ESSAYS IN FAMILY ECONOMICS

by **Ashley Burdett**

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy in ECONOMICS

Department of Economic

UNIVERSITY OF ESSEX

September 30, 2023

Declaration of Authorship

I, Ashley Burdett, declare that this thesis titled, and the work presented in it are my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

I declare that the work in this thesis was carried out in accordance with the requirements of the University's Regulations and that it has not been submitted for any other academic award.

Chapter 1 of this thesis is a sole-authored paper. Chapter 2 is a paper co-authored with Prof. Melvyn Coles (University of Essex). Chapter 3 is a joint paper with Dr. Ben Etheridge (University of Essex), Dr. Yikai Wang (University of Essex) and Dr. Li Tang (Middlesex University).

Signed: Ashley Burdett

Date: September 30, 2023

Abstract

This thesis contains three stand-alone papers. Each paper addresses a timely research question that is relevant for ongoing policy debates.

The first paper examines whether the "housing crisis" deters the transition to adulthood in England. Specifically, I consider whether the increase in local house prices reduces leaving the family home and/or the transition to a partnership. Using an instrumental variable approach, I find causal evidence that greater house prices increase the probability of living with parents but do not influence relationship formation.

The second paper explores the reason for the high level of instability of young people's relationships. We present an equilibrium partnership model with endogenous separations and frictions with sorting by match quality. Our estimates show that on-the-job is the major channel of partner separations and that pairwise exclusivity is important: well-matched partners, on average, jointly reduce outside contact rates by 77%. Self-enforced commitment explains how cohabitation thrives as an institution even as UK divorce law weakens and marriage rates decline.

Reflecting the emergence of the Covid-19 pandemic during my study the final paper concentrates on the impact of working from home on worker productivity during the pandemic. We explore how workers adapted to working conditions and which type of workers thrived in each work location. We find that as the pandemic progressed, those who previously performed well in a work location were more likely to remain there. We also estimate factors affecting productivity outcomes across locations controlling for endogenous selection. We find that those in 'good' jobs were advantaged in the home environment. Key personality traits – agreeableness and conscientiousness – impact productivity across locations.

Acknowledgements

First and foremost, I would like to thank my supervisors, Professor Melvyn Coles and Professor Marco Francesconi, for their continuous support and guidance over the past several years. The many insightful discussions regarding research have been invaluable to my development as a researcher. Having the opportunity to co-author with Melvyn opened my eyes to how research should be done and the standards that must be maintained to be the world-leading researcher he is.

I have also benefited greatly from co-authoring with several valuable colleagues including Dr. Ben Etheridge, Dr. Yikai Wang, Dr. Li Tang, Dr. Lisa Spantig and Dr. Apostolos Davillas. It has truly been an inspiring process and you have been very generous in your patience and in sharing your rich knowledge and experience.

I would also like to thank the faculty members at Essex for creating such a vibrant research and open research environment that encouraged me to develop my ideas. In particular, I would like to express my gratitude to Dr. Alex Clymo for his generosity and patience in showing me how to approach and develop my computational skills, and to Dr. Lukas Mayr for his guidance and advice.

In addition, I want to thank my fellow PhD students at Essex for their camaraderie and support throughout my studies. Although we experienced many unexpected circumstances, we got through it together.

I was fortunate enough to receive financial support in the form of a studentship from the South East Network for Social Sciences (SeNSS). The generous provision of extended funding given the onset of the pandemic was invaluable.

Finally, my final thanks go to my family. Without your everlasting support and the inspiration I take from each of you.

Table of Contents

List of Figures

Chapter 1 Appendix

Chapter 2 Appendix

Chapter 3 Appendix

List of Tables

Page

Introduction

This thesis is composed of three chapters, each of which can be read as a stand-alone academic paper. The first two chapters focus on the household turnover in the UK. There have been vast changes in households since the mid-twentieth century. No longer do people on average leave education, find a job and quickly settle down with a partner for life to have children with. We are observing increased instability and heterogeneity throughout the life course characterized, for example, by divorce and repartnering with stepchildren, more living outside of families and increased interdependence across generations. Whilst evidence suggests these changes have significant welfare implications, they are yet to be fully understood and have significant consequences for the sustainability of welfare systems across the world. It is therefore a timely and important task for researchers to understand these patterns and their potential consequences. The first two chapters of this thesis seek to add to this research agenda.

The final chapter is a step change from the previous chapters. As the Covid-19 pandemic spread across the world, in the middle of my PhD studies, this not only shaped my daily life but also the direction of my research. I was strongly motivated to contribute to the growing Covid-19 literature that continues to seek to understand the implications of these very unusual circumstances. The third chapter is on the Covid-19 papers I wrote during this period. Co-authored with Dr. Ben Etheridge, Dr. Yikai Wang and Dr. Li Tang, the final paper explores the important policy question of how working from home impacts productivity.

In the following I present each chapter in more detail:

Chapter 1: The first chapter is a purely empirical paper exploring whether housing market conditions are contributing to the delay in the transition to adulthood we observe in England. The UK housing market is characterized by comparatively high house prices, high private sector rents and limited support for young people with low incomes. Has this contributed to the increase in parental co-residence and delay in partnership formation documented in the UK? After providing some descriptive analysis, I seek to answer this question by obtaining causal estimates of the impact of house prices on the probability of a young person (20-29 years old) living with a partner or with parents as well as on the probability of transitioning out of the parental home or forming a cohabiting relationship. I adopt an instrument variable approach, constructing an instrument based on local planning regulatory restiveness. The estimates find that increases in house prices increase living with parents and deter living in the parental home, whilst there is limited evidence that this plays a role in the relationship decisions of young people.

Chapter 2: The second chapter, co-authored with Professor Melvyn Coles, was inspired by the descriptive work undertaken during the writing of Chapter 1 in which we found a large amount of relationship turnover in the UK particularly early on in a relationship. In particular, we document that the relationship separation hazard in the UK is surprisingly steep: the hazard in the first year is 15 per cent and declines to a long-run average of 2 percent. What can explain these turnover patterns? Why would so many couples form just to break up shortly after? We show that the existing standard equilibrium marriage models are unable to capture these patterns and propose that on-the-job search is necessary.

We develop an equilibrium partnership model in which partners can jointly agree to reduce their outside contact rates if it increases their joint surplus. We think of this as the couple agreeing to spend more time together, so that both partners know that the other is not searching, therefore overcoming trust issues. We estimate the model using simulated method of moments and find that on-the-job search is crucial to match the steep decline in the separation hazard. We also find evidence of a quantitatively important learning process; cohabitation is an experience good. Empirical evidence shows that learning is particularly important for explaining the behaviour of young adults.

Chapter 3: The final chapter pivots in terms of topic, but also addresses a timely question. The Covid-19 pandemic emerged during my time as a PhD student and, given its unprecedented nature many, many important new research questions emerged. Wanting to understand the world around me, my research focus expanded to also include the very strange circumstances we found ourselves in. The third chapter is one of the Covid-19 papers.

Co-authored with Dr. Ben Etheridge, Dr. Yikai Wang and Dr. Li Tang, the final paper explores the important policy question of how working from home impacts productivity. As companies continue to develop their working-from-home policies in the aftermath of the pandemic, understanding this relationship is crucial not only for businesses but also for governments seeking to take advantage of potential new opportunities. We aim to add to the growing but mixed body of evidence exploring the impact of working from home on productivity, using the widespread experience caused by the pandemic. We use a self-reported measure of productivity change, validated using external measures, to conduct a series of empirical exercises exploring how worker productivity and working from home evolved across the pandemic. We find evidence of significant heterogeneity across work types and job types in terms of productivity and that workers who had positive experiences working from home are more likely to select to continue to work from home. Building on these results, we construct a selection model of location choice to estimate the impact of individual characteristics, home environment and personality traits on productivity when working from home and at the workplace separately. We find that a productivity advantage is experienced by those in 'good jobs' (in large firms, with managerial duties and high earnings) when working from home. We also find that those high in agreeableness and conscientiousness performed better across locations, but those with higher cognitive ability experienced worse productivity growth while at home. Overall our results provide rich insights on which factors affected productivity deferentially across locations during the pandemic.

Chapter 1

The Transition to Adulthood in England: The Role of the Housing Market*

1 Introduction

Evidence demonstrates the period of transition to adulthood is being drawn out as more young people are residing in the parental home at older ages (Knipe, 2017), investing more in education (Blundell et al., 2016), and postponing many life cycle events such as marriage and having children (Beaujouan and Ní Bhrolcháin, 2011; Neels et al., 2017). The existing literature highlights the important roles played by individual and parental resources (Belloc, 2009), labour market conditions (Rosenzweig and Wolpin, 1993; Kaplan et al., 2009; Ermisch, 1999; Becker et al., 2010) and the welfare state (Aassve et al., 2002; Chiuri and Del Boca, 2010; Low et al., 2018) in explaining these developments. However, whilst there is much anecdotal and survey evidence[1] suggesting that housing costs also play a significant role in making these decisions, there is relatively little research focusing on this relationship. This paper seeks to add

*This paper draws on data from Understanding Society, distributed by the UK Data Service. Understanding Society is an initiative funded by the Economic and Social Research Council and various Government Departments. All errors remain the responsibility of the authors.

[1]A number of recent non-scientific surveys have found such a link. For example, SpareRoom found that "24% of respondents said they would consider moving in with their partner earlier than planned because of the cost-of-living crisis" (https://theface.com/life/inflationship-sex-relationships-money-moving-in-together-cost-of-living-crisis-politics-society) and YouGov found that "4 out of 10 young adults have said they will not settle down until they can buy their own house" (https://www.theguardian.com/money/2010/mar/21/house-prices-young-couples-marriage).

insight by investigating the causal impact of housing costs on the transition to adulthood, in particular on the probability of leaving the parental home and transitioning to a partnership in England.

Popular media and politicians often declare the UK has long been experiencing a housing crisis.[2] This is borne out in the data. House prices in the UK have been rapidly increasing compared to similar countries. Indeed, since 1980 the UK has seen the greatest increase in house prices of the G7 countries, exhibiting a 300 percent real terms increase between 1980 and 2017 (Miles and Monro, 2021). This growth has significantly outpaced that of incomes making access to homeownership increasingly financially difficult (Hilber and Vermeulen, 2016). For example, between 1995 and 2007 as real house prices increased by 168 percent real income per household only increased by 27 percent (Kuenzel and Bjørnbak, 2008). Whilst the aftermath of the Great Financial Crisis (GFC) of 2008 temporarily cooled the market, benefits were limited as the terms of financing home ownership became more challenging; both the size of the required deposits and interest rate spreads increased (Mulheirn, 2019).[3] Unsurprisingly in these conditions the private rental sector grew. Due to the relatively inelastic supply of rental properties, this created simultaneous upward pressure on private sector rents.[4] Over this period low-income households also faced significant reductions in the amount of social support on offer from UK governments in terms of available social housing and housing benefits for young adults as part of the Austerity policy regime (Berrington and Stone, 2014). Thus overall the availability and affordability of housing and access to home ownership in the UK has become increasingly challenging, particularly for young people who are lower down in the income distribution.

[2]For example in a White Paper in 2017 titled 'Fixing Our Broken Housing Market' (DCLG, 2017), Prime Minister Theresa May describes the housing market as "one of the greatest barriers to progress in Britain today".

[3]The average size of the deposit increased from 10 per cent to 25 per cent of the purchase price of the property between 2007 and 2009. The average deposit for a first-time buyer is over £26,000 (representing 79% of the average annual income from which the mortgage is paid) (Corlett and Odamtten, 2021).

[4]To illustrate the extent, in 2019 the average rent in each region was unaffordable for a woman on the median earnings (Reis, 2019), and in 2022, private sector rents grew twice as fast as wages, with the average renter spending 35% of their income on rent (Zoopla, 2022).

The particularly challenging housing market in the UK makes it an ideal case to consider the role of house prices in the delay to adulthood. Straightforwardly, high housing costs may prohibit forming independent households and prevent relationship formation often tied to home ownership. While a number of papers (Dettling and Kearney, 2014; Aksoy, 2016) have applied a similar logic to explore the causal impact of house prices on fertility, the causal impact of housing markets on leaving the parental home and relationship formation has been relatively unexplored.

This intuitive idea has been set out formally in the seminal work considering the various life cycle transitions. Choices are posited to be the outcome of a constrained utility comparison problem in which the outcome is the living arrangement with the greatest indirect utility (for the decision to leave the parental home see McElroy, 1985 and for the decision to marry see Becker, 1974). Naturally, the housing market enters this set-up through the budget constraint, such that higher prices lead to a tighter budget constraint reducing the choice set.

Building upon this general framework, in a series of prominent papers Ermisch formalizes the impact of housing costs on the choice to leave the parental home. (Ermisch and Di Salvo, 1997; Ermisch, 1999, 2016) In this framework, parents with altruistic preferences choose to support their children through private transfers conditional on their living arrangements. Housing is assumed to be a public good within the family and a less expensive way of supporting a child. It is shown that the impact of an increase in the cost of housing crucially depends on the parents' price elasticity of housing demand. Whilst an increase in the cost of housing always reduces the utility of living independently, the model also allows parents to adjust their housing consumption in response to a price change, thus the relative utility across options is not obvious. He shows that if the parents' housing demand is elastic, then higher prices make young people better off living independently. However, the high transaction costs in the UK housing market suggest that parents have inelastic housing demand and so the prediction is that higher house prices encourage co-residence.[5]

[5]Important related models of co-residence include Rosenzweig and Wolpin (1993, 1994); Kaplan (2012); Manacorda and Moretti (2006), these papers however do not directly address the cost of housing

Predictions concerning the impact on relationship formation however are less clear. A key difference between models is the number of living arrangement options available to young people and the assumed different costs of each. For example, an individual who lives alone may find it advantageous to live with a partner (or flatmates, or parents) when costs are high to split the costs and enjoy the economies of scale (Salcedo et al., 2012), whilst following on from the above, living with a partner is more expensive than living with parents.[6] Therefore the total effect will depend on the size of flows from different original states and the impact of house prices on that particular transition to a relationship. In addition, the demography literature finds support for a norm in which higher quality housing/home ownership is expected when forming a partnership increasing the financial barrier to relationship formation. (Forrest et al., 1999; Mulder, 2006; Bowmaker and Emerson, 2015). Aassve et al. (2002) also highlights that the increased co-residence with parents due to high house prices may potentially have a knock-on effect on relationship formation by reducing young people's ability to search for a partner. [7]

The related empirical work has not often gone beyond analyzing correlations. The lion's share of the applied research considering co-residence with parents focuses on the role of the labour market and income in determining co-residence both utilizing cross-sectional variation (Rosenzweig and Wolpin, 1993; Manacorda and Moretti, 2006; Ermisch, 1999; Aassve et al., 2002) and panel variation to focus on transitions (Ermisch and Di Salvo, 1997; Kaplan et al., 2009; Kaplan, 2012). They typically find that young people's income/labour market success has a positive impact on the probability of independence, whilst the impact of parental income is mixed. A small subset of the empirical literature that considers the role of the housing market has generally established a significant negative relationship between the probability of living indepen-

but account for privacy costs and utility shocks. Aassve et al. (2002) takes a fundamentally different approach by utilising a matching model akin to that used in the labour market to explain the living arrangements of young people.

[6] By similar logic, house prices have been found to be an important determinant of divorce timing in the short run (Rainer and Smith, 2010; Farnham et al., 2011; Klein, 2017).

[7] Other indirect mechanisms linking relationship formation and housing demand have also been proposed: signalling status, particularly in the Chinese context (Wei et al., 2012) or by increasing the insurance provided by the marriage contract (Lafortune and Low, 2020)

dently/departure and local housing costs. (Haurin et al., 1993; Ermisch and Di Salvo, 1997; Ermisch, 1999; Bayrakdar and Coulter, 2018; Andrew and Meen, 2003)[8] However, the measure of the housing market in the UK studies is typically very crude (regional level) (Ermisch and Di Salvo, 1997; Ermisch, 1999) and these papers do not address the potential endogeneity caused by reverse causality and omitted variables. One related paper that seeks to obtain causal estimates focuses on the co-residence of couples with their older parents in China (Li and Wu, 2019). Using land supply as an instrument, they find a causal relationship between house prices and parental co-residence and that baseline estimates are biased towards zero. I address these concerns by utilizing more local housing conditions (at the local authority district level (LAD)) and adopting a different instrumental variable approach in an attempt to obtain unbiased estimates. Specifically, I exploit variation in housing market regulatory restrictiveness as an instrument for local house prices. [9]

Considering the transition to a partnership, earlier research focusing on correlations establishes a significant negative association between marriage and housing costs. (Ermisch, 1981; Ermisch and Di Salvo, 1997; Lauster, 2006; Bowmaker and Emerson, 2015; Andrew and Meen, 2003) However, unlike the coresdience research, there have been a number of attempts to address endogeneity concerns. Similar to here these papers typically adopt an instrumental variable approach however, none focus on a Western country in a modern time period as I do. For example, Hill (2014) estimates the effect of additional building permits (as a proxy for housing supply) at the US city level on marriage rates in the baby boom period. He constructs an exogenous measure of predicted permits which is interacted with a time-invariant measure of geographical constraints on developable land and a national series of permit issuance to use as an instrument. Focusing on Iran, Gholipour and Farzanegan (2015) utilize

[8] Results in Bleemer et al. (2014) establish a positive relationship. They attribute their estimate to the net effect after accounting for the fact that wealthy families might not only live in areas with higher house prices but also may be more likely to offer private transfers to their children to support independent living.

[9] Related, Martins and Villanueva (2006) focus on young people's access to mortgage markets. They show that differences in mortgage market imperfections within Europe can explain up to 20 percent of the cross-country variance of establishing a new household.

lagged measures of marriage and housing market conditions as instruments and find a negative relationship between marriage rates and house prices. In the modern Chinese context, Wrenn et al. (2019) use geographic constraints to construct their instrument, while using Taiwanese data Chu et al. (2020) exploit tax reforms and changes in the stock index level to construct multiple instruments. All of these papers find evidence that higher house prices decrease marriage for young people who do not own property.

The logic of the instrument adopted here has been widely used in the US context (Dettling and Kearney, 2014; Mian et al., 2013; Chetty et al., 2017) and draws upon the ideas in Saiz (2010). Housing markets in which the housing supply is restricted be it by geography or by regulatory restrictiveness will have inelastic housing supplies. Therefore house prices are likely to be higher on average than the housing markets in which the housing supply can easily change (Gyourko, 2009). Indeed, Hilber and Vermeulen (2016) demonstrates that regulatory restrictiveness is a key determinant of local house prices in England (also see Cheshire and Sheppard, 2002; Barker, 2006). Similar to Aksoy (2016), which investigates the relationship between house prices and fertility in England, I exploit the exogenous variation in regulatory restrictiveness across the UK to investigate departure from the family home and marriage.

This paper is organized as follows. The data sources are detailed in Section 2. I then present aggregate trends of interest. In Section 3, the empirical strategy is set out for both the static and dynamic analyses. I also discuss the endogeneity concerns and how they will be addressed. The results are presented in Section 4 first addressing living with parents and then cohabiting with a partner. In section 5 a series of robustness checks are presented before the conclusion.

2 Data

2.1 Data Sources

This paper utilizes the following data sources:

United Kingdom Household Longitudinal Survey (UKHLS), 1991 - 2019

The British Household Panel Survey (BHPS) and the United Kingdom Household Longitudinal Study (UKHLS) are annual nationally (UK) representative longitudinal surveys. The BHPS spans 1991–2008, whilst the UKHLS is an ongoing survey that began in 2009. (I will refer to USoc when discussing both surveys.) I utilise information from the first ten waves of the UKHLS stopping at 2019 to avoid complications from the Covid-19 pandemic period. The BHPS began with a sample of approximately 5,500 randomly selected private households containing around 10,000 individuals. In each wave, these original sample members are surveyed along with any new members of their household or members of newly formed households. In 2009, the UKHLS effectively replaced the BHPS, adopting a similar structure but with a significantly larger baseline sample (around 40,000 households). Across both surveys, a large range of individual and household-level information is collected including details about demographic and socioeconomic characteristics.

For this analysis, I include all observations of young people (20-29 years old) from the original sample between 1995 and 2019 to match the housing data detailed below.[10] Full-time students are omitted because university education is often tied to a location and sometimes a living arrangement. Whilst many individual characteristics such as age, race, parental status and employment status can be obtained from the main survey, an individual-level living arrangement variable is not included. I therefore determine living arrangements using reported marital status information and the house grid which records the relationship between each possible pair of individuals in the household.[11] This data structure also enables me to map parental characteristics to their adult child if they reside at the same address. I also have information about LAD of residence enabling me to map individuals to local housing market conditions. There are 296 different LADs observed in the sample used.

[10]The BHPS and UKHLS periodically increased the samples to facilitate research regarding smaller populations. I omit these from the sample to preserve the representativeness.

[11]By relationship/partnership I am grouping together both a cohabiting relationship and marriage.

Land Registry

For house price information, I use an annual house price index (HPI) provided by the Land Registry covering the period 1995 to 2019.[12] This index is based on actual residential sales transactions of new and resold houses. The measure accounts for seasonality and controls for housing stock quality to allow for like-for-like comparison through time. It smooths out short-term price fluctuations by taking a three-month rolling average. I construct yearly averages at the LAD level from the underlying monthly data. The base year is 2015 (=100). I utilize the HPI at the LAD level.

The analysis also utilizes planning application refusal rates, data for which is obtained from the Department of Housing, Communities and Local Government (DCLG). To build a new property in the UK, the developer must apply to the Local Planning Authority (LPA) for planning permission. These data capture the percentage of various types of planning applications that are rejected by the LPA. The data set covers 1995 to 2019 with a small amount of missing data.[13] I map the LPAs to the LADs using the 2023 boundaries.

Summary statistics are provided in Table A.1 in the Appendix.

2.2 Aggregate Trends

A: House Prices

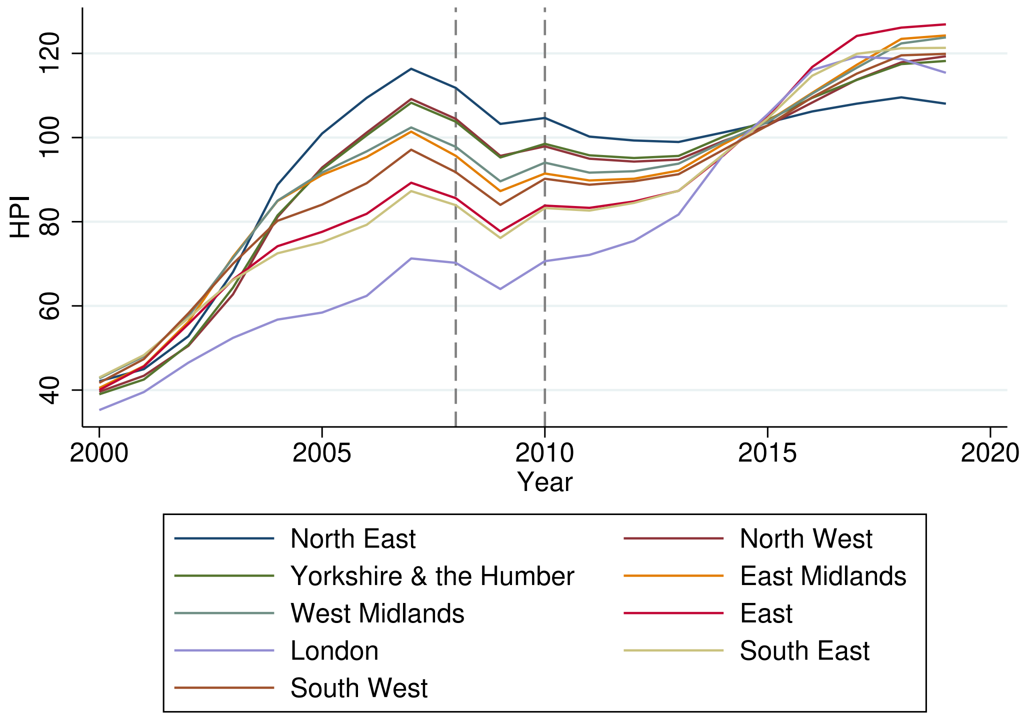
Figure 1 plots the regional HPI from 2000 to 2019. Overall we can see that since 2000 house prices have been increasing across regions of England. More precisely low interest rates and increased access to credit fuelled a period of substantial house price expansion across all regions until 2008. After the GFC, the average house decreased in 2008 and 2009, before starting to recover in 2010. While the recovery was timid in the less expensive regions of the country (North East, North West, West Midlands) until

[12]Instead I could have used data from the ONS, Nationwide and Halifax. I chose the Land Registry data because it is derived from a larger underlying sample since it also includes cash transactions.

[13]This is an unbalanced panel. 29 LPAs entered the data set after 1995 and 2 LPAs are missing data between 1999 and 2009. This issue appears to be due to the creation of unitary authorities creating issues with data recording. The main analysis includes the full set of LPAs. As a robustness check, I also examine the results excluding the LPAs that have missing data.

2014, the more expensive regions (London, South East, East) experienced consistently strong growth in house prices. Behind this trend, however, lies a great degree of heterogeneity within regions and the index obfuscates the price level difference across regions. This is demonstrated in Table A.2 in the Appendix.

Figure 1: HPI for English Regions

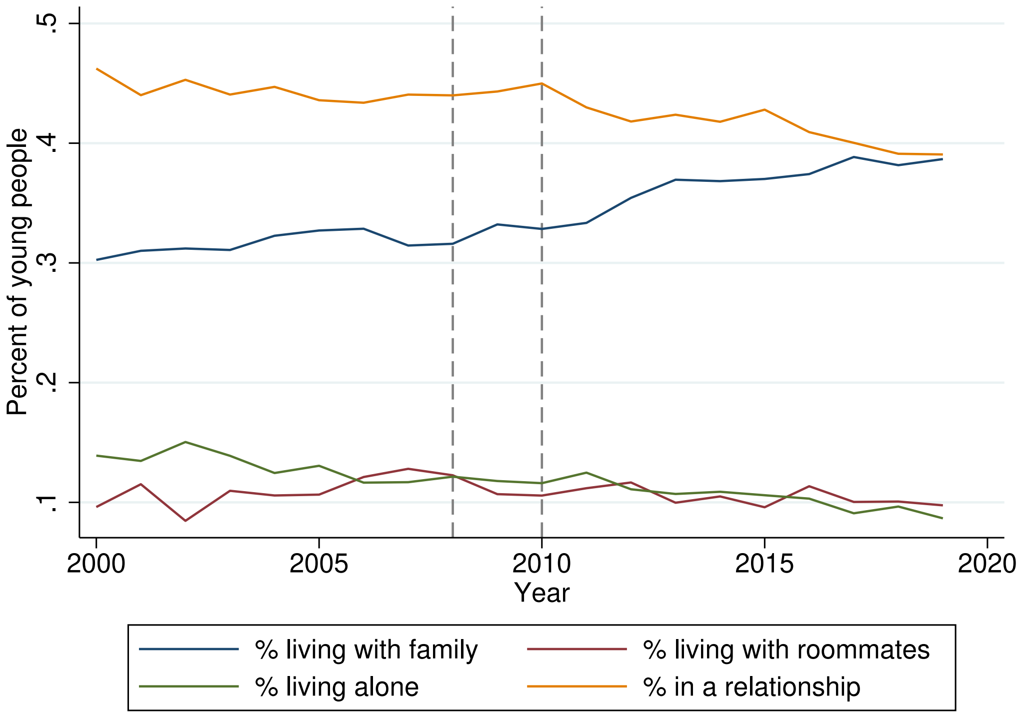


Note: Data is from the Land Registry. The regions correspond to the Government Office of Regions (GOR) categorization.

B: Living arrangements of young people

What was happening to the living arrangements of young people over this period? Figure 2 presents estimates of the percentage of young people (20-29) that are out of full-time education living in each living arrangement. The two vertical grey lines represent the GFC and 2010 when the austerity measures began to be implemented in the UK. The figure demonstrates that the percentage of young people living with their parents has been on an upward trend since 2000. There was an increase in the trend after 2010 as the economic downturn and austerity measures took hold. By 2018

Figure 2: Living Arrangements of Young People (20-29 years old) in England



Note: Data is from the Quarterly UK Labour Force Survey (LFS), 2000 - 2019. The living arrangements are constructed using the household grid. Survey weights were used to construct the percentages.

almost two-fifths of young people were living with their parents.[14] It is also clear that whilst most young people continue to live in a relationship, there is a clear decline during the economic downturn following the GFC; the percentage of young people living with a partner dropped by 3 percentage points between 2010 and 2012. The percentage of young people living with flatmates remains around 10% throughout the period, whilst the size of the minority of young people that live alone has been steadily decreasing from 14% to 8% between 2000 and 2018. Interestingly around the GFC and the immediate economic downturn the percentage living alone was flat potentially reflecting the sticky nature of housing contracts.

Although not presented here, if the age range included in the sample is altered, whilst there is a level effect, the patterns through time remain similar. The lower the age the greater the percentage living with parents, the greater living with roommates, and the smaller the percentage living in a relationship and alone. The impact of the post-GFC period is also generally magnified as the age decreases. Overall the patterns

[14] Knipe (2017) obtains a similar figure using the same underlying data.

suggest the vast majority of young people either live with their parents or with a partner. In addition, there has been a steady growth in co-residence with parents among young people in England at the same time there has been a decrease in living with a partner.

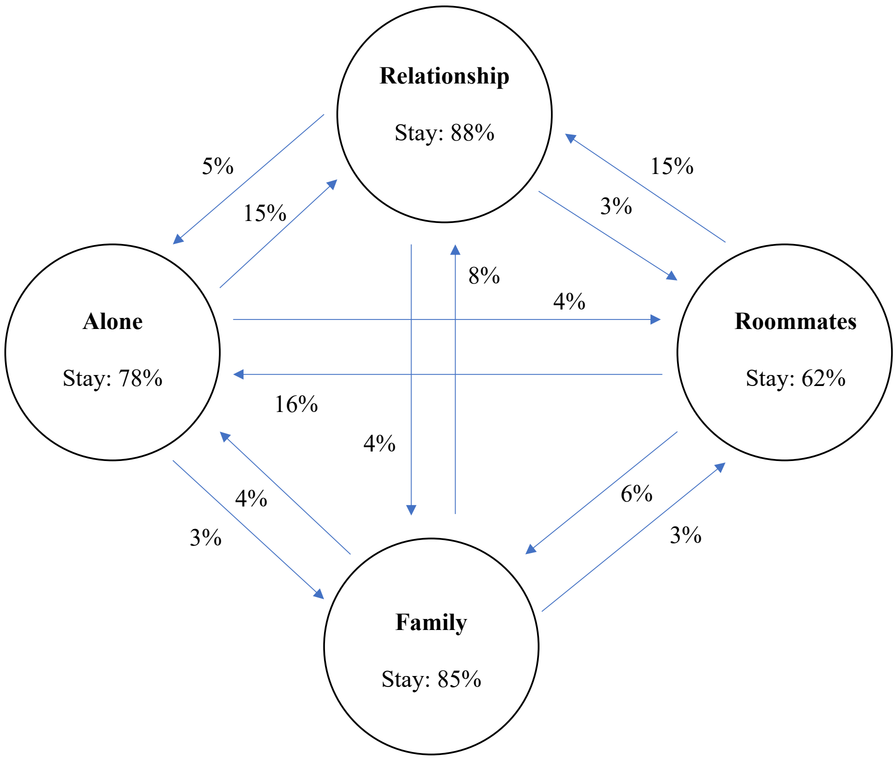
C: Average gross flows between living arrangements

Of course, levels only provide a partial picture. Is it that fewer young people are leaving the parental home, or is it that they are increasingly moving back in with their parents? Are young people delaying forming cohabiting relationships or are these relationships more fragile?

Figure 3 summarises the average annual flows between living arrangements of young people over the 2000 - 2019 period. It reports the average percentage that remained in each living arrangement and the percentage that transitioned to each possible alternative. The numbers do not account for young people moving between different forms of each living arrangement i.e. a young person living with a partner moving directly in with another partner in a given year would not be recorded as a transition.

The flow diagram demonstrates that the modern transition to adulthood is very diverse and multi-dimensional. Every possible flow is populated suggesting the transition to adulthood is not in one direction. Further, there is a surprising amount of instability in some types of living arrangements. In particular living in a flat share appears very unstable with 38% of individuals leaving flat sharing each year on average. It appears that the majority of those who leave a flat share progress to a more "adult" living arrangement - living with a partner or to lone accommodation. The percentage of those flowing out of living alone is also sizeable; on average one-fifth (22%) of people living alone leave this living arrangement each year. The vast majority of transitioning lone livers move in with a partner. Living with a partner is the most common and most stable living arrangement for young people. When a relationship ends, most young people transition to living alone or living with family. These patterns suggest, as expected, that living with roommates and living alone are used as transitory living arrangements, between living with the family and forming a relationship. The path between the two traditional life cycle states is no longer prescribed.

Figure 3: Average Annual Flow Between Living Arrangements of Young People (20-29 years old) in England



Note: Data is from the UKHLS, 2000 - 2019. Survey weights were used.

Considering the time series of outflows, the ranking remains the same across the observation window and so is reflected in Figure 3; flat-sharing is the least stable and a relationship is the most stable living arrangement. The data reveals however that the outflow from living with family has been on a downward trajectory since 2000. The outflow from a relationship appears relatively stable suggesting that any change in the share of young people living in relationships comes from a decline in the inflow. This echoes findings from elsewhere (Stone et al., 2011; Pelikh et al., 2022) which suggest that young people are delaying leaving the parental home and postponing forming their first cohabiting relationship.

3 Empirical Strategy

I follow Ermisch (1999) in organising the analysis by first considering a static discrete choice model, and then allowing for dynamics and focusing on transitions. Because the vast majority of young people live with parents or with a partner I place primary focus here. I use the same underlying empirical models to consider the choice to live with parents and the choice to live with a partner. Regarding the transitions, I am interested in transitions towards adulthood, therefore I focus on leaving the parental home and transitioning to a relationship (cohabiting or marriage).

