupation as their parents. Parental rank is the average of the percentile income ranks for fathers and mothers. The latter are calculated by cohort for their average income between 50 and 60 years old.

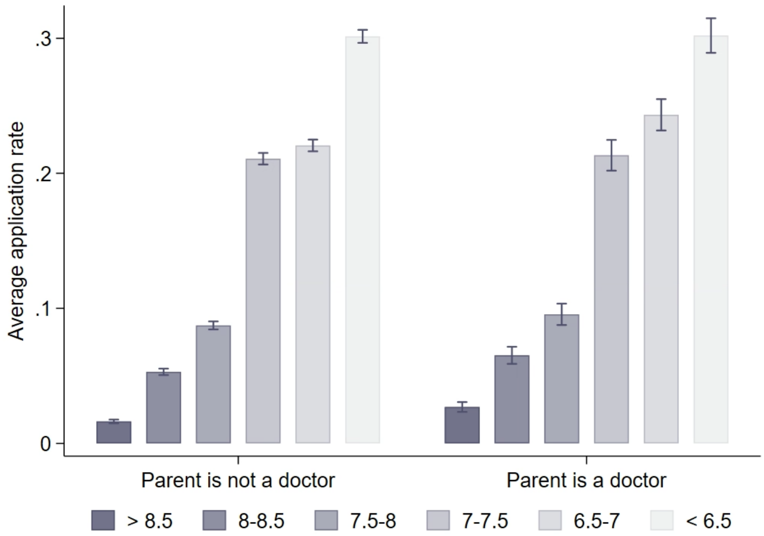
2.12 Appendix Figures

Figure 2.11: Probabilities of winning the lottery



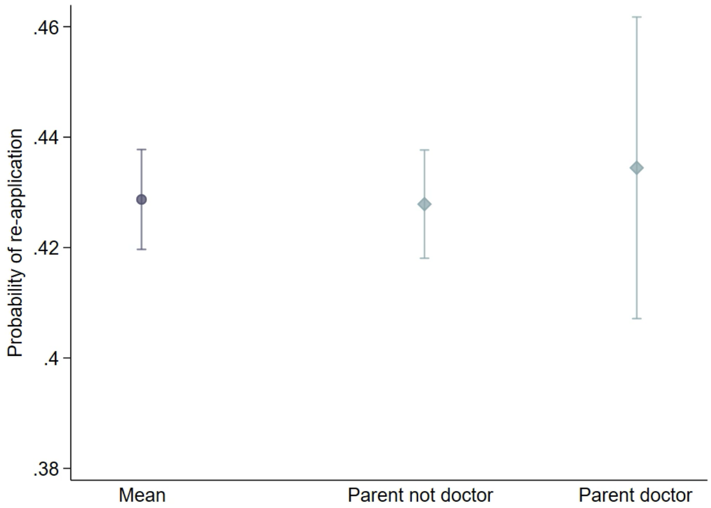
Notes: The plot displays the empirical probability of a successful application to medical school via the lottery, for the period considered in the analysis. The GPA groups are the ones used to weight the lottery. Probabilities change according to the number of applicants and of available places each year. From 1999, applicants with a GPA above 8 are automatically admitted to medical school.

Figure 2.12: Application rates by GPA and parental status



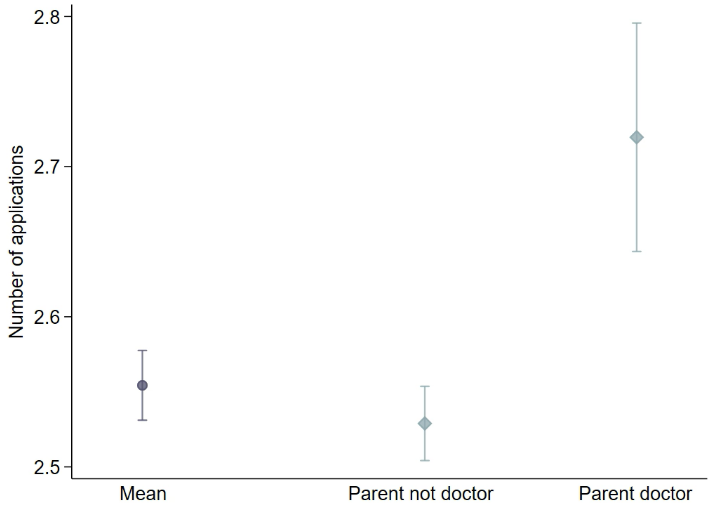
Notes: The plot displays the average number of applications to the medical school lottery with 95% confidence intervals. The estimates are obtained as linear predictions from an OLS regression of number of applications submitted on the dummy for having a parent in the medical profession. The specification also controls for Dutch nationality, cohort, gender, year of application, GPA category and income of both parents. Parents' income is measured as described in Section 2.3.1. The weighted estimate uses the inverse of the probability of being rejected, given the year and GPA category, to correct for over-representation of low GPA individuals.

Figure 2.13: Re-application rates, by parental status



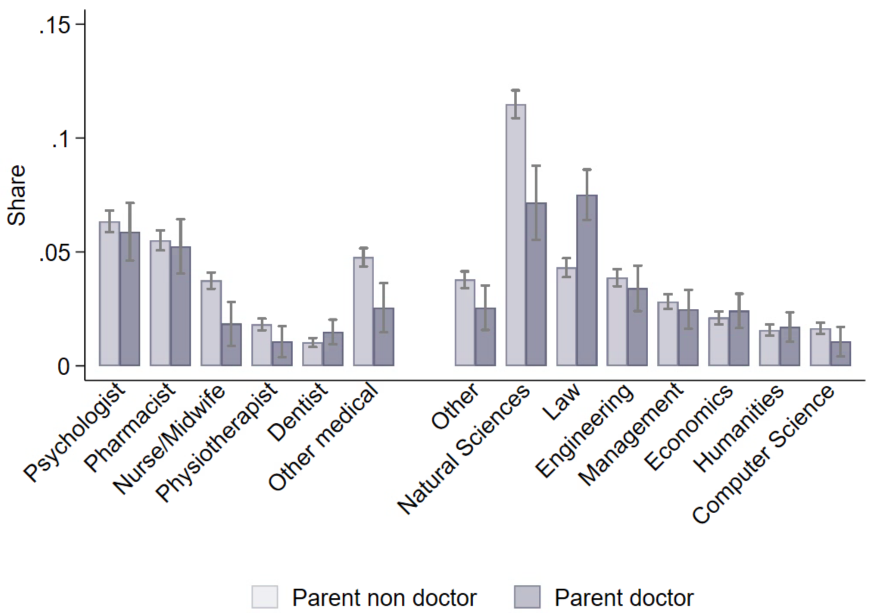
Notes: The plot displays the average probability of applying again when not admitted by the lottery. Spikes show the 95% confidence intervals. The estimates are obtained as linear predictions from an OLS regression of whether the student has applied more than once on the dummy for having a parent in the medical profession. The specification also controls for Dutch nationality, cohort, gender, year of application, GPA category and income of both parents. Parents' income is measured as described in Section 2.3.1. The weighted estimate uses the inverse of the probability of being rejected, given the year and GPA category, to correct for overrepresentation of low GPA individuals.

Figure 2.14: Number of applications, by parental status



Notes: The plot displays the average number of applications submitted, conditional on re-applying after losing the lottery. Spikes show the 95% confidence intervals. The estimates are obtained as linear predictions from an OLS regression of number of applications submitted on the dummy for having a parent in the medical profession. The specification also controls for Dutch nationality, cohort, gender, year of application, GPA category and income of both parents. Parents' income is measured as described in Section 2.3.1. The weighted estimate uses the inverse of the probability of being rejected, given the year and GPA category, to correct for overrepresentation of low GPA individuals.

Figure 2.15: Outside options - other fields only



Notes: The graph reports the average shares of individuals choosing each of the considered fields when losing the lottery.

2.13 Appendix Tables

Table 2.8: Foreign medical degrees

	(1)	(2)
	Has completed medical degree abroad	
Parent is a doctor	-0.012**	-0.011*
	(0.0054)	(0.0059)
Controls		X
Mean of dep. variable for non persistent	0.028	0.028
Observations	6,831	6,831
R-Squared	0.001	0.026

Notes: * $p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Standard errors in paretheses. The sample includes only lottery applicants who eventually register as doctors. Controls include gender, year of birth, Dutch nationality, GPA category, year of application and paternal and maternal incomes.

Table 2.9: Field coding

Code	Field	Code	Field
10	Generic and basic education	713	Electrical and energy technology
30	Personal skills	714	Electronics and industrial automation
100	Educational science	715	Mechanical engineering and metal working
200	Art	716	Vehicle, ship and aircraft engineering
211	Audiovisual engineering and media production	719	Other technology and technical services
212	Fashion, interior and industrial design	720	Food processing
215	Music and theatre	722	Woodworking, paper, plastic processing and ceramics
220	Humanities	723	Textile, clothing, footwear and leather manifacture
221	Theology and philosophy	724	Mineral extraction
222	History and archeology	731	Architecture and urban planning
231	Foreign languages	732	Construction and civil engineering
232	Native language, literature and linguistics	810	Crop and livestock production
300	Journalism, behaviour and society	812	Horticulture
311	Economics and econometrics	818	Interdisciplinary agriculture courses
312	Political and social sciences	820	Forestry and fishing
313	Psychology	841	Veterinary medicine and care
314	Sociology and cultural sciences	910	Healthcare
322	Library	911	Dentistry
410	Business administration	912	Medicine
412	Financial services	913	Nursing and obstetrics
413	Management business and personnel science	914	Medical diagnostics and technology
414	Marketing and public relations	915	Therapy and rehabilitation
415	Secretarial and administrative support	916	Pharmacy
416	Wholesale and retail	919	Other healthcare
417	Working skills	920	Wellbeing
420	Law	921	Care for disabled adults, elderly and family
541	Mathematics	922	Youth pedagogical work and childcare
511	Biology	923	Social work and career choice work
512	Biochemistry	1010	Personal services
521	Environment	1011	Home economics, facility services and cleaning
531	Chemistry	1012	Beauty and haircare
532	Earth sciences	1013	Catering
533	Physics	1014	Sport
542	Statistics	1015	Tourism and leisure
600	Computer sciences	1020	Hygiene and working condition
611	Computer use	1021	Public cleaning, water management and distribution
612	Database and network design and management	1022	Work safety and ergonomics
613	Software development and system analysis	1030	Public safety
619	Other computer science	1031	Armed forces
710	Technology and technical services	1032	Public order and safety
711	Chemical engineering and process engineering	1040	Transport and logistics
712	Environmental protection and technology		

Notes: Codes follow the ISCED-F 2013 system and courses are categorised by CBS. Numbering may slightly differ as a result of this assignment process. Moreover, a few similar fields have been subsequently aggregated during the data cleaning to avoid categories with too few observations.

Table 2.10: Matches with parental education

	Matched parental education data		
	(1)	(2)	(3)
Born after 1985	0.1983***	0.1968***	0.1956 ***
	(0.0006)	(0.0006)	(0.0006)
	[0.196]	[0.194]	[0.193]
Male	-0.0035**	-0.0035**	-0.0002
	(0.0006)	(0.0006)	(0.0007)
	[-0.0035]	[-0.0035]	[-0.0002]
Dutch nationality	-0.1015***	-0.1035***	-0.1031***
	(0.0013)	(0.0013)	(0.0013)
	[-0.0453]	[-0.0462]	[-0.0461]
Parental income		0.0000	0.0000
		(0.0000)	(0.0000)
		[0.0087]	[0.0094]
Fields FE:			
Generic programmes and qualification			0.0022
			(0.0071)
			[0.0002]
Education			0.0056**
			(0.0012)
			[0.0031]
Arts and Humanities			0.0200***
			(0.0015)
			[0.0088]
Social Sciences, Journalism and Information			0.0043***
			(0.0.0013)
			[0.0021]
Natural Sciences, Mathematics and Statistics			-0.0196***
			(0.0022)
			[-0.0056]
Information and Communication Technologies			0.0257***
			(0.0018)
			[0.0093]
Engineering, Manifacturing and Construction			-0.0022**
			(0.0010)
			[-0.0016]
Agriculture, Forestry, Fisheries and Veterinary			-0.00258***
			(0.0020)
			[-0.0081]
Health and Welfare			0.0157***
			(0.0009)
			[0.0123]
Services			0.0298***
			(0.0010)
			[0.0203]
Constant	0.5900***	0.5871***	0.5772***
	(0.0013)	(0.0013)	(0.0014)
Observations	2,605,439	2,605,439	2,605,439
R-Squared	0.040	0.040	0.041

Notes: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$. Standard errors in parentheses. Beta coefficients in squared brackets. The dependent variable is a dummy that equals 1 if at least one parent has a matched education level. Fields are coded according to the 1-digit ISCED-F 2013 system.

Table 2.11: Odds ratios of persistence by gender and parent

	(1)	(2)	(3)
	Chooses occupation j		
	Male	Female	Average
Parent is in occupation j	2.356***	1.850***	2.055***
	(0.0157)	(0.0120)	(0.0095)
Father is in occupation j	2.727***	1.977***	2.383***
	(0.0208)	(0.0202)	(0.0118)
Mother is in occupation j	1.700***	1.807***	1.763***
	(0.0221)	(0.0145)	(0.0144)
Observations	962,510	962,510	962,510

Notes: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$. Robust standard errors in parentheses. All specifications control for ISCED-F 2013 fixed effects, Dutch nationality, year of birth, education level and parental income.

Table 2.12: Returns associated with persistence in occupational choice

	(1)	(2)	(3)	(4)
	Log primary income			
Parent in same occupation	0.0833***	0.0648***	0.0468***	0.0278***
	(0.0184)	(0.0158)	(0.0094)	(0.0055)
Demographics		X	X	X
Education level			X	X
Parental characteristics				X
Observations	937,426	937,426	937,426	937,426
R-Squared	0.136	0.173	0.237	0.244

Notes: $*p < 0.1$, $**p < 0.05$, $***p < 0.01$. Robust standard errors in parentheses. All specifications control for 3 digit ISCED-F fixed effects. The demographics included in the second model are gender, Dutch nationality and year of birth. The Education level fixed effects refer to the highest education attained. Parental controls include parental occupation and income, separately for father and mother.

Table 2.13: Balancing tests by lottery outcome

	(1) Lost lottery	(2) Won lottery	(3) p-value
GPA: > 8.5			
Female	57.6%	53.8%	0.786
Age	18.7	18.8	0.990
Parent is a doctor	23.2%	18.2%	0.669
Observations	99	628	
GPA: 8-8.5			
Female	59.7%	60.6%	0.903
Age	19.0	18.9	0.908
Parent is a doctor	14.7%	14.6%	0.291
Observations	586	1701	
GPA: 7.5-8			
Female	62.1%	62.2%	0.921
Age	19.1	19.1	0.656
Parent is a doctor	13.3%	13.1%	0.572
Observations	1470	2229	
GPA: 7-7.5			
Female	59.6%	59.1%	0.647
Age	19.4	19.3	0.169
Parent is a doctor	12.2%	12.4%	0.176
Observations	4450	4483	
GPA: 6.5-7			
Female	58.0%	57.3%	0.507
Age	19.5	19.6	0.713
Parent is a doctor	13.4%	13.0%	0.724
Observations	5546	3919	
GPA: < 6.5			
Female	55.3%	55.6%	0.956
Age	19.8	19.8	0.302
Parent is a doctor	12.5%	11.7%	0.741
Observations	8094	4541	

Notes: The sample includes all applications to medical school in the years 1988-1999 (N = 54,900). Means are calculated across application years. P-values are obtained from linear regressions controlling for years of applications.

Table 2.14: Balancing tests by lottery outcome

	(1) Parent doctor	(2) Parent not doctor	(3) p-value
Female	50.5%	61.0%	0.000
Age	18.59	18.61	0.284
Dutch	94.0%	94.9%	0.143
Parental income	67,480	40,458	0.000
GPA: > 8.5	3.59%	1.80%	0.000
GPA: 8-8.5	7.70%	5.91%	0.010
GPA: 7.5-8	12.4%	11.4%	0.311
GPA: 7-7.5	24.3%	24.5%	0.840
GPA: 6.5-7	25.2%	24.9%	0.796
GPA: < 6.5	26.8%	31.4%	0.001
Observations	1364	9058	

Notes: The sample is the ones used for the final analysis. Means are calculated across application years. P-values are obtained from linear regressions controlling for years of applications. Age refers to the age at the time of the first application. Parental income is an average of paternal and maternal income.

Table 2.15: First stage

	(1) Enrolled Med school	(2) Completed Med school	(3) Registered doctor
Won lottery	0.4378***	0.4363***	0.3565***
	(0.0083)	(0.0083)	(0.0091)
F-stat	25.58	25.53	16.89
Observations	10,419	10,419	10,419
R-Squared	0.272	0.272	0.198

Notes: * $p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Robust standard errors in paretheses. All specifications control for gender, age, contry of birth, GPA category, year of application and the interactions of the latter two.

Table 2.16: Effects of persistence on income (OLS)

	(1)	(2)	(3)	(4)
	Income			
A. Parent is a doctor				
Enrolled in med school	16,464***	16,793***	16,441***	11,281***
	(4,333)	(4,227)	(4,357)	(4,224)
Mean of dep. variable	83,748	83,748	83,748	83,748
Observations	1,364	1,364	1,364	1,364
B. Parent is not a doctor				
Enrolled in med school	19,840***	20,015***	19,911***	18,159***
	(3,206)	(2,585)	(2,543)	(2,153)
Mean of dep. variable	77,936	77,936	77,936	77,936
Observations	9,055	9,055	9,055	9,055
C. Difference				
$\beta_P - \beta_{NP}$	-3,617	-3,222	-3,470	-3,878*
	(3,617)	(3,156)	(3,186)	(3,096)
Mean of dep. variable for non persistent	77,634	77,634	77,634	77,634
Demographics controls		X	X	X
Family controls			X	X
Field FE				X

Notes: * $p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Clustered standard errors at the 1-digit ISCED level in paretheses. All specifications control for year of application, GPA category and years in which income is measured. The second specification also controls for gender, Dutch nationality and year of birth. The third specification adds controls for parents cohort and income, separately for father and mother, which parent is a doctor and whether any other member of the family (among grandparents, uncles and aunts and older siblings) is also a doctor. The last specification also control for 2 digit ISCED-F fields.

Table 2.17: Effects of persistence on salary earnings and hours

	(1) Annual earnings	(2) Hourly earnings	(3) Monthly hours
A. Parent is a doctor			
Enrolled in med school	22,042***	16.73***	-11.85**
	(1,494)	(1.19)	(5.21)
Mean of dep. variable	52,296	31.51	149.3
Observations	1,180	1,180	1,180
B. Parent is not a doctor			
Enrolled in med school	18,988***	11.14***	6.29***
	(1,272)	(0.54)	(0.78)
Mean of dep. variable	49,531	30.36	144.3
Observations	8,061	8,061	8,061
C. Difference			
$\beta_P - \beta_{NP}$	3,054	5.59***	-18.14***
	(2,299)	(1.57)	(5.00)
Mean of dep. variable for non persistent	49,430	30.27	144.3
F-stat	70.4	139.7	31.11

Notes: * $p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Clustered standard errors at the 1-digit ISCED level in paretheses. All specifications control for year of application, GPA category, years in which income is measured, gender, Dutch nationality and year of birth, parents cohort and income, separately for father and mother, which parent is a doctor and whether any other member of the family (among grandparents, uncles and aunts and older siblings) is also a doctor and 2 digit ISCED-F fields.

Table 2.18: Effects of persistence on income by GPA

GPA categories:	$\beta_P - \beta_{NP}$
$>$ **8.5**	-232,467**
	(112,967)
Mean of dep. variable (non persistent)	93,283
Observations	271
8 - 8.5	-48,026*
	(24,587)
Mean of dep. variable (non persistent)	88,226
Observations	927
7.5 - 8	59,395**
	(26,758)
Mean of dep. variable (non persistent)	82,868
Observations	1,952
7 - 7.5	20,659*
	(12,413)
Mean of dep. variable (non persistent)	80,724
Observations	4,305
6.5 - 7	12,294*
	(6,971)
Mean of dep. variable (non persistent)	75,338
Observations	4,478
$<$ **6.5**	8,745
	(10,017)
Mean of dep. variable (non persistent)	71,890
Observations	5,793
F-stat	496.3

Notes: * $p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Clustered standard errors at the 1-digit ISCED level in paretheses. The specification controls for year of application, GPA category, years in which income is measured, gender, Dutch nationality, year of birth, parents cohort and income, separately for father and mother, which parent is a doctor, whether any other member of the family (among grandparents, uncles and aunts and older siblings) is also a doctor and 2 digit ISCED-F fields.

Table 2.19: Effects of persistence on income by gender

	(1) Male	(2) Female
A. Parent is a doctor		
Enrolled in med school	44,971***	55,849***
	(6,064)	(7,544)
Mean of dep. variable		
Observations	675	689
B. Parent is not a doctor		
Enrolled in med school	35,052***	31,155***
	(2,277)	(1,338)
Mean of dep. variable		
Observations	3,536	5,519
C. Difference		
$\beta_P - \beta_{NP}$	9,918	24,693***
	(7,162)	(8,360)
Mean of dep. variable for non persistent	90,906	68,958
F-stat	414.4	414.4

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the 1-digit ISCED level in paretheses. The specification controls for year of application, GPA category, years in which income is measured, gender, Dutch nationality, year of birth, parents cohort and income, separately for father and mother, which parent is a doctor, whether any other member of the family (among grandparents, uncles and aunts and older siblings) is also a doctor and 2 digit ISCED-F fields.

Table 2.20: Effects of persistence on income by parent

	(1) Father	(2) Mother
A. Parent is a doctor		
Enrolled in med school	47,728***	128,576
	(6,562)	(151,449)
Mean of dep. variable		
Observations	1,280	257
B. Parent is not a doctor		
Enrolled in med school	33,027***	34,125***
	(1,853)	(1,474)
Mean of dep. variable		
Observations	9,139	10,162
C. Difference		
$\beta_P - \beta_{NP}$	14,701*	94,450
	(7,892)	(150,195)
Mean of dep. variable for non persistent	77,663	78,613
F-stat	710.5	465.6

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the 1-digit ISCED level in paretheses. The specification controls for year of application, GPA category, years in which income is measured, gender, Dutch nationality, year of birth, parents cohort and income, separately for father and mother, which parent is a doctor, whether any other member of the family (among grandparents, uncles and aunts and older siblings) is also a doctor and 2 digit ISCED-F fields.

Table 2.21: Placebo analysis

	(1)	(2)	(3)
		Income	
Heterogeneity variable (H):	Male	Parental income	High GPA
$\beta_H - \beta_{NH}$	-4,625*	786	25,724
	(2,559)	(4,361)	(34,536)
Observations	10,419	10,419	10,419
F-stat	4.7	106.9	75.7

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Clustered standard errors at the 1-digit ISCED level in paretheses. Estimates are produced by simultaneously estimated regressions which follow the same model as in Equation 2.2, where the indicator for whether a parent is a doctor is substituted using variables for different heterogeneity margins. All specifications control for year of application, GPA category, years in which income is measured, gender, Dutch nationality, year of birth, parents cohort and income, separately for father and mother, which parent is a doctor, whether any other member of the family (among grandparents, uncles and aunts and older siblings) is also a doctor and 2 digit ISCED-F fields.

Table 2.22: Family practices and connections

	(1) Works for parent (pp)	(2) Same employer as parent (pp)
Parent is a doctor	0.96***	1.99***
	(0.16)	(0.24)
Mean of dep. variable	0.35	0.69
Observations	21,682	21,682
R-Squared	0.005	0.013

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in paretheses. All specifications control for gender, Dutch nationality, cohort fixed effects and year of application. The sample includes all individuals born on or after 1960 that have registered as doctors from 1999 and for which at least one parent can be observed.

Table 2.23: Likelihood of choosing a VWO high school

	Chose VWO school			
	(1)	(2)	(3)	(4)
Parent is a doctor	0.2250***	0.2219***	0.2250***	0.0568***
	(0.0119)	(0.0119)	(0.0119)	(0.0125)
Exam year FE		X	X	X
Household income control			X	X
Parental education level FE				X
Observations	70,906	70,906	70,906	70,906
Mean of Dep. variable	0.070	0.070	0.070	0.070
R-squared	0.031	0.035	0.031	0.110

Notes: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Robust standard errors in paretheses. All specifications control for gender, Dutch nationality and cohort fixed effects. The sample includes students finishing primary school in 2006 or 2007.

Table 2.24: Likelihood of choosing eligible tracks

	Chose eligible study track			
	(1)	(2)	(3)	(4)
Parent is a doctor	0.2367***	0.2354***	0.2356***	0.1934***
	(0.0252)	(0.0252)	(0.0252)	(0.0274)
Exam year FE		X	X	X
Household income control			X	X
Parental education level FE				X
Observations	4,345	4,345	4,345	4,345
Mean of Dep. variable	0.466	0.466	0.466	0.466
R-squared	0.022	0.025	0.025	0.039

Notes: Notes: * $p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Robust standard errors in paretheses. All specifications control for gender, Dutch nationality and cohort fixed effects. The sample includes students finishing primary school in 2006 or 2007 who chose a VWO high school.

Table 2.25: Effects of persistence on the probability of dropping out

	(1) Dropout (pp)
A. Parent is a doctor	
Enrolled in med school	-3.72**
	(1.57)
Mean of dep. variable	
Observations	1,364
B. Parent is not a doctor	
Enrolled in med school	-6.36***
	(0.44)
Mean of dep. variable	
Observations	9,058
C. Difference	
$\beta_P - \beta_{NP}$	2.64
	(1.66)
Mean of dep. variable for non persistent	1.17
F-stat	

Notes: * $p < 0.1$, **$p < 0.05$, ***$p < 0.01$. Clustered standard errors at the 1-digit ISCED level in paretheses. The specification controls for year of application, GPA category, years in which income is measured, gender, Dutch nationality, year of birth, parents cohort and income, separately for father and mother, which parent is a doctor, whether any other member of the family (among grandparents, uncles and aunts and older siblings) is also a doctor and 2 digit ISCED-F fields.

2.14 Theoretical Appendix

2.14.1 Decomposition of the difference in outcomes (Cunningham, 2021)

With π being the share of treated individuals, we have:

$$ATE = \pi ATT + (1-\pi)ATU =$$

$$\pi\mathbb{E}[Y_{i1}|D_i=1] - \pi\mathbb{E}[Y_{i0}|D_i=1] + (1-\pi)\mathbb{E}[Y_{i1}|D_i=0] - (1-\pi)\mathbb{E}[Y_{i0}|D_i=0] =$$

$$[\pi\mathbb{E}[Y_{i1}|D_i=1] + (1-\pi)\mathbb{E}[Y_{i1}|D_i=0]] - [\pi\mathbb{E}[Y_{i0}|D_i=1] + (1-\pi)\mathbb{E}[Y_{i0}|D_i=0]] =$$

$$\pi\mathbb{E}[Y_{i0}|D_i=1] + \mathbb{E}[Y_{i1}|D_i=0] - \pi\mathbb{E}[Y_{i1}|D_i=0] + \pi\mathbb{E}[Y_{i0}|D_i=1] +$$

$$\mathbb{E}[Y_{i0}|D_i=0] - \pi\mathbb{E}[Y_{i0}|D_i=0] + [\mathbb{E}[Y_{i1}|D_i=1] - \mathbb{E}[Y_{i1}|D_i=1]] +$$

$$[\mathbb{E}[Y_{i0}|D_i=1] - \mathbb{E}[Y_{i0}|D_i=1]] + [\mathbb{E}[Y_{i0}|D_i=0] - \mathbb{E}[Y_{i0}|D_i=0]]$$

$$\implies \quad \mathbb{E}[Y_{i1}|D_i=1] - \mathbb{E}[Y_{i0}|D_i=0] =$$

$$ATE - \pi\mathbb{E}[Y_{i1}|D_i=1] - \mathbb{E}[Y_{i1}|D_i=0] + \pi\mathbb{E}[Y_{i1}|D_i=0] + \mathbb{E}[Y_{i0}|D_i=1] +$$

$$\mathbb{E}[Y_{i0}|D_i=0] - \pi\mathbb{E}[Y_{i0}|D_i=0] + \mathbb{E}[Y_{i1}|D_i=1] - \mathbb{E}[Y_{i0}|D_i=1] +$$

$$\mathbb{E}[Y_{i0}|D_i=1] - \mathbb{E}[Y_{i0}|D_i=0]$$

$$\implies \quad \mathbb{E}[Y_{i1}|D_i=1] - \mathbb{E}[Y_{i0}|D_i=0] = ATE + [\mathbb{E}[Y_{i0}|D_i=1] - \mathbb{E}[Y_{i0}|D_i=0]] +$$

$$\mathbb{E}[Y_{i1}|D_i=1] - \pi\mathbb{E}[Y_{i1}|D_i=1] - \mathbb{E}[Y_{i1}|D_i=0] +$$

$$\pi\mathbb{E}[Y_{i1}|D_i=0] - \mathbb{E}[Y_{i0}|D_i=1] + \pi\mathbb{E}[Y_{i0}|D_i=1] + \mathbb{E}[Y_{i0}|D_i=0] - \pi\mathbb{E}[Y_{i0}|D_i=0]$$

$$\implies \quad \mathbb{E}[Y_{i1}|D_i=1] - \mathbb{E}[Y_{i0}|D_i=0] = ATE + [\mathbb{E}[Y_{i0}|D_i=1] - \mathbb{E}[Y_{i0}|D_i=0]] +$$

$$(1-\pi)\mathbb{E}[Y_{i1}|D_i=1] - (1-\pi)\mathbb{E}[Y_{i1}|D_i=0] +$$

$$(1-\pi)\mathbb{E}[Y_{i0}|D_i=1] - (1-\pi)\mathbb{E}[Y_{i0}|D_i=0]$$

$$\implies \quad \mathbb{E}[Y_{i1}|D_i=1] - \mathbb{E}[Y_{i0}|D_i=0] = ATE + [\mathbb{E}[Y_{i0}|D_i=1] - \mathbb{E}[Y_{i0}|D_i=0]] +$$

$$(1-\pi)\left[\mathbb{E}[Y_{i1}|D_i=1] - \mathbb{E}[Y_{i0}|D_i=1]\right] - (1-\pi)\left[\mathbb{E}[Y_{i1}|D_i=0] - \mathbb{E}[Y_{i0}|D_i=0]\right]$$

$$\implies \quad \mathbb{E}[Y_{i1}|D_i=1] - \mathbb{E}[Y_{i0}|D_i=0] =$$

$$ATE + [\mathbb{E}[Y_{i0}|D_i=1] - \mathbb{E}[Y_{i0}|D_i=0]] + (1-\pi)(ATT - ATU)$$

$$\implies \quad \mathbb{E}[Y_{i1}|D_i=1] - \mathbb{E}[Y_{i0}|D_i=0] = ATE + SB + (1-\pi)(ATT - ATU)$$

Because $ATE = \pi ATT + (1-\pi)ATU$:

$$\mathbb{E}[Y_{i1}|D_i = 1] - \mathbb{E}[Y_{i0}|D_i = 0] = \pi ATT + (1-\pi)ATU + SB + (1-\pi)(ATT - ATU)$$
$$\implies \quad \mathbb{E}[Y_{i1}|D_i = 1] - \mathbb{E}[Y_{i0}|D_i = 0] = ATT + SB$$

2.14.2 LATE and ATT

The exogenous instrument used in the analysis pushes into treatment those individuals that have been admitted to medical school at the time of their first application. It follows that the effects estimated are local average treatment effects (LATE). In other words, the obtained coefficients will estimate the treatment effects for *compliers*, i.e. those selected in at their first lottery round and who do enroll in medical school. This group will differ from the pool of all medical school enrollees, which would be classified as treated under the OLS specification, as it excludes *always takers*, i.e. those that have seen their first application rejected but still manage to enroll through subsequent applications. To understand the relationship between LATE and ATT, we can think of the following decomposition:

$$ATT = ATT(\text{compliers}) * Pr(\text{compliers}|D_i = 1) +$$
$$ATT(\text{always takers}) * Pr(\text{always takers}|D_i = 1)$$

Because $Pr(\text{compliers}|D_i = 1) + Pr(\text{always takers}|D_i = 1)$ and $ATT(\text{compliers}) = LATE$ this simplifies to

$$ATT = LATE * [1 - Pr(\text{always takers}|D_i = 1)] +$$
$$ATT(\text{always takers}) * Pr(\text{always takers}|D_i = 1) =$$
$$LATE + Pr(\text{always takers}|D_i = 1) * [ATT(\text{always takers}) - LATE]$$

As according to standard models of occupational choice we would observe a higher willingness to be in an occupation where higher returns are anticipated Roy (1951), it would be sensible to expect

$$LATE = ATT(\text{compliers}) < ATT(\text{always takers})$$

such that $ATT > LATE$. Similarly, for the selection bias, using the fact that $ATT \neq LATE$ we get

$$SB = ADO - LATE - [ATT - LATE]$$

which is going to be smaller than the estimated selection bias as long as $ATT > LATE$.