3.1 Static Analysis

The static analysis considers how house prices influence young people's living arrangement outcomes. I utilize a linear probability model to estimate the effect of house prices on the probability of living with parents and the probability of living with a partner. This approach exploits the time variation in LAD house prices conditional on individual characteristics to document the correlation with the living arrangements of young people. Whilst the shortcomings of the linear probability model are well-known (Horrace and Oaxaca, 2006), my primary interest is in establishing causal effects, which have been shown to be well approximated by linear models (Angrist and Pischke, 2009).[15] I estimate the following model:

$$Y_{i,d,t} = X_{i,t-1}\beta + HPI_{d,t-1}\gamma + \delta_t + \tau_d + \epsilon_{i,d,t}. \quad (1)$$

The dependent variable $Y_{i,d,t}$ is a dummy variable set equal to one if the young adult co-resides with their parents in period t for the analysis considering living with family and is set equal to one if the young adult lives with a partner for the relationship formation analysis. The measure of the HPI is at the LAD level and has a one-year lag. Whilst this is designated as a static model the lagged independent variables are

[15]The baseline estimates are qualitatively the same as the estimates obtained from the linear model. Results can be found in Appendix Table A.3 and Table A.4.

included as a first step to address endogeneity. Concerning the key independent variable, straightforwardly because the living arrangement at time t is the outcome of a utility comparison at time t, there is a clear threat from simultaneity bias. It is likely that the chosen living arrangements of young people influences the demand in the local housing market thus violating the strict exogeneity condition required to obtain unbiased estimates. This concern is exacerbated by the annual frequency of the data. Lagging the measure of house prices circumvents the simultaneity loop because a future choice is less likely to influence prices today. I lag all of the other time-varying controls for similar reasons. $X_{i,t-1}$ is a vector of individual controls from the previous year $t-1$ including age, sex, race, whether the young adult is a parent, education level, foreign-born dummy, an unemployed dummy and gross income less transfers.[16]

LAD dummies and year dummies are also included to absorb common heterogeneity across time and space that may bias the estimates. Including the LAD dummies is crucial to ensure that the estimated relationship between choices and house prices is not confounded by time-invariant differences in preferences for any particular living arrangement. For example, it is possible that young adults with more individualistic/less conservative views sort into districts with higher house prices to enjoy other amenities. If this is the case, regarding the partnership formation analysis, there will be a negative correlation between house prices and living with a partner that is driven by this sorting, not by a causal effect of house prices.

3.2 Dynamic Analysis

To analyze the impact of house prices on transitions (out of the parental home and into a partnership), I also estimate a series of related linear probability models of the form:

$$Y_{i,d,t} = X_{i,t-1}\beta + HPI_{d,t-1}\gamma + \delta_t + \tau_d + \epsilon_{i,d,t}. \tag{2}$$

[16] I utilize gross income instead of net income or wages because the definitions of these desirable variables in the harmonized version of the BHPS and USoc are not perfectly aligned.

Now the dependent variable $Y_{i,d,t}$ is a dummy variable set equal to one if a specific change in living arrangement (away from parents/into a cohabiting relationship) is observed for individual i between period $t-1$ and period t. Thus the samples are reduced accordingly.

When focusing on transitions out of the parental home I am also able to include parental characteristics in the specification which may play an important role as they are related to the quality of the parental home and the ability of parents to provide financial support (Rosenzweig and Wolpin, 1993). Specifically, I include total real parental gross income less transfers which includes the income of parents residing in the same household, a dummy set equal to one if at least one parent has a university degree and a dummy variable indicating that the young adult lives with both biological parents.

3.3 Endogeneity

Estimates of both equations (1) and (2) may be biased due to endogeneity created by reverse causality and/or omitted variable bias. Reverse causality might occur due to demand-driven price changes: either moving out of the parental home or forming a relationship increases the demand for housing and puts upward pressure on house prices. This would generate a positive correlation between the variables of interest, causing the baseline estimates to underestimate the impact of house prices on probabilities. As mentioned above, lagging the independent variables goes some way to addressing this concern, but because these are dynamic processes with forward-looking individuals the threat of reverse causality is not eliminated.[17]

Second, obtaining unbiased estimates requires that house prices are uncorrelated with unobserved factors that are absorbed in the error term (omitted variables). Whilst the inclusion of LAD dummies controls for unobserved time-invariant LAD characteristics, they do not account for local shocks that are correlated with both house prices

[17]For example, local housing development companies make their plans partially based on forward-looking demographic projections, which in turn influence local house prices. Hill (2014)

and living arrangement outcomes. For example, a local large business shutdown may be negatively correlated with both conditions in the local housing market and the probability of moving in with a partner (Autor et al., 2019), leading to a downward bias of the baseline estimates. Unlike with reverse causality, however, I am unable to sign the potential bias because other types of shocks may cause a positive correlation between house prices and young people's living arrangement choices.

To obtain causal estimates I, therefore, adopt an instrumental variable approach similar to Dettling and Kearney (2014) and Aksoy (2016).[18] The idea of the instrument builds upon Glaeser and Gyourko (2002), which establishes that zoning restrictiveness is a key driver of house prices in the US. Gyourko et al. (2008) formalized this idea by creating an index for regularity restrictiveness which was subsequently used in Saiz (2010) to develop an objective index of housing supply elasticity capturing the ease with which new housing can be built in an area accounting for local topology and regulatory restrictiveness. Mian and Sufi (2009) show that this measure of housing supply elasticity is strongly related to house price growth in the US, such that more inelastic areas exhibit stronger growth, which in turn is related to house price levels.

In the UK context, Hilber and Vermeulen (2016) find that the restrictiveness of the planning system influences house prices, although topology seems to play a limited role. Therefore whilst no formal indices are available for the UK, similar to Aksoy (2016), I capture regulatory restrictiveness as the percentage of major dwelling development applications rejected by the LPA in the given year.[19] Precisely, using this information at the LAD level, I construct a 3-year moving average of the fraction of large developments rejected. I anticipate that the greater the fraction rejected the more restrictive the LAD, thus the greater the local HPI in that year.

Validity of this instrument requires that living arrangement choices are only correlated with local regulatory restrictiveness through house prices. This condition is threatened on a number of fronts. Firstly, it is possible that local shocks are corre-

[18]Other examples of papers that utilize a similar instrument are: Kaplan et al. (2020); Mian et al. (2013).

[19]A major dwelling development is defined as one that consists of at least 10 separate homes.

lated with planning authority decision-making and young people's living arrangement choices. LPAs are often composed of committee members and councillors drawn from the local community with vested interests in the local areas and therefore naturally their decision-making may be influenced by local political pressures and/or self-interest. The highly centralized and time-intensive nature of the UK planning system however limits potential local influence. In the UK, planning permission from a local authority is required for every housing development. These local authorities, however, are set strict targets by centralized bodies as part of five-year plans constructed using backward-looking information. Further, the remaining discretion the LPAs do have is also likely tied to decisions made by central government via the allocation of funding for local services which typically drives local pressure. Therefore whilst local officials do have a say in planning decisions, the scope to respond to local shocks is limited and highly influenced by decisions taken outside of the local community.

I also account for short-run shocks in the construction of the instrument. I utilize a 3-year moving average of the local rejection rate to reduce the potential correlation with local shocks in time t. Taking the average also overcomes the concern that developers may strategically submit multiple applications in any given year.

An additional concern is raised by Davidoff (2015) who highlights the role of sorting which may invalidate housing elasticity type instrument. He argues that highly productive people are often attracted to high-growth, inelastic areas for more desirable amenities etc. If these individuals have characteristics that are correlated with the dependent variable of interest this violates the exogeneity condition invalidating the instrument. As mentioned above, however, including the LAD area dummies should address these concerns based on longer-term conditions that are in place throughout the panel. However, shorter-term changes in sorting-driven local changes that take place during the panel remain are not addressed with this strategy.

4 Results

4.1 Living with Parents

I begin by focusing on the factors associated with living with parents. Table 1 presents the estimates of the baseline static specification given in equation (1). The sample contains all observations of young people and the dependent variable is equal to one if the young adult resides with their parents. I suppress the estimates of the demographic controls but present those related to the labour market. The first column includes just the main independent variable of interest, lagged HPI for the LAD of residence. This very sparse specification yields a large positive estimate of 0.0249, which is significant at the 1 per cent level. This is consistent with a positive price effect of house prices on parental co-residence as predicted. Column 2 adds individual controls to the specification. The point estimate of local LAD HPI remains positive and highly significant, however the magnitude of the coefficient is significantly smaller. Whilst the negative and weakly significant point estimate on unemployment may appear puzzling, this is in line with the existing literature (Kaplan et al., 2009) which highlights the important role dynamics and persistence in labour market characteristics.[20] Income, however, is found to facilitate emancipation: the point estimate of real gross income is negative and highly significant.

The final column adds year and LAD dummies to control for common year and LAD unobserved factors. Whilst this accounts for a considerable amount of variation, again LAD HPI has a positive and statistically significant estimate. The estimated β is 0.00399 suggesting that a 10 percent increase in local house prices increases the probability a young person resides with their parents by 4 percentage points. This is greater than the estimates in Andrew and Meen (2003), which might be explained by the greater house prices during my window of observation.

[20]The significance of unemployment is not due to multicollinearity, because the correlation between unemployment and gross income is very strong (-0.129).

Table 1: Living With Parents, Static Linear Probability Model Estimates

	(1)	(2)	(3)
$\text{HPI}_{d,t-1}$	0.0249***	0.00990***	0.00399**
	(0.00235)	(0.00124)	(0.00193)
$\text{Unemployed}_{i,t-1}$		-0.0138	-0.00670
		(0.00847)	(0.00819)
Gross $\text{income}_{i,t-1}$ (00s)		-0.00211***	-0.00243***
		(0.000324)	(0.000298)
N	26,779	26,779	26,779
Adj. R^2	0.0214	0.657	0.674
Individual controls		✓	✓
LAD dummies			✓
Year dummies			✓

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Note: The sample contains young people (20-29 year olds) living in England in USoc between 1995 and 2019. The dependent variable is an indicator of living with parents. The base year of the HPI is 2015. The real gross income is measured in hundreds of GBP. Individual controls include age, age squared and a series of dummies capturing if the individual is non-white, whether they have a degree, whether they are foreign-born, if they are a parent and gender. All of the control variables are lagged by one year. Sample weights are utilized. Standard errors are presented in parenthesis and are clustered at the PSU level.

Next, Table 2 presents the estimates of the dynamic linear probability baseline model specified in equation (2). The set-up of the results is similar to those above in Table 1. Overall the estimates support a similar narrative. Across all three specifications, the point estimate for lagged LAD HPI is negative and highly statistically significant. Focusing on the saturated model in column 3, the estimate suggests a 10 percent increase in local house prices decreases the probability of leaving the parental home by 1.6 percentage points. Regarding labour market variables, the effect of being unemployed on transitioning away from parents is still negative and insignificant, however, the sign is now in line with basic intuition. Income is correlated with household formation: the estimates for gross income are positive and highly statistically significant.

Table 2: Living With Parents, Dynamic Linear Probability Model Estimates

	(1)	(2)	(3)
$\text{HPI}_{d,t-1}$	-0.00111***	-0.00107***	-0.00157**
	(0.000192)	(0.000155)	(0.000767)
$\text{Unemployed}_{i,t-1}$		-0.00479	-0.0128
		(0.0116)	(0.0119)
Gross $\text{income}_{i,t-1}$ (00s)		0.00205***	0.00186***
		(0.000588)	(0.000580)
Both biological $\text{parents}_{i,t-1}$		-0.493***	-0.511***
		(0.0177)	(0.0178)
Total parental gross $\text{income}_{i,t-1}$		-0.000650***	-0.000672***
		(0.000177)	(0.000182)
Parental $\text{degree}_{i,t-1}$		0.0361***	0.0339**
		(0.0114)	(0.0140)
N	6,778	6,778	6,778
Adj. R^2	0.00778	0.461	0.498
Individual controls	✓	✓	✓
LAD dummies			✓
Year dummies			✓

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Note: The sample contains young people (20-29-years old) living in England in USoc between 1995 and 2019. An observation is included if the young adult is observed in the previous year living with their parents. The dependent variable is an indicator set equal to one if the young adult is observed not to be living with their parents. The base year of the HPI is 2015. The real gross income is measured in hundreds of GBP. The total parental income captures the same measure for all parents living in the home. Parental degree is a dummy variable set equal to one if at least one person has a university degree. Individual controls include age, age squared and a series of dummies capturing if the individual is non-white, whether they have a degree, whether they are foreign-born, if they are a parent and gender. All of the control variables are lagged by one year. Sample weights are utilized. Standard errors are presented in parenthesis and are clustered at the PSU level.

Given the sample used to estimate the dynamic model only contains young adults who lived with their parents in the previous year, I am also able to include controls for parent characteristics. Having both biological parents at home greatly reduces the probability of leaving the paternal home. Parental income is also significantly associated with a lower probability of leaving the family home. This supports the theoretical claims of Ermisch (2016); the positive impact of higher parental income on the consumption of housing services is larger than the impact of private transfers which facilitate independence. Having more educated parents is also correlated with a transition to independence.

Overall these estimates suggest that higher local house prices are associated with higher parental co-residence, which (at least in part) is explained by a reduction in the probability of transitioning out of the parental home. The size of the estimates is relatively modest throughout. As explained above, these estimates should not be given a causal interpretation because they likely suffer from bias due to omitted variables and reverse causality. Table 3 presents the IV estimates in which the 3-year moving average of the local planning application rejection rate is used as an instrument.

Before discussing the IV results, it is important to acknowledge the reduction in the sample sizes compared to the previous tables. This is due to missing data in the application rejection data set. These gaps are created for some LADs due to the restructuring of the local authorities. Specifically, during this period a number of LADs merged into unitary authorities. Since LADs had responsibility for housing, this also meant a reorganisation of local planning authorities. Consequently, this process created some data loss in the data set used here. To provide reassurance that this does not bias the IV estimates, I present the baseline OLS estimates using this restricted sample first.

Table 3 is organized into two blocks of three columns. The first block presents results for the static specification (equation (1)) and the second block for the dynamic specification (equation (2)). The first column in both blocks corresponds to estimates of the baseline saturated OLS models from column 3 in Tables 1 and 2 using the reduced sample. Reassuringly these estimates are quantitatively similar to the full sample baseline results; same in sign, magnitude and level of statistical significance as those using the full sample. The second column in each block presents the first-stage estimates. The refusal rates are used to capture the regulatory restrictiveness of the planning system and therefore are anticipated to be positively correlated with prices. In both instances, the estimated first-stage relationship between refusal rates and house prices is as expected: the three-year moving average refusal rates are statistically significantly associated with house prices at the 1 percent level. The results of the first stage F statistics are both above the standard criteria for a strong instrument.

Given the validity of the instrument, turning to the second stage results. Across both models, the IV estimates maintain the same sign as the baseline estimates and are statistically significant at the 5 percent level. Noteworthy is the magnitude of the IV estimates. In both cases the estimates more than double in absolute value. This is consistent with the idea that the OLS estimates suffer from reverse causality, and has been observed in other studies that use similar instruments (Aksoy, 2016; Hill, 2014; Dettling and Kearney, 2014; Wrenn et al., 2019; Li and Wu, 2019). For the static model, a greater probability of living with parents decreases local house prices, thus imposing a downward bias on the OLS estimates. Regarding the dynamic model, the probability of leaving the parental home positively impacts house prices, thus also imposing a downward bias on the OLS estimates. To interpret the estimates, the IV point estimate for the static model is 0.00862 suggesting that a 10 percent increase in house prices increases the probability of living with parents by around 8.5 percentage points. The estimated effect in the dynamics models suggests that a 10 percent increase in house prices reduces the probability of leaving the parental home by 3.6 percentage points.

Table 3: Living With Parents, Instrumental Variable Model Estimates

	Static			Dynamic		
	Baseline	First Stage	IV	Baseline	First Stage	IV
$\text{HPI}_{d,t-1}$	0.00418**		0.00862**	-0.00163**		-0.00362**
	(0.00201)		(0.00402)	(0.000751)		(0.00163)
3 yr MA Rejection $\text{rate}_{d,t-1}$		0.241***			0.381***	
		(0.0432)			(0.0849)	
N	25,014	25,014	25,014	6,527	6,527	6,527
Adj. R^2	0.673	0.962	.	0.498	0.964	.
F Statistic		14.58		13.43		
Individual controls	✓	✓	✓	✓	✓	✓
LAD dummies	✓	✓	✓	✓	✓	✓
Year dummies	✓	✓	✓	✓	✓	✓

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Note: The sample contains young people (20-29 year olds) living in England in USoc between 1995 and 2019. The dependent variable in the static analysis is set equal to one if living with parents. In the dynamic analysis, an observation is included if the young adult is observed in the previous year living with their parents. The dependent variable is an indicator set equal to one if the young adult is observed not to be living with their parents. The base year of the HPI is 2015. The real gross income and the parental gross income are measured in hundreds of GBP. The total parental income captures the same measure for all parents living in the home. Individual controls include age, age squared and a series of dummies capturing if the individual is non-white, whether they have a degree, whether they are foreign-born if they are a parent and gender. All of the control variables are lagged by one year. The instrumental variable used is the 3-year moving average local planning application rejection rate for major dwelling projects (10 plus homes). Sample weights are utilized. Standard errors are clustered at the PSU level.

Combined these estimates support the popular idea that high house prices suppress the independence of young adults as captured by forming an independent household, however, the impact is again relatively modest. This is in line with many of the previous findings including Ermisch and Di Salvo (1997); Ermisch (1999); Li and Wu (2019); Andrew and Meen (2003); Bayrakdar and Coulter (2018). I also find evidence throughout that income is important. In addition, the difference between the static and dynamic estimates suggests that there is a considerable amount of "false starting" going on among young people i.e. returning to the parental household, which recent studies have documented Stone et al. (2014); Kaplan (2012). Next, I investigate whether a similar relationship-inhibiting effect can be established regarding forming a cohabiting relationship.

4.2 Living with a Partner

In this section I present the results of the analysis that considers the association between the relationship formation of young people and house prices. Andrew and Meen (2003) and Ermisch and Di Salvo (1997) find that demographic transitions are key to understanding leaving the parental home therefore one would anticipate a similar impeding impact of house prices to the analysis above. However, as demonstrated in Figure 3, one transition no longer implies the other, therefore it might be the case that relationship formation is related to individual and environmental characteristics differently. The presentation of the results is similar to that in the previous section.

First, Table 4 offers the OLS estimates of equation (1) using a dummy variable set equal to one if the young adult resides with a partner and zero otherwise. Across the three specifications the statistically significant point estimates suggest a negative association between house prices and living with a partner. Focusing on the saturated model (column 3), a 10 percent increase in local house prices relates to a 1.4 percentage point increase in the probability of living with a partner, significant at the 10 percent level. Labour market characteristics enter significantly with the anticipated sign: being unemployed is negatively related to living with a partner and income is positively related.

Table 4: Living With a Partner, Static Linear Probability Model Estimates

	(1)	(2)	(3)
$HPI_{d,t}$	-0.0221***	-0.00404***	-0.00136*
	(0.00231)	(0.00134)	(0.000806)
$Unemployed_{i,t}$		-0.0249**	-0.0168*
		(0.00964)	(0.00919)
Gross $income_{i,t}$		0.00390***	0.00403***
		(0.000365)	(0.000367)
N	26,779	26,779	26,779
Adj. R^2	0.0164	0.598	0.615
Individual controls	✓	✓	✓
LAD dummies			✓
Year dummies			✓

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Note: The sample contains young people (20-29 year olds) living in England in Usoc between 1995 and 2019. The dependent variable is an indicator set equal to one if the young adult lives with a partner. The base year of the HPI is 2015. The real gross income and total parental gross income are measured in hundreds of GBP. Parental degree is a dummy variable set equal to one if at least one person has a university degree. Individual controls include age, age squared and a series of dummies capturing if the individual is non-white, whether they have a degree, whether they are foreign-born, if they are a parent and gender. All of the control variables are lagged by one year. Sample weights are utilized. Standard errors are presented in parenthesis and are clustered at the PSU level.

Estimates of the dynamic specification are provided in Table 5. These results echo the story from the static model. The estimated coefficients on LAD HPI are negative across specifications. Labour market characteristics again enter as expected; employment and more income facilitate forming a relationship. However, the estimates of unemployment status are now not statistically significant reflecting again the complex relationship suggested in the parental co-residence analysis.

Table 5: Living With a Partner, Dynamic Linear Probability Model Estimates

	(1)	(2)	(3)
$\text{HPI}_{d,t-1}$	-0.00777***	-0.00751***	-0.00476**
	(0.00116)	(0.00118)	(0.00209)
$\text{Unemployed}_{i,t-1}$		-0.00871	-0.00934
		(0.00759)	(0.00789)
$\text{Gross income}_{i,t-1}$		0.00206***	0.00239***
		(0.000397)	(0.000417)
N	13,232	13,232	13,232
Adj. R^2	0.00516	0.175	0.310
Individual controls	✓	✓	✓
LAD dummies			✓
Year dummies			✓

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Note: The sample contains young people (20-29 year olds) living in England in USoc between 1995 and 2019. An observation is included in the sample if they are observed not living with a partner in the previous year. The dependent variable is an indicator set equal to one if the young adult lives with a partner in the current year. The base year of the HPI is 2015. The real gross income and total parental gross income are measured in hundreds of GBP. Parental degree is a dummy variable set equal to one if at least one person has a university degree. Individual controls include age, age squared and a series of dummies capturing if the individual is non-white, whether they have a degree, whether they are foreign-born, if they are a parent and gender. All of the control variables are lagged by one year. Sample weights are utilized. Standard errors are presented in parenthesis and are clustered at the PSU level.

Turning to the IV estimates in Table 6. First, again the reduction in the sample due to data loss does not appear to bias the estimates of the baseline OLS specifications. Further, the F statistics provide evidence that the instrument is strong. Interestingly, in this case, the IV estimates reported in columns 3 and 6 are not statistically significant, suggesting that the baseline estimates were capturing spurious correlations. That is, once I isolate the exogenous variation in house prices, a statistically significant negative correlation between house prices and relationship formation is no longer detected. Thus, I do not find evidence that house prices influence relationship formation among young adults in England.

Comparing the results, one interpretation is that housing affordability inhibits young people's transition to independence, but, on average, is not a determining factor when thinking about partnering up. This reflects the results in Bayrakdar and Coulter (2018), using a subset of the data I use here. Not addressing endogeneity, they find a weaker and smaller impact of house prices on partnership formation than on transitions to other destinations from living with parents. In spite of the popular discourse, this is not too surprising given the prominent role of the match quality/love draw in driving relationship turnover in the marriage literature.

Table 6: Living With a Partner, Instrumental Variable Model Estimates

	Static			Dynamic		
	Baseline	First Stage	IV	Baseline	First Stage	IV
$HPI_{d,t-2}$	-0.00139*		-0.00516	-0.00464**		-0.00825
	(0.000831)		(0.00407)	(0.00221)		(0.00681)
3 yr MA Rejection $rate_{d,t-1}$		0.241***			0.217***	
		(0.0432)			(0.0643)	
N	25,014	25,014	25,014	12,797	12,797	12,797
Adj. R^2	0.609	0.962	.	0.316	0.958	.
F Statistic		14.14			11.92	
Individual controls	✓	✓	✓	✓	✓	✓
LAD dummies	✓	✓	✓	✓	✓	✓
Year dummies	✓	✓	✓	✓	✓	✓

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Note: The sample contains young people (20-29 year olds) living in England in USoc between 1995 and 2019. The dependent variable in the static analysis is set equal to one if living with a partner. In the dynamic analysis, an observation is included if the young adult is observed in the previous year not living with a partner. The dependent variable is an indicator set equal to one if the young adult is observed not to be living with a partner in the current year. The base year of the HPI is 2015. Individual controls include age, age squared and a series of dummies capturing if the individual is non-white, whether they have a degree, whether they are foreign-born, if they are a parent and gender. All of the control variables are lagged by one year. The instrumental variable used is the 3-year moving average local planning application rejection rate for major dwelling projects (10 plus homes). Sample weights are utilized. Standard errors are clustered at the PSU level.

5 Heterogeneity and Robustness

In this section I implement various robustness checks on the model specification and sample to explore whether this impacts the estimates. For the sake of space, I only present the robustness checks for analysis that considers co-residence with parents since the relationship formation IV estimates yielded insignificant results. Further when I ran the same checks, in general, the null effect prevailed.

5.1 Alternative Control Variables

I start by presenting estimates of alternative specifications of both the static and dynamic models in Table 7 which explore the interpretation and specification of the main independent variable of interest. The top four rows present the estimates of versions of the saturated baseline OLS model (column 3 in Tables 1 and 2) and the bottom four rows contain the corresponding IV estimates. Column 1 in both blocks of Table 7 includes an additional dummy variable set equal to one if the individual reports they are in a difficult financial position.[21] The inclusion of local house prices and the control for income captures the idea of an affordability constraint influencing choices. However, this additional subjective measure is related not only to sentiment about financial position today but also about the future which perhaps more accurately reflects the individual's state of mind when decision-making. In the static model, the estimates are negative and statistically significant at the 1 percent level in both the baseline and IV versions. Initially, this may seem unexpected; young people who feel under difficult financial conditions are less likely to live with their parents. However, when living with parents, it is likely that young people pay little, if any, rent therefore when surveyed they are, on average, benefiting from the cost saving: it captures the outcome of the arrangement rather than the motivation. The estimates for LAD HPI remain similar in magnitude and are statistically significant. Thinking about transitions out of the parental home, self-reported financial difficulties enter with negative coefficients which are not statistically significant. The main estimates of interest remain similar. These estimates therefore provide some evidence that young people utilise their parental home to alleviate financial stress.

[21] This variable is derived from a survey question that allows the respondent to select one of 5 options describing their financial situation. I transformed this information into a dummy variable set equal to one if the individual reports either "Finding it quite difficult" or "Finding it very difficult".

Table 7: Living With Parents: Alternative Independent Variable Estimates

	Static			Dynamic		
	(1)	(2)	(3)	(1)	(2)	(3)
OLS:						
$\text{HPI}_{d,t-1}$	0.00435**			-0.00162**		
	(0.00213)			(0.000788)		
HPI - Terraced $\text{houses}_{d,t-1}$		0.00282***			-0.00206**	
		(0.00101)			(0.000952)	
HPI - 3 yr $\text{MA}_{d,t-1}$			0.00368**			-0.00127*
			(0.00172)			(0.000754)
Financial $\text{difficulty}_{i,t-1}$	-0.0231***			-0.00638		
	(0.00782)			(0.0155)		
IV:						
$\text{HPI}_{d,t-1}$	0.00732**			-0.00349***		
	(0.00336)			(0.00129)		
HPI - Terraced $\text{houses}_{d,t-1}$		0.00337**			-0.00465**	
		(0.00165)			(0.00211)	
HPI - 3 yr $\text{MA}_{d,t-1}$			0.00651*			-0.00347*
			(0.00364)			(0.00201)
Financial $\text{difficulty}_{i,t-1}$	-0.0227***			-0.00532		
	(0.00779)			(0.0137)		
N	25,014	25,014	25,014	6,527	6,527	6,527
Individual controls	✓	✓	✓	✓	✓	✓
Parent controls				✓	✓	✓
LAD dummies	✓	✓	✓	✓	✓	✓
Year dummies	✓	✓	✓	✓	✓	✓

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Note: The sample contains young people (20-29 year olds) living in England in USoc between 1995 and 2019. The dependent variable in the static analysis is set equal to one if living with parents. In the dynamic analysis, an observation is included if the young adult is observed in the previous year living with their parents. The dependent variable is an indicator set equal to one if the young adult is observed not to be living with their parents. The base year of the HPI of all houses and just terraced houses is 2015. Individual controls and parent controls are the same as for the baseline analysis. All of the control variables are lagged by one year. The instrumental variable used is the 3-year moving average local planning application rejection rate for major dwelling projects (10 plus homes). Sample weights are utilized. Standard errors are clustered at the PSU level.

The next two columns in both blocks consider specifications that utilize alternative measures of LAD HPI. The first is the LAD HPI of terraced houses only. Because these are smaller, lower-priced houses, they are perhaps more likely to be part of young people's feasible set. Unsurprisingly, due to the high correlation between house prices of different types of houses, the estimates remain qualitatively unchanged. In both columns 3, I use a 3-year moving average of the LAD HPI, again similar results are obtained although the statistical significance of the IV estimate has reduced to the 10 percent level.

5.2 Sub-sample Analysis

Next, I consider how co-residence with parents varies across different sub-samples. Results are presented in Table 8. Given the distinct characteristics of the London housing market (very high prices and greater price volatility), I begin by estimating the model on the young people outside of the London region and then the young people who live in London. Due to the high price level in the capital city, the estimated price effects are anticipated to be stronger in this sub-sample. This is indeed what the static estimates suggest; the magnitude of the estimated effects is larger for the London sub-sample than for the rest of England. Interestingly, the impact on the probability of leaving the parental home however is smaller in magnitude than in the rest of England. One possible explanation for this is that affluent parents living in the capital facilitate independence by compensating higher house prices with higher transfers as found in the US (Bleemer et al., 2014). The difference for both analyses however is not statically significantly different from zero.

I also estimate both the static and dynamic specifications for males and females separately. The estimated effect of LAD HPI is statistically significant across specifications for males, but statistical significance varies for females. Men on average leave the parental home at an older age than women, the estimates suggest this may in part be influenced by their greater sensitivity to house prices. Note, however, the difference is statistically significant only in the static model.

Table 8: Living With Parents: Sub-sample Analysis

	Static				Dynamic			
	No London	London	Males	Females	No London	London	Males	Females
OLS:								
$\text{HPI}_{d,t-1}$	0.00242**	0.00792***	0.00648**	0.00210*	-0.00183**	-0.000204***	-0.00181*	-0.00128**
	(0.00108)	(0.00277)	(0.00318)	(0.00109)	(0.000864)	(0.0000712)	(0.000990)	(0.000504)
IV:								
$\text{HPI}_{d,t-1}$	0.00487**	0.0103*	0.00988**	0.000283	-0.00212*	-0.000172**	-0.00320**	-0.00285*
	(0.00211)	(0.00545)	(0.00466)	(0.000221)	(0.00126)	(0.0000775)	(0.00163)	(0.00163)
N	22,624	2,390	11,161	13,853	5,895	632	3,529	2,998
Individual controls	✓	✓	✓	✓	✓	✓	✓	✓
Parent controls					✓	✓	✓	✓
LAD dummies	✓	✓	✓	✓	✓	✓	✓	✓
Year dummies	✓	✓	✓	✓	✓	✓	✓	✓

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Note: The sample contains young people (20-29 year olds) living in England in USoc between 1995 and 2019. The dependent variable in the static analysis is set equal to one if living with parents. In the dynamic analysis, an observation is included if the young adult is observed in the previous year living with their parents. The dependent variable is an indicator set equal to one if the young adult is observed not to be living with their parents. The base year of the HPI of all houses is 2015. Individual controls and parent controls are the same as for the baseline analysis. All of the control variables are lagged by one year. The instrumental variable used is the 3-year moving average local planning application rejection rate for major dwelling projects (10 plus homes). Sample weights are utilized. Standard errors are clustered at the PSU level.

5.3 Booms and Busts

For the period under examination the housing market in England experienced periods of boom and bust. Figure 1 illustrates that these periods were relatively synchronous across regions. Estimates here average across all parts of the housing market cycle, however, it is possible that the effects are asymmetric i.e. effects of house prices on living arrangement choices in different ways during booms than busts. For example, given the housing market is pro-cyclical, it might be that during booms, young people's earnings are less able to keep up with the growing prices and therefore are more likely to live with their parents or perhaps, during booms young people are optimistic about labour market prospects and so the negative impact of house prices is diluted.

I therefore estimate the models separately for housing market booms and busts. The UK experienced two relevant housing booms in the data set from 2000 until the GFC in 2008, and then from 2013 until 2019. The aftermath of the GFC was characterised as a bust period (2009 - 2012). The estimates of this exercise are presented in Table 9. The results suggest that house prices are only related to transitions during busts, while the main results stand during booms. Papers focusing on the labour market suggest labour market conditions are more important in bust periods (Kaplan, 2012), I find some evidence that housing markets are more important during booms. However, the difference is not statistically significant across the specifications.

Table 9: Living With Parents: Booms and Busts

	Static		Dynamic	
	Boom:	Bust:	Boom:	Bust:
OLS:				
$HPI_{d,t-1}$	0.00672**	0.00155	-0.00179*	-0.00123**
	(0.00332)	(0.00135)	(0.000944)	(0.000585)
IV:				
$HPI_{d,t-1}$	0.00886**	0.00202	-0.00383**	-0.00163
	(0.00418)	(0.00230)	(0.00195)	(0.00175)
N	18,968	6,046	5,285	1,242
Individual controls	✓	✓	✓	✓
Parent controls			✓	✓
LAD dummies	✓	✓	✓	✓
Year dummies	✓	✓	✓	✓

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Note: The sample contains young people (20-29 year olds) living in England in USoc between 1995 and 2019. The boom columns sub-sample contains observations from 2000-2008 and 2013-2019, whilst the bust columns sub-sample contains observations from 2009-2012. The dependent variable is an indicator set equal to one if the young adult is observed not to be living with their parents. The base year of the HPI of all houses is 2015. Individual controls and parent controls are the same as for the baseline analysis. All of the control variables are lagged by one year. The instrumental variable used is the 3-year moving average local planning application rejection rate for major dwelling projects (10 plus homes). Sample weights are utilized. Standard errors are clustered at the PSU level.

5.4 Alternative Instruments

Finally, I investigate whether the estimated relationships withstand the use of alternative but related instruments. As mentioned above, using a time-varying measure rejection rates may not be exogenous because it is still related to contemporaneous shocks. Taking the lagged 3-year moving average, attempts to overcome this problem, however, it remains a possible threat to validity. Therefore I also use the 5-year moving average as well as the average refusal rate over the observation window interacted with the HPI for England similar to Dettling and Kearney (2014). Table 10 presents the results. In both instances, the results are weaker but the interpretation remains intact. These results should be interpreted with some caution however because the F-statistics when using the average rejection rate interacted with the national HPI is close to/just under 10.

Table 10: Living With Parents: Alternative Instruments

	Static		Dynamic	
	5 Yr MA Rejection Rate	Avg Rejection Rate × HPI	5 Yr MA Rejection Rate	Avg Rejection Rate × HPI
HPI_{dt}	0.00422**	0.00463*	-0.00151**	-0.00275
	(0.00206)	(0.00271)	(0.000735)	(0.00214)
N	23,906	25,014	6,257	6,527
F Statistics	12.33	10.26	11.32	9.32
Individual controls	✓	✓	✓	✓
Parent controls			✓	✓
LAD dummies	✓	✓	✓	✓
Year dummies	✓	✓	✓	✓

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Note: The sample contains young people (20-29 year olds) living in England in USoc between 1995 and 2019. The dependent variable in the static analysis is set equal to one if living with parents. In the dynamic analysis, an observation is included if the young adult is observed in the previous year living with their parents. The dependent variable is an indicator set equal to one if the young adult is observed not to be living with their parents. The base year of the HPI of all houses is 2015. Individual controls and parent controls are the same as for the baseline analysis. The instrumental variable used in columns 1 and 3 the 5-year moving average local planning application rejection rate for major dwelling projects (10 plus homes). The instrumental variable used in column 2 and column 4 is the average rejection rate between 1995 and 2019 interacted with the English HPI. Sample weights are utilized. Standard errors are clustered at the PSU level.

6 Conclusion

The aim of this paper was to investigate the popular conception that the ”housing crisis” in the UK has led to young people delaying their transition to adulthood. Work elsewhere has tried to obtain causal estimates of the impact of house prices on fertility in England (Aksoy, 2016). In this paper, I considered two other life cycle transitions towards adulthood leaving the parental home and transitioning to a cohabiting relationship. An IV approach was adopted to address clear endogeneity concerns (omitted variable bias and reverse causality), the likely bias of the estimates obtained in the existing literature. The local average rejection rate is used as an instrument for local house prices capturing the idea that more regulatory restrictiveness leads to higher prices in a given district. The results suggest that while high local house prices are related to increasing financial dependency on parents through co-residence, there is limited causal evidence of a similar effect of house prices on relationship formation. Thus, I do not find any evidence in these data ”new, money-shaped shadow is looming over millennials' dating lives and relationships”. (The Guardian, 2023)[22]

A number of future research possibilities arise from this paper. Firstly, I utilized house price information to capture the cost of housing even though increasingly young adults have to turn to the private rental sector for housing. The reason for this is that good quality data on rents for England is not available. However, increasingly this information is being made available on online property listing websites and therefore may be used for future research as longer panels become available.

One dimension not explored here is how public policy influences choices. An obvious policy area is public support for housing. Housing low-income individuals in England shifted from a regime focused on housing provision to a system based on subsidising private renting in the 1980s. It would be interesting to investigate whether this shift in the type of housing support offered plays a role in the delay to adulthood. Another natural extension is to consider the role and generosity of housing benefits, since the GFC this has been part of a stated agenda to reduce the welfare benefits bill by transferring responsibility for young adults from the state back onto families (Cameron, 2012). Housing benefit caps and changes to eligibility criteria have made living independently more challenging and hence less attractive for young people

[22] Quote taken from an op-ed piece which can be found herehttps://www.theguardian.com/lifeandstyle/2023/jul/22/housing-crisis-killing-romance-modern-dating-jane-austen.

(Clair, 2022).[23] It would be interesting to explore the role of limited social housing provision and reduction in the generosity of government support for young people. Utilizing regional exposure instruments could be one useful approach to adopt.

Finally, the results in this paper might help contribute to thinking about important policy questions. In particular, whilst the initial descriptive analysis in this paper highlights that leaving the parental home no longer implies starting a family, the results indicate the transitions considered here are clearly related to living with parents significantly reducing the probability of moving in with a partner. Many Western governments are currently grappling with below-replacement fertility rates, the housing market might be one policy tool for governments to seek to alleviate this problem.

[23]For example the coalition government restricted housing benefits for those aged under 35, to the level of a room in a shared house. The Shared Accommodation Rate (SAR) was originally introduced as the Single Room Rate, capping housing benefits for those aged under 25, but was extended to those aged under 35 in April 2012.

Chapter 2
Seeking the One: Love, Betrayal and the Desire for Exclusivity*

1 Introduction

"I often cried in public after my husband left me for someone else. Once I saw his car parked in a place we used to visit regularly and burst into tears. Sometimes I'd think of the horrors of being dumped while out shopping and just cry. I reached a point where I didn't care if people stared - sometimes they did and sometimes they didn't. So what?" — HattieB90[1]

> *"Thank you for the days*
> *Those endless days, those sacred days you gave me*
> *I'm thinking of the days*
> *I won't forget a single day, believe me."* — Ray Davies (The Kinks)

Almost everyone desires to meet their "One": a soulmate with whom they will spend the rest of their lives. But around half of new UK cohabiting partnerships turn out to be less than ideal and end in separation. What leads to this change of heart? Why do many people form relationships only to separate (shortly) after? Is it that partners fall out of love or, facing economic adversity together, choose to separate into singleness? Or is it instead that one partner meets someone else (betrayal via on-the-job search), or perhaps there is learning, that after living together the couple realizes

*This chapter was co-authored with Professor Melvyn Coles. We are grateful to Jan Eeckhout, Marco Francesconi and seminar participants at the University of Essex, the Dale T. Mortensen Conference and the Essex/RHUL/Bristol Junior SaM Workshop 2022, Bristol SaM Workshop for their invaluable comments. This paper draws on data from Understanding Society, distributed by the UK Data Service. Understanding Society is an initiative funded by the Economic and Social Research Council and various Government Departments. All errors remain the responsibility of the authors.

[1] Anonymous post, Guardian online, 29/7/21. https://www.theguardian.com/commentisfree/2021/jul/28/have-you-cried-with-despair-in-public-there-is-nothing-braver-or-better

it is a poor match and separate (cohabitation is an experience good)? Because match outflows are large - one third of new partnerships fail within 7 years - the aim of this paper is to identify match turnover in the UK. We find high turnover rates of new partners are mainly due to betrayal where, like the labour market, relatively few quit a match to become unmatched but many quit for a preferred partner. Despite the decline in formal marriage rates, however, we also estimate strong partner commitment effects. For example the average separation rate of all partnerships is only 2.1% p.a., just one seventh of match turnover rates in the UK labour market,[2] and suggests average match durations of around 50 years; i.e. 'til death do they part.

Following Becker (1973) a large literature focuses on the matching margin, on who marries whom (see recent surveys Chade et al., 2017; Chiappori and Salanié, 2023). A small but important search literature more recently considers equilibrium matching with endogenous separations, where the match distribution depends on both partnership outflows and inflows; most recently see Goussé et al. (2017); Holzner and Schulz (2019); Ciscato (2019); Shephard (2019) and the literature survey below. These papers, however, rule out on-the-job search, and so separation instead implies both (ex-)partners become single. But Devereux and Turner (2016) finds that on-the-job search is necessary to explain the rapid repartnering rates (of at least one ex-partner) following separation, while McKinnish (2007) establishes that U.S. divorce rates are directly affected by the gender composition of a partner's workplace. Given the importance of on-the-job search in explaining match turnover in the labour literature (e.g. Mortensen, 2003), it is perhaps surprising this re-sorting mechanism has been little explored in the partnership literature.

Partner contacts with others, possibly at work, maybe at the gym, even at the school gate, are inevitable. Different to the labour literature, we consider two-sided on-the-job search, where either partner might meet someone else. Analogous to the Prisoner's Dilemma, trust issues arise where each partner would prefer the other to credibly promise to "forsake all others". But without recourse to handcuffs, commit-

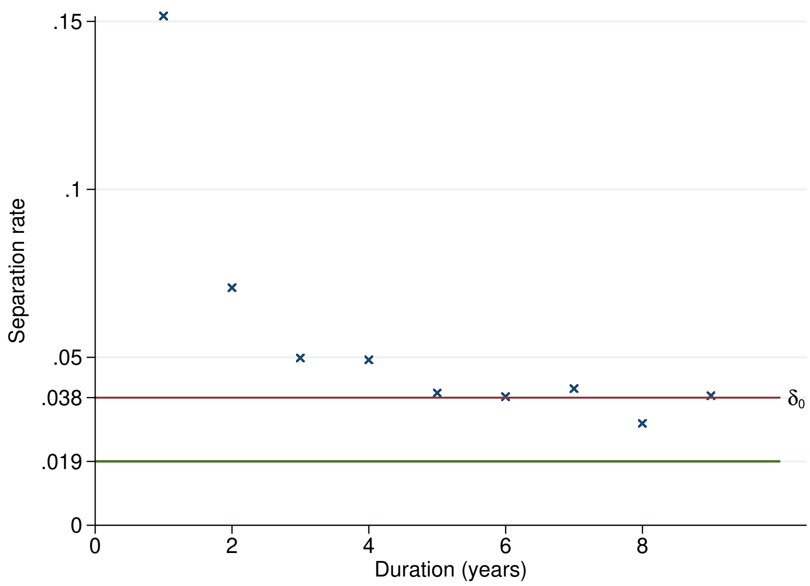
[2]The average separation rate in the UK labour market is 16% (Table 2 Lehmann and Wadsworth, 2000)

ment is imperfect. Here we extend the on-the-job search approach of Cornelius (2003); Burdett et al. (2004); Masters (2008); Gautier et al. (2009) to allow for endogenous commitment, not necessarily by marriage. Our central insight is that when outside contacts are inevitable, cohabiting with a lovely partner (married or not, with or without children) is not only an enjoyable activity, a romantic evening in together is also an excellent monitoring device: it is common knowledge that neither partner is currently out meeting others. Using an equilibrium search framework with sorting by match quality, we show the very well-matched have a negative joint return to on-the-job search, for each is unlikely to find a much preferred partner, while the cost of being dumped is large. Reflecting the very low average UK separation rate (2.1% p.a.), the estimated model finds partner exclusivity effects are large: by becoming more exclusive, the best-matched (endogenously) reduce their outside contact rates by an (average) 77%. High turnover rates across new partners instead reflect a relationship ladder structure where, analogous to job ladder turnover in the labour literature (e.g. Burdett and Mortensen, 1998), new partners in relatively poor matches remain open to meeting others.

The importance of our equilibrium partnership approach goes beyond explaining the observed turnover patterns. Following the UK Divorce Reform Act (1969), easing divorce law finds that UK divorce rates almost doubled between 1970 and 2008 (ONS, 2019), and that cohabitation without marriage became commonplace (Svarer, 2004; Berrington and Stone, 2015). Yet we show commitment effects, even across the unmarried, remain strong. This has significant implications for marriage and fertility choice. Greater exclusivity (via jointly reduced outside contact rates) not only makes a match longer lived it also increases the return to match specific investments. For example partners might be much more likely to start a family if the match is (likely) forever rather than just for a few years; see for example Quercioli (2005) in a job-ladder context. Indeed choosing to be more exclusive is itself a match-specific investment. In the Conclusion we discuss the likely linkages between finding the ”One”, selection into marriage and/or starting a family, noting that the exclusivity margin is further re-enforced should new parents be too exhausted to meet others. Most notably, in our view, exclusivity is self-enforced, which explains why cohabitation thrives as an institution even as formal marriage rates decline.

Taking a closer look at the empirical turnover patterns in the UK demonstrates why including on-the-job search is vital. Figure 1 describes the UK partnership separation hazard function for 1991-2008.[3]

Figure 1: UK Separation Hazard Function



Notes: The separation hazard function is estimated using the main survey waves of the British Household Panel Survey (BHPS), 1991-2008 (see Section 5.1 for further details). The top horizontal line describes GJR's estimated $\delta_0 = 0.038$, the lower line is the lower bound of the separation rate across all UK partnerships (equal to 0.019).

The separation hazard function describes the average UK dissolution rate of cohabiting partnerships by the duration of their match (heterosexual partners, married or cohabiting, under 50 years old). An important empirical difficulty is explaining why this hazard function is so steep: it is initially very high, 15.2% of new partnerships fail in the first year, but noting the average dissolution rate across all partners under 50 is just 2.6% pa, dissolution rates at long durations are very low.[4] The standard equilibrium partnership approach with endogenous dissolutions assumes exogenous match

[3]The data is taken from the British Household Panel Survey (BHPS), a UK panel data set which is fully described in Section 5.1.

[4]See also Bruze et al. (2015) for similar evidence for Denmark.

quality shocks and no on-the-job search. For example the Goussé et al. (2017) approach (GJR from now on) is particularly important because, with unobserved match quality (i.e. partner love draws), it shows how data on the separation margin can identify sorting across types at the matching margin. Using this same data, GJR estimate match shocks occur at an annual rate $\delta_0 = 3.8\%$ which provides their upper bound on partnership separation rates. Furthermore the estimated love draw distribution has large variance because $2.6/3.8 = 68\%$ of these match shocks must cause break-up. Conditional on a match shock, their estimates find even the best matches (with positive assortative matching) still have a divorce probability of around 50%. This then implies a lower bound on partnership separation rates equal to 1.9%. With no on-the-job search, Figure 1 reveals the implied **range** of partner separation rates $s \in [0.019, 0.038]$ is much too narrow for this data set.

According to this data, the average UK hazard rate to first cohabitation (prime age singles, 24-28 years old) is $h^p = 0.15$. On-the-job search then implies the marginal partnership (one which yields no surplus) breaks up at rate $2h^p = 0.30$ because each subsequently leaves at rate h^p and there are two in a partnership. This larger upper bound generates a range of partner separation rates $s \in [0, 0.3]$ which successfully brackets the separation hazard function (see Figure 1). Importantly, however, we find that explaining the steepness of the hazard function additionally requires strong exclusivity effects, that well matched partners commit to much reduced outside contact rates.

Agent heterogeneity is clearly important for explaining individual separations, for "each family is unhappy in its own way"[5]; see for example Weiss and Willis (1997). The major advantage of our approach, however, is that we provide a clear picture of aggregate partner turnover. In particular our estimates find the hazard rate to meet and move in with a long-term committed partner is just 10% p.a.; i.e. it takes an average 10 years to find the "One". This slow process reflects both the low entry rate into first cohabitation ($h^p = 0.15$) and its high failure rate, where only two-thirds of new

[5]Taken from the opening sentence of Leo Tolstoy's great novel 'Anna Karenina' (1878): "All happy families are alike; each unhappy family is unhappy in its own way".

partnerships survive 7 years. This view of matching is also consistent with Table 1 which describes separation rates instead by age. Specifically the estimated separation rates of the over-30s (who have likely found their "One") are appreciably lower than those of the under-30s (who likely have not). The estimated model finds that around two-thirds of new cohabiting partners have found their "One" but, consistent with the fact that around one-half of all new matches ultimately fail, commitment is imperfect. Our estimates find an average 8 year itch where, even for committed partnerships, a possibly tempting outside contact occurs, on average, around once every 8 years. Heartbreak, where one previously committed partner instead moves out for someone else, is not rare.

Table 1: Partner Separation Rates by Age of the Female Partner

Age of the female partner	20-25	26-30	31-40	41-50	50+
Average separation rate (pa)	15.1%	6.5%	3.1%	1.7%	0.4%

Notes: This table presents estimates of the average partner separation rate (δ_0) for subsamples of couples grouped by the age of the female partner using main survey wave data from the BHPS, 1991 - 2008. Survey weights are used.

We also find evidence of a quantitatively important learning process; cohabitation is an experience good. Specifically a good fit of the separation hazard function requires a "premature break-up" process where an estimated 8% of new cohabiting partners find they are in a "bad match" which has an expected duration of just 5 months; i.e. it does not take long to discover you're in a bad match. We use probit regressions to identify those partner characteristics which best predict premature break-up, measured as those new partnerships which break-up within the first year. Consistent with the idea that premature break-up is due to learning (for who would volunteer to enter a bad match?), we find that the education and racial characteristics of partners do not predict premature break-up. Age, however, does predict premature break-up and in a very specific way: premature break-up is more likely when the female partner is below 25 years old and that her male partner is some years older (the estimated maximal break-up rate occurs with him around 4 years older). Reflecting that starting a family is a valuable match specific investment (and so a signifier of an expected long-lived match), we also find premature break-up is much less likely if partners have started a family.

Following the literature survey, Section 3 considers an equilibrium matching free-for-all where a marital oath to ”forsake all others” is assumed not credible and on-the-job search presumes partners cannot avoid meeting others. It formally establishes the existence and uniqueness of a matching equilibrium with a relationship ladder property. Section 4 extends this framework by defining and fully describing a cohabitation equilibrium with endogenous partner search intensity. Section 5 uses simulated method of moments to estimate model parameters and so identifies the turnover statistics described above. Section 6 uses a probit analysis to identify partnership characteristics that are most associated with premature break-up. The conclusion discusses possible directions for future research, including endogenous fertility outcomes and marriage.

2 A Brief Literature Survey

There are three principal types of sorting. Following Becker (1973) the larger literature considers **vertical sorting** where match payoffs increase with partner attributes, such as ability or pizzazz. Whether there is positive or negative assortative matching then depends critically on the structure of match payoffs and the transferability of utility (e.g. Burdett and Coles, 1997; Shimer and Smith, 2000; Smith, 2006; Eeckhout and Kircher, 2010 and see Chade et al., 2017 for a recent survey). Important contributions with on-the-job search include Postel-Vinay and Robin (2002); Cahuc et al. (2006); Menzio and Shi (2011); Lise and Robin (2017); Hagedorn et al. (2017) among many others.

Homophily instead supposes partners prefer to match with similar types, say someone with the same political beliefs or level of education. Teulings and Gautier (2004) consider equilibrium sorting where payoffs decrease with the distance between types, while a larger empirical literature describes sorting across types consistent with homophilic tastes (Marimon and Zilibotti, 1999; Gautier et al., 2010b; Gautier and Teulings, 2015).[6]

Here instead we consider sorting by **idiosyncratic match (love) draws**. Burdett and Wright (1998) were the first to consider equilibrium matching with search frictions and random love draws. Important hybrid models with vertical sorting and iid match draws but with no on-the-job search include Fernandez et al. (2005); Seitz (2009); Gemici and Laufer (2011); Bruze et al. (2015), GJR. Recent extensions also introduce partner level shocks, say one partner might become unemployed (Holzner and Schulz, 2019), or experience a wage shock (Ciscato, 2019), or consider life cycle effects with aging (Shephard, 2019). An interesting parallel is that on-the-job search also describes a partner-level shock process, but rather than one partner becoming unemployed, here instead one partner meets someone else. Our model might be interpreted as a hybrid of homophily with iid match draws, where all seek a greater match draw within own-type matches.

Our approach is closely related to the equilibrium job ladder literature where workers quit for higher waged employment (e.g. Burdett and Mortensen, 1998; Moscarini and Postel-Vinay, 2008, 2012, 2013, 2016). An important difference is that this literature assumes only the worker can quit, while here there are two-sided separations where either partner might leave. Cornelius (2003); Burdett et al. (2004) show in a non-transferable utility context that two-sided on-the-job search instead generates additional trust issues (also see Gautier et al., 2010a; Masters, 2008). For example a betrayal equilibrium may arise where, say, blue and green partners separate for their own colour simply because each fears the other will similarly betray them. With im-

[6]It is interesting that the sufficient conditions for positive assortative matching in Shimer and Smith (2000) requires some homophily: that the lowest productivity types find match payoff decreases with more productive partners.

perfectly transferable utility and bargaining, we formally establish a matching equilibrium exists which satisfies a particular identifying restriction (one which ensures there is a single (trust) outcome between partners), we find the resulting equilibrium is unique and implies a relationship ladder whereby a partner will only betray/quit for a better match draw. We also find that endogenous search intensity (commitment) is important. In the labour framework Lentz (2010) argues that higher skilled workers search with greater intensity for jobs, while Christensen et al. (2005) find that better paid workers choose lower on-the-job search effort. Bagger and Lentz (2019) consider a similar issue with mismatch between heterogeneous firms and workers. Postel-Vinay and Robin (2004) instead consider how the firm's response to an employee's outside offer (to match or not to match) might give employees a greater incentive to seek outside offers.

The equilibrium bargaining problem with two-sided on-the-job search and imperfectly transferable utility is complex. Using a one-period bargaining framework, the highly influential Browning and Chiappori (1998) approach establishes that the negotiated household terms of trade are consistent with efficient bargaining. But in a dynamic framework offering your partner an extra cookie changes the probability the partner leaves for someone else. Indeed perhaps the key insight of Burdett and Mortensen (1998) is that even when firms have all the bargaining power, firms offer wages above the worker reservation wage so as to reduce employee quit rates. In this extended approach however, the bargaining set of match values is not typically convex and the standard Nash bargaining approach is deeply problematic; see for example (e.g. Shimer, 2006; Bonilla and Burdett, 2010; Fujita and Ramey, 2012; Elsby and Gottfries, 2022; Gottfries, 2018).[7] With also two-sided on-the-job search, the additional trust issues imply the bargaining set of match values has pathological properties. Here instead we adopt the Kalai and Smorodinsky (1975) bargaining protocol which not only ensures household agreements are always jointly efficient, it does not require the set of partner match values to be convex.

[7] An alternative approach uses sequential auctions (Postel-Vinay and Robin, 2002; Dey and Flinn, 2005; Cahuc et al., 2006).

There is a small literature that considers the rise of cohabitation and decline in marriage rates. Lundberg et al. (2016) documents the uneven retreat from marriage and the increase in cohabitation across education groups in the U.S. It suggests that, for low-educated couples, cohabitation has become an alternative arrangement under which to have children. Berrington and Stone (2015) document similar patterns in the UK. Other studies address specific factors explaining the choice of cohabitation and the different household choices that result, such as learning about the quality of new partners (Brien et al., 2006), differences in commitment which lead to different degrees of specialization (Gemici and Laufer, 2011), how institutional differences between marriage and cohabitation affected marital choices (Calvo, 2022). In the conclusion we consider the likely linkages between finding the "One", cohabitation/marriage and starting a family.

3 The Basic Model: Equilibrium matching without an exclusivity margin

Time is continuous, has an infinite horizon and throughout we only consider steady state. To be consistent with the larger part of the matching literature, we assume utility between partners is imperfectly transferable and the terms of trade are determined by bargaining. Although the following allows for the general case, that bargaining outcomes may be asymmetric, for the empirical application we focus on a gender-free partnership framework.

There is a unit measure of ex-ante identical agents who meet, say, on the same "island". There are search frictions on the island and contacts are random. For clarity of exposition, we later consider endogenous exclusivity but here begin with the simplest case, that all have the same contact rate $\overline{\lambda}$, both those single and those in partnerships. There is ex-post match heterogeneity: conditional on a contact there is an idiosyncratic match (love) draw θ which is considered an iid draw from cdf $F(.)$, whose density F' exists and has finite support $[\underline{\theta},\overline{\theta}]$. Both potential partners observe θ but outside parties do not. Marital contracts are incomplete: private actions within the household are not verifiable by courts and so it is not possible to write an enforceable contract on the household terms of trade. Matched partners do not respond to partner outside offers because the quality of the outside contact θ' is unobserved by the potentially deserted partner and partners cannot anyway contract to enforce different marital terms of trade. Because here we also assume no children and no accumulation of assets, there is nothing to divide on separation. And so if a partner prefers an outside offer, the partner simply leaves (with no recall). The terms of trade are determined by bargaining over match values where, for reasons made clear below, we adopt the Kalai and Smorodinsky (1975) bargaining protocol.

There is agent turnover where all die at the same rate $\delta > 0$, while δ also describes the inflow of new agents who are initially unmatched. For simplicity we assume partner death is perfectly correlated; i.e. if a death occurs in the partnership, then both die.[8] All have the same subjective time rate of preference $\rho \geq 0$ and so $r = \rho + \delta > 0$ will describe each agent's total discount rate.

In any θ -partnership, flow payoffs potentially depend on a vector of partner contributions $\underline{z}_i$, $i = 1, 2$ to the household good. The following straightforwardly generalises to payoff functions $U_i(\underline{z}_i, \underline{z}_{-i}; \theta)$ which are strictly increasing in partner contributions $\underline{z}_{-i}$, strictly increasing in the match draw θ, differentiable and quasiconcave in $(\underline{z}_i, \underline{z}_{-i})$. This matching framework can thus allow a large variety of household structures. To focus on the matching process, however, we adopt a simple illustrative example where z_i, z_{-i} are scalars and agent $i's$ flow match payoff is

$$U_i = u(y - z_i) + \theta[z_i + z_{-i}],$$

where $u(.)$ is increasing, twice differentiable and strictly concave with $u'(0) = \infty$ and $y > 0$ describes individual flow earnings. The first term describes flow utility from private consumption while matched, the second describes the flow value derived from

[8]Otherwise partner death is an additional source of becoming single which is a trivial, though slightly more cumbersome, variation of the model.

individual contributions (z_i, z_{-i}) to the household public good. Assuming $u(y)$ also describes flow utility while single, note that if neither contributes to the household public good, then $z_i = 0$ implies each has flow payoff $U_i = u(y)$ and there is no match surplus. For ease of exposition, we adopt the tie-break rule that agents reject the match whenever there is no strict gain to trade and that a partner leaves when indifferent. We will however relax this assumption in the empirical application.

3.1 Efficient Agreement on Home Production

It is important to consider asymmetric agreements for it may be the case that partner 1 is better off increasing z_1 to make partner 2 better off and so reduce 2's separation rate; e.g. Burdett and Mortensen (1998). Following Browning and Chiappori (1998), we consider the set of efficient household agreements; i.e. partners choose z_1, z_2 to solve:

$$\max_{z_1,z_2} U_1 = u(y - z_1) + \theta[z_1 + z_2] \text{ s.t. } \begin{array}{l} \text{(i) } u(y - z_2) + \theta[z_1 + z_2] \geq U_2 \\ \text{(ii) } z_1 + z_2 \geq 0 \\ \text{(iii) } z_1, z_2 \leq y, \end{array}$$

where, at this stage, partner 2's payoff U_2 is an exogenous parameter. Contributions z_i are not constrained to be positive which allows agents to make side payments for partners to "spend on themselves". Restriction (ii), however, requires that total investment in the household public good must be non-negative while (iii) requires private consumption $y - z_i$ cannot be negative. The assumption $u'(0) = \infty$ ensures constraint (iii) is never binding in any equilibrium agreement. Given an arbitrary value for U_2 the second constraint may bind. The matching equilibrium, however, yields an important simplification, that constraint (ii) cannot be binding in any equilibrium partnership. The simple logic is if the constraint binds, the partners make no investment in the household public good and because match surplus is then zero the tie-break assumption implies no such matches form in equilibrium. Lemma 2 below thus describes the set of efficient household agreements for all equilibrium $\theta-$partnerships, those with strictly positive match surplus. Lemma 1 first identifies θ where a strict gain to trade exists.

Lemma 1. *A strict gain to trade exists if and only if $\theta > \theta^R = 0.5u'(y)$.*

Proof. Proof of Lemma 1 is in Appendix A.

Lemma 1 reflects the public good aspect of the household good, where both enjoy a strictly positive return to a household investment, and joint investment is optimal if θ is large enough. Consider then an equilibrium $\theta-$partnership and so $z_1 + z_2 \geq 0$ is not a binding constraint. For all such partnerships, the efficient household agreement reduces to solving

$$\max_{z_1,z_2} U_1 = u(y - z_1) + \theta[z_1 + z_2] \text{ s.t. } u(y - z_2) + \theta[z_1 + z_2] \geq U_2.$$

Because the partner utility constraint is necessarily binding (otherwise increase z_2), the Kuhn-Tucker necessary conditions for optimality yield Lemma 2, where concavity of $u(.)$ ensures this describes the maximum.

Lemma 2. *Efficient household trade in any equilibrium match θ implies optimal z_1^*, z_2^* given by*

$$\begin{aligned} u'(y - z_1^*) &= \theta(1 + \mu) && (1) \\ u'(y - z_2^*) &= \theta(1 + \frac{1}{\mu}) && (2) \\ U_2 &= u(y - z_2^*) + \theta[z_1^* + z_2^*], \end{aligned}$$

where $\mu \geq 0$ is the associated Lagrange multiplier.

Proof. Follows from standard arguments.

The standard household approach is to describe the Pareto frontier of household payoffs $U_1 = \Psi^H(U_2; \theta)$ and consider how partners bargain over that frontier. However with on-the-job search, agents instead bargain over match values where they take into account that extracting too many match rents increases their partner's likelihood of leaving. Given the z_i^* defined by (1), (2), let $\{z_1^*(\mu;\theta), z_2^*(\mu;\theta)\}_{\mu\geq 0}$ denote the set of efficient household agreements, now indexed by $\mu \geq 0$, with corresponding partner payoffs

$$\begin{aligned} U_1(\mu;\theta) &= u(y - z_1^*) + \theta[z_1^* + z_2^*]; \\ U_2(\mu;\theta) &= u(y - z_2^*) + \theta[z_1^* + z_2^*]. \end{aligned}$$

Standard algebra establishes $U_1(.)$ is continuously differentiable in μ and strictly decreasing (partner 1 contributes more as μ increases, partner 2 contributes less). U_2 is instead continuously differentiable and increasing in μ, where the Envelope Theorem implies the utility trade off along the household frontier is

$$\frac{dU_1}{dU_2} = -\mu. \tag{3}$$

It is immediate that the Pareto frontier $\Psi^H(.)$ is concave and symmetric around the 45^o line.[9] We refer to μ as the household terms of trade where $\mu = 1$ corresponds to the symmetric agreement.

3.2 On-the-Job Search and Lifetime Values

Although each agent contacts a potential partner at rate $\overline{\lambda} > 0$, the contacted person's willingness to match depends on their current status: single or partnered. Let λ denote the (reduced form) rate at which each agent contacts a potential partner *who is willing to match* and let $H(V)$, with finite support $[\underline{V}, \overline{V}]$, describe the probability that matching with this willing partner yields an expected lifetime value no greater than V. Of course, λ, $H(.)$ are endogenously determined in the matching equilibrium. Anticipating the properties of a matching equilibrium we focus on the case that λ , $H(.)$

[9] As μ increases, U_1 falls continuously, U_2 increases continuously with a slope equal to -μ which becomes progressively more negative.

satisfy the following regularity condition:

Regularity Condition (RC): λ, $H(.)$ satisfy:

$$\frac{r}{\lambda} + 2[1 - H(V)] - H'(V)[V - V^S] > 0 \,\forall\, V \geq V^S,$$

where V^S, the value of being single, is defined by

$$rV^S = u(y) + \lambda \int_{V^S}^{\overline{V}} [V - V^S] H'(V) dV. \tag{4}$$

Note that RC requires that $H(.)$ is differentiable and so rules out mass points in $H(.)$ (except perhaps at V^S). A standard contradiction argument in the equilibrium price dispersion literature would argue that if there were a mass point at $V > V^S$, then partner 1 would prefer to offer an extra cookie because the resulting discrete fall in partner 2's separation rate makes partner 1 strictly better off (and so contradicts a mass point of partnerships at V). However such arguments cannot be applied here. For example, the extra cookie argument fails should partner 2 believe partner 1 is instead worse off (by giving up the marginal cookie) and their separation rates then increase. Indeed rather than bribe partner 2 with an extra cookie (a price deviation argument), why don't these partners simply agree that neither will quit for an outside offer with the same value V? With a mass point in $H(.)$ at V, the resulting (discrete) fall in both separation rates implies both are then strictly better off with a match value strictly greater than V (and there can be no mass point at V). With two sided separations, equilibrium arguments must not only consider deviating μ but also consider how partners co-ordinate separation choices.

The important role played by RC is it guarantees unique equilibrium separation choices in all possible partnerships (see Lemma 3 below), where Figures 4 and 5 demonstrate the difficulties which arise if RC is not satisfied. To simplify the exposition we always assume the density H' exists. Claim 1 now describes additional information due to RC.

Claim 1. *RC guarantees*

$$\Delta = \left[\frac{r}{\lambda} + [1 - H(V_1)] + [1 - H(V_2)]\right]^2 - H'(V_1)H'(V_2)[V_1 - V^S][V_2 - V^S] > 0$$

for all $V_1, V_2 \geq V^S$.

Proof.

$$\begin{aligned}
\left[\frac{r}{\lambda} + [1 - H(V_1)] + [1 - H(V_2)]\right]^2 &= \left[\frac{r}{\lambda} + 2[1 - H(V_1)]\right]\left[\frac{r}{\lambda} + 2[1 - H(V_2]\right] + [H(V_1) - H(V_2)]^2 \\
&\geq \left[\frac{r}{\lambda} + 2[1 - H(V_1)]\right]\left[\frac{r}{\lambda} + 2[1 - H(V_2]\right] \\
&> H'(V_1)[V_1 - V^S]H'(V_2)[V_2 - V^S]
\end{aligned}$$

by RC. □

Now consider any partnership with a strict gain to trade, $\theta > \theta^R$, and define the set of *incentive compatible household trades*

$$M(\theta) = \{\mu : U_1(\mu;\theta) \geq u(y), U_2(\mu;\theta) \geq u(y)\},$$

where $\mu \in M(\theta)$ implies both agents prefer (at least temporarily) a match (μ, θ) with $U_i(\mu;\theta) \geq u(y)$ than being single with $u(y)$. $M(.)$ is non-empty for all $\theta > \theta^R$ (because $\mu = 1 \in M(\theta)$) and is connected (because $U_1(.), U_2(.)$ are monotone). For any $\theta > \theta^R$ and $\mu \in M(\theta)$, standard recursive arguments imply partner values (V_1, V_2) are the solution to the pair of implicit functions

$$\begin{aligned}
rV_1 &= U_1 + \lambda \int_{V_1}^{\overline{V}} [V' - V_1] dH(V') + \lambda[1 - H(V_2)][V^S - V_1]; \\
rV_2 &= U_2 + \lambda \int_{V_2}^{\overline{V}} [V' - V_2] dH(V') + \lambda[1 - H(V_1)][V^S - V_2].
\end{aligned}$$

In words, flow match value equals own household payoff U_i plus the expected gain by receiving an outside offer from a willing partner, while the last term describes the expected loss should one's partner leave for a preferred match. Integration by parts yields

$$
\begin{aligned}
rV_1 &= U_1 + \lambda \int_{V_1}^{\overline{V}} [1 - H(V')] dV' + \lambda [1 - H(V_2)][V^S - V_1]; &(5)\\
rV_2 &= U_2 + \lambda \int_{V_2}^{\overline{V}} [1 - H(V')] dV' + \lambda [1 - H(V_1)][V^S - V_2], &(6)
\end{aligned}
$$

where the discounted value of being single is

$$
rV^S = u(y) + \lambda \int_{V^S}^{\overline{V}} [1 - H(V')] dV'. \tag{7}
$$

Standard arguments imply V^S exists, is unique and $V^S \geq u(y)/r$.