Chapter 3

UK Self-employment in the Twenty-first Century

This chapter is jointly co-authored with Robert Blackburn and Stephen Machin.

3.1 Introduction

Self-employment rates have increased in the United Kingdom for the 45 years preceding the Covid-19 outbreak. The trend has been, for long a time, an exception among the OECD countries (Blanchflower, 2000) and has caused self-employed workers to nearly double their representation in the labour force since the 1970s. Yet, growth has come with a significant change in the composition of this group. In particular, the decrease in self-employed with employees has been more than compensated by an important rise in the number of solo-self employed, especially under new alternative working arrangements and, more recently, the advent of the gig economy (Giupponi and Xu, 2020).

The Covid-19 pandemic outbreak, with its violent and long-lasting impact on the global and UK economy, has posed an unprecedented challenge to this renewed category of workers. The negative demand shocks triggered by the lockdowns have particularly affected self-employment-intensive sectors, such as hospitality, retail, construction and artistic activities. Despite the government effort to put in place new forms of support targeted specifically at businesses, self-employed workers have been hit extremely hard and have struggled recovering since (Blackburn, Machin, et al., 2021; Blackburn, Ventura, et al., 2022; Blundell and Machin, 2020; Blundell, Machin, and Ventura, 2020, 2021). On top of this, Brexit and the cost-of-living crisis have further exacerbated the economic situation of both households and businesses, with self-employed being once again disproportionately affected (Blackburn, Machin, et al., 2023a; Blackburn, Machin, et al., 2023b).

In this paper, we ask three main questions: how has the development of self-employment in the UK differed across ex-ante distinct demographic groups (i.e., according to gender, education and age)? How heterogeneous were the economic consequences of Covid-19 for these groups? Finally, how is the pool of self-employed workers changing as a consequence of these macro developments, and which categories of self-employed display greater resilience to these shocks? The novelty of our results stems from a set of newly collected survey data, the LSE-CEP Survey of Self-employment, which allows us to explore additional topics and issues that are usually not covered by standard sources and official labour market statistics. This is paired with results obtained using the UK Labour Force Survey, which helps us analyse in greater detail self-employment trends for a longer period of time.

We answer the questions above by first giving an overview of the recent trends in UK self-employment, with a particular attention to the heterogeneity in this group of workers. We show that, as previously documented, self-employment has significantly grown in the country in the last twenty years, with the biggest increase in solo self-employment. This is in stark contrast with the numbers for salaried employees, which have remained virtually unchanged in the same period. We also show that most of the increase is explained by a rise in women self-employed and explore how this change has differed across various sectors.

We then turn to how the Covid-19 crisis and its aftermath have affected the financial situation of self-employed workers. We show that the shares of self-employed individuals experiencing financial difficulties and with low income have remained stable in the years after the pandemic shock, and above their pre-emergency levels. Solo self-employed generally appear less affected than those with employees, but the heterogeneity analysis uncovers further differences across groups. Women, less educated, and younger workers have, on average, suffered a greater hit, but differences across these demographics are thinner when considering solo self-employed, while they appear more pronounced among those with employees. Similar patterns are also reflected in the rates at which these groups have claimed financial support during the Covid-19 crisis. Finally, there is significant demand from self-employed workers for a guaranteed form of emergency financial support, with around 45% willing to sacrifice up to 10% of their income for it. There is generally higher support for this type of policy from solo self-employed as well as some slight differences across demographic groups.

The paper tackles the third question by exploring post-pandemic flows of workers entering or exiting self-employment. It finds that the big drop in self-employed workers during and after the Covid-19 shock is attributable to a substantial rise in outflow which has not been matched by an equal growth in inflow in the following years. Industries have been differentially impacted, with the additional challenge of Brexit disproportionally affecting some of them. Upon leaving self-employment, solo workers and those with employees tend to sort differently into salaried jobs and unemployment or inactivity, with some significant dissimilarity across demographic groups. Finally, a willingness to pay experiment implemented in the survey reveals substantial interest from self-employed workers to move into employee jobs, even when if it requires an income sacrifice.

Related literature. At the beginning of the millennium, Blanchflower (2000) documented downward trends in self-employment in most OECD countries, with the UK being one of few notable exceptions. The same pattern is documented for more recent years by Taylor et al. (2017), Boeri et al. (2020) and Giupponi and Xu (2020) among others, with the latter two also highlighting a change in the composition of this group. As a consequence, the increase in freelancing and business creations has often been central in the policy discussion, with a specific focus on differential tax treatment (Gifford, 2023; Peachey, 2017), the regulation of new forms of self-employment, especially within the gig economy (Partington, 2019; Taylor et al., 2017), and the lack of some appropriate social insurance (Cribb, Emmerson, et al., 2023; Zafar, 2023). This paper provides new evidence of the recent trends in UK self-employment, with a special focus on the aftermath of the Covid-19 pandemic. The results will not only speak to the levels and changes of the self-employment population, but will also investigate variations in hours worked and origins (destinations) of new (former) self-employed workers. The newly collected data will also allow us to explore the latest tendencies in incomes.

Recent literature on self-employment has also made salient the vast heterogeneity present in this group of workers. The main distinction that has been brought forward by these contributions is that of the so-called "opportunity" and "necessity" self-employed. While the former term usually denotes proper entrepreneurs, with a defined business idea and growth ambitions, the latter indicates those workers whose best alternative to unemployment in the labour market is to be self-employed (Margolis, 2014). As this conceptual difference is usually unobservable, it is often proxied by self-employed with and without employees or, in the USA, self-employed in incorporated and unincorporated businesses. In this context, Levine and Rubinstein (2017a,b) show that these two types of self-employed workers differ both in their ex-ante characteristics and the type of businesses they engage in, and tend to entry at different times of the business cycle. Block and Sandner (2009) further show that this differential selection also reflects in opportunity self-employed remaining longer into self-employment, while Cowling and Wooden (2021) find that solo self-employed have a small probability of converting to employers. In the UK, Boeri et al. (2020) define solo self-employment as an intermediate status between employment and unemployment, with large scope for underemployment, while Giupponi and Xu (2020) consistently find them at the bottom of the earnings distribution. At the same time, in spite of the adverse economic consequences of going solo, Cribb and Xu (2020) show that this group experience on average higher level of wellbeing, likely associated with the higher independence and flexibility brought by the type of work arrangement. The distinction between entrepreneurs and other business owners is particularly relevant in the context of developing countries, where occasional solo self-employment is often prevalent and represents an alternative to unemployment and poverty, rather than conscious creation of a new businesses, and tends to disappear with development (Gindling and Newhouse, 2014; Pietrobelli et al., 2004; Temkin, 2009). Additional investigation of the heterogeneity in self-employment has concerned gender, education level and age. Findings from previous literature around these traits have highlighted that self-employment appears to be an alternative to part-time work for many women (Georgellis and Wall, 2005) who, consistently, earn usually less than men (Hundley, 2001; Williams, 2000). Furthermore, workers with higher education are more likely to succeed in self-employment (Kangasharju and Pekkala, 2002; Robinson and Sexton, 1994) and the probability of leading a business rises with age (Blanchflower, 2000; Karoly and Zlsslmopoulos, 2004). As these characteristics are usually analysed in isolation, we contribute to this stream of literature by further decomposing self-employment and investigating whether workers with and without employees, and that are ex ante different by the traits above, display differential patterns in the labour market.

In the last years, self-employed workers have also been widely discussed among the public due to the strong negative impact they suffered from the recent pandemic, both in the UK (Blundell and Machin, 2020; Blundell, Machin, and Ventura, 2020, 2021; Yue and Cowling, 2021) and around the world (Beland et al., 2020; Graeber et al., 2021; Richter and Patel, 2022). In this paper, we build on our previous policy work and use novel survey data to explore how the Covid-19 shock differentially affected various types of self-employed. Specifically, our ad hoc survey modules allow us to investigate uncharted aspects of workers' plans and aspirations in this new post-pandemic phase, giving us hints of what this group may look like in the near future.

The rest of the paper is organized as follows. Section 3.2 provides details on the different sources of data used in the analysis. Section 3.3 presents the descriptive results from the Labour Force Survey for recent trends in level and changes in the self-employed workforce and introduces the different types of heterogeneity considered in the paper. Section 3.4 makes use of the newly collected data on self-employment to document the workers' situation during and after the Covid-19 shock. Section 3.5

employs a mix of primary and secondary data to investigate the current flows out of self-employment as well as workers' intentions to leave. Section 3.6 concludes and highlights directions for future research.

3.2 Data

This chapter uses data from two main sources, combining both secondary and primary data. This section explains how each of them is utilised and thoroughly describes the newly collected data.

3.2.1 The LSE-CEP Survey of UK Self-Employment

Covering around 1,500 individuals, the LSE-CEP Survey of UK Self-employment (henceforth, LCSUSE) was first used by Blundell and Machin (2020) in May 2020 to collect information on British self-employed workers. Since then, we have regularly released additional follow-up rounds, up to its seventh in June 2023. The data resulting from these survey waves are a series of repeated cross-sections, made representative of the population using weights drawn from the UK Labour Force Survey (henceforth, LFS). The project originated with the purpose of monitoring the self-employed living and working conditions following the unprecedented Covid-19 shock, at a time when official statistics were yet to be released. The focus on this specific category of workers is justified by both the exceptional diversity of this group, which made predictions on their outcomes potentially ambiguous, as well as the significant growth observed in their population over the last two decades (ibid.). Additionally, the crisis had further highlighted the lack of appropriate social insurance for the self-employed. Excluded from the furlough scheme, they were offered novel forms of economic support, whose popularity and uptake rates required new data to be studied.

We launched six additional rounds of the LCSUSE throughout the following three years. The core set of questions remained constant and collected information on demographics, income and profit levels, labour supply, working habits and willingness to pay for extra support from the government. The remainder was edited at each round, to reflect different topics, salient at the time of release, such as inflation and the cost of living crisis, access to pensions, voting intentions and willingness to leave self-employment. Approximately 1,500 respondents were interviewed at each round, and the survey was distributed online by an external company. Once valid responses are retained and weighted the complete seven-waves dataset includes 9,414 observations.

The strength of LCSUSE lies in its ad hoc modules, which have been designed to delve into specific aspects of the economic life of the UK self-employed, that are both relevant to the moment the survey is released and have not been covered by alternative sources. In this paper, we use these information to provide insights on the workers' government support uptake, willingness to pay for additional support and intentions to leave self-employment. The appendix in Section 3.10 reports the survey questions utilized in this chapter.

3.2.2 UK Labour Force Survey

The UK LFS allows us to explore long-term trends in self-employment as well as flows in and out of this group. To this purpose, we use the five-quarter longitudinal survey from January 2000 to June 2023, so that workers entering or leaving self-employment can be defined based on their status in the previous year. Observations are weighted to be representative of the national population.

3.3 Trends in UK self-employment

It is a known fact that self-employment in the UK has grown significantly since the 1970s, making it among the OECD countries with the highest levels and growth in this group of workers (Giupponi and Xu, 2020). Figure 3.1 shows how up to the end of 2019 the number of self-employed had increased by almost 60% with respect to the first quarter of 2000. In contrast, the increase in employees was three times smaller in the same period. After that, the Covid-19 pandemic has brought a large decrease in 2020, with self-employed dropping by over 10% (Blundell, Machin, and Ventura, 2021), but leaving employees largely unaffected.

This unprecedented increase was accompanied by an important shift in the composition of self-employed. It is evident, in the first place, when looking at how weekly hours worked changed for this group of workers (Figure 3.2). This aspect is likely to reflect the recent rise of alternative work arrangements, such as zero hour contracts (Datta et al., 2019), as well as of the opportunities for flexible working provided by the gig economy. The latter has effectively transformed working relationships of those operating through digital platforms, which often obey to the the idea of a "on-demand economy" (Todolí-Signes, 2017).

At the same time, recent literature has highlighted that the increase in self-employment was driven by a fast growth in solo self-employed, as opposed to self-employed who are also employers (Giupponi and Xu, 2020). The second category has instead experienced a 46% decrease until the end of 2019, as confirmed by Figure 3.3, with further drops after the pandemic. Figure 3.4 shows novel evidence of the decrease in hours worked having been equally led by solo self-employed with a 20% reduction just before the Covid-19 shock compared to its 2000 levels (in contrast to 10% for those with employees). This is consistent with this group being the most likely to engage in many of the newly available type of work opportunities, such as those on digital platforms.

The distinction between solo self-employed and self-employed with employees is possibly the most used in the literature when exploring heterogeneities in self-employment. It typically refers to the contrast between small, individually led businesses with little potential to grow, and proper entrepreneurs, with the employer role of the latter representing creation of economic value. In the context of the US, this discrepancy is sometimes proxied by unincorporated and incorporated businesses. We argue that the pool of self-employed workers may, in fact, encompass broader heterogeneity, which might be correlated, although not fully captured, by this first distinction. For this reason, we build on previous literature to select a few candidate characteristics and test whether they can more accurately describe different groups existing among the self-employed. In particular, we choose to explore the role of gender, education and age, in conjunction with the "solo" and "with employees" categories.

We start by investigating whether these demographic groups have displayed differential trends in the last two decades. The three panels in Figure 3.5 show the quarterly shares of workers that are self-employed in each category, by whether they are solo or with employees. Panel A uncovers interesting differences: while the share of solo self-employed men has increased by 27% and that of women self-employed with employees has decreased by 32%, the other two groups have experienced even more substantial changes. In particular, the share of women that choose to be solo self-employed has increased by 137% between January 2000 and its peak in December 2019, while the share of men in self-employment who hire other workers has dropped by 63% in the same period. On the other hand, Panel B shows that having a degree seem to be a less relevant distinction, with differently educated workers entering solo self-employment at similar rates and growing over time. More educated self-employed with employees have instead decreased by two thirds, against the 35% of those with no degree, up to the first quarter of 2020. Finally, Panel C reports differences by age. We choose to use 40 years old as a significant threshold, although results are similar if splitting the sample differently around it. In this case, although trends are the same for each age category, we observe slightly more evident changes for older solo workers, increased by 67% (23% for those below 40 years old) until the end of 2019. Self-employed with employees have dropped in the same period by 56% and 70% for workers above and below 40%, respectively.

Looking at hours worked in Figure 3.6, we observe that self-employed women with employees have historically worked nearly as many hours as male solo self-employed. In recent years, the level increased, with women catching up just before the Covid-19 pandemic. Female solo self-employed, however, consistently work fewer hours than their male counterparts. Panel B shows that for both solo self-employed individuals and those with employees, having a degree has not led to significant differences in hours worked over the past years. However, among solo self-employed, convergence occurred due to a decrease in hours worked by highly educated workers. Finally, age does not seem to cause discrepancies in labor supply.

Overall, we find that the increase in solo self-employment has been largely driven by rapid growth in the number of women joining this group, along with a significant rise among older solo self-employed. Conversely, the decline in "entrepreneurial" self-employment can be attributed to a decrease in men and highly educated workers choosing this business type over the years. Next, we deepen this analysis to better understand the differences among these groups. To this end, we document the trajectories of workers entering self-employment, tracking their labor market status prior to self-employment and their industry of destination. In setting up this analysis, we draw on literature that distinguishes between different types of self-employment (i.e., solo and with employees), which in turn attract workers with distinct backgrounds, abilities, and outside options. Table 3.1 begins by examining workers whose entry into self-employment is recorded in the LFS data, reporting the distribution of new self-employed by their employment status in the previous year. We use data up to the end of 2019 to document the pre-pandemic situation, thereby avoiding potential outliers in recent years. First, the average distribution of self-employed individuals with and without employees shows that those with employees are 39% more likely to have a background as an employee. In contrast, solo self-employed individuals are about 3.5 times more likely to have been unemployed and 1.8 times more likely to have been inactive prior to joining self-employment. Zooming into the finer groups, gender comes up again as one of the most relevant margins. Men are 24% more likely to have been employees and 60% more likely to have been unemployed compared to women before becoming self-employed without employees. For those with employees, these figures are 16% and 100%, respectively. Conversely, women entering solo self-employment are 1.4 times more likely to come from economic inactivity, and those starting with employees are 2.7 times more likely to do so. It is also notable that women are not only more likely to be inactive prior to self-employment than men, but they are also much more likely to have been inactive rather than unemployed. This suggests that while some men may be using self-employment (especially when going solo) to escape unemployment, for women it may be indeed a way to transition from outside to inside the labour market. Although education and age are less influential than gender in determining self-employment origins, it is noteworthy that among employers workers with a degree are 11% more likely to have an employee background compared to those without a degree. Conversely, those without a degree are 1.45 times more likely to have been inactive.

Where do these groups land once in self-employment? Figure 3.7 displays the distribution of self-employed individuals across industries, broken down by the selected characteristics. While there are few common features across groups, such as self-employed individuals with employees being typically concentrated in Distribution and Hospitality, some categories show clear industry specialization. Unsurprisingly, women self-employed are concentrated in Education and Health but, when having employees, they are also widely represented in Distribution and Hospitality. On the other hand, Construction is dominated by men. The latter is also common among workers with no degree, as is Distribution and Hospitality, but Education and Health usually employ highly educated individuals. Finally, workers above the age of 40 are slightly more likely than younger ones to be in Finance.

3.4 Self-employment and the Covid-19 crisis

In the spring of 2020, the UK and global economy were hit by the Covid-19 pandemic. Since the very beginning, the self-employed have been strongly affected by the new crisis (Blundell and Machin, 2020), in particular because of the social distancing measures and the physical constraints imposed by the government on businesses. The emergency has also produced long lasting consequences for the self-employed who, for the following years, have experienced financial difficulties and low incomes (Figure 3.8)[1]. Table 3.2 displays the results of regression analyses delving into the heterogeneous experiences of different categories of self-employed workers. The first line highlights that solo self-employed have, on average, unambiguously been less affected financially than self-employed with employees, reporting trouble paying

[1]Having financial difficulties is defined as reporting trouble paying for basic expenses. Low incomes are below £1,000.

for basic expenses 15 to 44% less often than average. This can be rationalised in light of the greater flexibility solo freelancers have to organize their work, as well as their usually lower costs of running a business, which together may have attenuated financial hardships. Conditional on industries, men are 3.9 percentage points less likely than women to have had financial issues in the considered period. Interestingly, the sign reverses when interacting gender with the type of self-employment, although the coefficient is less precisely estimated, suggesting that gender differences are smaller for solo self-employed (3.3-3.7 percentage points) than for those with employees (7.4-8.3 percentage points). Workers with higher education have also fared better (5.5 percentage points) than those without a degree, and this is not significantly different for self-employed with and without employees. Finally, individuals above 40 years old are on average 16.7 percentage points less likely than younger once to experience troubles paying with basic expenses, but the difference increases to 27 percentage points for self-employed who also have employees. Overall, while self-employed with a degree seem to have benefited from it regardless of being solo or having employees, differences across gender and age are more pronounced for employer self-employed while decreases, at times considerably, among solos. Among the considered groups, older workers, both with and without employees, seem to have been the least affected. Interestingly, this remains true when controlling for how long they have been self-employed for.

In the UK, the emergency has also highlighted the challenges brought by extending social insurance to self-employed, with new temporary sources of economic support made available for this group. The Self-Employment Income Support Scheme (SEISS) has provided, over five rounds, grants up to £7,500 to those reporting losing some business due to the Covid-19 crisis. Payments were pegged to previous year's profit, rather than linked to the actual financial loss experienced, and businesses with profits above £50,000 were excluded. Crucially, newly self-employed were excluded, as trade in the previous tax year was one of the requirements. According to the UK government, this has made eligible approximately 3.36 million individuals (HMRC, 2022), i.e. circa 80% of self-employed. Our survey data show that only 36% applied for this new fund (Figure 3.9). This should not be interpreted as support not being needed. In fact, it comes with relevant lack of awareness regarding the policy, with 38% of those who had not applied as of September 2020 being not sure of their eligibility and 45% of self-employed unsure of its generosity level (Blundell, Machin, and Ventura, 2020). On the other hand, total claims for Universal Credit have doubled in the first six month of 2020 (Department for Work and Pensions, 2024), with 20% of self-employed reporting applying and 16%, roughly corresponding to 670,000 individuals, claiming for the first time. An additional 28% also applied for other types of support (Figure 3.9). Table 3.3 gives some insights on how self-employed groups differentially applied for financial support. Note that any type of support is considered in this case. Moreover, heterogeneous application rates may be a joint consequence of the degree to which the specific group has been affected by the crisis, their awareness of the potential sources of support and eligibility criteria, as well as the group's actual likelihood to be eligible. The estimates mostly track those displayed in Table 3.2, with solo self-employed being generally less likely to claim, men demanding support more rarely, although less so when without employees, and more educated and older workers also more prone to claim.

The evidence above prompts the question of whether a regularly established form of emergency coverage, comparable to employee's furlough, may be desirable for self-employed workers. In order to tackle this issue, which is currently unanswered by the literature, the LCSUSE includes a choice experiment that elicit self-employed willingness to pay for financial support in times of crisis. To this end, the questions utilize the generosity of the coverage offered by the SEISS support package[2] to assess how much (pre-tax) income workers would be prepared to sacrifice to be assured income support in the face of future pandemics and other economic shocks. Figure 3.10 displays the average willingness to pay this type of insurance, across all survey rounds, by whether self-employed have employees. The graph appears striking as, regardless of the type of self-employed, around 45% of workers would be willing to sacrifice up to 10% of their income to be sustained by the government in times

[2]Note that this was lower in the second round of SEISS funding. The second survey round has therefore used the coverage rate of the time.

of crisis. Moreover, demand appears steeper for solo self-employed, with a greater share of them willing to pay a small fee for emergency support. Figure 3.11 further decomposes willingness to pay by gender, education and age. The main differences across categories appear to be for solo self-employed between those with and without a degree, with the latter showing higher demand for coverage even for high income sacrifices, and for those with employees between workers above and below 40 years old, with older workers being more willing to give up on small shares of their income to obtain emergency support. Finally, Figure 3.18 in the appendix of Section 3.9 also shows the same figure for the first and last survey rounds only, in order to explore the role of salience of potential emergencies in the early Covid-19 months. In the three years between May 2020 and May 2023 the demand for additional support has flattened, with the marginal workers, who would have only chosen to sacrifice income for low shares, slightly dropping, but same or higher willingness to pay for more substantial income sacrifices.

Overall, we uncover a greater impact of the Covid-19 pandemic on self-employed with employees, women, less educated and young workers, with smaller differences across this traits among solo self-employed. This pattern is reflected in applications for financial support, except in the case of older workers who, although being less dramatically affected, exhibit higher rates of claims for funding. This could reflect a greater awareness around the availability of resources brought by their experience. The willingness to pay experiment obtained using the LCSUSE also reveals substantial and widespread demand for an institutionalised form of emergency support for self-employment.

3.5 A new exodus from self-employment

What does the post-Covid world look like for the self-employed? As shown in Figure 3.1, recent months have not yet seen the number of self-employed pick up again after the drop caused by the pandemic. This is in contrast with the stable trend exhibited by salaried employees, for whom the measures readily implemented, such as the furlough scheme, seem to have largely absorbed the shock. Diving deeper into the flow composition uncovers two main aspects, depicted in Figure 3.12. First, the drop in self-employed workers observed at the beginning of the Covid-19 pandemic derived from both a drop in inflow as well as a sizeable rise in outflow from self-employment. Secondly, while the outflow has return to pre-crisis levels, this has not been compensated by growth in inflow, which instead remained low.

The post-pandemic change in the number of workers engaged in self-employed did not affect all fields equally. Figure 3.13 displays the percentage change by industry between the end of 2019 (before Covid-19 hit) and the end of 2022. Three years after the shock, no sector has returned to its previous levels, manifesting once again the severity of the hit taken by self-employment. The most affected workers appear to be the ones in Education and Health, consistently with the constraints placed on these industries by the pandemic. It is also important to notice that the timing of Covid-19 partly overlapped that of the official execution of Brexit. This is likely to have discouraged foreign workers from starting a business in sectors that were traditionally highly populated by immigrants, such as Construction and Banking and Finance (Migration Watch, 2016).

We further investigate the path of workers moving out of self-employment by looking at their status after leaving. Figure 3.14 exploits LFS data to show the full trends in destinations after the change for self-employed with and without employees[3]. The first takeaway from the graphs is that outside options for these two groups starkly differ. While the share of employer self-employed that go into employee jobs fluctuates around 80%, this is more than 25% lower for solo self-employed, who instead go into unemployment or inactivity close to 50% of the time. This should not be surprising in light of solo self-employment being often assimilated to "necessity entrepreneurship" in the literature, as well as displaying different cyclicality with respect to those with employees (Levine and Rubinstein, 2018). Table 3.4 confirms the tendency of solo self-employed to be less likely to become employees when leaving self-employment, up to 83% less than average in the most complete specification. The regressions capture no differences across gender, but solo self-employed with a

[3]Note that we exclude for simplicity the destination "Other" including, among others, family unpaid work or government schemes.

degree appear significantly more likely than less educated workers to obtain employee jobs. Finally, workers above 40 years of age are less likely to move into salaried employment than their younger counterparts.

The different outside options faced by the heterogeneous group of the self-employed opens once again the question of whether working conditions that are closer to those of employees may be more desirable, and at which cost, to business owners. In order to gauge their willingness to pay for a salaried position, the LCSUSE frames a hypothetical experiment with randomised earnings streams which asks respondents to think about: i) whether they could work in an employee job; and ii) their willingness to pay for a move through an income sacrifice or move for an enhanced income. Figure 3.15 shows the basic results of the experiment by whether the respondent is a solo self-employed or has employees, based on the assumption that their financial stability, commitment to running an enterprise and self-employment experiences may differ. The vertical axis shows the percentage of the self-employed who would be prepared to change their job and the horizontal axis shows how much of their current income they would need to allow a move. While for both types about 60% of workers would switch to salaried employment for a significant increase in income, self-employed with employees are also more likely to be willing to sacrifice income in order to obtain an employee position. On average, 40% of self-employed would switch for the same income they currently earn. Figure 3.16 delves deeper into the question by exploring heterogeneity in demand. We do not observe significant differences across gender or education, but the age margin appears relevant in this case. While willingness to pay is generally higher for younger workers, this is particularly the case for self-employed with employees, where 40% would agree to the move for a 30% salary sacrifice and 60% for no change. Although the reasons for it are not clear in the data, this may reflect dissatisfaction of youths with their conditions in self-employment.

Overall, the results show a significant share of the self-employed would be willing to change their employment status, even if that came with a reduction in income. Why don't they then? In the survey, we ask respondents whether they believe it would be hard to find a job as a salaried employee. The answers reveal that only 25% deem it easy, while 47% of them believe it would be hard or very hard. A few reasons emerge for this, as displayed in Figure 3.17. The highest obstacle appears to be the lack of similarly paid jobs in the employee sector. Interestingly, for solo self-employed concerns regarding the lack of skills, old age and health issues are also higher than for self-employed with employees. These findings may, once again, reflect solo workers' lower levels of income security and access to the social security infrastructure (Cieślik and Stel, 2024), and confirm the views that have identified employer self-employed as more motivated by their enterprise and career ambitions (Boeri et al., 2020). Table 3.5 completes the picture by showing the results of a few descriptive regressions that investigate which groups are more likely to believe they would struggle finding an employee job. We find that solo self-employed appear more concerned regarding a potential move, although the coefficient become insignificant once controlling for age, and changes sign for males. Workers with a degree are generally more optimistic about their chances, while older self-employed are 21% more likely than average to think it would be hard, regardless of whether they have employees. Taken together, these last findings suggest a situation of distress for self-employed workers, who long for a more stable financial situation and working conditions, but lack what it takes to make a move.

3.6 Conclusions

This chapter documents recent trends of the UK self-employed and presents novel evidence on their situation during and after unprecedented times of crisis. The newly collected survey data allows us to explore uncharted questions, such as workers' demand for regular emergency support and their willingness to pay for employee jobs.

The analysis uncovers new features of the pre-pandemic growth in UK self-employment, which has been particularly fueled by the rise in women-led solo businesses. Exploiting the richness of UK LFS data, it has also highlighted differences across education levels and age.

The LCSUSE data additionally lets us investigate the experiences of different groups of self-employed during the Covid-19 crisis. We confirm previous findings claiming that businesses were hit particularly hard by the pandemic, and delve further into the question by looking at their claim rate for financial support, as well as their desire for an institutionalized form of emergency fund. In this context, we find that 45% of workers would be willing to sacrifice up to 10% of their current income for this type of coverage.

The last part of this chapter explores post-pandemic flows in and out of self-employment, as well as workers' desires for a more stable, employee-like type of status. The findings suggest the implementation of Brexit policies may have contributed to the stall in inflow into self-employment, which has in turn maintained the number of self-employed substantially lower than its pre-crisis levels. They also highlight that some groups of self-employed, and in particular those with employees, are more likely to move into salaried employment when changing their status. Finally, the survey found large willingness to move to employee jobs, with four in ten workers prepared to move for the same income. While this is likely to reflect a desire for income stability and a guaranteed social security net, an actual move is deemed hard by virtually half of the self-employed, who remain trapped in their current conditions.

By highlighting some main heterogeneity margins, this work has put forward the idea that the pool of self-employed workers, in particular in the UK, is very diverse and heterogeneous, over and beyond the standard distinction of opportunity and necessity entrepreneurs. While the chapter has provided a first attempt at disentangling some of this complexity, more work is required to fully understand how to effectively classify different types of self-employment. Given the financial insecurity faced by this group, understanding this aspect will be of paramount importance to direct policy towards interventions that can guarantee stability as well as new foundations for growth and innovation.

References

Beland, Louis-Philippe, Oluwatobi Fakorede, and Derek Mikola (2020). “Short-term effect of COVID-19 on self-employed workers in Canada”. In: *Canadian Public Policy* 46.S1, S66–S81.

Blackburn, Robert, Stephen Machin, and Maria Ventura (2021). “Covid-19 and the self-employed-18 months into the crisis”. In.

— (2023a). “The self-employment trap?” In.

Blackburn, Robert, Steve Machin, and Maria Ventura (2023b). “From Covid-19 to collapse? The self-employed and the cost of living crisis”. In.

Blackburn, Robert, Maria Ventura, and Stephen Machin (2022). “Covid-19 and the self-employed-A two year update”. In.

Blanchflower, David G (2000). “Self-employment in OECD countries”. In: *Labour economics* 7.5, pp. 471–505.

Block, Jörn and Philipp Sandner (2009). “Necessity and opportunity entrepreneurs and their duration in self-employment: evidence from German micro data”. In: *Journal of industry, competition and trade* 9, pp. 117–137.

Blundell, Jack and Stephen Machin (2020). *Self-employment in the Covid-19 crisis*. Tech. rep. Centre for Economic Performance, London School of Economics and Political ...

Blundell, Jack, Stephen Machin, and Maria Ventura (2020). *Covid-19 and the self-employed: Six months into the crisis*. Centre for Economic Performance, London School of Economics and Political ...