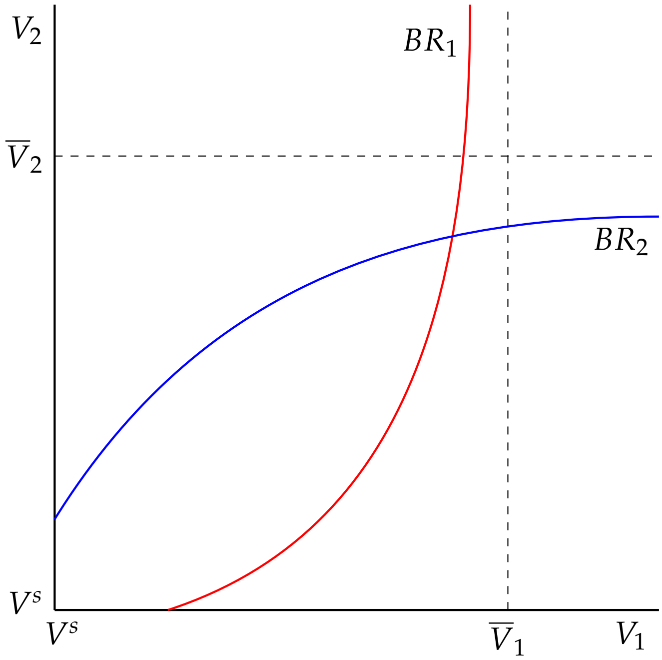
Figure 2 graphs the two implicit functions (5), (6), labelled BR_1, BR_2 respectively. Consider first (5) which determines V_1 for any $V_2 \geq V^S$. Standard arguments[10] establish a solution for V_1 exists, is unique and $U_1 \geq u(y)$ implies $V_1 \geq V^S$. Furthermore, no mass points in $H(.)$ implies V_1 increases continuously with V_2 as drawn in Figure 2 and partner 1's maximal value $\overline{V}_1$ is given by

$$
r\overline{V}_1 = U_1 + \lambda \int_{\overline{V}_1}^{\overline{V}} [1 - H(V')] dV',
$$

where $U_1 \geq u(y)$ implies $\overline{V}_1 \geq V^S$. Similarly (6) determines V_2 for any $V_1 \geq V^S$ and the same arguments establish a solution for V_2 exists, that $V_2 \geq V^S$ increases continuously with V_1 and is bounded above by an equivalent $\overline{V}_2 \geq V^S$. Continuity of BR_1, BR_2 then implies a solution to (5), (6) exists and that $(V_1, V_2) \in [V^S, \overline{V}_1] \times [V^S, \overline{V}_2]$.

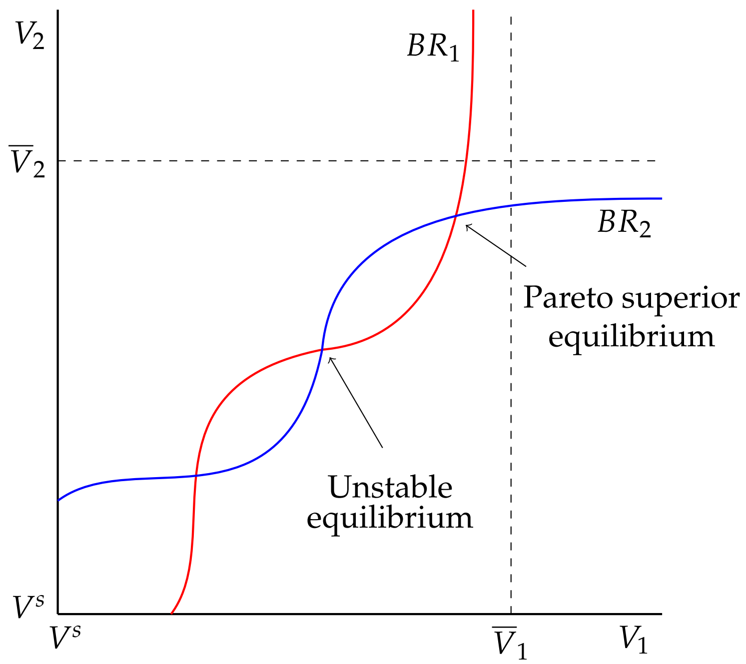
[10]The LHS of (5) is continuous and strictly increasing with V_1, the RHS is continuous and decreasing with V_1 and bounded below by $U_1 \geq u(y)$. Hence a single intersection exists (which determines V_1) and standard arguments establish the claims made.

Figure 2: Existence of (V_1, V_2)



Because both BR_1, BR_2 are increasing functions, there may be multiple solutions as illustrated in Figure 3. We will refer to the middle solution in Figure 3 as the *unstable equilibrium*. Multiplicity arises because partner separation rates are strategic complements: a higher partner separation rate reduces own value of the match which then increases own separation rate (and vice versa). The Pareto superior outcome corresponds to jointly lowest separation rates. Cornelius (2003) and Burdett et al. (2004) interpret this outcome as trust: if both partners trust the other to stay, then the match has greater value and both are then more likely to stay.

Figure 3: Multiple Solutions for (V_1, V_2)



It is easy to show that $\Delta > 0$, as defined in Claim 1, describes the single crossing property: that BR_1 is strictly steeper than BR_2 at any intersection (V_1, V_2) and so, as drawn in Figure 2, any solution for (V_1, V_2) is unique. Hence by Claim 1, RC guarantees unique equilibrium partner separation rates for any terms of trade $\mu \in M(\theta)$ and we have established Lemma 3.

Lemma 3. *If there are no mass points in $H(.)$, then lifetime match values $(V_1(\mu;\theta), V_2(\mu;\theta))$ defined by (5), (6) exist and satisfy $V_1, V_2 \in [V^S, \overline{V}_1] \times [V^S, \overline{V}_2]$ for all $\theta > \theta^R$ and $\mu \in M(\theta)$. If RC also holds then $(V_1(\mu;\theta), V_2(\mu;\theta))$ are unique.*

3.3 Value Frontiers and Bargaining

Given any $\theta > \theta^R$, we can identify the Pareto frontier of match values (V_1, V_2) by tracing out

$\{V_1(\mu;\theta), V_2(\mu;\theta)\}_{\mu \in M(\theta)}$. Consider then a marginal increase in $\mu \in int\ M(\theta)$. Assuming the density function $H'(.)$ exists, total differentiation on (5), (6) implies variations (dV_1, dV_2) satisfy

$$rdV_1 = dU_1 - \lambda[1-H(V_1)]dV_1 - \lambda[1-H(V_2)]dV_1 + \lambda H'(V_2)[V_1 - V^S]dV_2; \quad (8)$$

$$rdV_2 = dU_2 - \lambda[1-H(V_2)]dV_2 - \lambda[1-H(V_1)]dV_2 + \lambda H'(V_1)[V_2 - V^S]dV_1, \quad (9)$$

with (dU_1, dU_2) describing the underlying variation in household payoffs. Because $dU_1 = -\mu dU_2$ at any efficient agreement μ, solving yields the pair of differential equations

$$\begin{bmatrix} dV_1/d\mu \\ dV_2/d\mu \end{bmatrix} = \frac{\partial U_2(\mu,\theta)/\partial\mu}{\Delta} \begin{bmatrix} \text{-}\mu[r+s] + \lambda H'(V_2)[V_1 - V^S] \\ [r+s] - \mu\lambda H'(V_1)[V_2 - V^S] \end{bmatrix}, \quad (10)$$

with Δ as defined in Claim 1, $\partial U_2(\mu,\theta)\partial\mu > 0$ and $s = \lambda[1-H(V_1)] + \lambda[1-H(V_2)]$ which is the partner separation rate. The slope of the value frontier is thus

$$\frac{dV_2}{dV_1} = \frac{dV_2/d\mu}{dV_1/d\mu} = -\frac{1}{\mu}\left[\frac{1 - \frac{\lambda}{r+s}\mu H'(V_1)[V_2 - V^S]}{1 - \frac{\lambda}{r+s}\frac{1}{\mu}H'(V_2)[V_1 - V^S]}\right]. \quad (11)$$

(11) reveals how the slope of the value frontier $\{V_1(\mu;\theta), V_2(\mu,\theta)\}_{\mu \in M(\theta)}$ not only depends on the household terms of trade μ, but also on marginal quit incentives, $\lambda H'(.)$, and the cost of being dumped, $[V_i - V^S]$. Figures 4 and 5 demonstrate examples of possible value frontiers.[11] It is straightforward to show at the symmetric agreement $\mu = 1$ and regardless of the sign of Δ, that $dV_2/d\mu = -dV_1/d\mu > 0$; i.e. partner 2 is strictly better off with an increase in μ, partner 1 is strictly worse off and the slope of the value frontier always equals -1.[12] With symmetric partner threat points, it is immediate that $\mu = 1$ solves the necessary conditions for the Nash bargaining solution, though there might also exist asymmetric solutions. Partner symmetry im-

[11] If density $H'(.)$ is not continuous, then the value frontier has kinks (its slope is not continuous).

[12] (11) implies the slope is -1 at $\mu = 1$ because the solutions $V_2 = V_1$ are necessarily symmetric. A simple contradiction argument then suffices: suppose $dV_1/d\mu \geq 0$. Because this implies $dV_2/d\mu > 0$ (partner 2 enjoys improved terms of trade and is less likely to be dumped) the value frontier is then upward sloping which contradicts its slope equals -1. Hence $dV_1/d\mu < 0$ and so $dV_2/d\mu = -dV_1/d\mu > 0$.

plies the frontier is symmetric across the 45^o line and is increasing at the corners for sufficiently large θ.[13]

Figure 4: Unique Value Frontier (with RC)

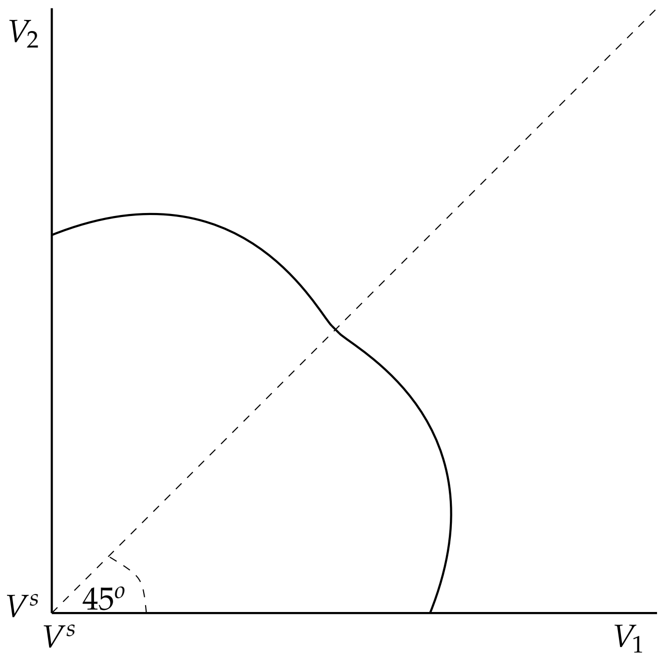
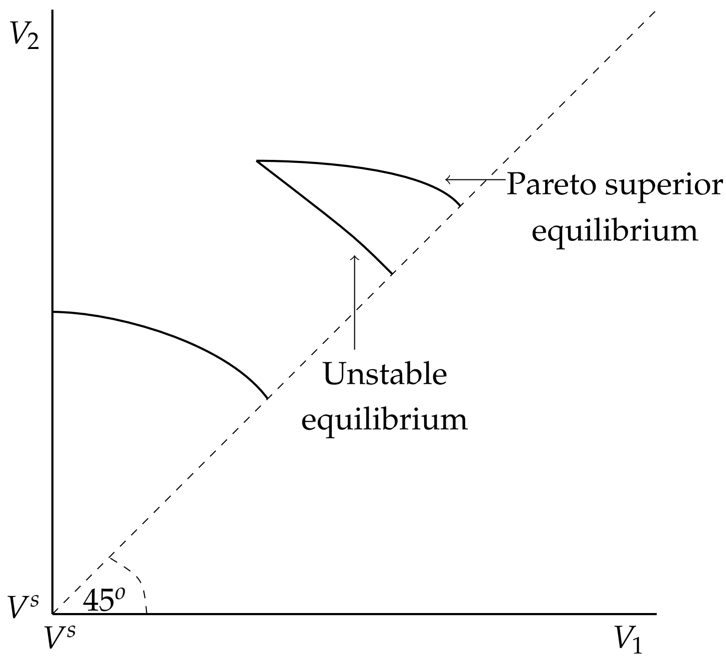


Figure 5 describes the case of multiple equilibria (3) which can arise only when RC does not hold. As demonstrated in Figure 3, these equilibria are Pareto rankable and the unstable (middle) equilibrium has $\Delta < 0$. Figure 5 graphs the 3 value frontiers for $\mu \geq 1$ (noting the case $\mu \leq 1$ is the mirror image). Because an increase in μ implies BR_2 shifts up and BR_1 shifts left, Figure 3 implies a point may be reached where BR_2 forms a tangent with BR_1 (where $\Delta = 0$). At this tangency point, the unstable equilibrium necessarily converges to one of the other equilibria, and both then fail to exist thereafter. Figure 5 depicts the case where the unstable equilibrium converges to the Pareto superior equilibrium. In this case, asymmetric terms of trade allow partners in the unstable equilibrium to coordinate to lower separation rates and so increase match surplus. It might instead be that the unstable equilibrium converges to the Pareto inferior equilibrium in which case the reverse logic applies.

[13]It can be established that when match surplus is large, the frontier is upward sloping at the corners. This occurs for Burdett and Mortensen (1998) reasons: when the surplus is (sufficiently) large, it is payoff increasing to share some of that surplus to reduce your partner's separation rate and so enjoy the (sufficiently) large surplus for even longer.

Figure 5: Multiple Value Frontiers ($\mu \geq 1$)



Although it is tempting to assume partners coordinate separation rates to the Pareto superior equilibrium, this requires that partners are able to coordinate beliefs on each others' behaviours, but beliefs (i.e. trust) are not contractible; e.g Cornelius (2003), Burdett et al. (2004).[14] To proceed from now on we only consider matching equilibria which satisfy RC and so have a unique value frontier as described in Figure 4. The standard matching approach then typically assumes the terms of trade are determined by the Nash Bargaining Solution. This approach is problematic, however, because the bargaining set of match values is not necessarily convex. An alternative approach, popular in a static framework, is to allow agents to negotiate over the convex hull using lotteries. Unfortunately that approach is not dynamically consistent with incomplete contracts. For example suppose the value frontier is not concave and, by symmetry, suppose agents agree on a lottery where with probability $\frac{1}{2}$ the agreement is $\mu_1 < 1$ with partner values (V^1, V^2) where $V^1 \neq V^2$, and with probability $\frac{1}{2}$ the agreement is instead $\mu = \frac{1}{\mu_1}$ with reversed partner values (V^2, V^1). Note first that the lottery does not imply partners mix over household production for mixing on

[14] A contract which specifies full partner compensation on desertion would solve this coordination problem because similar to the sequential auctions approach, separations would then be jointly efficient. Divorce law instead only describes a legal division of assets, child access agreements and maintenance payments. Without full compensation to the deserted partner, separations are often acrimonious.

production is inefficient (because $\Psi^H(.)$ is strictly concave). Rather the lottery requires perfect commitment to permanent household terms of trade $\mu \in \{\mu_1, \frac{1}{\mu_1}\}$ where the lottery determines the winner. Without enforceable contracts on the household terms of trade, commitment to these terms of trade by the loser is already problematic (for μ_1 does not satisfy the Nash Bargaining Solution on the actual frontier). But suppose the lottery outcome was enforceable and that partner 1 wins the lottery and partner 2 obtains loser value $V^2 < V^1$. Unobserved outside offers then imply partner 2 will leave for any outside offer $V > V^2$, while the lottery requires partner 2 only quits to an outside offer $V > \frac{1}{2}[V^1 + V^2] > V^2$. Without complete contracts, the loser quits the partnership too quickly and the lottery approach is both unenforceable and dynamically inefficient.

Given lotteries over the household terms of trade are not empirically relevant, we do not go this route. Instead we adopt the Kalai and Smorodinsky (1975) bargaining protocol, not only because it does not require a convex bargaining set, but because it also satisfies strong symmetry. Specifically the critical simplification of the gender neutral approach with Kalai and Smorodinsky (1975) is that symmetric partners then negotiate the symmetric terms of trade $\mu = 1$ and so enjoy equal household payoffs

$$U_i = U(\theta) \equiv U_i(1; \theta).$$

RC and Lemma 3 then imply unique match values $V_i(\theta) = V(\theta)$ given by the functional equation

$$rV(\theta) = U(\theta) + \lambda \int_{V(\theta)}^{\overline{V}} [1 - H(V')]dV' + \lambda[1 - H(V(\theta))][V^S - V(\theta)]. \qquad (12)$$

The equilibrium analysis is now straightforward, the difficulty being to establish the resulting equilibrium does indeed satisfy RC.

3.4 The Matching Equilibrium with RC

Let $1 - G(\theta')$ denote the steady state measure of agents who are in a partnership with match draw $\theta \geq \theta'$ and so $G(\theta^R)$ is the steady state measure of single agents.

Definition 1. *A* ***Matching Equilibrium*** *is the set* $\{V(.), V^S, G(.), \lambda, H(.)\}$ *where* λ, H *satisfy RC and:*

(i) household match values $V(.)$ *satisfy (12) for all* $\theta \geq \theta^R$ *with* V^S *given by (7);*

(ii) the distribution of matched partners $G(.)$ *is consistent with individually optimal turnover and steady state;*

(iii) offers $\lambda, H(.)$ *are consistent with individually optimal turnover and steady state* $G(.)$*.*

Theorem 1 describes the unique Matching Equilibrium.

Theorem 1. *A Matching Equilibrium satisfying RC exists, is unique and has an analytic solution:*

$$G(\theta) = \frac{1}{\left\{1 + \frac{3\overline{\lambda}}{\delta}[1 - F(\theta)]\right\}^{1/3}}; \tag{13}$$

$$\lambda[1 - H(V(\theta))] = \overline{\lambda} \int_{\theta}^{\overline{\theta}} G(\widetilde{\theta}) F'(\widetilde{\theta}) d\widetilde{\theta}; \tag{14}$$

$$V^S = \frac{u(y)}{r} + \int_{\theta^R}^{\overline{\theta}} \frac{U'(x)}{r} [1 - [1 + \frac{s(x)}{r}]^{-1/2}] dx; \tag{15}$$

$$V(\theta) = V^S + \int_{\theta^R}^{\theta} \frac{U'(x)}{r} \left\{[1 + \frac{s(x)}{r}][1 + \frac{s(\theta)}{r}]\right\}^{-1/2} dx, \tag{16}$$

where $s(\theta) = 2\overline{\lambda} \int_{\theta}^{\overline{\theta}} G(\widetilde{\theta}) F'(\widetilde{\theta}) d\widetilde{\theta}$ *is the equilibrium partner separation rate.*

Proof. Proof of Theorem 1 is in Appendix A.

The proof establishes RC implies a second important simplification, that equilibrium match values $V(\theta)$ are necessarily strictly increasing in $\theta \geq \theta^R$. Any equilibrium satisfying RC thus generates a **relationship ladder** where $\theta-$partners use on-the-job search to find better matches $\widetilde{\theta} \geq \theta$. Equilibrium separations then only occur whenever a partner θ meets an outside contact with match draw $\widetilde{\theta} > \theta$ *and the contact is willing*; i.e. the outside agent is also in a match no better than $\widetilde{\theta}$ which, with random contacts and in steady state, occurs with probability $G(\widetilde{\theta})$. A $\theta-$partnership thus breaks down at rate $s(\theta) = 2\overline{\lambda}\int_{\theta}^{\overline{\theta}} G(\widetilde{\theta})F'(\widetilde{\theta})d\widetilde{\theta}$ depending on which partner is the first to meet a willing contact $\widetilde{\theta} > \theta$. (14) describes the equilibrium arrival rate of outside offers in V-space. Using (14) in (12) yields the equilibrium functional equation for $V(\theta)$:

$$rV(\theta) = U(\theta) + \overline{\lambda}\int_{\theta}^{\overline{\theta}}[V(\widetilde{\theta}) - V(\theta))]G(\widetilde{\theta})F'(\widetilde{\theta})d\widetilde{\theta} + \overline{\lambda}\int_{\theta}^{\overline{\theta}}[V^S - V(\theta)]G(\widetilde{\theta})F'(\widetilde{\theta})d\widetilde{\theta}, \tag{17}$$

where the first integral is the personal benefit of receiving a preferred outside offer $\widetilde{\theta} > \theta$, the second is the loss if instead your partner meets a willing $\widetilde{\theta} > \theta$. The proof establishes (16) is its closed form solution. Note that (15) implies $V(\theta)$ depends on flow payoff $u(y)$ while unmatched and on the transition rates $s(.)$ between partnership states. The proof is completed by showing λ, $H(.)$, defined by (14), do indeed satisfy RC.

The proof establishes there is a unique matching equilibrium satisfying RC and that it generates a relationship ladder. This does not imply there are no other equilibria. For example, it is possible to construct equilibria where blue and green partners do not "trust" each other, even though household payoffs $U_i(\theta, \mu)$ are colour invariant. For example, a green might separate from a blue partner for a possibly worse $\theta-$match but with a green partner who is trusted more. Because the added trust increases match value, it is consistent that blue and green partners tend to separate for their own colour simply because each does not trust the other not to similarly betray them. The equilibrium relationship ladder instead implies partners only separate according to the payoff relevant variable θ.

This relationship ladder is not unrelated to the job ladder literature where workers quit for better paid employment. In that literature, however, positive sorting implies the distribution of wages paid across employed workers, typically denoted $G(w)$, first order stochastically dominates the wage offer distribution $F(w)$. This does not occur here: (13) does not imply first order stochastic dominance - it is not the case that $G < F$ for all $F \in [0,1]$. The essential difference is the break-up process. In the job ladder literature only the worker can choose to leave the match. Instead in a partnership framework either partner may separate and if two already partnered agents form a new partnership, then two ex-partners are dumped into singleness. This separation process not only generates the trust issues described above, it also undermines the positive sorting result. And because partner turnover is not jointly efficient (with incomplete contracts), we now show this has important implications for exclusivity choice.

4 Romantic Evenings and the Desire for Exclusivity

The remainder of the paper (and the empirical application) focuses on the issue of partner exclusivity. In principle each partner would prefer (i) greater search freedom for him/herself, for there is always the possibility of meeting someone even more wonderful, but (ii) less search freedom for their partner because being dumped is costly. Because seeking to climb the relationship ladder may not be jointly efficient, here we suppose partners can increase exclusivity by enjoying romantic evenings together and so jointly reduce outside contact rates. The issue is how does such exclusivity, interpreted as a joint fall in outside contact rates $\lambda \leq \overline{\lambda}$, affect match values? Theorem 2 describes the preference for exclusivity for the general case.

Consider any efficient match (μ, θ) with $\theta > \theta^R$, where RC and Lemma 3 imply unique partner values $V_1(\mu, \theta)$, $V_2(\mu, \theta)$. Consider Φ defined by:

$$\Phi = \left\{ \int_{V_1}^{\overline{V}} [1 - H(V')]dV' + [1 - H(V_2)][V^S - V_1] \right\} + \mu \left\{ \int_{V_2}^{\overline{V}} [1 - H(V')]dV' + [1 - H(V_1)][V^S - V_2] \right\}. \tag{18}$$

Theorem 2. *(Household preference for exclusivity)*

Given RC, then for any efficient match (μ, θ) with $V_1(\mu; \theta)$, $V_2(\mu; \theta)$ defined by Lemma 3:

(i) if $\Phi > 0$, variation $d\lambda > 0$ is joint value increasing (partners desire less exclusivity);

(ii) if $\Phi < 0$ variation $d\lambda < 0$ is joint value increasing (partners desire more exclusivity).

Proof. Proof of Theorem 2 is in Appendix A.

Consider the first large bracketed term in (18). The first term in that bracket describes partner 1's expected increase in value by receiving an outside offer, the second term in that bracket is the expected loss if instead partner 2 receives the next outside offer and so dumps partner 1. If the total effect is negative, partner 1 would prefer a more exclusive relationship with fewer outside offers for both. The second bracketed term is analogous to the first but for partner 2. Thus, if the second bracketed term is negative, partner 2 agrees and both prefer greater exclusivity $d\lambda < 0$. However suppose $\Phi < 0$ so greater exclusivity is jointly efficient but, say, the second bracketed term is positive. In that scenario partner 2 does not, ceteris paribus, wish for greater exclusivity. But because μ also describes the shadow price of partner 2 utility in the household match, then $\Phi < 0$ implies it is joint value increasing to agree greater exclusivity.

The following focuses on the symmetric bargaining solution with $\mu = 1$ and so symmetric partner values $V_1 = V_2 = V$. In that case, Φ simplifies to

$$\Phi(V) = 2\left\{ \int_V^{\overline{V}} [V' + V^S - 2V] H'(V') dV' \right\}.$$

The preference for exclusivity reduces to asking whether the expected joint continuation value $[V' + V^S]$ through on-the-job search and separation exceeds $2V$, the joint value of continuing their match. If $\Phi < 0$ then both prefer neither to receive outside offers; i.e. they prefer full exclusivity. Importantly for what follows there is equilibrium sorting on exclusivity: marginal partners with match value $V = V^S$ necessarily have $\Phi > 0$, while well matched partners with $V > \frac{1}{2}[\overline{V} + V^S]$ instead have $\Phi < 0$.

4.1 Equilibrium Matching with Sorting by Exclusivity

Anticipating the empirical application, we consider sorting by equilibrium exclusivity only for the symmetric bargaining case, where negotiated $\mu = 1$, and with the relationship ladder property, that partner values $V(\theta)$ are increasing in θ. For reasons that will become clear, we adopt the following notation:

(i) singles now contact potential partners at rate λ_0;

(ii) $\theta-$partners can costlessly implement any joint outside contact rate $\lambda^*(\theta) \in [\phi\lambda_0, \lambda_0]$ where exclusivity parameter $0 \leq \phi \leq 1$;

(iii) there is exogenous breakdown of the match: a match shock arrives at an exogenous rate $\delta_0 \geq 0$ whereupon the match is destroyed (say new $\theta < \theta^R$) and the partners separate into singleness.

We assume romantic evenings are not costly (aside from foregone match opportunities) and so partners can costlessly self-enforce any $\lambda^*(\theta) \in [\phi\lambda_0, \lambda_0]$. Although search effort is costless, Theorem 2 shows greater search effort is only joint surplus increasing when $\Phi > 0$. We refer to chosen $\lambda^*(\theta)$ as partner search intensity and define aggregate search intensity

$$\Lambda = \lambda_0 G(\theta^R) + \int_{\theta^R}^{\overline{\theta}} \lambda^*(\theta) G'(\theta) d\theta. \tag{19}$$

The assumption λ_0 is fixed presumes constant returns to matching.

Taking endogenous search intensity into account, let $1 - P(\widetilde{\theta})$ denote the probability a randomly contacted agent is in a partnership better than $\widetilde{\theta}$. Random search with endogenous search intensity implies

$$1 - P(\widetilde{\theta}) = \frac{\int_{\widetilde{\theta}}^{\overline{\theta}} \lambda^*(x) G'(x) dx}{\Lambda}, \tag{20}$$

which is the fraction of aggregate search intensity due to those agents in matches $x \geq \widetilde{\theta}$.

In any symmetric partnership equilibrium with the relationship ladder property, a $\theta-$partner receives a preferred outside offer $\widetilde{\theta} \geq \theta$ at rate $\lambda^*(\theta) \int_\theta^{\overline{\theta}} P(\widetilde{\theta})F'(\widetilde{\theta})d\widetilde{\theta}$ where $P(\widetilde{\theta})$ now describes the probability the (random) outside contact is willing. For $\theta \geq \theta^R$, the functional equation for equilibrium $V(\theta)$ becomes

$$(r+\delta_0)V(\theta) = U(\theta) + \delta_0 V^S + \lambda^*(\theta) \int_\theta^{\overline{\theta}} [V(\widetilde{\theta}) + V^S - 2V(\theta))]P(\widetilde{\theta})F'(\widetilde{\theta})d\widetilde{\theta}, \tag{21}$$

which reflects partner symmetry, that when either partner receives and quits to a preferred outside match $\widetilde{\theta} > \theta$ the other is dumped with value V^S. Theorem 2 now determines the jointly optimal exclusivity decision

$$\lambda^* = \begin{cases} \lambda_0, & \text{if } \int_\theta^{\overline{\theta}} [V(\widetilde{\theta}) + V^S - 2V(\theta))]P(\widetilde{\theta})F'(\widetilde{\theta})d\widetilde{\theta} > 0 \\ \lambda^* \in [\phi\lambda_0, \lambda_0], & \text{if } \int_\theta^{\overline{\theta}} [V(\widetilde{\theta}) + V^S - 2V(\theta))]P(\widetilde{\theta})F'(\widetilde{\theta})d\widetilde{\theta} = 0 \\ \phi\lambda_0, & \text{if } \int_\theta^{\overline{\theta}} [V(\widetilde{\theta}) + V^S - 2V(\theta))]P(\widetilde{\theta})F'(\widetilde{\theta})d\widetilde{\theta} < 0. \end{cases}$$

The functional equation (21) now yields the **equilibrium exclusivity rule**

$$\lambda^* = \begin{cases} \lambda_0, & \text{when } V(\theta) > [U(\theta) + \delta_0 V^S]/(r+\delta_0) \\ \lambda^* \in [\phi\lambda_0, \lambda_0], & \text{when } V(\theta) = [U(\theta) + \delta_0 V^S]/(r+\delta_0) \\ \phi\lambda_0, & \text{when } V(\theta) < [U(\theta) + \delta_0 V^S]/(r+\delta_0). \end{cases} \tag{22}$$

The equilibrium exclusivity rule reveals a very simple intuition: in any match with $V(\theta) < [U(\theta) + \delta_0 V^S]/(r+\delta_0)$, both partners would strictly prefer to be fully committed ('til death do they part) with corresponding match payoff $[U(\theta) + \delta_0 V^S]/(r+\delta_0)$. They thus agree on full exclusivity with minimal outside contact rate $\lambda^* = \phi\lambda_0$. Conversely partners with $V(\theta) > [U(\theta) + \delta_0 V^S]/(r+\delta_0)$ prefer not to be exclusive.

Closing the model requires jointly determining $G(.)$ and $P(.)$. With endogenous search intensity, the partnership separation rate is

$$s(\theta) = 2\lambda^*(\theta) \int_\theta^{\overline{\theta}} P(\widetilde{\theta})F'(\widetilde{\theta})d\widetilde{\theta}.$$

If $dG(\theta) = G(\theta + d\theta) - G(\theta)$ denotes the measure of partners in matches $\widetilde{\theta} \in [\theta, \theta + d\theta]$ then standard steady state arguments imply

$$dG(\theta)[\delta + \delta_0 + s(\theta)] = \left\{ \lambda_0 G(\theta^R) + \int_{\theta^R}^{\theta} \lambda^*(x) G'(x) dx \right\} P(\theta) F'(\theta) d\theta,$$

where the LHS describes the flow out of partners (through death, destruction or separation), and the RHS describes the inflow which is the measure of those in a worse match $x < \theta$ along with their corresponding search intensity where, given a random contact, each forms a new partnership $\widetilde{\theta} \in [\theta, \theta + d\theta]$ when the contacted partner is willing. Hence $G(.)$ is described by the differential equation

$$\frac{dG}{d\theta} = \frac{P(\theta)^2}{\delta + \delta_0 + s(\theta)} \Lambda F'(\theta), \tag{23}$$

where $\Lambda F'(\theta)$ describes aggregate flow contacts with match draw θ and $P(\theta)^2$ describes the double coincidence of wants, that neither party is already in a better match. Hence taking endogenous search intensity into account, $G(.)$ and $P(.)$ are jointly determined by (23) and (20) along with boundary condition $G(\overline{\theta}) = 1$.

Definition 2. *A* ***Cohabitation Equilibrium*** *is the set* $\{V(.), G(.), P(.), \lambda^*(.), \Lambda, \theta^R, V^S\}$ *which satisfy the above equilibrium conditions, where the value of being single satisfies*

$$rV^S = u(y) + \lambda_0 \int_{\theta^R}^{\overline{\theta}} [V(\theta) - V^S] P(\theta) F'(\theta) d\theta, \tag{24}$$

and $V(\theta^R) = V^S$.

Although endogenous search intensity implies there is no longer a closed form solution, a *Cohabitation Equilibrium* has two useful simplifications. First, it is block recursive: the equilibrium values and choices $\{V(.), P(.), \lambda^*(.), \theta^R, V^S\}$ are determined separately from distributional outcomes $\{G(.), \Lambda\}$. A potential complication is that a mixed strategy region $\Theta \subset [\theta^R, \overline{\theta}]$ might exist with optimal choices $\lambda^*(\theta) \in (\phi\lambda_0, \lambda_0)$ for $\theta \in \Theta$. The numerical results, however, find the equilibrium exclusivity rule always has the single crossing property: there is a unique threshold $\theta^c \in (\theta^R, \overline{\theta}]$ where cohabiting partners $\theta > \theta^c$ have $[U(\theta) + \delta_0 V^S]/(r + \delta_0) > V(\theta)$ and so choose to be exclusive, while partners $\theta < \theta^c$ have $[U(\theta) + \delta_0 V^S]/(r + \delta_0) < V(\theta)$ and choose $\lambda^* = \lambda_0$. The cohabitation equilibrium finds $V(\theta)$ is continuous and strictly increasing in θ but is not continuously differentiable at θ^c (though its right and left derivatives exist).[15] Partner separation rate $s(\theta) = 2\lambda^*(\theta) \int_\theta^{\overline{\theta}} P(\widetilde{\theta})F'(\widetilde{\theta})d\widetilde{\theta}$ is strictly decreasing in θ (partners in better matches are less likely to leave) but is not continuous at θ^C where partners switch to being exclusive.

Because $\lambda^*(\theta)$, $s(\theta)$ are not continuous variables, we solve numerically for a Cohabitation Equilibrium by defining auxiliary variables

$$
\begin{aligned}
A(\theta) &= \int_\theta^{\overline{\theta}} P(\widetilde{\theta})F'(\widetilde{\theta})d\widetilde{\theta}; \\
S(\theta) &= \int_\theta^{\overline{\theta}} V(\widetilde{\theta})P(\widetilde{\theta})F'(\widetilde{\theta})d\widetilde{\theta}.
\end{aligned}
$$

Differentiation implies the dynamical system $\{P, A, S\}$ evolves according to the set of ordinary differential equations

$$
\begin{aligned}
P'(\theta) &= \frac{\lambda^*(\theta)P(\theta)^2F'(\theta)}{\delta + \delta_0 + s(\theta)}; && (25) \\
A'(\theta) &= -P(\theta)F'(\theta); && (26) \\
S'(\theta) &= -V(\theta)P(\theta)F'(\theta), && (27)
\end{aligned}
$$

with initial values $P = 1$, $A = 0$, $S = 0$ at $\theta = \overline{\theta}$. We solve this system using backward iteration. Given current values for $\{P(\theta), A(\theta), S(\theta)\}$ at $\theta \leq \overline{\theta}$, $V(\theta)$, $s(\theta)$ are simultaneously determined by

$$
\begin{aligned}
V(\theta) &= \frac{U(\theta) + \delta_0 V^S + \lambda^*(\theta)S(\theta) + \lambda^*(\theta)A(\theta)V^S}{r + \delta_0 + s(\theta)}; \\
s(\theta) &= 2\lambda^*(\theta)A(\theta),
\end{aligned}
$$

[15]The discrete change in search intensity λ^* at threshold θ^c implies $V(.)$ is continuous but is not continuously differentiable at θ^c. However at any crossing point θ^C where V$(\theta) = [U(\theta) + \delta_0 V^S]/(r + \delta_0)$, it can be shown that if $V(\theta)$ is strictly flatter than $[U(\theta) + \delta_0 V^S]/(r + \delta_0)$ on the left then it is also strictly flatter on the right. Similarly if $V(\theta)$ is strictly steeper than $[U(\theta) + \delta_0 V^S]/(r + \delta_0)$ to the left then it is also strictly steeper to the right.

with $\lambda^*(\theta)$ given by the equilibrium exclusivity rule (22). Thus given a candidate value for V^S, backward iteration from $\theta = \overline{\theta}$ yields $\{P, A, S, V, s, \lambda^*\}$ for all $\theta \in [\theta^R, \overline{\theta}]$. The equilibrium boundary condition $V^S = V(\theta^R)$ then ties down V^S. Computing the cohabitation equilibrium reduces to a simple one dimensional grid search for V^S and, in all numerical solutions, the switch-point θ^c, where $V(\theta^c) = [U(\theta^c) + \delta_0 V^S]/(r + \delta_0)$, is unique.