— (2021). *Covid-19 and the self-employed: Ten months into the crisis*. Centre for Economic Performance, London School of Economics and Political ...

Boeri, Tito et al. (2020). “Solo self-employment and alternative work arrangements: A cross-country perspective on the changing composition of jobs”. In: *Journal of Economic Perspectives* 34.1, pp. 170–195.

Cieślik, Jerzy and André van Stel (2024). “Solo self-employment—-Key policy challenges”. In: *Journal of Economic Surveys* 38.3, pp. 759–792.

Cowling, Michael Leith and Mark Wooden (2021). "Does solo self-employment serve as a 'stepping stone'to employership?" In: *Labour Economics* 68, p. 101942.

Cribb, Jonathan, Carl Emmerson, et al. (2023). *Challenges for the UK pension system: the case for a pensions review*. R255. IFS Report.

Cribb, Jonathan and Xiaowei Xu (2020). *Going solo: how starting solo self-employment affects incomes and well-being*. Tech. rep. IFS working papers.

Datta, Nikhil, Giulia Giupponi, and Stephen Machin (2019). "Zero-hours contracts and labour market policy". In: *Economic Policy* 34.99, pp. 369–427.

Department for Work and Pensions (2024). *Universal Credit statistics, 29 April 2013 to 9 May 2024*. Tech. rep. URL: https://stat-xplore.dwp.gov.uk/webapi/jsf/login.xhtml.

Georgellis, Yannis and Howard J Wall (2005). "Gender differences in self-employment". In: *International review of applied economics* 19.3, pp. 321–342.

Gifford, Charlotte (Nov. 4, 2023). "The election 'open goal' that the Tories have ignored". In: *The Telegraph*. URL: https://www.telegraph.co.uk/money/tax/red-tape-crisis-choking-britains-self-employed/.

Gindling, Thomas H and David Newhouse (2014). "Self-employment in the developing world". In: *World development* 56, pp. 313–331.

Giupponi, Giulia and Xiaowei Xu (2020). "What does the rise of self-employment tell us about the UK labour market". In: *London: Institute of Fiscal Studies*, pp. 1–29.

Graeber, Daniel, Alexander S Kritikos, and Johannes Seebauer (2021). "COVID-19: a crisis of the female self-employed". In: *Journal of Population Economics* 34, pp. 1141–1187.

HMRC (2022). *Self-Employment Income Support Scheme (SEISS)*. Tech. rep.

Hundley, Greg (2001). "Why women earn less than men in self-employment". In: *Journal of labor research* 22.4, pp. 817–829.

Kangasharju, Aki and Sari Pekkala (2002). "The role of education in self–employment success in Finland". In: *Growth and change* 33.2, pp. 216–237.

Karoly, Lynn A. and Julie Zlsslmopoulos (2004). "Self-employment among older US workers". In: *Monthly Labor Review*, p. 25.

Levine, Ross and Yona Rubinstein (2017a). "Entrepreneurship and liquidity constraints over the business cycle". In: *Unpublished manuscript*.

— (2017b). "Smart and illicit: who becomes an entrepreneur and do they earn more?" In: *The Quarterly Journal of Economics* 132.2, pp. 963–1018.

— (2018). *Selection into entrepreneurship and self-employment*. Tech. rep. National Bureau of Economic Research.

Margolis, David N (2014). "By choice and by necessity: Entrepreneurship and self-employment in the developing world". In: *The European Journal of Development Research* 26, pp. 419–436.

Migration Watch (2016). *Migration Watch UK – Written evidence (LMT0007)*. Tech. rep. URL: https://committees.parliament.uk/writtenevidence/77918/html/.

Partington, Richard (June 28, 2019). "Gig economy in Britain doubles, accounting for 4.7 million workers". In: *The Guardian*. URL: https://www.theguardian.com/business/2019/jun/28/gig-economy-in-britain-doubles-accounting-for-47-million-workers.

Peachey, Kevin (Mar. 9, 2017). "Spring Budget 2017 tax rise: What's the fuss about?" In: *BBC News*. URL: https://www.bbc.com/news/business-39217655.

Pietrobelli, Carlo, Roberta Rabellotti, and Matteo Aquilina (2004). "An empirical study of the determinants of self-employment in developing countries". In: *Journal of International Development* 16.6, pp. 803–820.

Richter, Jack I and Pankaj C Patel (2022). "Impact of the COVID-19 pandemic on the hours lost by self-employed racial minorities: evidence from Brazil". In: *Small Business Economics* 58.2, pp. 769–805.

Robinson, Peter B and Edwin A Sexton (1994). "The effect of education and experience on self-employment success". In: *Journal of business Venturing* 9.2, pp. 141–156.

Taylor, Matthew et al. (2017). *Good work: The Taylor review of modern working practices*. Vol. 116. Department for Business, Energy & Industrial Strategy London.

Temkin, Benjamin (2009). “Informal self-employment in developing countries: entrepreneurship or survivalist strategy? Some implications for public policy”. In: *Analyses of Social Issues and Public Policy* 9.1, pp. 135–156.

Todolí-Signes, Adrián (2017). “The ‘gig economy’: employee, self-employed or the need for a special employment regulation?” In: *Transfer: European Review of Labour and Research* 23.2, pp. 193–205.

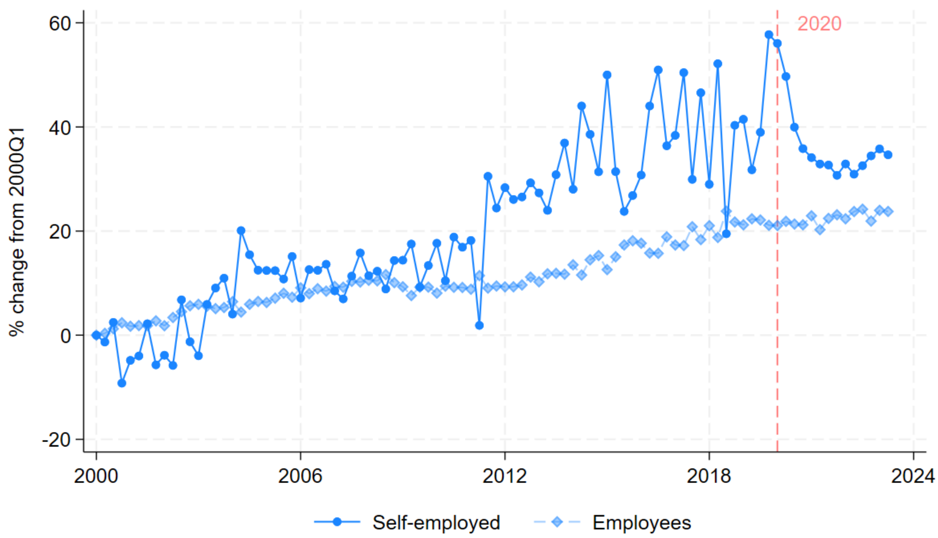
Williams, Donald R (2000). “Consequences of self-employment for women and men in the United States”. In: *Labour Economics* 7.5, pp. 665–687.

Yue, Wei and Marc Cowling (2021). “The Covid-19 lockdown in the United Kingdom and subjective well-being: Have the self-employed suffered more due to hours and income reductions?” In: *International Small Business Journal* 39.2, pp. 93–108.

Zafar, Aamina (Oct. 10, 2023). “Worry self-employed are ’being left behind’ as third have no pension”. In: *Financial Times*. URL: https://www.ftadviser.com/pensions/2023/10/10/worry-self-employed-are-being-left-behind-as-third-have-no-pension/.

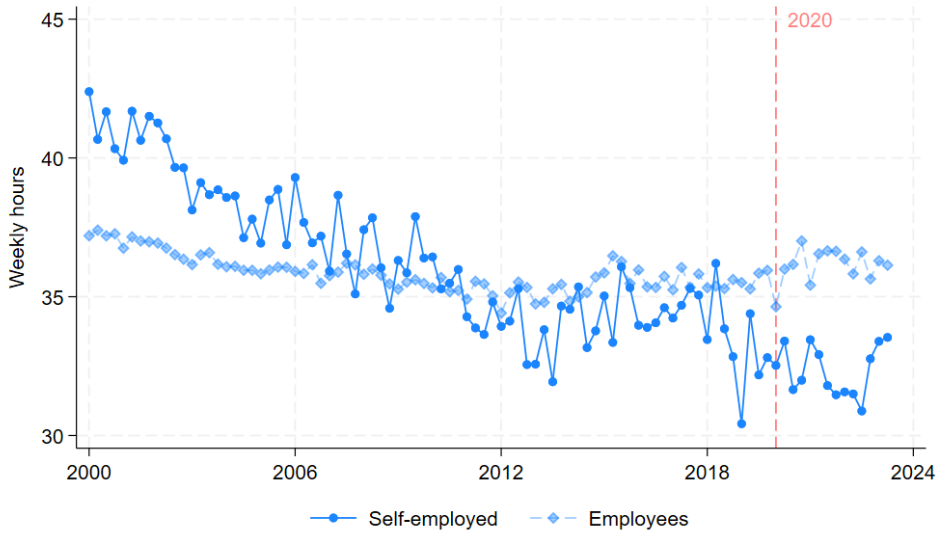
3.7 Figures

Figure 3.1: Change in self-employment vs. employees



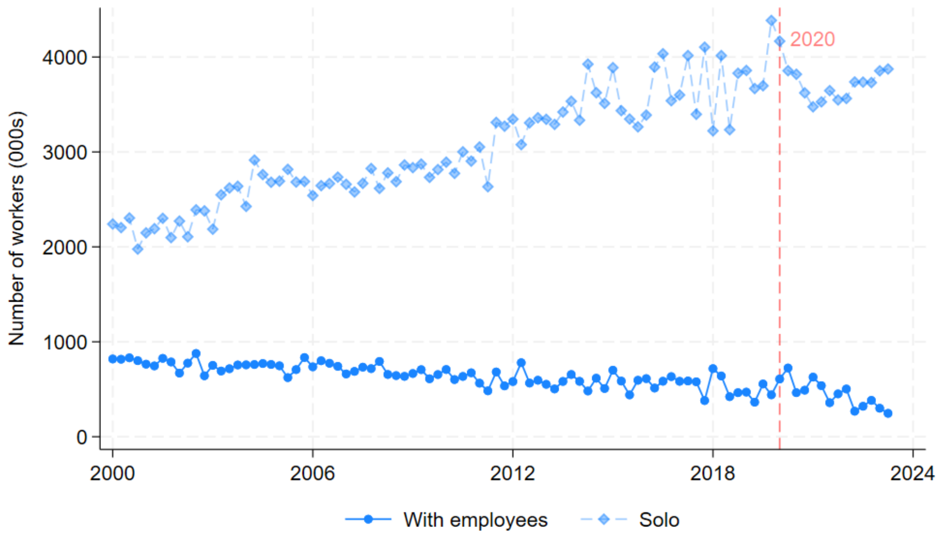
Notes: The plot displays the percentage change in the number of workers for employees and self-employed since 2000Q1. Data source: Five-quarter longitudinal UK LFS 2000Q1 - 2023Q2.

Figure 3.2: Weekly hours self-employment vs. employees



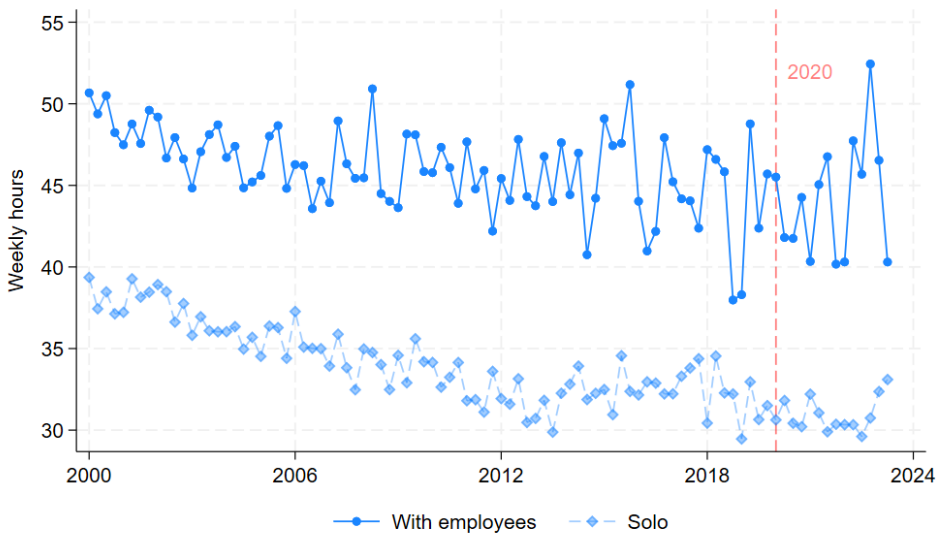
Notes: The plot displays weekly hours worked for employees and self-employed since 2000Q1. Data source: Five-quarter longitudinal UK LFS 2000Q1 - 2023Q2.

Figure 3.3: Workers in solo vs. with employees self-employment



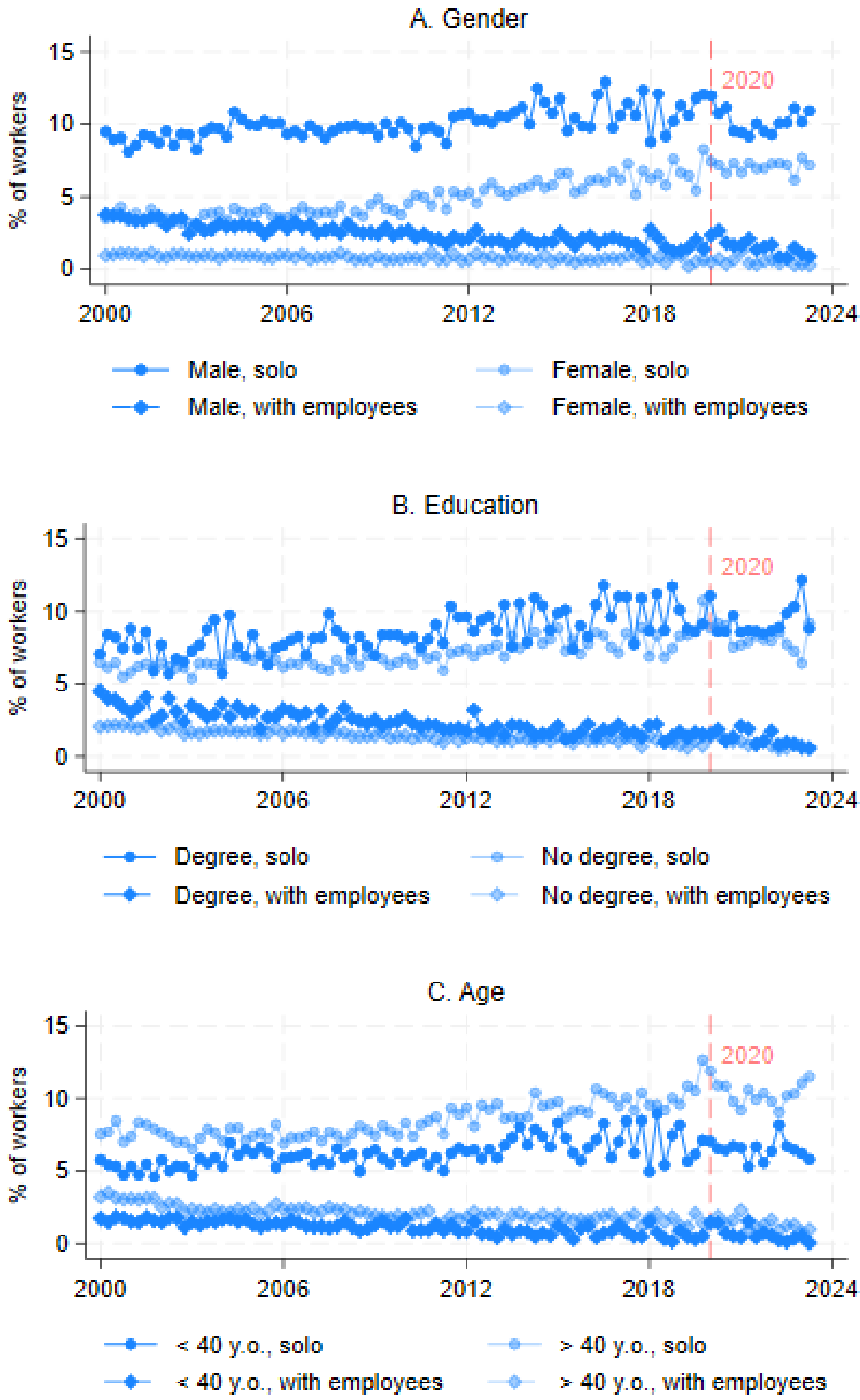
Notes: The plot displays number of workers for solo self-employed and self-employed with employees since 2000Q1. Data source: Five-quarter longitudinal UK LFS 2000Q1 - 2023Q2.

Figure 3.4: Weekly hours for solo vs. with employees self-employed



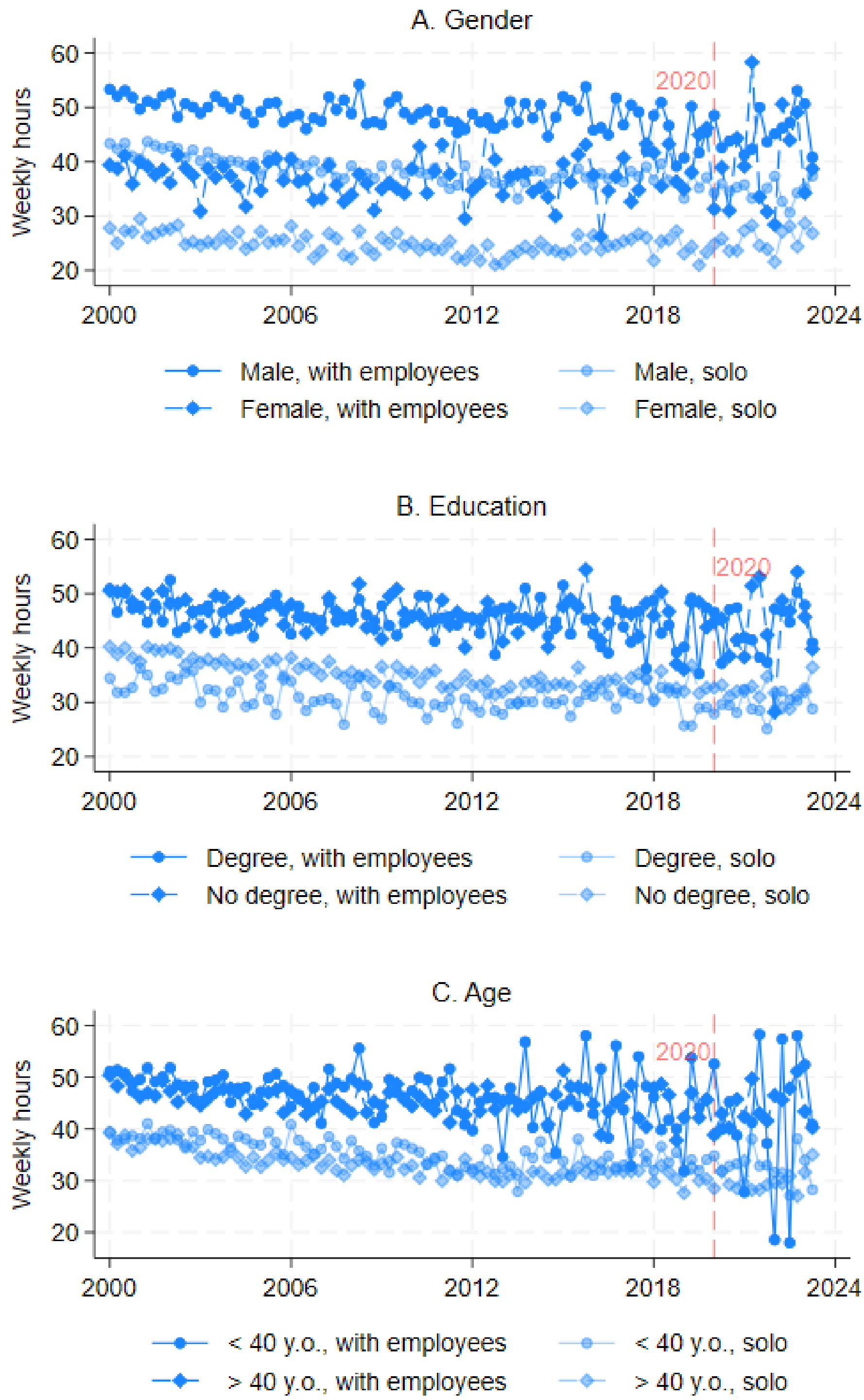
Notes: The plot displays weekly hours worked for solo self-employed and self-employed with employees since 2000Q1. Data source: Five-quarter longitudinal UK LFS 2000Q1 - 2023Q2.

Figure 3.5: Shares of self-employed by gender, education and age



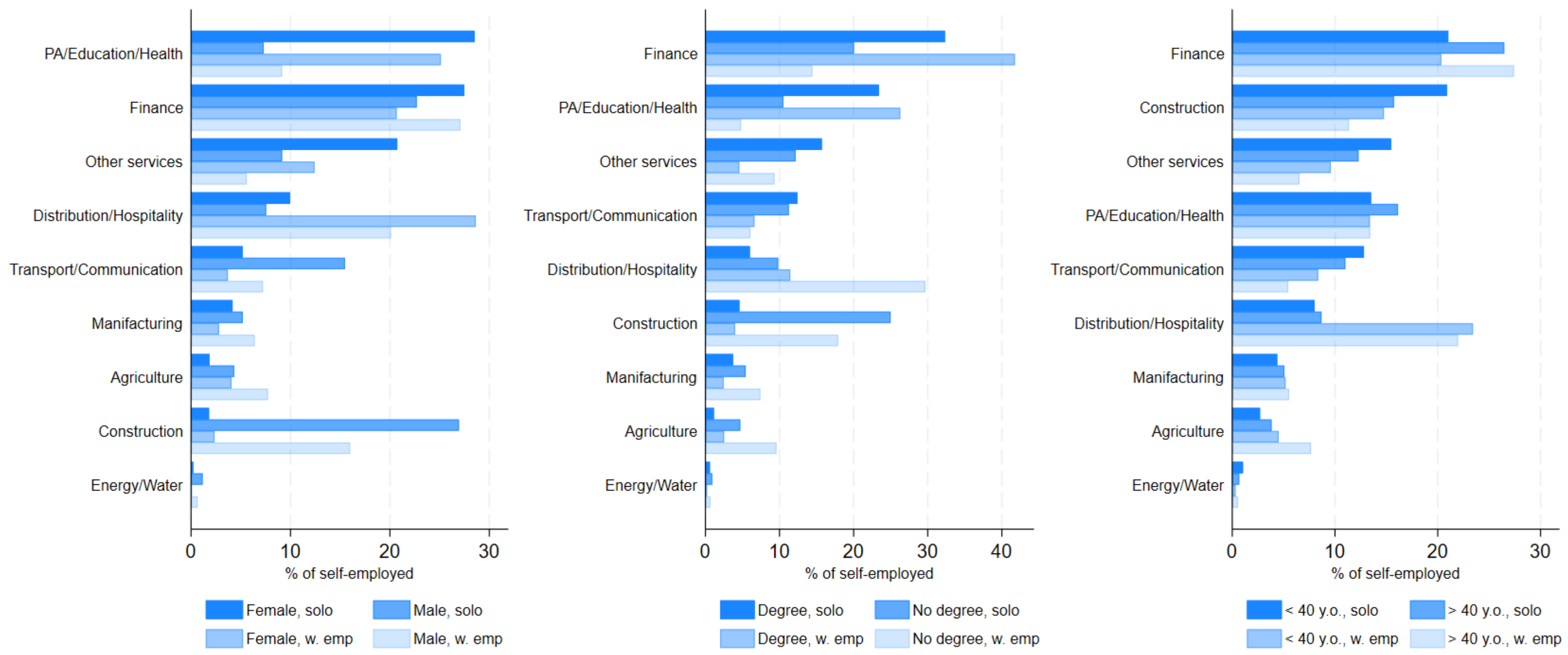
Notes: The plot displays the share of workers that are self-employed for different segments of the population since 2000Q1. Data source: Five-quarter longitudinal UK LFS 2000Q1 - 2023Q2.

Figure 3.6: Hours worked by self-employed by gender, education and age



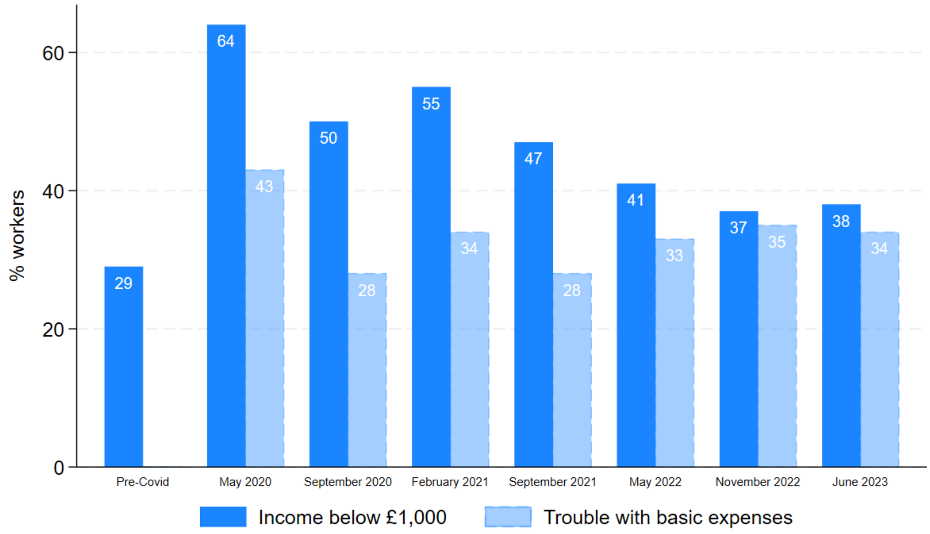
Notes: The plot displays the share of workers that are self-employed for different segments of the population since 2000Q1. Data source: Five-quarter longitudinal UK LFS 2000Q1 - 2023Q2.

Figure 3.7: Shares of self-employed by gender, education and age



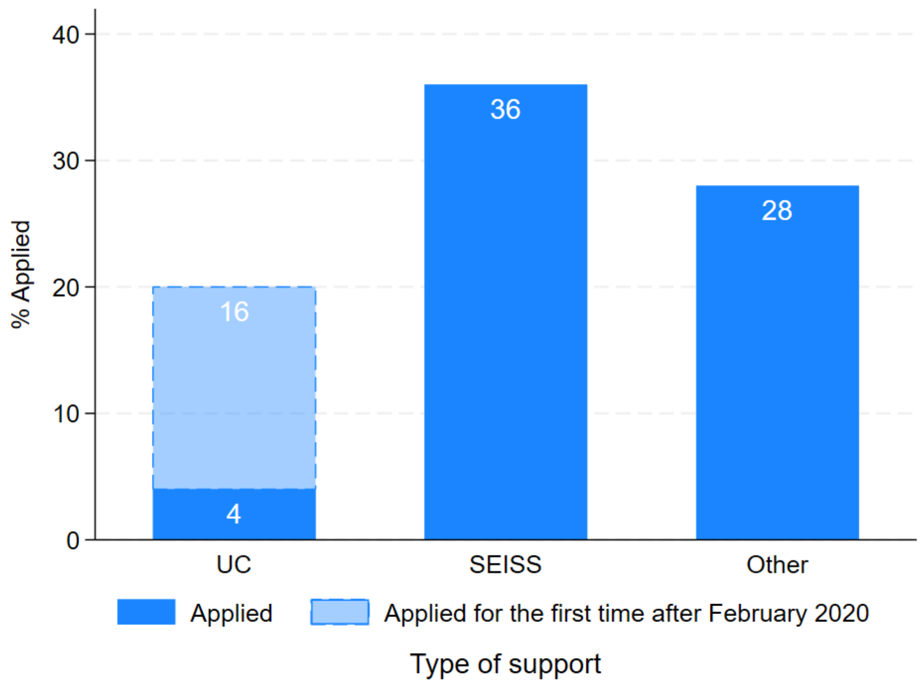
Notes: The plot displays the share of workers that are self-employed for different segments of the population since 2000Q1. Data source: Five-quarter longitudinal UK LFS 2000Q1 - 2023Q2.

Figure 3.8: The financial situation of the self-employed



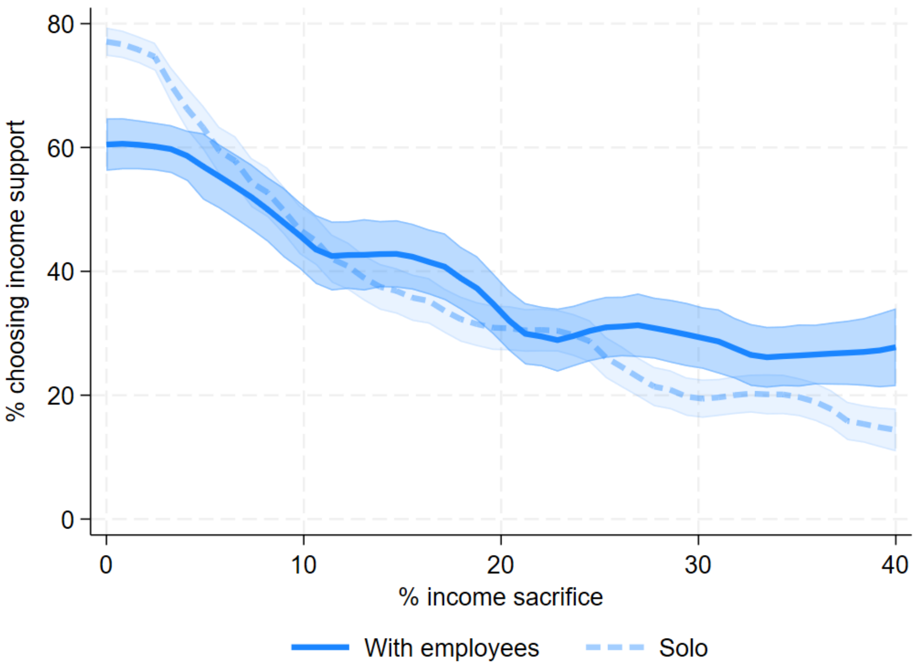
Notes: The plot displays the shares of self-employed workers that report financial difficulties (Survey question: *Over the last month, have you experienced trouble paying for basic expenses such as rent, mortgage repayments, bills and essentials?*) and incomes below the lowest threshold of £1,000. Data source: LSE-CEP Survey of UK Self-Employed.