The system is block recursive and equilibrium $\{G(.), \Lambda\}$ are determined separately. First, choose a candidate value for Λ. Given initial value $G(\overline{\theta}) = 1$, equilibrium $G(.)$ is found using backward iteration on (23) and equilibrium Λ is tied down by the boundary condition

$$\Lambda = \lambda_0 G(\theta^R) + \int_{\theta^R}^{\overline{\theta}} \lambda^*(\theta) G'(\theta) d\theta.$$

Computing the cohabitation equilibrium numerically is straightforward and we now use simulated method of moments to estimate parameters.

4.2 Identification of the Cohabitation Equilibrium

Because there is no data on the dating process, for identification purposes we must reduce the model to essentials. Consistent with the previous section we suppose flow partner payoffs are more generally given by

$$U(\theta, y) = u(y) + \Psi(\theta),$$

where $\Psi(\theta)$ describes the (shared) flow surplus to any match $\theta \geq \theta^R$ and is a strictly increasing function of θ. This specification yields two useful normalisations. First given $\Psi(.)$ is arbitrary, there is no further loss in generality by assuming $F(.)$ is uniform on $[0, 1]$. Second, because payoffs in all states are additive in $u(y)$, there is no further loss in generality by normalising $u(y) = 0$ (e.g. see (15), (16) in Theorem 1). The reservation match value is then given where $\Psi(\theta^R) = 0$ and normalising $\Psi(0) = 0$ implies $\theta^R = 0$. The interpretation for what follows is that λ_0 describes the single's meeting rate with a desired partner θ such that cohabiting would yield positive match surplus $\Psi(\theta)$.

We set the death rate $\delta = 1/50$ per annum, a subjective time rate of preference $\rho = 2\%$ and so discount rate $r = \delta + \rho = 4\%$ per annum. We further restrict attention to a geometric surplus function $\Psi(\theta) = \overline{\Psi}\theta^{\gamma}$ with $\gamma > 0$. Because $u(y) = 0$, however, $\overline{\Psi}$ only rescales utilities and so we can normalise $\overline{\Psi} = 1$. We thus arrive at the following parsimonious set of parameters to be estimated:

Table 2: Parameters to be Estimated

λ_0	Meeting rate with a desired potential partner
ϕ	Reduced contact rates through full partner exclusivity
δ_0	Exogenous breakdown of the match
γ	Skewness of match surplus Ψ

Preliminary results, however, find the high 15% failure rate of new partnerships requires an additional premature break-up process (M, μ). The most natural interpretation is that cohabitation is an experience good, where steady state $M \geq 0$ partners learn at rate μ they are incompatible and separate. Rather than complicate the cohabitation equilibrium by introducing additional learning dynamics (see for example Jovanovic, 1982; Anderson and Smith, 2010; Marinescu, 2016), we instead incorporate and consistently estimate a "reduced form" premature break-up process as follows.

4.3 The Extended Model: Premature break-up (M, μ)

For calibration purposes only, we now allow a (steady state) mass $M \geq 0$ matches with $\theta = 0$ (zero surplus). We return to using $\overline{\lambda}$ as the contact rate with desirable partners but now with probability p the match draw $\theta = 0$ and the match is "bad" (with zero surplus) and with complementary probability $1 - p$ the match is "good" with $\theta \sim U(0, 1]$ as previously described. Because forming a bad match with surplus $\Psi = 0$ is payoff equivalent to remaining single, the extended framework is payoff equivalent to the cohabitation equilibrium with $\lambda_0 = \overline{\lambda}(1 - p)$. But we can further allow partners in bad matches to separate into singleness at an excess rate $\mu \geq 0$

because they are indifferent to doing so. Assuming only singles form zero surplus matches (given contact), Lemma 4 describes the resulting steady state measure M of "bad matches" given excess breakdown rate $\mu \geq 0$.

Lemma 4. *Given the underlying cohabitation equilibrium with $\lambda_0 = \overline{\lambda}(1-p)$, premature break-up $\mu \geq 0$ implies $M \in [0, G(0)]$ solves*

$$\frac{\overline{\lambda}p}{G(0) + \int_0^{\overline{\theta}} \phi^*(\theta)G'(\theta)d\theta} [G(0) - M]^2 = M \left\{ \delta + \delta_0 + 2\overline{\lambda}(1-p) \int_0^1 P(\theta)d\theta + \mu \right\},$$

where $\phi^*(\theta) \equiv \lambda^*(\theta)/\lambda_0$ in the calibration equilibrium.

Proof. Proof for Lemma 4 is in Appendix A.

It is easy to show that a solution for steady state $M \in [0, G(0)]$ exists and is unique given any $\mu \geq 0$. We consistently quantify the premature break-up process (M, μ) by estimating the following extended set of parameter values:

Table 3: Extended Parameters

$\overline{\lambda}$	Meeting rate with a (mutually) desired potential partner
ϕ	Partner exclusivity.
δ_0	Exogenous breakdown of the match.
γ	Skewness of match surplus.
p	Fraction of contacts a bad match ($\theta = 0$)
μ	Excess break-up rate of bad matches ($\theta = 0$)

where the cohabitation equilibrium is as previously described with $\lambda_0 = \overline{\lambda}(1-p)$ and Lemma 4 determines the steady state measure M of bad matches.

5 Estimation

5.1 The Data

To be consistent with GJR, we use data from the BHPS from all 18 waves, covering 1991 to 2008 (UK data). The BHPS collects information annually from a sample of 5,505 households containing approximately 10,000 original sample members (OSMs). The BHPS allocates survey weights to each household so that the data set is representative of the whole UK. Each year OSMs and the adults they lived with were interviewed on a wide range of topics including demographic characteristics, labour market status, income, marital status and fertility. Using this information we observe relationship (cohabitation and marriage) transitions based on an individual's reported de facto marital status and their partner's unique identification number, and so can directly measure transitions into and out of relationships. To preserve representativeness, we focus entirely on the OSMs where survey weights are automatically adjusted to account for sample attrition.[16] Whenever an OSM is in a relationship, we also have information on partner characteristics. Because our analysis does not require as many control variables (e.g. wages, family values, hours of household work per week), we do not have the same data selection issues which arise in GJR. The above restrictions and definitions leave us with a baseline sample containing 1,750 partnership events, 622 of which (i.e. 35%) were observed to separate.

An important challenge with the data, however, is there is information on the death of a partner when formally married, but not with unmarried cohabitation. Thus when calculating separation rates, for married couples an observed death of a partner is not counted as a separation, though it does mean the end of the partnership, while an unobserved death of a cohabiting partner is counted as a separation. Of course Table 1 reveals measured separations are heavily skewed towards the young where this mortality issue would appear less important. Nevertheless, when estimating the separation hazard function below, we mitigate this mortality issue by following GJR

[16]Specifically, we weight the estimates using individual response weights, variable xrwght, and adjust as necessary when estimating using subsamples

and restricting the sample to those couples under the age of 50. As a robustness check, we have re-estimated the target moments below by instead counting a married partner death as a separation so that both partnership types are treated consistently. This does not alter the estimates of the separation hazard by duration and only slightly increases the measured average separation rate $\overline{h}^S$ (results are available upon request).

5.2 Target Moments

Transition rate to first relationship: To identify $\overline{\lambda}$, the meeting rate with a desired partner, we target the transition rate to a first relationship (h^p) for prime-age young people (24 to 28 years of age). We do this by computing the (weighted) fraction of those that transition to their first relationship within this age interval. To be included each individual has to have been observed at the previous age and confirmed they have not previously married or cohabited with a partner. Estimated $h^P = 0.148$.

Separation hazard: To compute the separation hazard by duration ($h^S(\tau)$) we restrict attention to relationships observed from their inception, and for comparability with GJR, we focus on couples in which both partners are between 22 and 50 years old. The hazard for each duration $\tau - 1$ is then computed as the (weighted) fraction of the number of relationships observed at duration $\tau - 1$ which is separated by duration τ. Spells with intermittent missing waves are right censored when the couple is first missing from the dataset. Figure 1 in the Introduction graphs the estimated hazard function and the actual values are reported in Table 5 below. The estimated hazard function implies the probability a new partnership survives 7 years is

$$\prod_{\tau=0}^{\tau=6}[1 - h^S(\tau)] = 0.63.$$

The average separation rate: We compute the average separation rate ($\overline{h}^S$) across all partners as the fraction of separations observed using survey weights. Observations are included if the couple reported being together in the previous wave regardless of their age. Direct transitions, i.e. when an individual transitions from one relationship to another between consecutive waves, are accounted for by checking partner ID numbers remain the same across waves. Estimated $\overline{h}^S = 2.1\%$.[17]

5.3 Simulated Method of Moments: Parameter estimates and discussion

Noting that flow in equals flow out in a steady state, standard arguments imply the equilibrium match rate of singles is

$$h^P = \frac{M\left\{\mu + \delta + \delta_0 + 2\overline{\lambda}(1-p)\int_0^1 P(\theta)d\theta\right\}}{G(0) - M} + \overline{\lambda}(1-p)\int_0^1 P(\widetilde{\theta})d\widetilde{\theta},$$

where the first term describes their flow into (short-lived) bad matches, the second into (longer-lived) good ones. The calibration sets $\overline{\lambda}$ to exactly match $h^p = 0.148$.

The average separation rate across all partners is

$$\overline{h}^S = \frac{[\delta_0 + s(0) + \mu]M + \int_0^1 G'(\theta)[\delta_0 + s(\theta)]d\theta}{M + \int_0^1 G'(\theta)d\theta},$$

where "bad matches" separate at rate $(\delta_0 + s(0) + \mu)$, good matches separate at rate $(\delta_0 + s(\theta))$ where $s(\theta) = 2\lambda^*(\theta)\int_\theta^1 P(\widetilde{\theta})d\widetilde{\theta}$. The calibration sets ϕ (exclusivity) to exactly match $\overline{h}^S = 0.021$.

The remaining parameters $(p, \mu, \delta_0, \gamma)$ are chosen to match the estimated hazard function $h^S(.)$ using a standard quadratic loss function. According to the model, the average separation rate at duration τ is

$$h^S(\tau) = \frac{(\delta_0 + s(0) + \mu)\left\{(\delta + \delta_0 + s(0) + \mu)Me^{-[\delta+\delta_0+s(0)+\mu]\tau}\right\} + \int_0^1[\delta_0 + s(\theta)]\Lambda P(\theta)^2 e^{-([\delta+\delta_0+s(\theta)])\tau}d\theta}{(\delta + \delta_0 + s(0) + \mu)Me^{-[\delta+\delta_0+s(0)+\mu]\tau} + \int_0^1 \Lambda P(\theta)^2 e^{-([\delta+\delta_0+s(\theta)])\tau}d\theta},$$

where the denominator is the measure of new partnerships which survive to duration τ, and the numerator is the corresponding population-weighted separation rates. Table 4 describes the resulting parameter estimates:

[17] If you truncate the sample to include only those up to the age of 50 as done in GJR, the estimate becomes 2.6%. This measure, however, excludes long and successful partnerships.

Table 4: Estimated Model Parameters

$\overline{\lambda}$	0.29	Meeting rate with a mutually desired potential partner
ϕ	0.23	Partner exclusivity
δ_0	0	Exogenous breakdown of the match
γ	0.30	Skewness of match surplus
p	0.10	Fraction of contacts a bad match
μ	2.11	Excess break-up rate of bad matches
θ^c	0.35	Threshold for partner exclusivity
λ_0	0.26	Meeting rate with a good match ($\theta > 0$)
M	0.0006	Steady state bad matches

Notes: This table presents the estimated parameters using the simulated method of moments.

The first four rows describe the main parameter estimates, the middle rows describe the additional premature break-up process, and the final 3 rows describe endogenous outcomes. The estimated meeting rate $\overline{\lambda}$ implies an average spell of 3.4 years to meet a mutually desired partner where cohabiting would generate a positive joint surplus. But because fraction $p = 10\%$ of these contacts yield zero surplus, the meeting rate to a (strictly) good match is instead $\lambda_0 = 0.26$; a wait of around 4 years. The choice of $\overline{\lambda}$ ensures the model exactly matches the h^p target, while exclusivity ϕ ensures the model exactly matches the 2.1% average separation rate.

The remaining parameter values $\{\delta_0, \gamma, p, \mu\}$ target the separation hazard function and Table 5 describes the resulting fit.

The extended model provides a very good match to the separation hazard function. The implied premature break-up process is fast: estimated $\mu = 2.1$ implies "bad" matches have an expected duration of less than 6 months. This high failure rate then implies steady state $M = 0.06\%$ is extremely small. This does not imply, however, that bad matches are rare, for estimated $p = 10\%$ of contacts are bad matches and together with their high failure rate ($\mu = 2.1$) this generates the required 15.2% failure rate of new partnerships. Although bad matches are not rare, few bad matches survive their

Table 5: Match to the Hazard Function

Hazard rate	Data	Estimates	
		With premature break-up	Good matches only
$h^S(0)$	0.152	0.152	0.068
$h^S(1)$	0.071	0.069	0.060
$h^S(2)$	0.050	0.054	0.054
$h^S(3)$	0.049	0.048	0.048
$h^S(4)$	0.039	0.043	0.043
$h^S(5)$	0.038	0.038	0.038
$h^S(6)$	0.041	0.035	0.035
$h^S(7)$	0.030	0.032	0.032

Notes: This table compares the estimated separation hazard functions from the simulated method of moments and the separation hazard in the data. Column 2 contains the moments constructed from the BHPS (1991-2008). Column 3 presents the estimated separation hazard function when we use the extended model with premature break-up. Column 4 describes the implied separation rates by duration for good matches only

first year and barely 1% survive two years. Table 5 demonstrates for durations $\tau \geq 2$ that the separation hazard function of the extended model (column 3) is the same as the hazard function for good matches only. The insight is that the premature break-up process is essential for explaining the high 15% failure rate of new partnerships, where this process requires a significant fraction of new partnerships are bad matches and a fast "learning" rate μ. Section 6 below examines the premature break-up process in greater detail, while the rest of this section focuses on the underlying cohabitation equilibrium.

A particularly interesting result is that the estimated match destruction rate $\delta_0 = 0$ (its corner value). This occurs because a strictly positive δ_0 yields a flatter separation hazard function (for example see Figure 1). The underlying empirical difficulty is explaining why the empirical hazard function is so very steep, which is why simulated method of moments selects the corner value $\delta_0 = 0$. We return to this issue in the conclusion.

The estimated structural model implies the average dissolution rate $s(\theta)$ of exclusive partnerships $\theta > \theta^C$ is just 1.6% per annum, which is below the (assumed constant) mortality rate of 2%; i.e. 'til death do they likely part. And because exclusive matches have high survival rates, steady state then finds most adults are indeed in exclusive relationships. Consistent with Table 1, it is the interaction of exclusivity choice and steady state composition effects which explain the low average separation rate 2.1% across all partnerships.

The interesting question then is how long does it take a young entrant to find their exclusive match? We formally define meeting the One as entering an exclusive partnership $\theta \geq \theta^C$. If λ^C denotes the rate at which anyone not already in an exclusive match successfully forms a partnership $\theta > \theta^c$, the model implies

$$\lambda^C = \overline{\lambda}[1-p] \int_{\theta^c}^{1} P(\widetilde{\theta}) d\widetilde{\theta},$$

which requires

(i) meeting a good match which, for anyone not already exclusive, occurs at rate $\lambda^* = \lambda_0 = \overline{\lambda}[1-p]$;

(ii) that the match is so good it is exclusive $\widetilde{\theta} \geq \theta^C$;

(iii) that the contacted potential partner is willing; i.e. not already better matched;

(iv) taking into account exclusive partners $\theta > \theta^c$ have much reduced contact rates ($\phi = 0.23$).

Putting the information together, the structural model calculates

$$\lambda^C = [0.287][1-0.100][0.382] = 0.099 \tag{28}$$

and so an expected duration equal to 10 years. At first sight, this might seem too slow but instead, it fully reflects the low match rate to first partnership (estimated $\mathrm{h}^p = 0.148$) and high partnership failure rates. Indeed if $\widehat{P}$ is the probability the first partnership is with the One, then estimated $\widehat{P} = \frac{\lambda^C}{h^p} = 67\%$, which closely reflects that only 63% of new partnerships survive 7 years. Except $\widehat{P}$ here is a structural estimate: it implies around $\frac{2}{3}$ of first time cohabiters have found their One.

Finally it is interesting to note that imperfect exclusivity ($\phi = 0.23$) implies exclusive partners still face outside temptation: estimated $\phi\lambda_0 = 0.06$ implies an exclusive partner is enticed by an outside contact around once every 17 years. This might not seem so much but there are two in the match: a potential partnership destroying shock instead occurs around once every 8 years which, in the context of a lifetime together, is a substantial risk. Of course those in the best matches $\theta \simeq 1$ have little to fear - they are unlikely to meet a preferred outside partner and the outside contact might not anyway be willing. But estimated $\theta^c = 0.35$ implies there are many exclusive matches $\theta > \theta^C = 0.35$ which, though good enough to be exclusive, are at great risk of succumbing to a willing outside contact $\widetilde{\theta} \in (0, 1]$. Although a slow process, heartbreak (where an exclusive partner finds their partner leaves for someone else) is not particularly rare. And so despite 67% of first partnerships being with the One, the 8 year itch explains why around one half of first partnerships fail within 20 years.

6 On the Determinants of Premature Break-up: A probit analysis

This section uses probit regressions to identify which partner characteristics if any, predict premature break-up, measured as those partnerships which break-up within their first year. A pure learning interpretation of premature break-up suggests observable partner characteristics should not predict premature break-up for who would voluntarily enter a bad match? But an important data issue is the estimated model predicts that around 7% of good matches also break-up in the first year (see column 4 of Table 5) and so around one half of first year break-ups are due to the betrayal process, the rest due to premature break-up.

The main survey waves of the BHPS provide information on the demographic characteristics of both partners (e.g. age, education, race, prior fertility (i.e. a single mum or dad)), and so we can consider the premature break-up rates by partnership characteristics. There is also information on partner employment status and even drug/alcohol abuse. As above, the sample includes all observed heterosexual couples in which both partners are between the ages of 22 and 50.

The dependent variable is the probability of a premature break-up, measured as a dummy variable equal to one if the couple separates in the first year of their partnership and zero otherwise.[18] Table 6 reports the probit estimates where only the following ex-ante individual characteristics are included as conditioning variables:

(i) age, measured as a female (and male) dummy variable which is set equal to 1 if the female (male) partner is under the age of 25;

(ii) their age gap (male partner age - female partner age) entered as a simple quadratic;

(iii) education composition by gender; partners are described as college graduates or non-college graduates;

(iv) racial composition, partners are described as white or BAME (not by gender due to small sample size in some cells);

(v) prior labour market status by gender; partners are either employed or non-employed;

(vi) others (first cohabiting relationship, single parent prior to the match, drug/alcohol dependency)

and also a full set of region of residence dummies. Table 6 describes the resulting set of probit estimates. Both the estimated coefficients and the average marginal effect are presented to ease interpretation.

Perhaps the most interesting feature is how few variables actually predict premature break-up which is consistent with the learning hypothesis. For example, being a single dad does not predict premature break-up, and there is very limited evidence that being a single mum predicts premature break-up (the estimate is only significant

[18] We therefore estimate the regression models using cross-section data containing one observation from the second year of each relationship.

at the 10% level), reflecting that this information was known when the partners chose to move in together. Similarly ex-ante characteristics such as the education and racial composition of partnerships are also not associated with premature break-up, where the results are unaffected by using finer partitions of individual race and education. We include ”first relationship” as a possible measure of partnership naivety but this also does nothing when all controls are included, nor do drug/alcohol dependency issues.

There are two robustly estimated effects. The first is that high break-up rates are strongly associated with the female partner being young (below 25 years old). On average a couple with a young female has between 8 and 11 percentage point increase in the probability of separating in the first year. Surprisingly the young male dummy variable (under 25 years old) is not significant. What is significant, however, is his age relative to hers: the quadratic age-gap terms imply the premature break-up rate reaches a maximum when the male partner is around 4 years older than her. The labour market status variable suggests being both non-employed has a significant effect on premature break-up rates compared to couples in which both partners are employed.

The second set of probit results (Table 7) omits those ex-ante partner characteristics found not to be significant (e.g. race and education composition effects, though for a robustness check they are included but not reported in column 4) and includes endogenous variables which are likely correlated with, and so a proxy for (unobserved) exclusivity choice. For example, it would seem highly likely that once partners have found their One, they might further choose to marry and/or start a family.

Not surprisingly Table 7 finds the fertility outcomes and the married dummy variable are highly significant: the match is much less likely to prematurely break-up when married with a new baby. This suggests an important direction for future research: to consider endogenous fertility outcomes along with finding the One. The labour market status variables find (male employed only) and (both unemployed) are significant but the difference is not statistically significant. This reveals that premature break-up is much more likely when a female is non-employed (prior to the match) and, surprisingly, unaffected by male employment status.

The age effects remain robust: that break-up in the first year is much more likely when she is young (below 25) and he is 3-4 years older.[19] Of course, this result is purely descriptive and the crucial question is why these correlations? Is it due to the learning process or betrayal? A simple way to distinguish is to follow Devereux and Turner (2016) and measure repartnering outcomes following a premature break-up. If a breakup is due to learning, then both separate and are single with re-matching probability $h^P = 0.148$ for the following year. But if the breakup is due to betrayal then a reasonable assumption is the betraying partner immediately moves in with someone else, while the other becomes single with match rate $h^P = 0.148$. According to the estimated model, around one half of premature break-ups are due to betrayal (see first row, Table 5, good matches only). If true, a ballpark figure for the probability of being repartnered within a year of a premature break-up is $\frac{1}{4}[1 + 3 \times 0.148] = 0.361$. Table 8 provides the estimated probabilities of repartnering in the first year after a premature break-up.

Table 8 presents some interesting insights. Similar to Ermisch (2002), which uses a shorter version of the BHPS, and to studies using US data (Devereux and Turner, 2016), we estimate a high initial repartnering rate after a separation, far above $h^P = 0.148$ establishing that many of these premature break-ups are indeed due to one partner moving in with someone else. There is no evidence of a gender gap in repartnering, which further tells us that male and female partners are equally likely to betray the other. The estimated repartnering probabilities, however, are below the 0.361 ballpark value implied by the model. An obvious possibility is that the betraying partner does not necessarily move in with the outside partner; i.e. there may instead be a betrayal by cheating. We could extend the model to allow two different types of betrayal, say

[19] See a graph of the predicted separation probabilities for the observed age gaps in the Empirical appendix (B) demonstrating the quadratic shape and age gap with the maximum predicted probability of separation.

there is also ”cheating but not both are willing to move in together”. However, with no direct information on cheating, we leave this issue to future research.

Table 6: Determinants of the Probability of a Premature Break-up: Couple Characteristics

DV = Separation in year 2 of partnership	(1)		(2)		(3)		(4)		(5)	
	β / (SE)	Mfx	β / (SE)	Mfx	β / (SE)	Mfx	β / (SE)	Mfx	β / (SE)	Mfx
Young female partner	0.354***	0.086***	0.411***	0.100***	0.427***	0.104***	0.347***	0.083***	0.451***	0.110***
	(0.126)		(0.130)		(0.136)		(0.129)		(0.140)	
Young male partner	-0.103	-0.022	-0.071	-0.015	-0.053	-0.011	-0.097	-0.021	-0.032	-0.007
	(0.147)		(0.151)		(0.151)		(0.150)		(0.154)	
Age gap (male - female age)	0.022	0.002	0.025	0.003	0.024	0.002	0.024	0.003	0.027*	0.003*
	(0.016)		(0.016)		(0.016)		(0.016)		(0.016)	
Age gap squared	-0.003**		-0.003**		-0.003**		-0.003**		-0.003**	
	(0.001)		(0.001)		(0.001)		(0.001)		(0.001)	
Education (Omitted: Both high)										
Male high, female low	0.233	0.054	0.189	0.044	0.221	0.051	0.240	0.055	0.195	0.045
	(0.156)		(0.158)		(0.156)		(0.154)		(0.156)	
Male low, female high	-0.020	-0.004	-0.047	-0.010	-0.033	-0.007	-0.009	-0.002	-0.041	-0.009
	(0.145)		(0.146)		(0.144)		(0.143)		(0.143)	
Both low	0.108	0.023	0.053	0.012	0.080	0.017	0.108	0.023	0.041	0.009
	(0.125)		(0.129)		(0.127)		(0.126)		(0.130)	
Race (Omitted: Both white)										
One partner white, other BAME	0.183	0.044	0.260	0.064	0.200	0.048	0.198	0.048	0.275	0.068
	(0.323)		(0.315)		(0.315)		(0.319)		(0.306)	
Both BAME	-0.361	-0.067	-0.319	-0.060	-0.342	-0.064	-0.351	-0.065	-0.300	-0.056
	(0.438)		(0.452)		(0.432)		(0.439)		(0.444)	
Labour market status (Omitted:Both employed)										
Male employed only	0.319*	0.075*	0.262	0.060	0.296*	0.069*	0.287	0.066	0.221	0.050
	(0.169)		(0.181)		(0.172)		(0.184)		(0.197)	
Female employed only	0.211	0.047	0.220	0.049	0.194	0.043	0.213	0.048	0.208	0.047
	(0.293)		(0.297)		(0.299)		(0.294)		(0.301)	
Neither employed	0.640***	0.171***	0.559***	0.146***	0.595***	0.156***	0.626***	0.167***	0.523***	0.134***
	(0.174)		(0.175)		(0.178)		(0.174)		(0.177)	
Single mum before partnership			0.277**	0.065**					0.233*	0.054*
			(0.120)						(0.127)	
Single dad before partnership			0.045	0.010					0.026	0.006
			(0.308)						(0.319)	
First partnership for at least one partners					-0.239**	-0.052**			-0.184	-0.040
					(0.121)				(0.134)	
At least one partner reported health problems due to drugs or alcohol							0.781	0.231	0.772	0.225
							(0.506)		(0.504)	
Observations	1,147		1,147		1,147		1,147		1,147	
Pseudo R^2	0.071		0.077		0.076		0.074		0.083	
Region Control	Yes		Yes		Yes		Yes		Yes	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Note: Probit regression coefficients (β) and average marginal effects (Mfx) are presented. Standard errors are clustered at the primary sampling unit level. Region controls correspond to the region (GOR) of residence.

Table 7: Determinants of the Probability of a Premature Break-up: Match specific investments

DV = Separation in year 2 of partnership	(1)		(2)		(3)		(4)	
	β / (SE)	Mfx	β / (SE)	Mfx	β / (SE)	Mfx	β / (SE)	Mfx
Young female partner	0.293***	0.068***	0.274***	0.065***	0.271***	0.062***	0.398***	0.091***
	(0.100)		(0.099)		(0.099)		(0.120)	
Age gap (male - female age)	0.031**	0.004**	0.026*	0.003*	0.031**	0.004**	0.034**	0.004**
	(0.015)		(0.015)		(0.015)		(0.016)	
Age gap squared	-0.003**		-0.003***		-0.003***		-0.004***	
	(0.001)		(0.001)		(0.001)		(0.001)	
Labour market status (Omitted: Both employed)								
Male employed only	0.597***	0.147***	0.370**	0.088**	0.585***	0.143***	0.424**	0.098**
	(0.174)		(0.161)		(0.162)		(0.188)	
Female employed only	0.212	0.044	0.176	0.038	0.175	0.036	0.157	0.032
	(0.295)		(0.296)		(0.298)		(0.306)	
Neither employed	0.747***	0.194***	0.652***	0.173***	0.730***	0.187***	0.591***	0.146***
	(0.183)		(0.172)		(0.187)		(0.192)	
Couple had child before partnership	-0.446*	-0.080*			-0.483*	-0.084*	-0.581*	-0.095*
	(0.269)				(0.275)		(0.305)	
Couple had child year partnership formed	-0.802***	-0.120***			-0.882***	-0.127***	-0.802**	-0.119**
	(0.306)				(0.315)		(0.330)	
Couple had child first year of partnership	-0.717***	-0.114***			-0.763***	-0.118***	-0.782***	-0.119***
	(0.234)				(0.238)		(0.243)	
Number of biological children in household	-0.055	-0.012			0.028	0.006	0.026	0.005
	(0.153)				(0.159)		(0.153)	
Married			-0.405***	-0.080***	-0.420***	-0.080***	-0.402***	-0.076***
			(0.150)		(0.129)		(0.130)	
Observations	1,147		1,147		1,147		1,147	
Pseudo R^2	0.093		0.075		0.102		0.118	
Region Control	Yes		Yes		Yes		Yes	
Full set of individual characteristics	.		.		.		Yes	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Note: Probit regression coefficients (β) and average marginal effects (Mfx) are presented. Survey weights are used. Standard errors are clustered at the primary sampling unit level. Region controls correspond to the region (GOR) of residence. The full set of individual controls includes all of the control variables used in column 5 of Table 6.

Table 8: Average Repartnering Probability in the First Year After a Premature Break-up

	All	Females	Males
Repartnering probability	0.303	0.316	0.290
	(0.025)	(0.032)	(0.039)

Notes: This table presents estimates of repartnering rates and the standard errors in parentheses using data from the BHPS main survey, 1991-2008. The average probability is measured as the ratio of the number of people who are observed in a premature breakup relationship in wave $t-2$, who are also observed to be single in $t-1$ and then in a new relationship in wave t, over the total number of people at risk in wave t who have broken up in a premature breakup in wave $t-1$. The sample and definitions follow those detailed in subsection 5.1. Survey weights are used.

7 Conclusion and Future Research

This paper has considered endogenous matching and separation in a partnership equilibrium with on-the-job search and endogenous exclusivity. We have shown how to reconcile the facts that 15% of new cohabiting partnerships in the UK fail in their first year, that 37% fail in the first 7 years, yet the average separation rate across all cohabiting partners is just 2.1% per year. The results find:

- it is important to consider on-the-job search with endogenous exclusivity to explain the steepness of the separation hazard function and the low average separation rate;
- clear evidence of a premature break-up process. The natural interpretation is (some) new partners learn they are in a "bad" match and the estimated learning rate is fast: a bad match has an expected duration of just 5 months. Probit results find premature break-up is uncorrelated with ex-ante partner characteristics (including education and race) but is correlated with age. Premature break-up is most likely when the female partner is below 25 years old, her male partner is around 3-4 years older, they are unmarried and have not started a family;
- reflecting the high failure rates of new partnerships, a new entrant to the singles market faces an average 10 year wait to find and settle down with an exclusive partner;
- but being exclusive does not solve all problems: exclusive partners face an average 8 year itch where one partner is potentially tempted by someone else. The best matches, of course, easily resist such temptations. Unfortunately there are many matches that, though good enough to be exclusive, are still at risk of succumbing to a willing outside contact. Although a slow process, heartbreak (where an exclusive partner is left for someone else) is not particularly rare and the cost of heartbreak is large (it takes an expected 10 years to find another One);

- once separations due to learning and betrayal are taken into account, separations through exogenous breakdown seem much less important.

For comparability with GJR we have used the same data set (BHPS). Our key contribution is to demonstrate the importance of on-the-job search and commitment effects for explaining the observed turnover patterns. Future research might consider other countries where for example Bruze et al. (2015); Brien et al. (2006) find qualitatively identical turnover patterns for the US and Denmark.

An important extension of our approach would be to endogenize investment in relationship-specific capital such as fertility or marriage. Indeed an important property of the relationship ladder is that higher quality matches are longer lived and so provide a greater return to match specific investments (e.g. Quercioli, 2005). Having children is an investment in a partnership in much the same way as is becoming exclusive and reducing outside contact rates. It seems natural that such choices are highly correlated; i.e. singles might wait till they have found their One before starting a family. Indeed the extended framework can then be used not only to examine endogenous fertility outcomes but also the modern prevalence of step-families. For example, in the US, 40% of married couples with children are in fact step-couples.[20]

Because we have ruled out ex-ante agent heterogeneity, we could not consider partner level shocks which some have found to be important. For example, Holzner and Schulz (2019) estimate that around 10% of break-ups are due to economic shocks, that one partner has become unemployed (also see Ciscato, 2019; Shephard, 2019). Such partner shocks are not inconsistent with the process considered here, where instead the partner shock is that one partner has met someone else rather than become unemployed. Indeed it would seem reasonable that disappointed partners might become more willing to move in with someone else. Extending and estimating the framework with ex-ante heterogenous agents, individual partner shocks and on-the-job search is clearly an important direction for future research.

[20]https://www.smartstepfamilies.com/smart-help/marriage-family-stepfamily-statistics

The estimated corner value $\delta_0 = 0$ reflects the difficulty of explaining the steep hazard function. An alternative approach might instead allow age dependent strategies, say the young are liable to early separation, see for example Shephard (2019) and our results on premature break-up. In our context instead, new singles might be born ″youthful″ where ″youthful″ types have a high cost to being exclusive and, say, are unable (unwilling) to commit to (long term) exclusive strategies. Suppose, however, ″youthful″ singles mature at some exogenous rate ζ, where ″mature″ adults can instead commit costlessly to being exclusive. The added heterogeneity in separation rates will then increase the steepness of the implied hazard function. Targeting the empirical hazard function, the extended framework might then estimate δ_0 positive consistent with the current literature. Unfortunately equilibrium sorting then depends on agent types, on being ″youthful″ or ″mature″, and so is much more complex. Future research might identify this extension by additionally targeting separation rates by age (see Table 1) and repartnering rates (see Table 8). The former information would seemingly capture age effects, while the latter data identifies the nature of the separation - do both partners becoming single or is one moving to a new relationship?