Figure 3.9: Applications for financial support



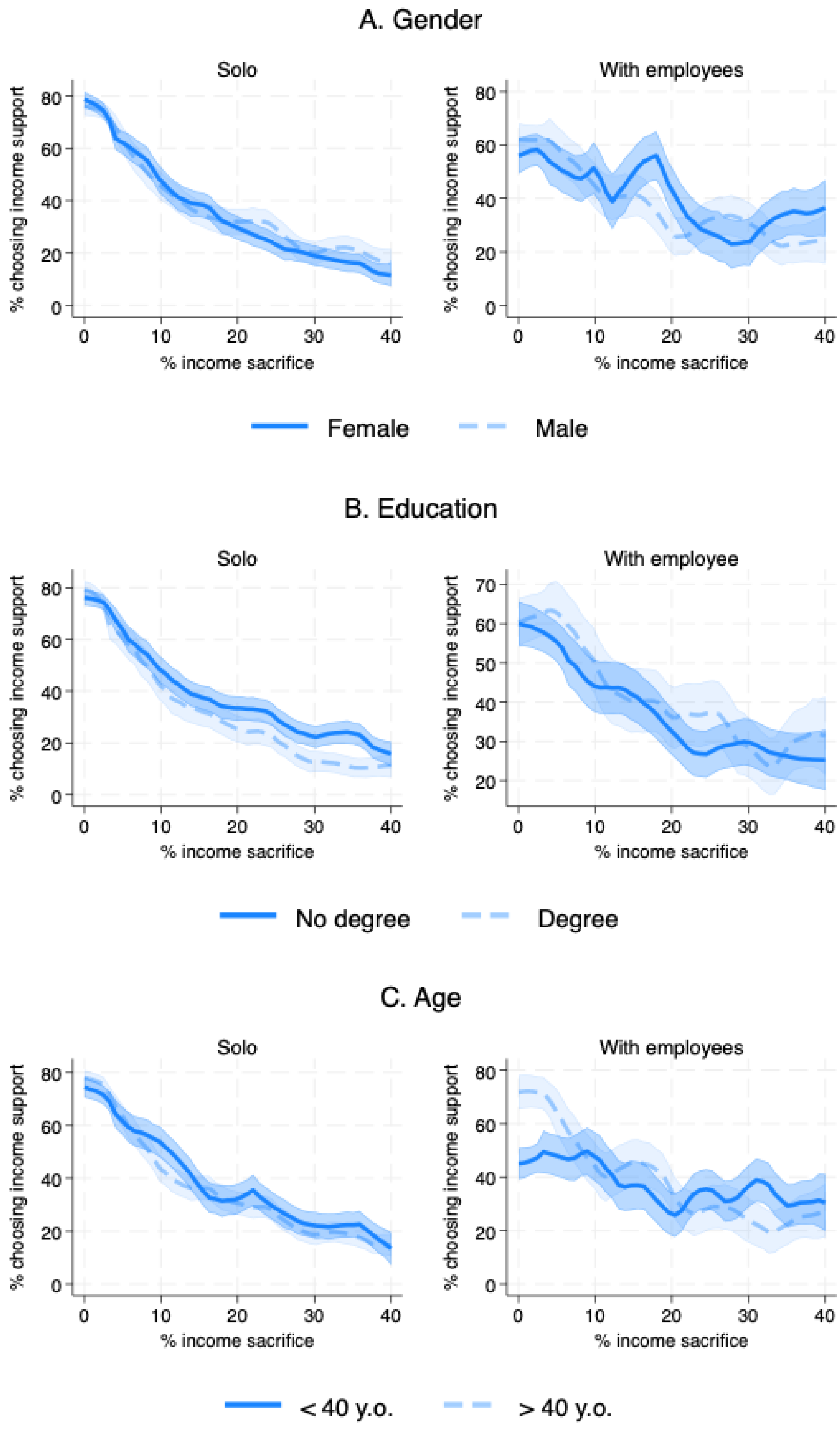
Notes: The plot displays the shares of self-employed workers that applied for different types of financial support between March 2020 and May 2023. "UC" and "SEISS" stand for Universal Credit and Self-Employed Income Support Scheme, respectively. Other includes the Coronavirus Business Interruption Loan Scheme, the Coronavirus Large Business Interruption Loan Scheme, the Covid-19 Corporate Financing Facility, the Bounce Back Loan Scheme, the Future Fund, the Job Retention Bonus, and the Eat Out Help Out Scheme. Data source: LSE-CEP Survey of UK Self-Employed.

Figure 3.10: Willingness to pay for emergency support, by type



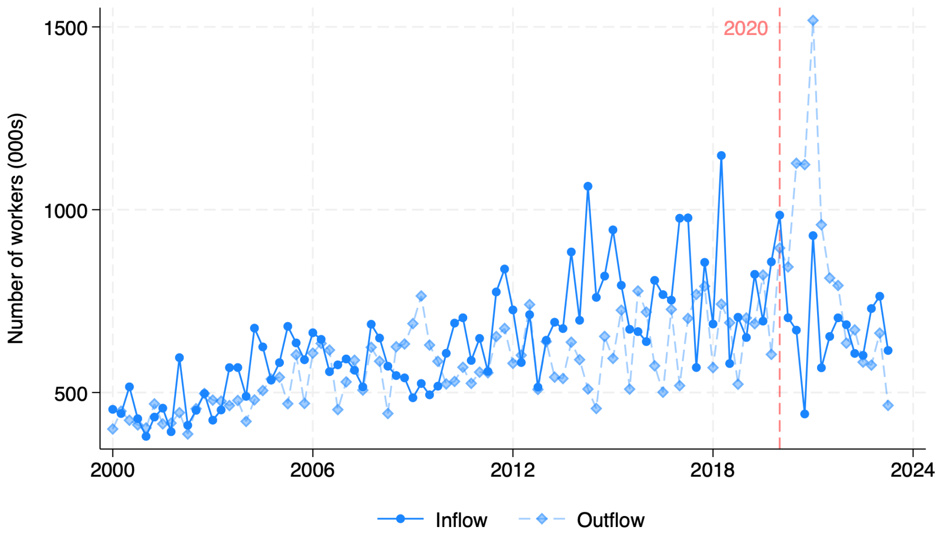
Notes: The plot displays the shares of self-employed workers that would be willing to change by a certain share their pre-tax income to receive financial support in times of crisis. The generosity of the support is the same as the most recent SEISS funding. The shaded areas represent 5% confidence intervals. Data source: LSE-CEP Survey of UK Self-Employed.

Figure 3.11: Willingness to pay for emergency support, by gender, education and age



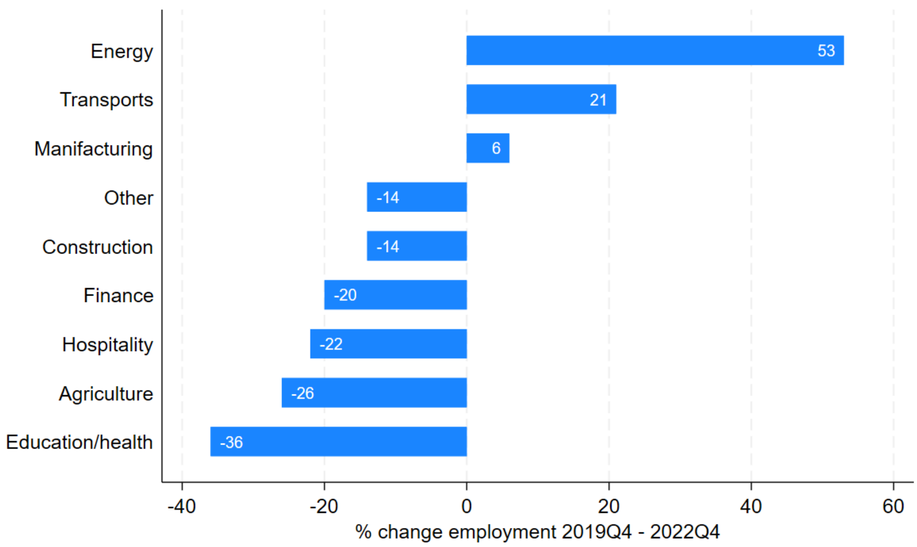
Notes: The plot displays the shares of self-employed workers that would be willing to sacrifice a certain share of their pre-tax income to receive financial support in times of crisis. The generosity of the support is the same as the most recent SEISS funding. The shaded areas represent 5% confidence intervals. Data source: LSE-CEP Survey of UK Self-Employed.

Figure 3.12: Flows in and out of self-employment



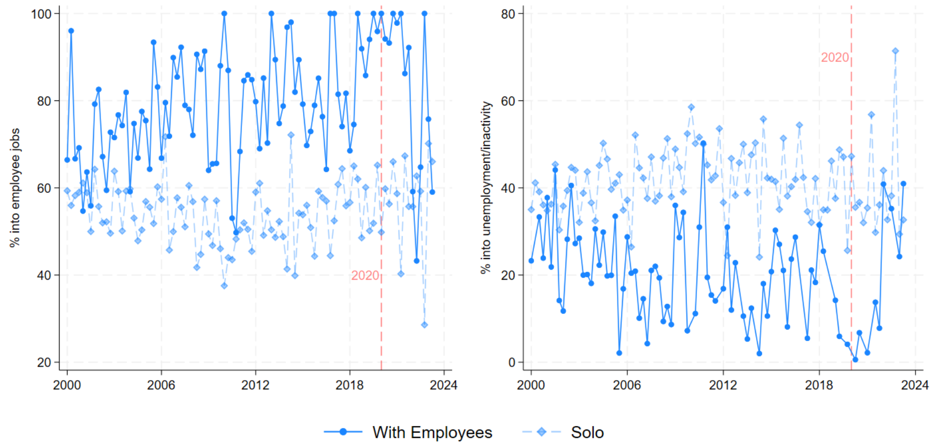
Notes: The plot displays the number number of workers moving in and out of self-employment since 2000 Q1. The status change refers to the same quarter in the previous year. Data source: Five-quarter longitudinal UK LFS 2000Q1 - 2023Q2

Figure 3.13: Post-pandemic change in self-employment, by industry



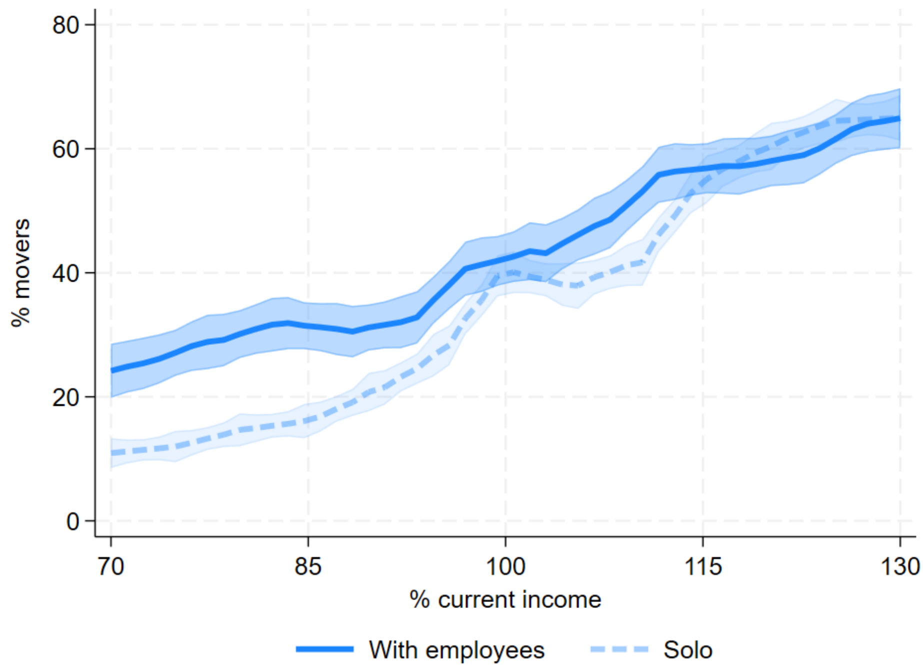
Notes: The plot displays the percentage change in the number number of self-employed workers between 2019Q4 and 2022Q4, for each given industry. Data source: Five-quarter longitudinal UK LFS 2001Q1 - 2023Q2

Figure 3.14: Self-employment outflow, by destination status and type



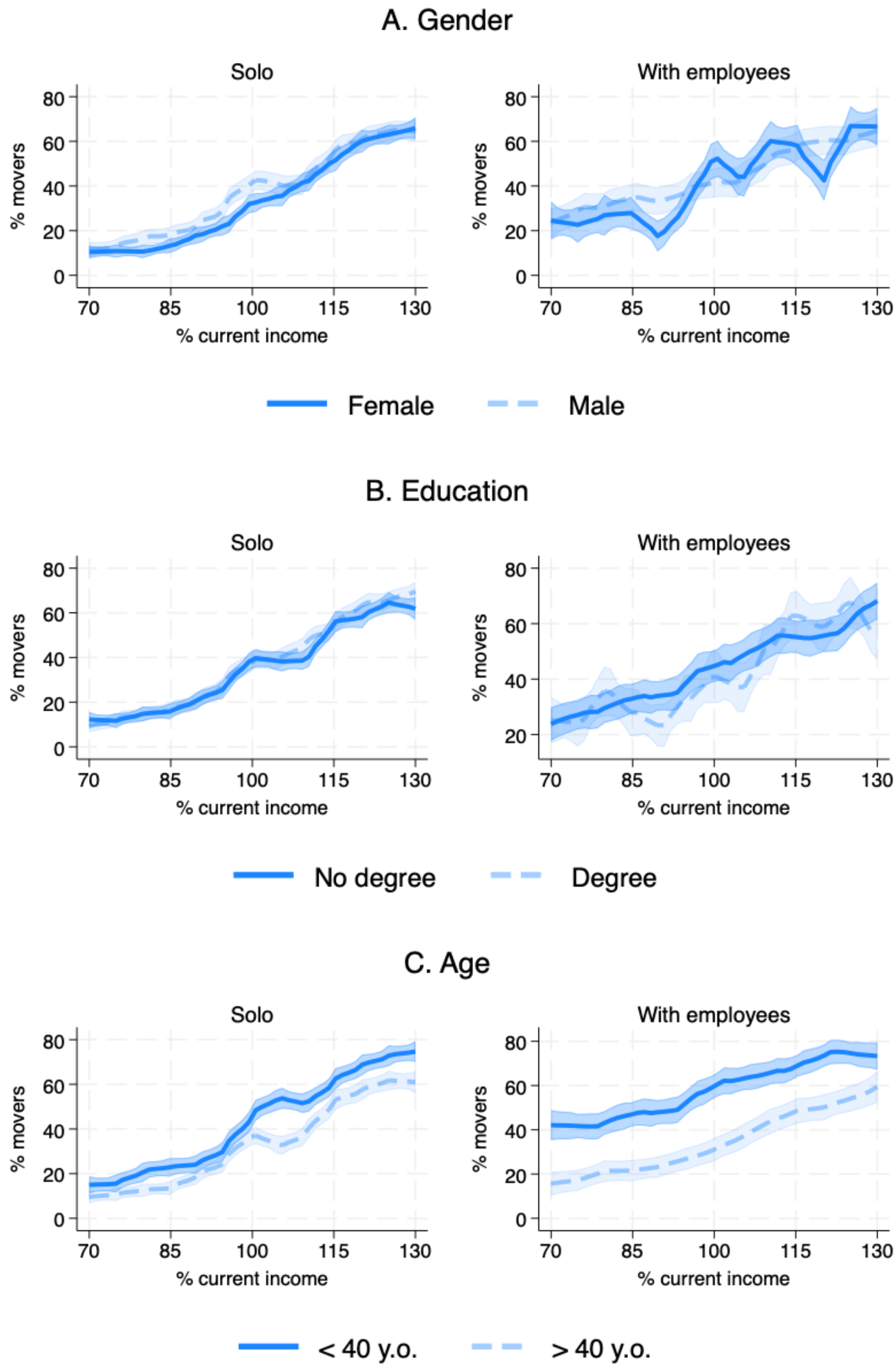
Notes: The plot displays the share of workers exiting self-employment going into either employee jobs or unemployment/inactivity. The status change refers to the same quarter in the previous year. Data source: Five-quarter longitudinal UK LFS 2001Q1 - 2023Q2.

Figure 3.15: Willingness to pay for an employee job, by type



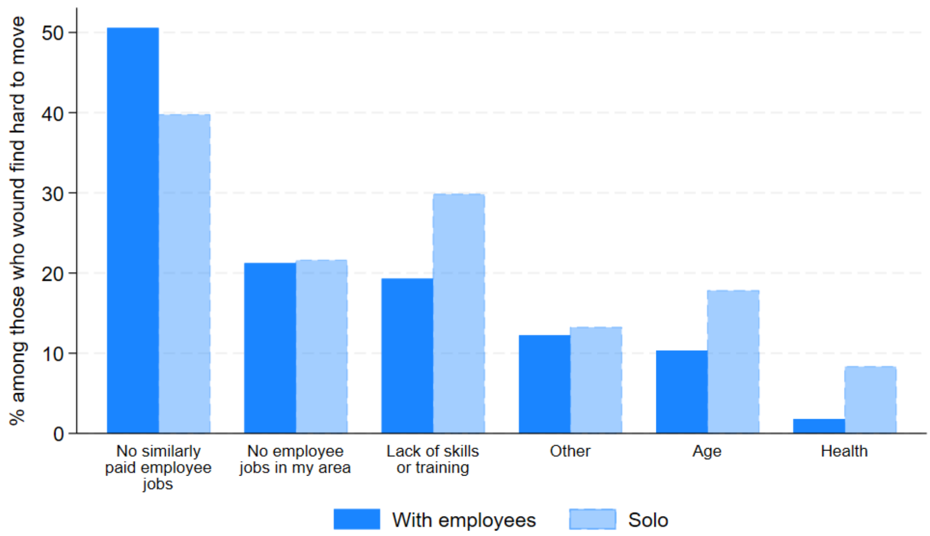
Notes: The plot displays the shares of self-employed workers that would be willing to change by a certain share their pre-tax income to move to an salaried employee job. The shaded areas represent 5% confidence intervals. Data source: LSE-CEP Survey of UK Self-Employed.