Because we have no information on dating, the empirical application has focused on cohabitation dynamics. To reveal the underlying nature of the UK partner search process, we reproduce a (non-scientific) survey of UK partners who believe they have finally found their One. Table 9 shows that many settled UK partners have indeed previously cohabited with someone else, and ″heartbreak″ and ″cheating″ are not uncommon. Consistent with our 67% measure of first cohabiting partnerships being the One and a slow 8 year itch process, those averaged (and rounded) numbers are reassuringly small. Nevertheless reflecting the long (average) 10 year wait prior to moving in with the One, the Table also reveals many have long and eventful dating histories.

Table 9: Pathways to Meeting the "One"

Event	Number of instances	
	Women	Men
Partners lived with	**1**	**1**
Fallen in love	**2**	**2**
Heartbreak	**2**	**2**
Times cheater	**1**	**1**
Times cheated on	**1**	**1**
Disaster dates	4	4
Blind dates	2	3
Stood up on a date	1	2
Online dates	2	3
Dating relationships (year or less)	3	4
Dating relationships (year or more)	2	2
Long distance relationships	1	1
Number kissed	15	16
Sexual partners	7	10
One night stands	4	6

Note: This table reproduces the outcome of a 2012 survey of 2,000 UK adults who believed they had met the "One". The survey commissioned by Graeme Simsion's publishers, as reported online in Her magazine:
https://www.her.ie/life/whats-your-number-study-finds-the-average-number-of-dates-and-relationships-before-we-find-the-one-90330

Chapter 3

Worker Productivity During Covid-19 and Adaptation to Working From Home*

1 Introduction

Across the world, the Covid-19 pandemic caused widespread disruption to working practices, including, most saliently, a vast increase in working from home (WFH). The share of the labour force working from home increased from around 5% to over 40% in the U.S. during the first lockdown of Spring 2020 (Bloom, 2020), with a similar change seen in the UK (Reuschke and Felstead, 2020). As the pandemic progressed, evidence accumulated that increased WFH will likely persist for the foreseeable future (Barrero et al., 2021b).[1] Indeed, in the UK by end-2022, 44% of the labour force worked from home at least partially (Office for National Statistics, 2023), even as the pandemic was largely over.

The shock of Covid-19 raises many questions on which evidence is still needed. For example, how did the change in working practices affect workers of different types, in

*This paper is co-authored with Ben Etheridge, Li Tang and Yikai Wang. We are grateful to discussions with Thomas Cornelissen, Tom Crossley, Emilia del Bono, Emma Duchini, Alex Whalley and seminar audiences at Sussex, Reading, Bank of Canada, Essex and the CES-Ifo 2022 Summer Institute Conference. Etheridge thanks support from the British Academy, grant number SRG22\221343 and the ESRC Centre for Micro-Social Change (MiSoC), award number ES/S012486/1. This paper draws on data from Understanding Society, distributed by the UK Data Service. Understanding Society is an initiative funded by the Economic and Social Research Council and various Government Departments. All errors remain the responsibility of the authors.

[1]For a wider discussion and extensive references see also the dedicated discussion of the literature below.

different jobs and with different household circumstances? Focusing more specifically on WFH, how did job experiences and performance during Covid-19 shift patterns of worker location as the pandemic progressed? And what factors affected this performance across locations?[2] These last two questions are particularly important for assessing the evolution of preferences for WFH (Aksoy et al., 2022). The rise in WFH has important implications for labour markets and economic geography, with evidence accumulating that its rise has already affected, for example, the distribution of house prices as well as wage inequality (Barrero et al., 2022).

In this paper we address these questions using the Covid-19 module from the UK Household Longitudinal Survey (UKHLS), which provides representative panel data for much of the pandemic in the UK, from April 2020 to September 2021. In this survey, all workers were asked about both their current working location as well as about changes in their productivity since a reference period before the pandemic's onset. These data allow us to examine how worker performance varied across job and worker types and was influenced by, for example, the presence of children, as well as housing characteristics. These data also allow us to track the joint evolution of productivity and worker location through various stages of the pandemic, both at times of strong restrictions, and when policies were more relaxed. Compared to other related datasets used in the literature, such as the Survey of Working Arrangements and Attitudes (Barrero et al., 2021b), these data allow us to track the *same* individuals over time.

We make three main contributions to the already large literature on inequality during Covid-19 and to the growing literature on working from home (WFH). First we provide the most systematic evidence of working location and productivity outcomes over the course of the pandemic using representative labour market data. Compared to papers using similar data to ours from self-reports (e.g. Deole et al., 2023; Aksoy et al., 2022; Felstead and Reuschke, 2021), we document more extensively inequalities in how productivity varied over time. For example, we find that workers in jobs that are less suitable for WFH reported lower productivity than before the pandemic.

[2] Throughout the paper we use the term 'location' to refer to the worker's physical location, either at home (WFH) or in a workplace away from home, such as an office or building site.

Consistent with this, and with the literature, females and low earners also reported worse productivity outcomes on average. The findings for females varied systematically with the presence of children in the house and the severity of restrictions; in fact the gap with males attenuated as the pandemic progressed. The opposite types of workers, e.g., those in the 'right' occupations and with high incomes, reported higher productivity than previously.

A particular strength of our analysis is that we incorporate external measures of both potential and realized productivity. Building on our earlier work (Etheridge et al., 2020) we examine: feasibility of home work (from Adams-Prassl et al., 2022); the need for physical proximity to others (Mongey et al., 2021), as well as realized output statistics at the industry level from the National Accounts. The sector-level correlations between our reported productivity changes and these external measures are always of the expected sign, which acts as a powerful validation of the survey data. The advantages of using individual-level reported productivity over these external measures on their own are that we can go beyond the characteristics of the job to look at the joint contribution of individual and job characteristics, as well as, for example, the role of the housing environment. An additional strength of our analysis over studies that use the same data as us (such as Deole et al., 2023) is that we go beyond using Likert-type responses and exploit the full quantitative implications of the survey: specifically we analyse in detail the answers to additional survey questions that elicit a quantitative assessment of productivity changes.

Our second contribution is to use the longitudinal aspect of our data to provide evidence on factors determining worker location as the pandemic progressed. Evidence on this front is important for understanding how preferences for WFH are continuing to evolve (Aksoy et al., 2022; Chen et al., 2023). We focus on productivity experiences and provide original evidence that workers positively selected into the home environment, based on previous productivity outcomes. In general terms, this evidence indicates that factors of production were better allocated as the pandemic progressed and provides a microfoundation for why the macroeconomy performed much better in the second lockdown than in the first. Interestingly, we also show that the marginal group - those who were most likely to change location in subsequent periods - evolved over the pandemic in intuitively credible ways. For example in the easing of September 2020, the group that were full-time WFH in June 2020 showed the greatest flexibility in whether they subsequently returned to the usual place of work, and depended most on their realized productivity in the earlier period. Alternatively in the return to lockdown in January 2021, it was the group that were part-time WFH in September 2020 that most responded to their earlier productivity outcomes: those who had previously WFH full-time naturally remained at home, however they previously performed.

Building on these results, our third contribution is to examine in detail factors affecting work performance *across* work locations. To do this rigorously we carefully formulate a selection model of location choice, and use it to estimate models separately at home and at the usual place of work. For exclusion restrictions we use pre-pandemic commuting patterns, which we show to be important in determining location during the pandemic. Here we go beyond examining the effect of standard job and individual characteristics only. We also examine the role of the home environment and, perhaps with most novelty, the role of cognitive ability and personality traits, about which the survey contains rich measures. Relating to our earlier results, we find that the productivity advantage experienced by those in 'good jobs' (in large firms, with managerial duties and high earnings) pertained particularly to the *home* environment. Those working for large firms, for example, did not fare better than those working for smaller firms while in the usual place of work. Among other results, we find that those high in agreeableness and conscientiousness performed better generally, while those with higher cognition experienced worse productivity growth while at home. We interpret this latter result as indicating that the advantage of high cognitive skills was blunted somewhat in the home environment. Overall our results provide rich insights on which factors affected productivity differentially across locations during the pandemic. These insights are useful for policy makers and planners within firms considering how to make WFH work in the future.

The paper proceeds as follows: We begin in Section 2 with a brief review of the related literature. Section 3 introduces the data, discussing how we use the questions on productivity and documenting basic trends in WFH and productivity across the pandemic. In Section 4 we investigate in further detail unequal outcomes in how productivity changes related to individual and job characteristics, as well as assessing dynamics in location choice. In Section 5 we use the selection model to examine outcomes within each location specifically. Section 6 concludes. Extensive Appendices provide further details of our analyses.

2 Related Literature

Our paper relates to three broad strands of literature. First it contributes to papers studying working from home as an 'alternative' practice. This literature has focused both narrowly on estimating the treatment effect of WFH on productivity, and more broadly on the long-term viability of WFH as a central component of working life, and its implications for labour markets and economic geography. Second our paper contributes to the literature documenting the complex movements in inequality across gender and socioeconomic groups both during and after Covid-19, as well as other recessions. Finally our estimates of outcomes by occupation and industry relate to the macro literature on sector-specific productivity changes and optimal policies during the Covid-19 pandemic.

First, how WFH impacts productivity has received increasing attention in recent years, especially since the Covid-19 outbreak, with mixed results. One approach to addressing this question has been to focus on a single inherently remotable job within a single firm. Bloom et al. (2015) study workers' productivity and attitude towards WFH using a randomized control trial of call-centre workers in a Chinese travel agency. They find that WFH led to a 13% performance increase and that, after the experiment, over half of the workers chose to switch to home-working. Recent research, however, finds more negative effects. Emanuel and Harrington (2023) examine work performance at a US call centre before and during the pandemic, using Covid-19 office closures to separately identify the impact of WFH and worker selection. Their estimates suggest that WFH has a negative impact on both the quality and quantity of output, and that home workers are negatively selected on baseline productivity. Conducting an experiment in the data-entry sector in India, Atkin et al. (2023) similarly find that randomly assigned home workers are 18% less productive than their office working colleagues. They also find a positive selection into home working, but also importantly a *negative selection on treatment effect*: those who select into the home would in fact gain most from being in the office. They explain this finding by arguing that those who are most constrained in terms of productivity at home, such as mothers, often have the strongest preference for home work. Focusing on the pandemic period, Gibbs et al. (2023) examine IT workers in Asia and also find detrimental effects of WFH. Thinking about the possible longer-term implications, Emanuel and Harrington (2023) and Gibbs et al. (2023) find that WFH is associated with a reduction in on-the-job training and coaching, which may eventually negatively impact worker productivity and worker retention.

While these papers all focus on particular narrow occupations, the Covid-19 outbreak and related lockdowns in many countries dramatically increased the prevalence of WFH in almost *all* occupations. Indeed, the above papers point towards heterogeneous outcomes across job types suggesting that the overall impact of WFH on productivity across industries/occupations/jobs requires closer investigation if we are interested in how a general shift to WFH will impact the economy. Specifically relating to these findings on productivity and selection, while we are not able to provide precise estimates of average treatment effects, our results do indicate that selection on treatment effect is, on average, *positive*. In contrast to Atkin et al. (2023) whose results come from asking workers for their own preferences, our results come from observed transitions, presumably resulting from a bargaining process between worker and employer. Overall, it seems sensible that employers would want the workers who adapt least well to WFH to return to the office.

Our results also relate to work on broader trends across the labour market. Using data similar to ours Felstead and Reuschke (2020) document the increase in WFH after March 2020. They find little effect of workers' productivity at home on average during the first lockdown. The same patterns — increasing home-working and not much change in workers' average productivity at home — are also found in Europe and North America (see Rubin et al., 2020 for the Netherlands; Eurofound, 2020 for Europe as a whole; and Brynjolfsson et al., 2020 for the US). Also using the UKHLS, Deole et al. (2023) report that average reported productivity was slightly higher at home as the pandemic progressed, but take no account of the endogeneity of work location as we do here. Complementing this evidence from individuals, Brinkley et al. (2020) provide evidence from a small survey of firms that also supports broadly non-detrimental effects of WFH during the pandemic. We go beyond these papers in providing richer evidence from across the pandemic: We use full quantitative information on productivity in the UKHLS Covid module and incorporate a wider array of evidence both from within the main UKHLS survey and from external sources.

More broadly still, the literature has begun to explore how persistent the move to home working will be, and effects on economic geography. Prior evidence indicates that most workers value the ability to WFH (Mas and Pallais, 2017). Barrero et al. (2021b) report survey evidence from individuals of their employers' stated intentions post-pandemic and find that 20% of working hours will be conducted from home in the medium term, compared to 5% pre-pandemic and a peak of around 40-50% at the pandemic's start. Their rule of thumb is that 50% of workers will be able to work an average of 2 days a week at home. Bick et al. (2021) and Felstead and Reuschke (2021) similarly provide evidence of workers' beliefs about future WFH. As the Covid-19 pandemic has drawn to a close, more direct evidence about WFH has begun to emerge. Utilizing natural language processing methods on vacancy data, Hansen et al. (2023) find that the percentage of new job postings continues to have a positive trend, with 18% of new jobs in the UK advertised as remote work in January 2023. Alternative evidence of long-term changes comes from house prices, with Gupta et al. (2021) and

Brueckner et al. (2023) finding changing patterns of inner-city and sub-urban prices, consistent with anticipated long-term shifts and Mondragon and Wieland (2022) finding an increase of remote work causing housing price increase, consistent with anticipated long-term shifts.[3] Augmenting these studies, our work provides evidence on which types of workers are most likely to persist with home working, and how this relates to, for example, housing conditions and commuting patterns.

Second our work contributes to the large literature on the complex heterogeneous effects of Covid-19, and implications for inequality that are still developing after the pandemic. Early in the Covid-19 pandemic, it was found that the economically disadvantaged groups, such as low-income groups and females, suffered larger declines in economic outcomes: for example, Adams-Prassl et al. (2020) document that female workers reported a lower ability to work from home, and also document that women were more likely to lose their jobs in the UK and in the US early in the pandemic, finding worse outcomes for lower earners. Alon et al. (2022) provide evidence that the Covid-19 recession was a "shecession" in many countries, attributing the heterogeneity to different industrial structure and variation in Covid related policies. However, patterns of inequality following the initial lockdown have been complex, and evidence is emerging that the tight labour market following the end of the pandemic has benefited low-wage workers in the US substantially (Autor et al., 2023). Our paper contributes to this strand of the literature by studying inequality of worker productivity across gender and socioeconomic groups, and across the whole of the pandemic. We find that females and mothers in particular suffered larger productivity declines during the lockdowns, but less so during the rest of the pandemic. Our work also naturally lends itself to future work assessing the role of WFH on the evolution of inequality post-pandemic.

Finally, our results can be used by the literature on sector-specific productivity of working from home, and optimal sectoral policies. Estimates of productivity changes by sector are important for macroeconomic models that try to capture the sectoral

[3] See also a survey by Garrote Sanchez et al. (2021) covering many of these issues. Additionally Gottlieb et al. (2021) assess possibilities for WFH across several developing countries.

and aggregate labor and output changes during the Covid-19 pandemic, such as that developed by Baqaee and Farhi (2022). Bonadio et al. (2021) study the impact of the Covid-19 pandemic on GDP growth and the role of the global supply chains. These papers typically discipline the labor supply shock across sectors using ex-ante measures of exposure, such as those provided by Dingel and Neiman (2020), Adams-Prassl et al. (2022), Mongey et al. (2021) or Alipour et al. (2023). However, there is space for improvement in these macro studies by using measures of realized labor productivity changes.

3 Data

We use data from the UKHLS (also known as 'Understanding Society'), a large-scale national household panel survey that covers a representative sample of UK households administered from 2009. In April 2020, the survey created the Covid-19 Study - an additional web survey fielded to collect information about survey members' experiences and behaviours during the pandemic. The Covid-19 module was initially conducted monthly from April 2020 until July 2020 and then at lower frequencies thereafter - in September and November 2020, and then in January, March and September 2021. The analysis makes specific use of the Covid-19 study waves three, five, seven and nine, conducted in June and September 2020, and January and September 2021, each of which include questions on self-reported productivity. To provide information on the early lockdown, we make use of data from the April and May 2020 waves of the Covid module. We also make extensive use of the '2019 wave' of the UKHLS main survey. This 2019 wave merges data collected in the main survey's waves 10 and 11. Additionally, we use further data from even earlier main survey waves, as discussed below.

Some background details on the UKHLS Covid-19 study are as follows: The underlying sampling frame consists of all those who participated in the UKHLS main survey's waves 8 and 9 (sampled over 2016-2018). To conduct the fieldwork, the sample was initially contacted using a combination of email, telephone, postal and SMS requests.[4] Of those eligible, and who responded to the main survey wave 8 or 9, the response rate was a little under 50%. To adjust our analysis for non-response, we use the survey weights provided. In addition, to allow for the stratification of the sample by post (zip) code, we cluster all regressions at the primary sampling unit level. For a further discussion of the Covid module and underlying UKHLS design see (Institute for Social and Economic Research, 2020).

The main variable of interest is self-reported productivity in the month of the interview and compared to a stated baseline from before the pandemic. To elicit this the survey includes some bespoke questions. Precisely, in the fifth, seventh and ninth waves (September 2020, January 2021, September 2021) all those in work are asked as follows:

> *"Please think about how much work you get done per hour these days. How does that compare to how much you would have got done per hour back in January/February 2020?"*

If the respondent did not work from home before the pandemic, then the question ends with:

> *"...when, according to what you have previously told us, you were not working from home?"*

Interviewees are then asked to respond on a Likert-type scale of 1 to 5 ranging from *"I get much more done"* to *"I get much less done"*.

Interviewees who report productivity changes to this qualitative question are asked additional quantitative questions regarding productivity changes:

> *"Would you say that what you can do in an hour now would previously have taken you:"*

If interviewees report a productivity gain, they select one choice from the following:

[4] The interviews in the fifth and seventh waves, for example, were conducted in the seven days from Thursday June 25 and September 24, with around 75% of interviews completed within the first three days.

"1 - Up to an hour and a quarter";

"2 - Between an hour and a quarter and an hour and a half";

"3 - More than an hour and a half".

If interviewees report a productivity decline, they are given equivalent choices.

In order to generate a continuous measure of productivity change, we fit a Pearson type VII distribution to these responses. We find this fits the data better than a Gaussian distribution, which does not allow for suitably thick tails (see Table A.2 for quantitative results, including goodness of fit measures). Using this fit we impute mean productivity changes for all the seven possibly banded quantitative answers. For example, for those who say that they can now do in an hour what used to take more than an hour and a half we impute a productivity increase of 78%. Full details and results of the estimation are provided in Appendix A.[5]

One important issue arises during this process. The information from June 2020 is more limited: only the qualitative question was asked, and only to those who were working from home at least some of the time. We exploit these responses by first estimating productivity change cutoffs for each of the qualitative questions in September 2020, using the shape parameters estimated from the coincident quantitative data. We then assume that these cut-offs apply equally to the June responses. Using these cut-offs we can impute mean productivity changes in the June wave for each choice category. Our estimated cut-offs imply similar conclusions for the June wave to those in Etheridge et al. (2020) where we 'semi-standardized' the data by cardinalizing the Likert responses as -2, -1 ,0, 1, 2, and scaling by the standard deviation. In that paper, we in turn also showed that similar results were given using ordered probit models. However, our approach here improves on that earlier analysis, as well as related papers (such as Deole et al., 2023), by providing fully quantified results.

[5]In the Appendix we also assess the internal validity of the data in several ways. Specifically we show that: a) the estimated cut-offs are very similar over time; b) qualitative and quantitative responses are highly correlated within waves (within groups who report positive experiences and negative experiences respectively); c) both qualitative and quantitative responses are highly correlated across waves.

Beyond the information on productivity, we make use of much auxiliary information contained in the UKHLS surveys and other sources. Of particular interest, all respondents were asked to report their baseline earnings and place of work just before the pandemic, in January/February 2020. The survey elicits industry of work both in the baseline period and currently.

An objective of our analysis is to validate our findings by making comparisons with job-level metrics obtained elsewhere in the literature, typically using data on occupation. Unfortunately, current occupation was not collected directly in the Covid survey. We therefore use occupational information from the 2019 wave. These data are based on the SOC 2000 classification. To link these to external metrics founded on the US-based O*NET classification, we use the cross-walk described in Appendix B.1. For additional validation, we also use aggregate production data from the UK Office for National Statistics; see Appendix B.2 for further discussion.

Finally, in Section 5 we make use of two additional bodies of data from the main survey collected before the pandemic. First, to examine selection into work location, we use data on patterns of commuting to work, including reports of travel mode and any travel difficulties. These were collected in main survey waves 10 (collected over $2018-19$), 8 ($2016-17$), 6, 4 and 2. To make as full use of the data as possible, we include individuals for which any of these reports is available, taking the most recent provided. Second, to examine individual characteristics potentially affecting work productivity during the pandemic, we use data on cognitive function and 'big-5' personality traits. These were collected over $2011-12$ in main survey wave 3. The cognitive assessment comprises scores from four tests - on completing number series, immediate word recall, delayed word recall, and verbal fluency (see McFall, 2013, for extensive documentation) - from which we take the first principle component. Personality traits were measured using averages of scales for responses to 3 questions for each of the big-5 traits, borrowing the methodology documented in John et al. (1999).[6]

[6]For example, to assess agreeableness, interviewees are asked to assess themselves on a scale of 1-7 on the following statements: 'I see myself as someone who is sometimes rude to others' (reverse coded), 'I see myself as someone who has a forgiving nature', and 'I see myself as someone who is considerate and kind to almost everyone'.

To give an example of sample sizes, our total number of adjusted interviews in the September 2020 wave, which is the first to provide full data on productivity, is $10,607$. Of these interviews, $5,794$ individuals were in work and reported information about working location; $5,717$ additionally answered the productivity question. Overall, we work with three main samples. The full sample, analyzed in Section 4, contains $19,293$ total observations across the four Covid waves. The sample containing information on pre-covid commuting patterns, analyzed in Section 5, contains $18,557$ person-wave observations. In Section 5 we also analyse the sample containing information on personality traits and cognition, for which $13,552$ person-wave observations are available. Full summary statistics are presented in Table C.1 in Appendix C.

3.1 Proportions in Work and at Home

Before moving on to the analysis behind our main contributions, we review patterns of working from home and reported productivity during the pandemic. Our evidence here follows up on Etheridge et al. (2020), who report findings from the first wave of data in June 2020, as well as, among others, Felstead and Reuschke (2020), Felstead and Reuschke (2021) and Deole et al. (2023).

We first show patterns of WFH over time in Table 1. It shows, in simple format, some of the characteristics of our sample and broad trends in both frequencies of WFH and productivity changes. It also shows, in the final column, the stage of the pandemic in terms of national policy on social distancing. In June 2020 and January 2021 strong distancing policies were in place, including the widespread closure of hospitality and restricted rules on even small-scale social interaction. September 2020 and September 2021 were in periods of far more relaxed rules, including, for example, availability of hospitality and restaurants.

The first row of Table 1 shows that 76% of the working-age population were in work just before the pandemic. The second column reports an estimate of the proportion of working hours spent at home. We calculate this simply by imputing 20% for those who say ‘sometimes’ and 60% for those who say ‘often’. We find that home work accounted for only around 12% of working hours prior to the pandemic, but around 38% of working hours in June 2020. The third and fourth columns show simple averages of our variable capturing change in productivity. In June 2020 this is available for the WFH sample only, and the fourth column shows that for this group reported productivity was roughly flat.

Table 1: WFH and Productivity Change During the Covid-19 Pandemic

		Proportion in work	Proportion WFH	%ΔProd	%ΔProd if WFH	Strong Social Distancing
January-February 2020	Mean	0.76	0.12			
	Sample Size	14,490	11,292			
June 2020	Mean	0.59	0.38	-0.90*	-0.90	Yes
	Sample Size	10,336	7,825	3,498*	3,498	
September 2020	Mean	0.67	0.32	5.40	8.64	
	Sample Size	9,267	6,903	5,533	2,849	
January 2021	Mean	0.64	0.40	0.08	0.69	Yes
	Sample Size	8,443	6,247	4,753	2,887	
September 2021	Mean	0.70	0.30	9.04	13.00	
	Sample Size	9,212	6,944	5,509	2,750	
Total	Mean	0.65	0.35	4.09	4.94	
	Sample Size	37,258	27,919	19,293	11,984	
	# Individuals	12,438	9,828	7,713	4,928	

Note: This table reports employment, WFH and productivity change by Covid module wave. The base sample comprises working-age individuals (17-65). The first column corresponds to the proportion of the sample in work. The second column reports the proportion of time in work spent WFH. Following Felstead and Reuschke (2021), we weight the 4 possible responses in the raw survey question as 0 = *never*, 0.2 = *sometimes*, 0.6 = *often*, 1 = *always*. The third column relates to the percentage change in productivity. In June 2020 this includes those with some WFH only. The final column corresponds to the change in productivity for those that report any WFH in the current period. The top row provides information for the baseline period, elicited using retrospective questions in the Covid module. Those on furlough or working less than one hour per work are treated as if they are out of work. The sample for the last two columns is restricted to include those with a full set of control variables (individual characteristics, employment variables and household characteristics) to be consistent with the sample used in the main analysis.
* Excludes those in the usual place of work full-time.

The second row of the middle block shows that, by September 2020, the number in employment had increased compared to June, while the proportion of hours WFH had declined. In this month, individuals reported an increase in productivity on average, and those working from home reported an increase that was even larger. The next row shows that the proportion in work decreased slightly going into the lockdown in January 2021, and unsurprisingly the proportion of hours spent working from home increased again by 8 percentage points. Notably, self-reported productivity fell again compared to the previous wave both for the sample as a whole and for those working at home. Finally, by September 2021, the proportion in work increased again, while the proportion of hours at home declined to its lowest since before the pandemic. In this month, workers reported the highest levels of productivity, indicating that they had adapted to work during the pandemic, either at home or in the office.

To show some of the wide variation during the first year of the pandemic, Appendix Tables C.2 and C.3 show breakdowns by industry and occupation respectively. Focussing on industry, the first column of Table C.2 reports baseline home work patterns in January/February, before the pandemic, and documents the proportion of workers who worked at home at least some of the time. The second column shows the proportion of workers in this category in April, at the height of the lockdown period. It shows a very large increase in the proportion working from home across almost all industries. The exceptions are industries (such as Accommodation and Food Service) for which the effect of the lockdown was seen not so much in an increase in home work, but rather widespread job losses. The third column then records the change in proportion of home workers from April to June. It shows there was little change in working patterns by this metric even as the lockdown eased. The fourth column demonstrates the change in proportion of home workers from June to September 2020 after the first lockdown was fully eased. While the remaining industries show marginal increases in the proportion spending at least some of the time WFH, significant decreases are shown in three particular industries: 'Electricity and Gas', 'Financial and Insurance', and 'Education'.

4 The Evolution of Working from Home and Productivity Through the Pandemic

4.1 Change in Productivity by Worker Characteristics

We now document in further detail variation in the self-reported changes in productivity by characteristics of the worker. Our evidence is presented in Table 2. The first column examines the relationship between productivity changes and earnings, with workers split into terciles according to take home pay across the whole labour force in the baseline period. The observations are pooled across survey waves. It seems the lowest earning group faced relatively worse productivity outcomes on average, while productivity change of top earners was roughly 5.5% more than before lockdown and at least 2 percentage points more than either of the other two groups. It is worth re-emphasizing here that, as discussed in Section 3, the productivity changes reported in this table come from the distributional imputation using quantitative and qualitative survey questions, as explained in Appendix A.

Despite the gradient by earnings, column two of Table 2 shows that on average productivity changes are not substantially dependent on degree holding itself, with both degree holders and non-degree holder showing similar increases in productivity. Although not shown here, productivity is also not noticeably different across age. The third to the sixth columns then illustrate gender gaps that differ across the stages of the pandemic and by demographic characteristics. The last two rows of this block show males and females without children, while the first two rows show those with at least one child aged under 16. In June 2020, females suffered productivity declines while males did not, with mothers suffering the most. This likely reflected the unequal burden of home work, childcare and other distractions (Andrew et al., 2020). Thereafter, in September 2020, as lockdown eased, all groups saw considerable productivity increases including women with children. Consistent with Table 1, self-reported productivity then declined broadly for most groups in the second lockdown in January 2021. Again, mothers experienced the worst reduction, experiencing a reduction in

Table 2: Percent Changes in Productivity During Covid-19 by Worker Characteristics

DV = %Δ Productivity	(1)	(2)	(3) June'20	(4) Sept.'20	(5) Jan.'21	(6) Sept.'21	(7)	(8)	(9)	(10)
Monthly net earnings terciles:										
Bottom	2.50***									
	(0.72)									
Middle	3.14***							0.49	0.56	0.59
	(0.60)							(0.93)	(0.94)	(0.93)
Top	5.51***							3.50***	2.57**	2.58**
	(0.50)							(0.98)	(1.09)	(1.04)
Education:										
No degree		3.92***								
		(0.46)								
Degree		4.33***						0.30	0.19	0.14
		(0.51)						(0.75)	(0.80)	(0.80)
Parenthood and gender:										
Parent × Female			-5.01***	6.51***	-3.46***	7.68***				
			(1.26)	(1.18)	(1.30)	(1.15)				
Parent × Male			0.36	5.24***	1.46	8.49***		0.63	0.53	0.41
			(1.34)	(0.91)	(2.09)	(1.31)		(1.18)	(1.12)	(1.12)
No children × Female			-1.48	5.06***	0.87	10.22***		1.76*	1.82*	1.44
			(1.30)	(0.87)	(1.05)	(0.69)		(0.92)	(0.93)	(0.95)
No children × Male			2.05*	5.21***	0.50	8.77***		1.25	1.81*	1.43
			(1.09)	(0.77)	(0.94)	(0.91)		(0.99)	(1.03)	(1.05)
Employment type:										
Self-employed							-0.38			
							(1.35)			
Employee							4.34***	4.39***	3.03**	3.23**
							(0.35)	(1.37)	(1.53)	(1.54)
Constant								-1.64	44.76	47.20
								(1.57)	(41.32)	(41.46)
Observations	19,293	19,293	3,498	5,533	4,753	5,509	19,293	19,293	19,293	19,293
Wave dummies								Yes	Yes	Yes
Individual controls									Yes	Yes
Employment controls									Yes	Yes
Housing controls										Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Note: The dependent variable is imputed productivity change measure. Columns (1) to (7) show group means for displayed characteristics. Columns 8 to 10 show results from multivariate regressions including additional controls. Presence of child is defined as living with a biological child which is under the age of 16. In columns (8) to (10) the omitted wave is September 2020. Additional background controls used in columns (9) and (10) are as follows: Individual controls - quartic in age, marital status, BAME status (binary), region of residence; Employment controls - managerial duties, log of the number of employee in firm of employment, industry of work, occupation; Housing controls - number of rooms in house per occupant, home ownership, whether the house has internet access. Survey weights are used throughout and standard errors are clustered at the primary sampling unit level.

productivity compared to the baseline. By September 2021, all groups were performing well, although mothers still appeared to lag the rest slightly.

More detail on parental productivity changes is provided in Appendix Table C.4, which shows that for those with the youngest children (under the age of 5), fathers performed better than mothers in June 2020, but as badly as mothers in January 2021, and substantially worse than fathers with older children during the second lockdown. This indicates that outcomes for parents with very young children equalized across the pandemic somewhat.

Moving on, the seventh column shows that employees had significantly better outcomes than the self-employed. The right hand side of Table 2 then shows the effects of these same characteristics when we combine them in a multivariate regression with and without additional controls. The first column of this panel shows the most basic specification, additionally including a constant and wave dummies only. In this column, as with the subsequent two, we have chosen as the omitted category the worst performing group in each domain. The relative sizes of most of the factors (earnings tercile, degree holding and employment status) remain similar to the raw group mean estimates. The results by gender and household composition, which are now averaged over the stages of the pandemic also confirm the impression from the left-hand side: women with children, who are the omitted category, had the worst productivity outcomes and those without the children the best. Notice, however, that the average gaps between the groups compressed considerably since the earliest estimates from June 2020, and the differences between demographic groups, when averaged across all the available waves, are only marginally significantly different.

To further put the heterogeneity in experiences into perspective, the estimate on the constant therefore implies that the worst performing group (low-skilled, low-educated, self-employed mothers) experienced an average productivity decline of around 1.5%, referenced to September 2020. By comparison, adding up the effects on the groups with the best performing outcomes implies that employed, top-earning, female degree holders, without children reported an average increase in productivity of over 8% or 10 percentage points more, on average.

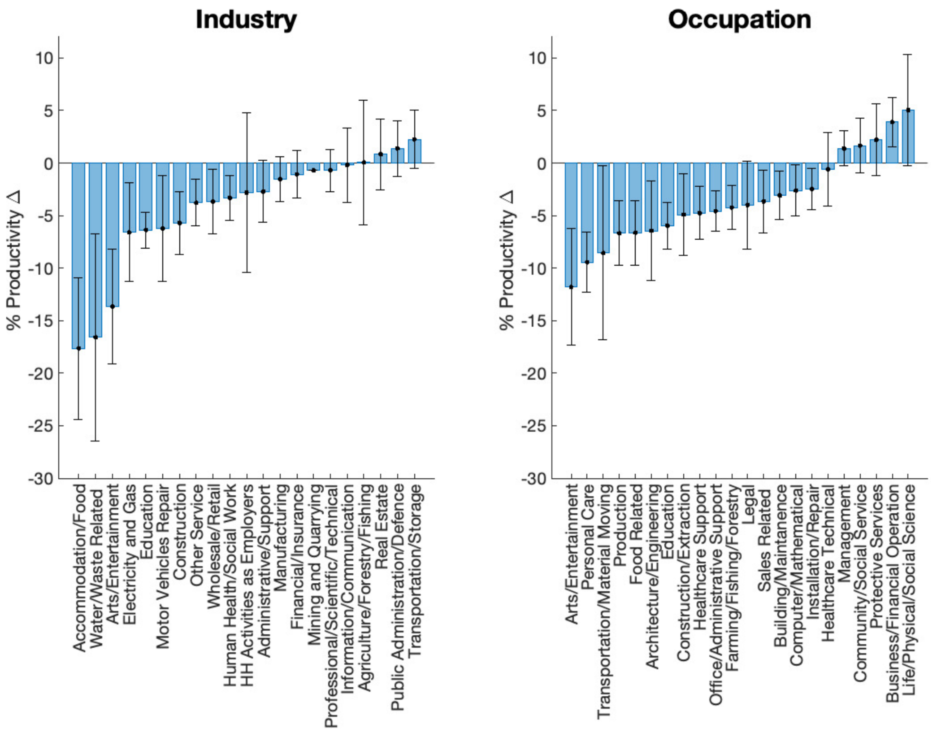
The remaining columns introduce additional controls, specifically dummies for industry, occupation, age and housing conditions. Interestingly, when we control for the housing environment, including the presence of spare rooms, a garden and adequate desk space, we find that these controls do little to explain away effects, apart from the coefficients on gender and parenthood. However it should be noted that these controls are fairly coarse, and we presume that fine occupational detail and a detailed treatment of the housing environment would explain a larger fraction of these productivity differences.

4.2 Changes in Productivity by Job Characteristics

A noticeable feature of the pandemic was differential performance and outcomes across different job types. For example, industry-specific policies were exploited during the pandemic, such as the prominent 'Eat Out to Help Out' policy instigated in the UK in August 2020, which successfully stimulated demand in the restaurant sector (Fetzer, 2022). More generally, commentators and researchers have observed the wide differential impacts by sector. Baqaee and Farhi (2022), for example, examine changes in hours by industry and show that such sector-specific supply shocks, together with demand shocks, are necessary for capturing the disaggregated data on GDP, inflation and unemployment.

We document some of this heterogeneity in Figure 1 by showing average productivity changes across industries and occupations in January 2021 compared to the baseline. Focussing on industries (left panel), the figure shows that during the second lockdown the majority of industries experienced a productivity loss compared to the pre-pandemic level. Indeed, 'Real Estate', 'Public Administration/Defence' and 'Transportation/Storage' were the only industries that exhibited productivity gains. As we show shortly, the degree of change in productivity across industries depends on job characteristics, as measured by external metrics. The ordering of industries is intuitive with the in-person services (such as 'Motor Vehicles Repair', 'Accommodation/Food' and 'Arts/Entertainment') experiencing the sharpest reductions in productivity while industries more suitable for home-work such as IT and finance sectors, performed reasonably well although the lockdown still created some productivity loss.

Figure 1: Mean Productivity Change in January 2021, by Industry and by Occupation



Note: This figure depicts the mean percentage productivity change by industry (left) and by occupation (right) from January/February 2020 to January 2021 using UKHLS Covid-19 module data. The lines correspond to the 95% confidence interval. Occupation information is taken from the 2019 UKHLS main survey responses and is converted into the 2-digit O*NET codes. See Appendix B.1 for additional details.

The right sub-plot of Figure 1 shows average productivity changes by occupation, also in January 2021. Here we take reported occupation stated in the 2019 wave of the UKHLS main survey as baseline and categorize workers using the 22 two-digit O*NET codes.[7] This panel shows that the occupation with the largest productivity increases were 'Life, Physical, and Social Science' and 'Business and Financial Operations'. These occupations require less physical contact than some of the other occupations and are relatively easier to be done at home than some other occupations, such as 'Healthcare Support'. The worst performing O*NET occupations include 'Personal Care', 'Education' and 'Arts/Entertainment'. For completeness, similar plots for June 2020, September 2020 and September 2021 can be found in Appendix C.

[7] As explained in Appendix B.1 and discussed above, the two-digit O*NET codes are derived by using a cross-walk to convert the 3-digit SOC 2000 codes contained in the UKHLS.

We next examine how our self-reported productivity changes relate to important job characteristics examined in the literature, again focusing on variation across occupations and industries. To this end, Figure 2 shows variation for January 2021 for three important metrics.

The top left sub-figure plots our measure of productivity change against average feasibility of WFH by occupation, taken from Adams-Prassl et al. (2022) who obtain their measure by asking workers to report the fraction of job tasks that can be performed from home. As such, we would expect this feasibility measure to be a key input into observed productivity during the lockdown period. Indeed we find a positive, albeit moderate correlation (weighted by occupation size) between this feasibility measure and reported productivity changes, $corr = 0.48$.

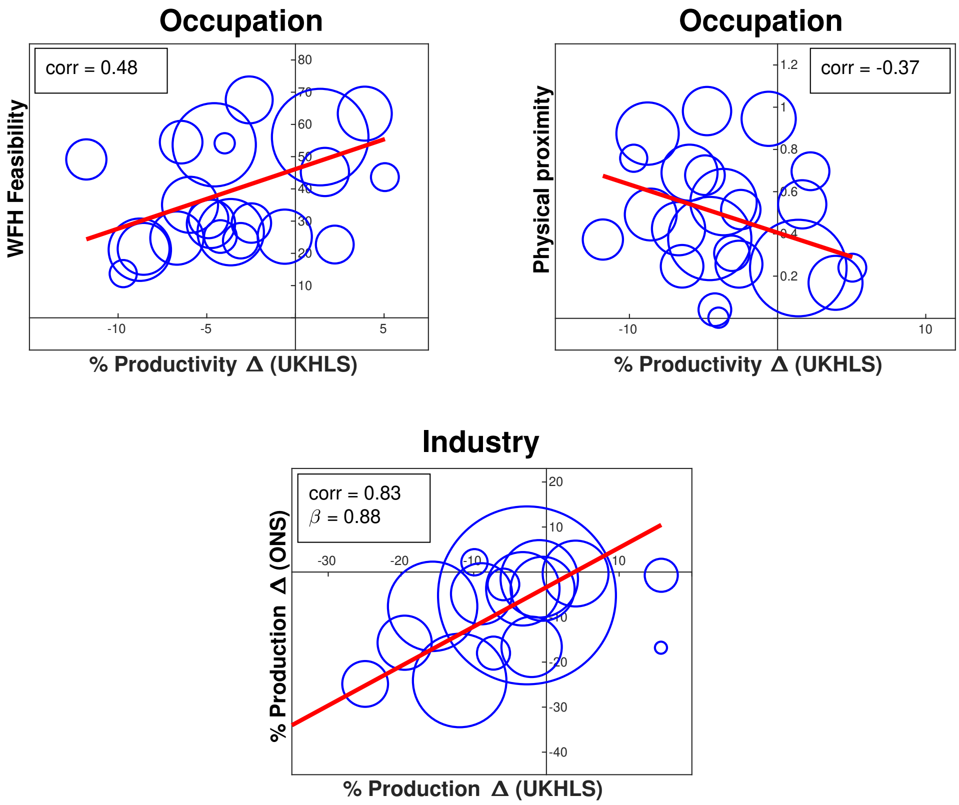
The top right sub-figure plots our self-reported productivity change against a measure of need for physical proximity with others, derived by Mongey et al. (2021), again using occupational O*NET descriptors. We expect a negative correlation between change in productivity and the need for physical proximity if our measure is capturing a similar underlying trait of occupations. Indeed, those occupations which are indicated to require close physical interaction between workers, such as 'Personal Care' and 'Arts and Entertainment' show the largest productivity declines during the lockdown. In fact, the correlation here is -0.37, indicating that individual productivity is just as much affected by this factor as pure feasibility of home work.

The bottom sub-figure compares our measure of productivity against aggregate output (value added) data from the ONS, which is provided at a relatively coarse industry division code level. For this plot we aggregate our individual level measure of productivity change into an implied sectoral-level change in total *output*, additionally using data on employment size and individual-level earnings and hours levels. We use output level rather than industry-level change in *productivity*, because a comparison with output change is in fact more straightforward to implement. We discuss this issue in further detail in Appendix B.2, where we show the calculations used to make either comparison.

This subplot shows that, in January 2021, the two measures have a strong correlation of 0.83. We also report that the beta on a weighted regression is 0.88, showing that the measures line up strongly in terms of quantitative magnitudes. We consider this relationship as remarkably strong given that there remain a few conceptual differences between our aggregated measure of output change and the change in sectoral output from the national statistics: in particular the measure on the horizontal axis accounts only for real productivity experienced by employees, while, for example, changes in profits due to shifts in output prices may also be important to changes in output at the sectoral level.

For completeness, we show the full set of comparable plots for each of these three measures additionally for June 2020, September 2020 and September 2021 in Appendix C, figures C.5, C.6 and C.7, with similar implications.

Figure 2: External Validation of Productivity Change Data for January 2021



Note: This figure depicts scatter plots of the UKHLS productivity change measure against alternative measures related to WFH used in the literature. The top two sub-figures compare the measures by occupation and the bottom sub-figure by industry. Bubble sizes are proportional to occupation/industry employment. The straight lines are the (weighted) lines of best fit. All statistics are weighted by employment. The top left sub-figure plots the UKHLS mean productivity change by occupation against the average WFH feasibility measure from Adams-Prassl et al. (2022). The top right sub-figure plots UKHLS mean productivity change by occupation against the measure of physical proximity from Mongey et al. (2021). The bottom sub-figure plots the UKHLS percentage change in output by industry against the ONS percentage change in output measure. For a discussion of the aggregation process see Appendix B.2. UKHLS occupation information is taken from the 2019 UKHLS main survey responses and is converted into the 2-digit O*NET codes. See the main text and Appendix B.1 and Appendix B.2 for a fuller discussion.

4.3 The Dynamics of Location Over the Pandemic

Section 3 showed that the proportion of hours spent WFH waxed and waned during the pandemic as various restrictions were tightened and relaxed. We have also shown that productivity during the pandemic varied systematically by characteristics of the individual and of the job. An interesting and natural question, therefore, is whether productivity experiences influenced location decisions as the pandemic progressed. We explore this question here.

To do this, we run dynamic regressions of the choice of location at time t during the pandemic on current characteristics, as well as past location outcomes. We additionally interact these past location outcomes with reported productivity change. The idea is that this interaction picks up the possibility of positive selection into WFH over time. When individuals were exposed to WFH early in the pandemic, those who reported productivity increases since the baseline should be more likely to continue WFH when restrictions were lifted in the autumn of 2020: Presumably both individuals would be more persuasive in asking for continued WFH, and firms would be more happy to carry on the arrangement. Likewise, those who reported productivity declines early in the pandemic would be more likely to be brought back into the workplace.

Table 3 reports the results of this exercise. Each column shows the estimates of a multinomial logit model of WFH in a separate wave of data, with successive addition of controls. The first column shows results for September 2020 with a full set of demographic controls, but not yet controlling for job or housing characteristics. Here the lagged observations of WFH come from June 2020 when, recall, we observe productivity outcomes only for those at least sometimes at home, and not those who remained full-time in the workplace. Our base omitted category in the lagged period is those who 'sometimes' or 'often' (which we refer to as 'part-time') WFH. Our prior belief is that this group is generally most likely to be the margin of moving between work locations. However, as we shall see, the group most on the margin differs from period to period.

The first column shows while there is a positive estimated coefficient on productivity for those who were part-time WFH in June 2020, it is not statistically significant. Neither is the full-time group significantly different from this part-time group. However, the bottom of the table shows that the marginal effect for those working full-time at home in June 2020 ('Sum: (1) + (3)') is 0.92 and is statistically significant. This implies that for those full-time at home in June there was a strong effect of reported productivity on later work location. This is intuitive: as restrictions were lifted, those who were full-time at home often had varied options of location in September. Their employers may have required them to come into work or kept them at home depending on the most productive outcome. The second column shows that this relationship remains when employment and housing controls are included.

The middle rows of Table 3 also show the pure effects of WFH status in the baseline period, from just before the pandemic, and in the previous period. All of these estimates have the expected sign. As seems intuitive, lagged WFH is much more important in predicting WFH status in September 2020 than the baseline WFH status. The point estimates imply that, conditional on full controls, an individual who was otherwise marginal and who was at home in June 2020 was 20 percentage points more likely to WFH in September than someone who was previously in the workplace.

The third and fourth columns of Table 3 show results when we examine location choice during the second main lockdown in January 2021. In this set of regressions we can now also examine not only those who WFH part-time or full-time in September 2020, but those who never worked from home in this preceding period. The evidence presented in these columns is overall weaker, but again we see some revealing patterns. The top row of column four shows that, when we include a full battery of controls, there is some evidence that subsequent work location depended on productivity experiences for those who were part-time at home in September 2020: those who performed better were more likely to be at home in January 2021. On the other hand, for the other groups (full-time or never) there is no evidence of any effect of productivity. The contrast with June-September 2020, however, is important. Compared to that previous interval, as the economy transitioned back into lockdown in January 2021 then those who were full-time WFH in September 2020 were no longer marginal candidates for location choice, and their productivity experiences were no longer important. In fact, and although not shown explicitly in the table, among the group who WFH full-time in September 2020 we see very little variation in location outcomes in January 2021, which explains the larger standard errors.

Finally, we examine the interval from January 2021 to September 2021. Again the difference in results compared to the earlier intervals is instructive. Now the stand-

Table 3: Dynamics of WFH: Effect of Past Productivity Outcomes

DV= WFH_t	Sept. 2020	Sept. 2020	Jan. 2021	Jan. 2021	Sept. 2021	Sept. 2021
(1) $\Delta Prod_{t-1}$	0.43	0.17	0.39	0.54*	0.10	0.07
	(0.30)	(0.32)	(0.32)	(0.32)	(0.41)	(0.42)
(2) $\Delta Prod_{t-1} \times WFH_{t-1} = \text{No}$			-0.58	-0.71	-1.46**	-1.57**
			(0.67)	(0.59)	(0.63)	(0.67)
(3) $\Delta Prod_{t-1} \times WFH_{t-1} = \text{Full-time}$	0.49	0.56	-1.10	-1.13	0.25	0.29
	(0.44)	(0.47)	(0.92)	(0.82)	(0.49)	(0.51)
WFH_{base} = No	-0.87***	-0.84***	-0.54***	-0.49***	-0.51***	-0.44***
	(0.10)	(0.12)	(0.15)	(0.16)	(0.13)	(0.14)
WFH_{base} = Full-time	0.58*	1.11***	-0.50	-0.48	0.45	0.54
	(0.34)	(0.36)	(0.54)	(0.59)	(0.32)	(0.34)
WFH_{t-1} = No			-1.61***	-1.68***	-2.08***	-2.17***
			(0.16)	(0.17)	(0.25)	(0.25)
WFH_{t-1} = Full-time	2.66***	2.38***	2.89***	2.77***	0.75***	0.74***
	(0.12)	(0.14)	(0.28)	(0.28)	(0.18)	(0.18)
Sum: (1) + (2)			-0.19	-0.16	-1.36***	-1.50***
			(0.59)	(0.51)	(0.48)	(0.53)
Sum: (1) + (3)	0.92***	0.73**	-0.71	-0.59	0.35	0.36
	(0.33)	(0.35)	(0.87)	(0.77)	(0.30)	(0.30)
Observations	2,789	2,789	3,845	3,845	3,435	3,435
Lagged WFH status (full set)	Yes	Yes	Yes	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes	Yes	Yes	Yes
Employment controls		Yes		Yes		Yes
Housing controls		Yes		Yes		Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

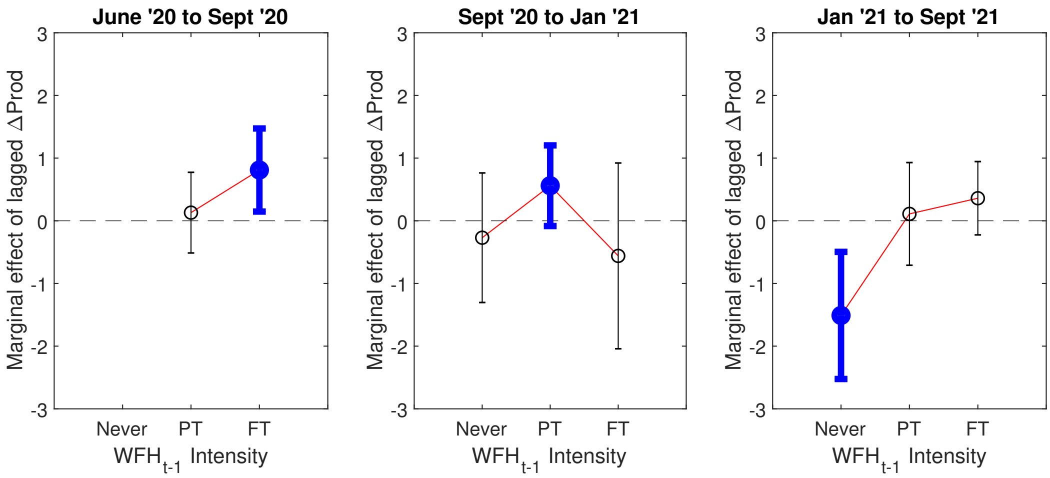
Note: This table reports the estimates of an ordered logit model. The dependent variable is a trichotomous WFH variable valued 0 if never WFH, valued 1 if WFH part-time (sometimes or often) WFH, valued 2 if WFH full-time (always). The omitted category for lagged dependent variables on the right hand side is part-time WFH. The background control variables used are the same as those in the final column of Table 2, together with a full set of indicators for lagged WFH status. Survey weights are used throughout. Standard errors are clustered at the primary sampling unit level.

out estimate is for those who were in the workplace in January 2021 (WFH_{t-1}=No). For these individuals, those who were more productive in the office were more likely to stay there and *less* likely to return home. In terms of quantities, for an otherwise marginal worker, being 10 percentage points more productive in the office translates to a 3 percentage points higher chance of staying away from home.

We view this 'negative' result for those not at home at all as a good test of our framework. To add to this, we hypothesize that for those not at home at all in June 2020, the effect of productivity experiences on subsequent WFH status would also be strongly negative. Unfortunately, however, the data are not available to test this.

The marginal effects from Table 3 (given by rows '(1)', 'Sum: (1) + (2)' and 'Sum: (1) + (3)') are also shown in Figure 3. We see clearly, and as just described, that the strongest effects on subsequent WFH status are for those who were full-time at home in June 2020 (positive effect) and never at home in January 2021 (negative effect), with some evidence of positive effects for those part-time at home in September 2020 and going into the subsequent lock-down.

Figure 3: Marginal Effect of Lagged $\Delta Prod$ Across Lagged WFH Status



Note: The above figures plot point estimates, together with 95% confidence intervals, of the marginal effects of lagged change in productivity, by lagged WFH status, on current WFH status. The raw estimates can be found in the second, fourth and final columns of Table 3. Solid and bolded points show effects that are significant at the 10% significance level. See text for more details.

We finish this section by examining again the middle rows of Table 3. The table shows that lagged and baseline WFH status continued to have a strong effect on current WFH status throughout the pandemic, even conditional on labour market controls such as industry and occupation and housing controls. In line with our main point, the coefficient on lagged WFH status also reflects the accumulation of previous experiences in specific work locations. As a nuance, it is worth noting that during the second lockdown of January 2021, baseline WFH status was less important in determining the location of work. For example, the coefficient on $WFH_{base}=Full\text{–}time$ is negative, and almost identical to $WFH_{base}=No$. This is not only compared to within-pandemic lagged WFH status, but also compared to the effect of baseline WFH status on locations in September 2020 and September 2021.[8] Clearly, in these periods of eased

[8] Although not shown here, the coefficient on $WFH_{base}=Full\text{–}time$ is significantly different from that on $WFH_{base}=No$ at the 1% level in September 2021.

restrictions baseline WFH status was more indicative of workers' propensity to be at home.

5 Factors Affecting Productivity Across Locations

5.1 Empirical Framework

Section 4 showed that productivity changes since before the pandemic have varied systematically by individual characteristics, household circumstance and, importantly, characteristics of the job. While this evidence provides important insights into unequal outcomes during the pandemic, and the evolution of WFH, it doesn't answer perhaps the key questions for individuals, businesses and policy makers. These include: what is the effect on productivity of WFH, and how does this depend on these characteristics? These are the raw questions addressed by Bloom et al. (2015), Atkin et al. (2023), Emanuel and Harrington (2023), and Gibbs et al. (2023). We now discuss our approach to answering these questions using a simple model of self-selection as in Heckman (1979) and French and Taber (2011). Intuitively to obtain selection-free estimates of key parameters we exploit instruments that affect preferences for work location during the pandemic but do not affect productivity. As we shall see in our application, unfortunately our empirical setting does not provide precise estimates of average treatment effects, but we can provide empirical rigour in identifying and estimating the marginal effect of characteristics across locations. In this way we contribute new evidence that is missing from studies that focus on narrower subsets of the population.

We lay out a full empirical framework in reasonable detail in Appendix D. Here we provide an intuitive discussion of the approach and discuss in further detail the elements that are non-standard. In particular, when considering selection into home/workplace, it is the difference in contemporaneous productivity across work locations that matters, but, in our data, we only observe productivity *changes*. Here we show that the model can be re-stated in terms of productivity changes naturally.

Our basic setup is as follows. Let productivity in each setting be given by:

$$\begin{aligned} prod_{it}^{h} &= g^{h}\left(X_{it}\right) + \epsilon_{it}^{h} \\ prod_{it}^{f} &= g^{f}\left(X_{it}\right) + \epsilon_{it}^{f} \end{aligned} \tag{1}$$

such that $prod_{it}^{j}$ is productivity in some suitable units (e.g. the logarithm of monetary units per hour), for individual i at time t, during the pandemic, in location j, with $j \in \{h, f\}$ denoting WFH or working from the office, respectively. X_{it} captures the bulk of characteristics that are relevant in either or both work locations, and which could be time-varying, such as work sector or infection status, or fixed, such as education, baseline WFH status, or the presence of children. ϵ_{it}^{j} is an unobserved mean-zero disturbance capturing idiosyncratic factors in each location.

Now define the extra utility effect of WFH compared to being located in the usual workplace as:

$$V_{it}^{h} = k\left(z_i, X_{it}\right) + \nu_{it} \tag{2}$$

where, importantly, z_i captures individual characteristics that affect utility but *not productivity* and v_{it} captures unobserved disturbances. The existence of z_i is key for identification. It is worth emphasizing that this model allows for decision making about location equally by the firm as much as by the individual. We use the term 'utility' broadly to capture all these factors, which might include strong employer preferences (even requirements) to be at home or in the office.

Given this set-up the decision rule is simple, individuals choose to work from home if there is an overall net gain in terms of productivity and utility. This is specified as:

$$j_{it}^{*} = \begin{cases} h & \text{if } prod_{it}^{h} - prod_{it}^{f} + V_{it}^{h} > 0 \\ f & \text{otherwise} \end{cases} \tag{3}$$

where j_{it}^{*} denotes the optimal work location choice for individual i at pandemic time t.

As mentioned above, in our data we only have access to productivity change information relative to a common baseline period. Therefore, to fit the data we have available, we next define *quasi*-differences in productivity as follows:

$$\tilde{\Delta} prod_{it}^{j} \equiv prod_{it}^{j} - prod_{i0}^{j^*}$$

This, importantly, captures the change in productivity at time t in each location j compared to the *observed* location j_0^* at time zero. Pre-pandemic work location is treated as given. The model could be enriched in this regard, but this would require additional instruments and is thus not pursued here. See Appendix D for further discussion.

Building on (3), it's the case that:

$$\begin{aligned}
& prod_{it}^{h} - prod_{it}^{f} + V_{it}^{h} > 0 \\
\iff & \left(prod_{it}^{h} - prod_{i0}^{j^*}\right) - \left(prod_{it}^{f} - prod_{i0}^{j^*}\right) + V_{it}^{h} > 0 \\
\iff & \tilde{\Delta} prod_{it}^{h} - \tilde{\Delta} prod_{it}^{f} + V_{it}^{h} > 0.
\end{aligned}$$

Therefore,

$$j_{it}^{*} = \begin{cases} h & \text{if } \tilde{\Delta} prod_{it}^{h} - \tilde{\Delta} prod_{it}^{f} + V_{it}^{h} > 0. \\ f & \text{otherwise.} \end{cases} \tag{4}$$

Thus we can rewrite the decision rule for location during the pandemic in terms of the quasi-differences, lending itself naturally to the data on productivity changes that are available.

In terms of identification, we observe j_{it}^{*}, $\Delta prod_{it} \equiv \tilde{\Delta} prod_{it}^{j^*}$ and the full array of covariates, including instruments z_i that affect the selection rule, but do not affect productivity. With these we can identify factors that affect productivity changes across locations. Again see the Appendix D for a more formal discussion.

Our candidates for instruments are variables affecting travelling to work in the pre-covid period: mode of travel, distance from work and reported travel difficulty. Our arguments for using these are twofold. First, we rely on a temporal argument: these variables are determined prior to the pandemic, and so they are not endogenous to work choices and outcomes during the Covid-19 outbreak. Second commuting difficulty should *prima facie* not affect productivity in any working location. As we will see, however, these variables clearly affected location choices.[9] One obvious reason for this is that how an individual travels to work impacts their exposure to infection and so their willingness to work away from home. As a final point, note that these variables are clearly not available for those who WFH full-time prior to the pandemic. We therefore exclude this 5% of the population, and base our conclusions on the sub-population of the workforce who previously worked away from home at least part time.

5.2 How Did Productivity Vary Across Work Location?

We now implement the selection framework presented above using a standard two-stage Heckman procedure. We start by presenting the first-stage probit regression of location choice on individual, employment and housing characteristics, and our excluded variables, the results for which are shown in Table 4. Here, and for the remainder of this section, we use a binary outcome for location choice, combining as the WFH group those who are at home 'always' or 'often', and as the non-WFH group those who report 'sometimes' or 'never'. Across the dataset this splits the sample roughly in half.

The first column of Table 4 shows results for a model which includes pre-covid mode of transport interacted with distance to work. Most saliently, and as suspected, it shows that pre-pandemic commuting distance is strongly related to within-pandemic WFH for those who previously used public transport. It is likely that, for these workers, alternative routes to work were less available, and the danger of infection in transit

[9] As discussed, we treat baseline WFH status as given, or exogenous in the model. In effect we argue that outcomes in the pre-pandemic period do not depend on commuting mode. As discussed in Appendix D our formal argument for this is that idiosyncratic productivity disturbances do not vary across work locations in the baseline period. Intuitively, the argument is that factors affecting productivity across locations before the pandemic were not nearly so heterogeneous, so selection issues are not such a concern.

was higher. On the other hand, distance does not seem so important for car users or users of other modes (mainly walking or cycling).

In the second column, we omit distance from work but include a binary indicator for reporting pre-covid travel difficulties. This indicator is only applicable to those who travelled by car or public transport. Overall the results show that, of all the groups, those who previously commuted by car, and without difficulties (the omitted category), were the most likely to continue visiting the workplace, and significantly more so than those who walked or cycled (see the 2nd row). Those who previously travelled to work by car and *did* have travel difficulties were also significantly more likely to WFH during the pandemic than the base group. This result suggests that commuting by car didn't become much easier during the pandemic, and that those whose commute was difficult took the opportunity to WFH when it was presented to them.

Finally, in the rightmost column, we include all of our instruments. Most of the insights remain, except that the role of travel difficulties is less significant when controlling for distance to work. Looking at the bottom of the table, we also notice that the chi-squared statistic on the excluded instruments is high across specifications indicating that these instruments have good explanatory power.

We now use these exclusion restrictions to explore factors that affect productivity, both at home and in the office. Of particular interest are the range of characteristics, such as features of the home environment, that are provided in the UKHLS survey, but difficult to find evidence on elsewhere. Results are shown in Table 5, where, as discussed above, we combine those who are 'always' or 'often' at home into the WFH group. It shows a range of factors across both locations, for two main specifications. The first two columns correspond to the broadest sample available. In all the regressions shown we use extensive controls, including for age, education, together with occupation and industry dummies. In Table 5 we report results for those factors which have previously been shown to be generally important to productivity during the pandemic, or which we think *a priori* might affect productivity differentially across locations.

The first three rows of Table 5 show the role of key and relevant individual characteristics. Columns 1 and 2 show that, as might be expected, parenthood had a negative effect on productivity while WFH, but not on productivity in the workplace. The third column, which shows p-values on the differences between the first two columns, confirms this conclusion. The second row then shows the coefficient on a gender dummy. Here males reported relatively better productivity outcomes than females when in the office. Given that we control extensively for job and demographic characteristics we find this result somewhat surprising. Nevertheless, it may reflect the fact that the workplace environment changed substantially during the pandemic, and this affected different types differentially. We return to this point later in the discussion. Finally, among individual characteristics, we examine the effect of BAME status, for which unequal outcomes have been documented elsewhere during the pandemic (e.g. Crossley et al., 2021). Here, however, we find no evidence of differential productivity outcomes.

The next block of rows of Table 5 show the roles of job characteristics. Concentrating still on the results presented in columns 1 and 2, we find that those with managerial duties performed better than those without while in the home environment. Column 2, however, shows that this difference was not apparent in the workplace. These results suggest that managerial duties were positively impacted by the enforced introduction of remote working technology. The second row in the block echoes the findings from Table 2, and shows that the self-employed performed particularly badly away from the home, although the difference compared to the home environment is not significant. Moving on, the third row shows that those working for larger firms performed better at home than those working for smaller firms, and that this gap was significantly smaller in the workplace. This result confirms the natural suspicion that large firms were better able to adapt to a home working environment. Finally, we re-examine the association of productivity with position in the earnings distribution, shown previously in Table 2, where we documented that those in the top tercile of the earnings distribution performed significantly better than those on the lowest wages. The point estimates suggest that those with top earnings performed better than those in the bottom tercile when WFH, but overall, we lack the power to say anything more conclusive here. Nevertheless, in combination, the overall impression from the second block is that those in good jobs, with managerial duties, high earnings and working for large firms, enjoyed an advantage while WFH, and that, among employees, fewer differences arose in the workplace.

Rows 9-12 show three characteristics of the housing environment. Aside from providing substantive insights, these characteristics provide a validation of the data and framework, because they should not affect outcomes in the workplace. Indeed, column 2 shows that none of these characteristics are significant at the 10% level away from the home. In terms of the home environment, we find that the size of the house, as measured by rooms per person, actually had no noticeable effect on productivity. We next examine the presence or not of broadband connection. The prior here is of course that a good internet connection was crucial for home working (Barrero et al., 2021a). The point estimate on broadband is indeed large, but the proportion of people who report *not* having broadband is in fact tiny, and so the precision on this estimate is very low. Finally, we examine the effect of having deskspace for all members of the household who need it, which seems to have a substantial association with productivity changes when WFH. Of course, we should not overstate this result given that it is measured during the pandemic. Nevertheless, it does show that this is the type of characteristic blamed by those with adverse productivity experiences.[10]

We also report outcomes for those who previously had experience of WFH. Interestingly, we find no strong evidence that they performed better at home than those who were never at home just before the pandemic. Finally, at the bottom of the table we also report the coefficient on the inverse mills ratio, capturing the strength of selection. Although results here are not strong, the coefficients are of the anticipated signs

[10] Respondents were asked: 'Thinking about everyone in your household who is currently working from home or home schooling. Does everyone have their own quiet space at a desk or table to work at?'

and the point estimate in column one has a p-value of 0.11 (not shown). This is consistent with the message from Section 5 that selection into work location is important.

The right-hand side of Table 5 then shows effects when we include extra information on individual characteristics. Specifically we include measures of personality traits that have been found to relate strongly to outcomes during the pandemic, mainly in terms of mental health (See, for example, Proto and Zhang, 2021). It seems plausible that workplace performance has a role in this relationship. These measures of traits were collected in wave 3 of the main UKHLS survey, around a decade before the pandemic. We also include a derived cognitive test score from the same wave, that may also impact outcomes. We make additional use of the cognitive test score by trimming the bottom 5% of the score distribution in our base sample, in line with recent evidence that those with low scores are not able to formulate precise answers to the type of question we assess very well (D'Acunto et al., 2022). Accordingly, the sample size when using these data is somewhat smaller than in the results shown previously. In particular, the sample now includes very few individuals under age 30, for whom the cognitive tests and personality questionnaire was not administered.

The upper rows of the right-hand side of Table 5 repeat results for those characteristics shown on the left-hand side. Reassuringly, results are highly similar and only in a couple of instances do the reported levels of significance change.

Turning to the cognitive score, we see that cognitive function, as measured by the first principle component from a battery of cognitive tests, did not impact outcomes in the workplace. However, the fourth column shows that those with higher cognitive function had *worse* outcomes while at home. Given that we control extensively for occupational and industrial characteristics, we interpret this not in terms of the type of work that more intelligent individuals perform, but rather that, for a given work task, the advantage that higher cognitive function confers was dampened while WFH.

Focusing next on the effect of traits, we see that the most noteworthy results are for agreeableness and for conscientiousness. As background to the discussion it is worth noting first that conscientiousness is reliably shown to be strongly positively associated with earnings *level* (Almlund et al., 2011; Prevoo and ter Weel, 2015): It captures facets such as industriousness and orderliness that promote high productivity and the accumulation of human capital (Gensowski, 2018). Here, we find positive point estimates on productivity changes in both working environments, even if the estimate is significant only in the workplace itself. Overall, this result indicates that those high in conscientiousness were better able to adapt to a working landscape that was rapidly changing. Indeed, and although not shown here, an average of the two coefficients from the fourth and fifth columns is significant at the 5% level.

Among the other traits, agreeableness is also typically shown to be associated with earnings, but *negatively* (Mueller and Plug, 2006): The polar opposite of agreeableness is disagreeableness, which is aligned with competitiveness (Almlund et al., 2011), and which has been shown to be predictive of labour market success (Reuben et al., 2015). Interestingly, however, we find that agreeableness is associated with significantly better outcomes during the pandemic in both home and workplace environments. One interpretation of this result therefore, is that the conditions which enable better outcomes for those who are more competitive, such as proximity to colleagues, were absent, and those with softer interpersonal styles were better able to adapt to new ways of interacting.

To conclude this section, we provide an estimate of the treatment effect of WFH on productivity. As discussed above, this is a key parameter that has been the subject of recent work, such as in Bloom et al. (2015). However, as also discussed previously, the breadth of our empirical setting and data do not suit a precise analysis. Nevertheless, we present results in Table 6. Recall first that Table 1 showed a naive comparison of means indicating that WFH correlated with better productivity growth during the pandemic on average. Pushing this further, the first column of Table 6 shows OLS results when adding background controls. It shows that the estimate on WFH remains positive and highly significant. The second column presents the estimates of a model with individual fixed effects, and therefore examines effects for those who move in and out of the home. For this group, the positive effect of WFH disappears. The final column shows the IV estimate, for which precision is noticeably reduced. In the context of our instruments, it indicates little evidence for a positive treatment effect. To say anything more conclusive, however, would require larger sample sizes or a research design which provides more power, such as examining a narrower set of occupations.

Table 4: First Stage Estimates, WFH During Covid-19 and Pre-Pandemic Commuting Patterns

DV = WFH_t	(1)	(2)	(3)
Commuting mode (Base = Car)			
Public	-0.05	0.06	-0.04
	(0.08)	(0.09)	(0.09)
Other	0.23*	0.27***	0.27**
	(0.12)	(0.10)	(0.12)
Distance to work (Car)	0.02		0.01
	(0.02)		(0.02)
Distance to work × Public	0.23***		0.22***
	(0.05)		(0.06)
Distance to work × Other	-0.01		0.00
	(0.05)		(0.05)
Travel difficulties (Car)		0.09**	0.09*
		(0.05)	(0.05)
Travel difficulties × Public		0.11	0.01
		(0.12)	(0.13)
Observations	18,557	18,557	18557
χ^2 on displayed variables	29.09***	14.80***	34.23***
Wave dummy	Yes	Yes	Yes
Lagged WFH status	Yes	Yes	Yes
Individual controls	Yes	Yes	Yes
Employment controls	Yes	Yes	Yes
Housing controls	Yes	Yes	Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Note: This table presents estimates of a probit model of a WFH binary (= 1 if WFH full-time or WFH often) on the instruments displayed, and the same controls as those listed in Table 5. Distance to work is measured in the 10s of miles. Individual controls include region of residence, degree status, quartic age variable, earnings tercile, whether have a child under the age of 16, BAME status (binary), marital status and sex. Employment controls include occupation, industry, log of the number of employees in the firm the individual works for, whether the individual has managerial responsibilities, and whether the individual is self-employed. Housing controls include the number of rooms per person, home ownership binary variable, internet access, and whether everyone who works from home has sufficient desk space. Survey weights are used throughout. Standard errors are clustered at the primary sampling unit level.