Figure 3.16: Willingness to pay for an employee job, by gender, education and age



Notes: The plot displays the shares of self-employed workers that would be willing to change by a certain share their pre-tax income to move to an salaried employee job. The shaded areas represent 5% confidence intervals. Data source: LSE-CEP Survey of UK Self-Employed.

Figure 3.17: Reasons not to move to employee jobs, by type



Notes: The plot displays the shares of self-employed workers who claim they would like to move to an employee jobs but would find it hard to do so, by motivation. Data source: LSE-CEP Survey of UK Self-Employed.

3.8 Tables

Table 3.1: Distribution of new self-employed by past employment status

	Employment status one year earlier			
	Employee	Unemployed	Inactive	Other
Solo				
Average	61.82	13.07	22.42	2.69
Female	53.06	9.58	34.55	2.80
Male	67.60	15.38	14.40	2.61
Degree	62.85	11.68	22.58	2.89
No degree	61.31	13.76	22.34	2.59
$<$ 40 y.o.	60.14	11.05	25.14	3.67
$>$ 40 y.o.	61.82	13.07	22.42	2.69
With employees				
Average	86.20	2.92	8.06	2.81
Female	77.67	1.75	15.90	4.68
Male	90.50	3.51	4.11	1.87
Degree	92.07	2.16	4.15	1.62
No degree	83.04	3.33	10.17	3.45
$>$ 40 y.o.	84.53	2.98	8.05	0.68
$<$ 40 y.o.	86.20	2.92	8.06	2.81

Notes: Weighted averages, in percentage points. Other includes those engaged in unpaid family work, coming from a government scheme or classified as Other by the LFS. Data source: Five-quarter longitudinal UK LFS 2000Q1 - 2019Q4.

Table 3.2: Financial difficulties - heterogeneity

	Trouble paying for basic expenses in the last month?							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Solo	-0.0688***		-0.106***		-0.0567***		-0.150***	-0.170***
	(0.0148)		(0.0220)		(0.0190)		(0.0222)	(0.0290)
Male		-0.0388***	-0.0831***					-0.0744***
		(0.0128)	(0.0251)					(0.0243)
Male x Solo			0.0503*					0.0376
			(0.0283)					(0.0277)
Degree				-0.0550***	-0.0293			-0.0558**
				(0.0129)	(0.0250)			(0.0245)
Degree x Solo					-0.0406			-0.0277
					(0.0286)			(0.0283)
Age: > 40 y.o.						-0.167***	-0.270 ***	0.265***
						(0.0129)	(0.0248)	(0.0248)
Age: > 40 y.o. x Solo							0.144***	0.140***
							(0.0287)	(0.0287)
Constant	0.633***	0.607***	0.695***	0.593***	0.641***	0.663***	0.794***	0.843***
	(0.0613)	(0.0615)	(0.0634)	(0.0607)	(0.0621)	(0.0595)	(0.0610)	(0.0640)
Observations	9,111	9,111	9,111	9,111	9,111	9,111	9,111	9,111
Mean dep. var.	0.387	0.387	0.387	0.387	0.387	0.387	0.387	0.387
R-Squared	0.038	0.036	0.041	0.038	0.042	0.059	0.068	0.070

Notes: Ordinary least square regressions. All specifications control for age, native status, share of time worked from home, industry, survey round and region fixed effects. Dependent variable takes value of 1 if respondent answers positively to the question "*Over the last month, have you experienced trouble paying for basic expenses such as rent, mortgage repayments, bills and essentials?*". Data source: LSE-CEP Survey of UK Self-Employed.

Table 3.3: Application for financial support - heterogeneity

	Trouble paying for basic expenses in the last month?							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Solo	-0.0897***		-0.136***		-0.0764***		-0.230***	-0.257***
	(0.0165)		(0.0232)		(0.0211)		(0.0218)	(0.0291)
Male		-0.0438***	-0.0969***					-0.0887***
		(0.0140)	(0.0268)					(0.0258)
Male x Solo			0.0606**					0.0482
			(0.0307)					(0.0299)
Degree				-0.0700***	-0.0247			-0.0399
				(0.0146)	(0.0279)			(0.0269)
Degree x Solo					-0.0687**			-0.0353
					(0.0322)			(0.0312)
Age: > 40 y.o.						-0.128***	-0.277 ***	0.283***
						(0.0139)	(0.0261)	(0.0258)
Age: > 40 y.o. x Solo							0.221***	0.225***
							(0.0306)	(0.0301)
Constant	0.505***	0.465***	0.576***	0.387***	0.451***	0.432***	0.601***	0.746***
	(0.0658)	(0.0660)	(0.0684)	(0.0636)	(0.0652)	(0.0624)	(0.0637)	(0.0679)
Observations	8,053	8,053	8,053	8,053	8,053	8,053	8,053	8,053
Mean dep. var.	0.583	0.583	0.583	0.583	0.583	0.583	0.583	0.583
R-Squared	0.095	0.091	0.098	0.072	0.080	0.081	0.093	0.120

Notes: Ordinary least square regressions. All specifications control for age, native status, share of time worked from home, industry, survey round and region fixed effects. Dependent variable takes value of 1 if respondent has claimed any of the available business financial support. Data source: LSE-CEP Survey of UK Self-Employed.

Table 3.4: Post-Covid moves to employee jobs - heterogeneity

	Moved to an employee job after Covid-19 pandemic							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Solo	-0.397***		-0.335***		-0.491***		-0.392***	-0.455***
	(0.0395)		(0.0786)		(0.0513)		(0.0864)	(0.113)
Male		0.0359	0.0987					0.0810
		(0.0537)	(0.0731)					(0.0734)
Male x Solo			-0.0914					-0.0651
			(0.0909)					(0.0892)
Degree				0.158***	-0.0419			-0.0494
				(0.0496)	(0.0569)			(0.0574)
Degree x Solo					0.204***			0.197**
					(0.0767)			(0.0781)
Age: > 40 y.o.						-0.179***	-0.175**	-0.177**
						(0.0599)	(0.0737)	(0.0755)
Age: > 40 y.o. x Solo							-0.0117	0.00933
							(0.101)	(0.102)
Constant	0.820***	0.454**	0.754***	0.426**	0.865***	0.525***	0.878***	0.881***
	(0.161)	(0.192)	(0.166)	(0.198)	(0.156)	(0.192)	(0.148)	(0.148)
Observations	689	689	689	689	689	689	689	689
Mean dep. var.	0.547	0.547	0.547	0.547	0.547	0.547	0.547	0.547
R-Squared	0.119	0.051	0.120	0.073	0.141	0.072	0.143	0.162

Notes: Ordinary least square regressions. All specifications control for time spent in self-employment, industry and survey year and quarter fixed effects. Dependent variable takes value of 1 if respondent has moved from self-employment to an employee job. The data is restricted to the period 2019Q4 - 2022Q4. The status change refers to the same quarter in the previous year. Data source: Five-quarter longitudinal UK LFS 2001Q1 - 2023Q2

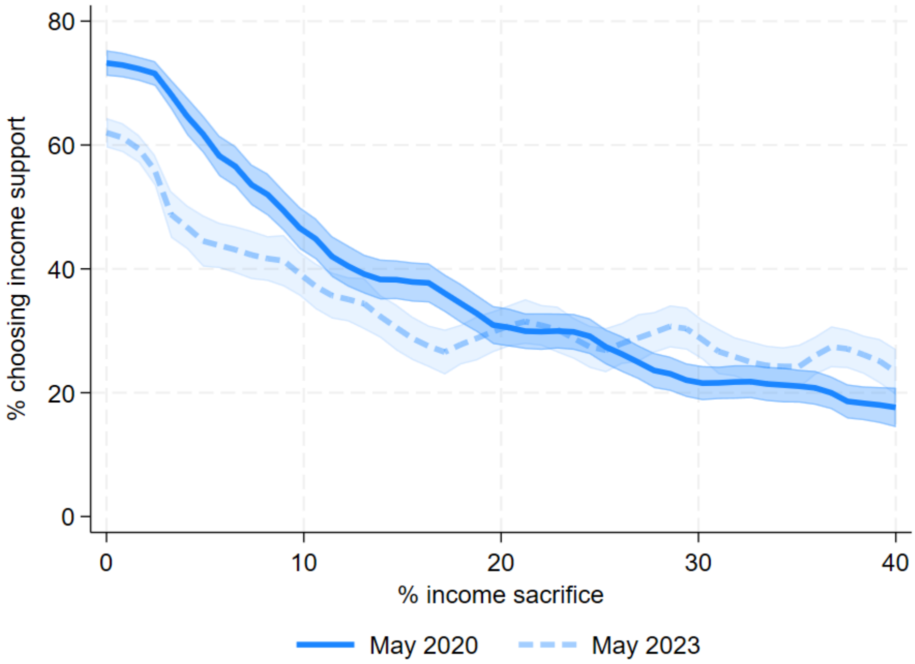
Table 3.5: Difficulties finding employee jobs - heterogeneity

	Hard to find an employee job							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Solo	0.0648***		0.112***		0.0492*		0.0223	0.0485
	(0.0194)		(0.0262)		(0.0256)		(0.0268)	(0.0361)
Male		-0.0471***	0.0142					-0.00136
		(0.0160)	(0.0316)					(0.0311)
Male x Solo			-0.0709**					-0.0614*
			(0.0355)					(0.0350)
Degree				-0.0401**	-0.0739**			-0.0618**
				(0.0161)	(0.0314)			(0.0311)
Degree x Solo					0.0489			0.0287
					(0.0354)			(0.0350)
Age: > 40 y.o.						0.147***	0.106***	0.106***
						(0.0158)	(0.0316)	(0.0315)
Age: > 40 y.o. x Solo							0.0477	0.0477
							(0.0362)	(0.0361)
Constant	0.386***	0.467***	0.379***	0.459***	0.423***	0.374***	0.366***	0.367***
	(0.0761)	(0.0755)	(0.0785)	(0.0740)	(0.0771)	(0.0723)	(0.0751)	(0.0789)
Observations	6,426	6,426	6,426	6,426	6,426	6,426	6,426	6,426
Mean dep. var.	0.506	0.506	0.506	0.506	0.506	0.506	0.506	0.506
R-Squared	0.025	0.024	0.027	0.022	0.025	0.037	0.040	0.044

Notes: Ordinary least square regressions. All specifications control for age, native status, share of time worked from home, industry, survey round and region fixed effects. Dependent variable takes value of 1 if respondent has claimed it would be hard or very hard for them to find a job as a salaried employee. Data source: LSE-CEP Survey of UK Self-Employed.

3.9 Appendix Figures

Figure 3.18: Willingness to pay for emergency support, by surey rounds



Notes: The plot displays the shares of self-employed workers that would be willing to sacrifice a certain share of their pre-tax income to receive financial support in times of crisis. The generosity of the support is the same as the most recent SEISS funding. The shaded areas represent 5% confidence intervals. Data source: LSE-CEP Survey of UK Self-Employed.

3.10 Survey questions

This section report the survey questions that have been used in the chapter.

Please report your (biological) sex:

Male

Female

Other (please specify):

What is your age?

In which industry is your (current or most recent) main job?

Agriculture, forestry and fishing

Mining and quarrying

Manufacturing

Electricity, gas, steam, and air conditioning supply

Water supply; sewerage, waste management and remediation activities

Construction

Wholesale and retail trade; repair of motor vehicles and motorcycles

Transportation and storage

Accommodation and food service activities

Information and communication

Financial and insurance activities

Real estate activities

Professional, scientific and technical activities

Administrative and support service activities

Public administration and defence; compulsory social security

Education

Human health and social work activities

Arts, entertainment and recreation

Other :

How long have you been continuously self-employed for?

If you were self-employed previously but not currently, please give the length of your most recent self-employment.

What is your highest education level?

No qualifications

Some GCSE / O-levels

5 or more GCSE / O-levels

Trade/technical/vocational training

A-level

Bachelor's degree

Masters degree

Doctorate

Other (please specify):

Were you born in the UK?

Yes

No

Which region do you live in?

East Midlands

East of England

London

North East

Northern Ireland

North West

Scotland

South East

South West

Wales

West Midlands

Yorkshire and the Humber

Do you have any employees?

Yes

No

Up to February 2020, what was your monthly personal income from your main job (pre-tax)?

Less than £1,000

£1,000 to £1,999

£2,000 to £2,999

£3,000 to £3,999

£4,000 to £4,999

£5,000 to £5,999

£6,000 to £6,999

£7,000 to £7,999

£8,000 to £8,999

£9,000 to £9,999

£10,000 to £14,999

£15,000 or more

What were your monthly profits from your main job in April 2020/August 2020/January 2021/August 2021/April 2022/October 2022/May 2022 (pre-tax)?

Less than £1,000

£1,000 to £1,999

£2,000 to £2,999

£3,000 to £3,999

£4,000 to £4,999

£5,000 to £5,999

£6,000 to £6,999

£7,000 to £7,999

£8,000 to £8,999

£9,000 to £9,999

£10,000 to £14,999

£15,000 or more

Over the last month, have you experienced trouble paying for basic expenses such as rent, mortgage repayments, bills and essentials?

Yes

No

How easy do you think it would be for you to leave self-employment and find a job as an employee?

Very hard

Moderately hard

Neutral

Moderately easy

Very easy

Why do you think it would be hard for you to find a job as an employee? Tick all the apply.

Lack of employee jobs in my local area

Lack of skills or training necessary for the available employee jobs

Hard to find similarly paid employee positions

Other (please specify):

Suppose you were offered an employee job that is similar to your job in self-employment.
Would you take the job if it paid x% less/more than your income * from self-employment?

Yes

No

In the last two years, the government has introduced a support package for the self-employed called the Coronavirus Self-employment Income Support Scheme.
Have you claimed a grant from this scheme? Claiming periods are in brackets. Please select all that apply.

I claimed the first grant (13 May to 13 July 2020)

I claimed the second grant (17 August to 19 October 2020)

I claimed the third grant (30 November 2020 to 29 January 2020)

I claimed the fourth grant (22 April to 1 June 2021)

I claimed the fifth grant (28 July to 30 September 2021)

No

Since the beginning of Covid-19, have you applied for any of the following support schemes?

Coronavirus Business Interruption Loan Scheme (closed 31 March 2021)

Job Retention Bonus (closed 31 March 2021)

Coronavirus Large Business Interruption Loan Scheme (closed 31 March 2021)

COVID-19 Corporate Financing Facility (closed 31 December 2020)

"Bounce back" loans (closed 31 March 2021)

Future Fund (closed 31 January 2021)

Recovery Loan Scheme (still running)

I have not applied for any business support

Have you made a claim for universal credit due to lost work since February 2020?

Yes

No

Was this your first time claiming Universal Credit?

Yes

No

Imagine that the government provides a one-time option for self-employed people to switch to a new type of self-employment in which they are given a guarantee of income support in the case of future pandemics or other significant economic shocks. Note that everything else (including your job/business activity) would remain unchanged. This will be called self-employment B. Those who do not join the scheme would not be eligible for any support in future crises. This will be called self-employment A. The generosity of the income support, should you opt for self-employment B, will be a one-off payment of approximately 80% of 3 months trading profits, up to £7,500.

If you could switch at no cost, which type of self-employment would you choose?

Self-employment A

Self-employment B (with income support)

Now assume that if you choose B, you will receive a x% reduction in your income in normal times via a tax. If you choose A, your taxes are unchanged.

Which type of self-employment would you choose?

Self-employment A

Self-employment B (with income support)