Table 5: Productivity Changes by Location: Controlling for Selection

DV = % Δ Productivity	WFH	Not WFH	p-value on difference	WFH	Not WFH	p-value on difference
Demographics						
Parent	-3.07**	0.07	0.02	-2.51*	-0.63	0.03
	(1.22)	(0.96)		(1.42)	(0.98)	
Male	-1.59	1.97**	0.01	-2.12*	1.75	0.00
	(1.16)	(0.86)		(1.34)	(1.03)	
BAME	-0.39	1.46	0.22	0.37	-0.79	0.12
	(1.96)	(1.42)		(2.24)	(1.60)	
Job Characteristics						
Managerial duties	2.99**	-0.24	0.02	4.15***	-0.40	0.00
	(1.22)	(0.84)		(1.33)	(0.98)	
Self-employed	-3.94	-5.14***	0.36	-0.92	-3.94**	0.00
	(2.89)	(1.73)		(3.01)	(1.97)	
Log size of firm	0.91**	-0.04	0.01	2.59***	0.34	0.01
	(0.36)	(0.24)		(0.81)	(0.58)	
Monthly net earnings: Middle tercile	0.69	0.58	0.48	0.19	-0.51	0.24
	(2.20)	(1.02)		(2.37)	(1.09)	
Monthly net earnings: Top tercile	3.22	0.71	0.01	2.63	-0.71	0.00
	(2.15)	(1.34)		(2.34)	(1.47)	
Housing characteristics						
Number of rooms in home, per person	0.85	0.54	0.38	0.69	0.86*	0.43
	(0.72)	(0.40)		(0.75)	(0.45)	
Home has internet access	8.82	4.45	0.00	6.23	6.11	0.45
	(7.61)	(4.37)		(8.54)	(4.51)	
All who WFH have desk space	4.82***	0.40	0.00	4.84***	-0.54	0.00
	(1.57)	(0.94)		(1.75)	(1.11)	
Baseline WFH						
Often/Sometimes	3.33	1.85	0.07	3.17	2.76	0.34
	(2.27)	(1.98)		(2.18)	(2.28)	
Cognition & Pers. Traits						
Cognition				-1.78**	0.01	0.04
				(0.70)	(0.48)	
Agreeableness				1.28**	0.86**	0.34
				(0.65)	(0.43)	
Conscientiousness				0.65	0.87**	0.41
				(0.60)	(0.44)	
Extraversion				0.54	-0.67	0.11
				(0.64)	(0.43)	
Openness				0.40	0.54	0.44
				(0.72)	(0.44)	
Neuroticism				-0.71	-0.46	0.40
				(0.65)	(0.41)	
$\widehat{\text{Inverse}}$ Mills	-4.52	2.55	0.00	-3.50	1.28	0.00
	(2.86)	(2.98)		(2.64)	(3.48)	
Observations	8,873	9,684		6,649	6,903	
Wave dummy	Yes	Yes		Yes	Yes	
Region of residence control	Yes	Yes		Yes	Yes	
Occupation and industry controls	Yes	Yes		Yes	Yes	
Additional individual controls	Yes	Yes		Yes	Yes	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Note: This table presents estimates of OLS regressions of percentage change in productivity controlling for selection effects. The columns headed by ”WFH” contains estimates when using the sub-sample of individuals who reported WFH as ”always” or ”often”. The columns headed by ”Not WFH” contain estimates when using the sub-sample of individuals who reported WFH as ”sometimes” or ”never”. Additional individual controls include: age up to and including the fourth power, marriage dummy, degree dummy, whether home is owned. when estimating the model controlling personality traits, the sample is trimmed at the bottom 5% of cognitive scores, corresponding to a threshold standardized score of -1.5. Survey weights are used throughout. Standard errors are clustered at the primary sampling unit level.

Table 6: Effect of WFH on Productivity Change

DV = % Δ Productivity	OLS	FE	IV
WFH_t	4.11***	-0.27	4.65
	(0.92)	(1.14)	(13.29)
Observations	18,557	18,557	18,557
Background controls	Yes	Yes	Yes
Individual fixed effects		Yes	
Commuting instruments			Yes

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Note: The table presents estimates of various models specified by the column titles. The dependent variable is productivity change (percent) since the baseline period. Background controls are those reported in Table 4. Survey weights are used throughout. Standard errors are clustered at the primary sampling unit level.

6 Conclusion

Across the world, the Covid-19 pandemic caused widespread disruption to working practices, including, most saliently, a vast increase in working from home (WFH). This increase in WFH seems certain to persist beyond the end of the pandemic. This change has important implications for labour markets and economic geography and raises many questions on which answers are still needed. Most pertinently, it is important to understand which types of workers perform well at home, and why, and what factors determine workers' choice of location.

In this paper we investigate these issues using representative panel survey data from the UK, spanning the pandemic. These data contain both information on workers' current working location as well as detailed reports on changes in their productivity since before the pandemic's onset. The survey also contains a host of additional information on individuals, their jobs and their background environment.

We present three broad findings: First, we show that productivity changes were heterogeneous across the workforce, and systematically related to factors associated with ease of WFH: overall job quality as measured by wage level; gender and the presence of children, and feasibility of WFH in terms of job tasks. Second, we show that, as the pandemic progressed, workers sorted into locations - WFH or working in the office - depending on their previous productivity experiences. Third, and building on these insights, we control for endogenous sorting and estimate factors affecting productivity *across* locations: We find direct evidence that those with better jobs and working for larger firms had better productivity outcomes *at home* in particular; outcomes were more equal in the office.

Our findings show that workers and firms are able to sort into locations to suit individual-specific productivity outcomes. Our findings also have important practical implications: large firms were better at making WFH work effectively, and so smaller employers should look for ways to mirror their structures. This information is also useful for policy makers looking to provide these smaller employers with support. Our findings also prompt further research: the survey we use here will in future enable an analysis of post-pandemic outcomes. These data are also highly suited for examining the potentially important interplay between WFH with health outcomes, which we do not address here.

Appendix A Tables

Table A.1: Summary Statistics

	N	Mean	Std. Dev.	Min.	Max.
Static Analysis:					
Female	26,779	0.554	0.497	0	1
Age	26,779	24.933	2.808	20	29
Non-white	26,779	0.101	0.302	0	1
Has children	26,779	0.271	0.449	0	1
Has degree	26,779	0.217	0.412	0	1
Unemployed	26,779	0.0891	0.285	0	1
Real Gross income (00s)	26,779	10.132	9.100	-67.217	150
Having financial difficulty	26,779	0.0906	0.287	0	1
Dynamic Analysis: Leaving the parental home					
Female	6,778	0.457	0.498	0	1
Age	6,778	23.465	2.579	20	29
Non-white	6,778	0.097	0.296	0	1
Has children	6,778	0.0347	0.183	0	1
Has degree	6,778	0.198	0.399	0	1
Unemployed	6,778	0.103	0.304	0	1
In a cohabiting relationship	6,778	0.0469	0.211	0	1
Real Gross income (00s)	6,778	9.378	7.438	-7.744	83.33
Lives with both biological parents	6,778	0.713	0.452	0	1
At least one parent has degree	6,778	0.212	0.409	0	1
Parental real gross income (00s)	6,778	28.297	23.525	-421.238	232.83
Having financial difficulty	6,778	0.0776	0.267	0	1
Dynamic Analysis: Leaving the parental home					
Female	13,232	0.521	0.500	0	1
Age	13,232	23.857	2.691	20	29
Non-white	13,232	0.115	0.319	0	1
Has children	13,232	0.140	0.329	0	1
Has degree	13,232	0.196	0.397	0	1
Unemployed	13,232	0.118	0.323	0	1
Lives with parents	13,232	0.778	0.416	0	1
Real Gross income (00s)	13,232	8.682	7.841	-7.744	130.6305
Having financial difficulty	13,232	0.107	0.309	0	1

Notes: The data is taken from the UKHLS between 2000 and 2019. Weights are used to calculate the summary statistics but not for the number of observations.

Table A.2: District Level House Price Variation, 2012

	GBP
Top 5 most expensive LADs in London:	
Kensington and Chelsea	1,077,366
City of Westminster	736,039
Hammersmith and Fulham	603,876
Camden	594,506
City of London	521,769
Top 5 most expensive LADs outside of London:	
Elmbridge	430,877
Mole Valley	360,784
St. Albans	357,248
Windsor and Maidenhead	347,748
Waverley	340, 991
Bottom 5 least expensive LADs:	
Burnley	76,641
Pendle	85,256
Stoke-on-Trent	88,304
City of Kingston upon Hull	89,022
Hyndburn	89,615

Notes: Data is an average house price for 2012. Data source: Land Registry

Table A.3: Living With Parents, Baseline Logit Estimates

	Static			Dynamic		
	(1)	(2)	(3)	(1)	(2)	(3)
$HPI_{d,t-1}$	1.111***	1.121***	1.100**	0.918***	0.831***	0.894***
	(0.0118)	(0.00140)	(0.0475)	(0.0124)	(0.0205)	(0.102)
N	26,779	26,779	26,779	6,778	6,778	6,778
Pseudo R^2	0.0161	0.6338	0.669	0.0092	0.1405	0.2867
Individual controls		✓	✓		✓	✓
LAD dummies			✓		✓	
Year dummies			✓		✓	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Note: The sample contains young people (20-29 year olds) living in England in USoc between 1995 and 2019. The dependent variable in the static analysis is set equal to one if living with parents. In the dynamic analysis, an observation is included if the young person is observed in the previous year living with their parents. The dependent variable is an indicator set equal to one if the young person is observed not to be living with their parents. The base year of the HPI is 2015. Individual controls include age, age squared and a series of dummies capturing if the individual is non-white, whether they have a degree, whether they are foreign-born, if they are a parent, a dummy for unemployment, real gross and gender. Parental controls capturing total gross income, whether at least one parent has a degree and whether both biological parents reside in the home are also utilized for the dynamic analysis. All of the control variables are lagged by one year. Sample weights are utilized. Standard errors are clustered at the PSU level.

Table A.4: Living With a Partner, Baseline Logit Estimates

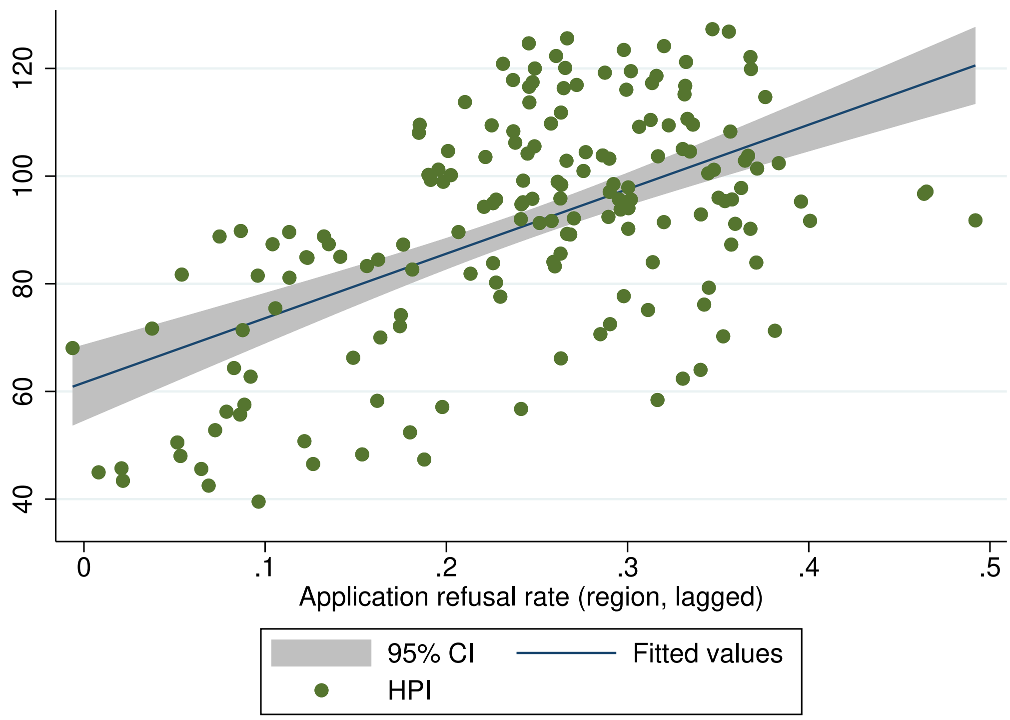
	Static			Dynamic		
	(1)	(2)	(3)	(1)	(2)	(3)
$HPI_{d,t-1}$	0.914***	0.892***	0.833**	0.921***	0.918***	0.905*
	(0.00963)	(0.0111)	(0.0210)	(0.0113)	(0.0108)	(0.302)
N	26,779	26,779	26,779	13,232	13,232	13,232
Pseudo R^2	0.0120	0.218	0.266	0.0078	0.0268	0.0289
Individual controls		✓	✓		✓	✓
LAD dummies			✓		✓	
Year dummies			✓		✓	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Note: The sample contains young people (20-29 year olds) living in England in USoc between 1995 and 2019. The dependent variable in the static analysis is set equal to one if living with parents. In the dynamic analysis, an observation is included if the young person is observed in the previous year living with their parents. The dependent variable is an indicator set equal to one if the young person is observed not to be living with their parents. The base year of the HPI is 2015. Individual controls include age, age squared and a series of dummies capturing if the individual is non-white, whether they have a degree, whether they are foreign-born, if they are a parent, a dummy for unemployment, real gross and gender. All of the control variables are lagged by one year. Sample weights are utilized. Standard errors are clustered at the PSU level.

Appendix B Figures

Figure B.1: HPI and Average Planning Application Rejection Percentage (2000-2019, by Region)



Notes: The average planning application rejection percentage for major dwellings projects. Observation for the 9 GORs in the UK are included. Data source: Land Registry.

A Proofs

Proof of Lemma 1:

Define match surplus $S = U_1 + U_2 - 2u(y)$ and consider the surplus maximisation problem

$$\max_{z_1,z_2} S(z_1, z_2) = u(y - z_1) + u(y - z_2) + 2\theta[z_1 + z_2] - 2u(y) \text{ subject to } z_1 + z_2 \geq 0.$$

Consider first $\theta > 0.5u'(y)$. Because $S(.)$ is strictly increasing in z_1, z_2 at (0,0), there exists $z_1 = z_2 = z > 0$ which is strictly surplus increasing and so a strict gain to trade exists for such θ.

Suppose instead $\theta \leq 0.5u'(y)$ and so $S(.)$ is decreasing in z_1, z_2 at (0,0). Standard arguments find the necessary conditions for optimality are satisfied at $z_1 = z_2 = 0$ and $S(.)$ strictly concave in (z_1, z_2) then implies $S(0, 0) = 0$ is a global maximum (Kuhn-Tucker Sufficiency Theorem). Hence there does not exist any z_1, z_2 satisfying $z_1 + z_2 \geq 0$ where $S(.) > 0$ and so no gain to trade exists for such θ. This completes the proof of Lemma 1. ■

Proof of Theorem 1:

Differentiating (12) wrt θ implies:

$$\frac{dV}{d\theta} = \frac{U_\theta(\theta)}{r + 2\lambda[1 - H(V(\theta)] - \lambda H'(V(\theta))[V(\theta) - V^S]}. \tag{A.1}$$

and RC now implies $V(\theta)$ is strictly increasing in $\theta > \theta^R$. Hence any equilibrium satisfying RC implies a partner in a match $\theta > \theta^R$ only quits to an outside offer $\theta' > \theta$. Given that, the first step is to characterise equilibrium $G(.)$.

Step 1: Steady state turnover implies $dG(\theta)$, the measure of agents in a partnership $[\theta, \theta + d\theta]$, must satisfy:

$$dG(\theta)[\delta + 2\overline{\lambda} \int_\theta^{\overline{\theta}} G(\widetilde{\theta})F'(\widetilde{\theta})d\widetilde{\theta}] = G(\theta)[\overline{\lambda}G(\theta)F'(\theta)d\theta]$$

where the LHS describes the flow out through death or separation (where either partner may leave), while the RHS describes the flow in of new agents who match in $[\theta, \theta + d\theta]$. Hence $G(.)$ solves the differential equation

$$\frac{dG}{d\theta} = \frac{\overline{\lambda} G(\theta)^2}{\delta + 2\overline{\lambda} \int_{\theta}^{\overline{\theta}} G(\widetilde{\theta}) F'(\widetilde{\theta}) d\widetilde{\theta}} F'(\theta) \tag{A.2}$$

with boundary value $G(\overline{\theta}) = 1$. Define $Z(\theta) = \int_{\theta}^{\overline{\theta}} G(\widetilde{\theta}) F'(\widetilde{\theta}) d\widetilde{\theta}$ and so $Z'(\theta) = -G(\theta) F'(\theta)$. Substituting out $\int_{\theta}^{\overline{\theta}} G(\widetilde{\theta}) F'(\widetilde{\theta}) d\widetilde{\theta}$ in (A.2) now implies:

$$\frac{1}{G} \frac{dG}{d\theta} = -\frac{\overline{\lambda} Z'(\theta)}{\delta + 2\overline{\lambda} Z(\theta)}$$

and integration yields:

$$\ln G(\theta) = A - \frac{1}{2} \ln[\delta + 2\overline{\lambda} Z(\theta)]$$

where A is the constant of integration. The boundary values $Z = 0, G = 1$ at $\overline{\theta}$ yield solution

$$G(\theta) = \left[\frac{\delta}{\delta + 2\overline{\lambda} Z(\theta)} \right]^{\frac{1}{2}}, \tag{A.3}$$

because (A.2) implies

$$\frac{dG}{d\theta} = \frac{G(\theta)^2}{\delta + 2\overline{\lambda} Z(\theta)} \overline{\lambda} F'(\theta)$$

we can now use (A.3) to eliminate $\delta + \overline{\lambda} Z(\theta)$ and so obtain:

$$\frac{1}{G(\theta)^4} \frac{dG}{d\theta} = \frac{\overline{\lambda}}{\delta} F'(\theta).$$

Integrating using boundary condition $F, G = 1$ at $\theta = \overline{\theta}$ now yields

$$\frac{1}{3} \left[1 - \frac{1}{G(\theta)^3} \right] = \frac{\overline{\lambda}}{\delta} [F(\theta) - 1]$$

which rearranges as the stated solution for $G(.)$ in Theorem 1. This completes Step 1.■

The equilibrium offer distributions are given by (14) where the LHS describes the arrival rate of willing partners whose match yields a value greater than $V(\theta)$, while the right-hand side describes the arrival rate of contacts $\widetilde{\theta} \geq \theta$ which yields value $V(\widetilde{\theta}) > V(\theta)$ and $G(\widetilde{\theta})$ is the probability the contacted agent is willing to match (i.e. not already in a match better than $\widetilde{\theta}$). Using (14) to substitute out $\lambda[1 - H(V)]$ in (12) yields the functional equation (21) described in the text. Differentiating wrt θ yields the differential equation:

$$[r + 2\overline{\lambda} \int_{\theta}^{\overline{\theta}} G(\widetilde{\theta})F'(\widetilde{\theta})d\widetilde{\theta}]\frac{dV}{d\theta} - \overline{\lambda}G(\theta)F'(\theta)[V - V^S] = U_\theta. \tag{A.4}$$

Integration using boundary condition $V(\overline{\theta}) = U(\overline{\theta})/r$ finds:

$$V(\theta) = V^S + \frac{r^{\frac{1}{2}}[U(\overline{\theta})/r - V^S] - \int_{\theta}^{\overline{\theta}} \frac{U_\theta(\theta')}{[r+2\overline{\lambda}\int_{\theta'}^{\infty} G(\widetilde{\theta})F'(\widetilde{\theta})d\widetilde{\theta}]^{1/2}} d\theta'}{\left\{[r + 2\overline{\lambda}\int_{\theta}^{\overline{\theta}} G(\widetilde{\theta})F'(\widetilde{\theta})d\widetilde{\theta}]\right\}^{1/2}}. \tag{A.5}$$

(simple check is to differentiate this solution). Putting $\theta = \theta^R$ and using the equilibrium boundary condition $V(\theta^R) = V^S$, standard algebra now establishes

$$V^S = U(\overline{\theta})/r - \int_{\theta^R}^{\overline{\theta}} \frac{U_\theta(\theta')}{[r^2 + 2r\overline{\lambda}\int_{\theta'}^{\infty} G(\widetilde{\theta})F'(\widetilde{\theta})d\widetilde{\theta}]^{1/2}} d\theta'$$

which can be rearranged as (15) in Theorem 1. Substituting out V^S in (A.5) and simplifying now yields the stated solution for $V(\theta)$.

The final step is to show this equilibrium satisfies RC. Differentiating (14) wrt θ implies:

$$\lambda H'(V)\frac{dV}{d\theta} = \overline{\lambda}G(\theta)F'(\theta). \tag{A.6}$$

(A.1) with $U_\theta > 0$ for all $\theta > \theta^R$ and noting $\lambda[1 - H(V(\theta))] = \overline{\lambda}\int_{\theta}^{\overline{\theta}} G(\widetilde{\theta})F'(\widetilde{\theta})d\widetilde{\theta}$ now implies

$$[r + 2\lambda[1 - H(V(\theta))]]\frac{dV}{d\theta} > \overline{\lambda}G(\theta)F'(\theta)[V(\theta) - V^S] \text{ for all } \theta > \theta^R. \tag{A.7}$$

Using the previous equation to substitute out $\frac{dV}{d\theta}$ now establishes RC is satisfied and so completes the proof of Theorem 1.■

Proof of Theorem 2: (V_1, V_2) solve:

$$rV_1 = U_1 + \lambda \int_{V_1}^{\overline{V}} [1 - H(V')]dV' + \lambda[1 - H(V_2)][V^S - V_1] \tag{A.8}$$

$$rV_2 = U_2 + \lambda \int_{V_2}^{\overline{V}} [1 - H(V')]dV' + \lambda[1 - H(V_1)][V^S - V_2], \tag{A.9}$$

Consider a joint variation in household values (dV_1, dV_2) due to a local variation $d\mu$ along with a variation $d\lambda$ in exclusivity. Total differentiation implies total variation:

$$\begin{aligned} rdV_1 =& dU_1 - \lambda[1 - H(V_1)]dV_1 - \lambda[1 - H(V_2)]dV_1 + \lambda H'(V_2)[V_1 - V^S]dV_2 \\ &+ d\lambda \left\{ \int_{V_1}^{\infty} [1 - H(V')]dV' + [1 - H(V_2)][V^S - V_1] \right\} \end{aligned} \tag{A.10}$$

$$\begin{aligned} rdV_2 =& dU_2 - \lambda[1 - H(V_2)]dV_2 - \lambda[1 - H(V_1)]dV_2 + \lambda H'(V_1)[V_2 - V^S]dV_1 \\ &+ d\lambda \left\{ \int_{V_2}^{\infty} [1 - H(V')]dV' + [1 - H(V_1)][V^S - V_2] \right\} \end{aligned} \tag{A.11}$$

where $dU_1 = -\mu dU_2$ Re-arrange as the matrix equation :

$$\begin{bmatrix} [r+s] & -\lambda H'(V_2)[V_1 - V^S] \\ -\lambda H'(V_1)[V_2 - V^S] & [r+s] \end{bmatrix} \begin{bmatrix} dV_1 \\ dV_2 \end{bmatrix} = \begin{bmatrix} \text{-}\mu & g \\ 1 & h \end{bmatrix} \begin{bmatrix} dU_2 \\ d\lambda \end{bmatrix}$$

where $s = \lambda[1 - H(V_1)] + \lambda[1 - H(V_2)]$ is the total partner separation rate, and $g = \int_{V_1}^{\infty}[1 - H(V')]dV' + [1 - H(V_2)][V^S - V_1], h = \int_{V_2}^{\infty}[1 - H(V')]dV' + [1 - H(V_1)][V^S - V_2]$.

Standard matrix algebra now implies:

$$\begin{bmatrix} dV_1 \\ dV_2 \end{bmatrix} = \frac{1}{\Delta} \begin{bmatrix} \text{a} & \text{b} \\ \text{c} & \text{e} \end{bmatrix} \begin{bmatrix} dU_2 \\ d\lambda \end{bmatrix}$$

where Δ is given by Claim 1 and

$$\begin{aligned} a =& -(r+s)\mu + \lambda H'(V_2)[V_1 - V^S] \\ b =& [r+s]g + \lambda H'(V_2)[V_1 - V^S]h \\ c =& [r+s] - \mu\lambda H'(V_1)[V_2 - V^S] \\ e =& (r+s)h + \lambda H'(V_1)[V_2 - V^S]g. \end{aligned} \tag{A.12}$$

Now consider a compensating variation $(dU_2, d\lambda)$ which holds V_2 constant; i.e. a variation where $dV_2 = 0$ and so restricts $c[dU_2] + e[d\lambda] = 0$. For any such variation, the compensated variational change in V_1 is then

$$\begin{aligned} dV_1 &= \frac{1}{\Delta}[a[dU_2] + b[d\lambda]] \\ &= -\frac{[ae - bc]}{c\Delta} d\lambda \end{aligned}$$

Now $c > 0$ at any efficient agreement (otherwise using Claim 1 it can be shown the value frontier must be upward sloping and such agreements are not efficient - see Figure 4). Hence the compensated variation implies $dV_1/d\lambda > 0$ if and only if det $A < 0$. Claim 2 now calculates that determinant.

Claim 2.

$$\det A = -\Delta \left\{ \int_{V_1}^{\infty} [1 - H(V')]dV' - [1 - H(V_2)][V_1 - V^S] + \mu \int_{V_2}^{\infty} [1 - H(V')]dV' - \mu[1 - H(V_1)][V_2 - V^S] \right\}$$

Proof. See below.

Hence by Claim 2, the compensated variation $dV_1/d\lambda > 0$ if and only if

$$\Phi = \left\{ \int_{V_1}^{\infty} [1 - H(V')]dV' - [1 - H(V_2)][V_1 - V^S] + \mu \int_{V_2}^{\infty} [1 - H(V')]dV' - \mu[1 - H(V_1)][V_2 - V^S] \right\} > 0.$$

as stated in Theorem 2, where the converse applies for $dV_1/d\lambda < 0$. Proving Claim 2 thus completes the proof of Theorem 2.

Proof of Clam 2: Requires straightforward algebra.

$$\begin{aligned}
det[A] \quad = \quad & \left\{ -(r+s)\mu + \lambda H'(V_2)[V_1 - V^S] \right\} (r+s) \int_{V_2}^{\infty} [1 - H(V')]dV' \\
& + \left\{ -(r+s)\mu + \lambda H'(V_2)[V_1 - V^S] \right\} (r+s)[1 - H(V_1)][V^S - V_2] \\
& + \left\{ -(r+s)\mu + \lambda H'(V_2)[V_1 - V^S] \right\} \lambda H'(V_1)[V_2 - V^S] \int_{V_1}^{\infty} [1 - H(V')]dV' \\
& + \left\{ -(r+s)\mu + \lambda H'(V_2)[V_1 - V_1^S] \right\} \lambda H'(V_1)[V_2 - V^S][1 - H(V_2)][V^S - V_1] \\
& + \left\{ -[r+s] + \mu\lambda H'(V_1)[V_2 - V^S] \right\} [r+s] \int_{V_1}^{\infty} [1 - H(V')]dV' \\
& + \left\{ -[r+s] + \mu\lambda H'(V_1)[V_2 - V^S] \right\} [r+s][1 - H(V_2)][V^S - V_1] \\
& + \left\{ -[r+s] + \mu\lambda H'(V_1)[V_2 - V^S] \right\} \lambda H'(V_2)[V_1 - V^S] \int_{V_2}^{\infty} [1 - H(V')]dV' \\
& + \left\{ -[r+s] + \mu\lambda H'(V_1)[V_2 - V^S] \right\} \lambda H'(V_2)[V_1 - V^S][1 - H(V_2)][V^S - V_2]
\end{aligned}$$

Cancel repeated terms:

$$\begin{aligned}
det[A] \quad = \quad & -(r+s)^2\mu \int_{V_2}^{\infty} [1 - H(V')]dV' - [r+s]^2 \int_{V_1}^{\infty} [1 - H(V')]dV' \\
& -(r+s)^2\mu[1 - H(V)][V^S - V_2] - [r+s]^2[1 - H(V_2)][V^S - V_1] \\
& +\lambda^2 H'(V_2)H'(V_1)[V_1 - V^S][V_2 - V^S] \int_{V_1}^{\infty} [1 - H(V')]dV' \\
& +\lambda^2 H'(V_2)H'(V_1)[V_1 - V^S][V_2 - V^S][1 - H(V_2)][V^S - V_1] \\
& +\mu\lambda^2 H'(V_2)H'(V_1)[V_2 - V^S][V_1 - V^S] \int_{V_2}^{\infty} [1 - H(V')]dV' \\
& +\mu\lambda^2 H'(V_2)H'(V_1)[V_2 - V^S][V_1 - V^S][1 - H(V_1)][V^S - V_2]
\end{aligned}$$

Collect appropriately

$$
\begin{aligned}
det[A] &= -(r+s)^2\left\{\mu\int_{V_2}^{\infty}[1-H(V')]dV' + \int_{V_1}^{\infty}[1-H(V')]dV' + \mu[1-H(V_1)][V^S-V_2] + [1-H(V_2)][V^S-V_1]\right\} \\
&\quad + \left\{\lambda^2 H'(V_2)H'(V_1)[V_1-V^S][V_2-V^S]\right\} \\
&\quad \times\left\{\int_{V_1}^{\infty}[1-H(V')]dV' + [1-H(V_2)][V^S-V_1] + \mu\int_{V_2}^{\infty}[1-H(V')]dV' + \mu[1-H(V_1)][V^S-V_2]\right\}
\end{aligned}
$$

and spot the common term

$$
\begin{aligned}
det[A] &= \left\{\lambda^2 H'(V_2)H'(V_1)[V_1-V^S][V_2-V^S] - (r+s)^2\right\} \\
&\quad \times\left\{\int_{V_1}^{\infty}[1-H(V')]dV' + [1-H(V_2)][V^S-V_1] + \mu\int_{V_2}^{\infty}[1-H(V')]dV' + \mu[1-H(V_1)][V^S-V_2]\right\} \\
&= -\Delta\left\{\int_{V_1}^{\infty}[1-H(V')]dV' + [1-H(V_2)][V^S-V_1] + \mu\int_{V_2}^{\infty}[1-H(V')]dV' + \mu[1-H(V_1)][V^S-V_2]\right\}
\end{aligned}
$$

which completes the proof of Claim 2 and so establishes Theorem 2. ■

Proof of Lemma 4: In the extended framework, fraction $1-G(0)$ are in partnerships $\theta > 0$, fraction M are in bad matches $\theta = 0$ and the rest $G(0)-M$ are single. Define aggregate search effort

$$
\overline{\Lambda} = \overline{\lambda}G(0) + \int_0^{\overline{\theta}} \phi^*(\theta)\overline{\lambda}G'(\theta)d\theta.
$$

Consider then steady turnover in bad matches. The entry flow is

$$
\text{Entry Flow } = [G(0)-M]\overline{\lambda}p\frac{\overline{\lambda}\,[G(0)-M]}{\overline{\Lambda}}
$$

where $[G(0)-M]S\overline{\lambda}$ describes the rate at which current singles contact a partner, where p is the probability it is a bad match, and taking search effort into account, $\frac{\overline{\lambda}[G(0)-M]}{\overline{\Lambda}}$ is the probability the contact is with another single (who are the only people who agree to such a match). The Exit flow is

$$
\text{Exit flow} = M\left\{\delta + \delta_0 + 2\overline{\lambda}(1-p)\int_0^1 P(\theta)F'(\theta)d\theta + \mu\right\}
$$

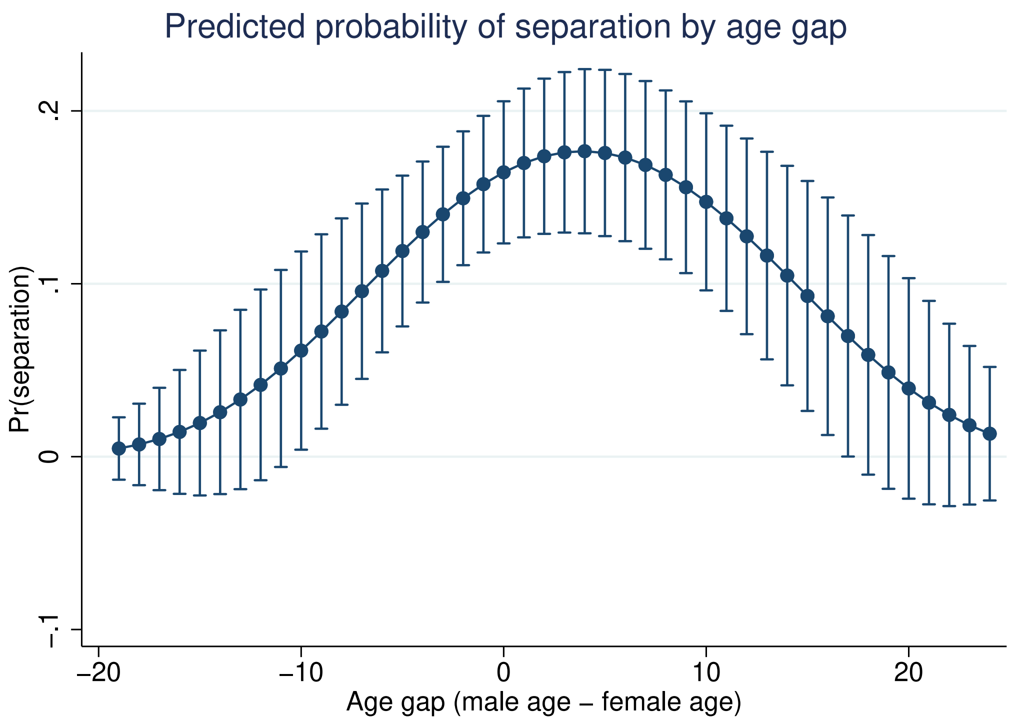
where M agents are in a bad match and each escapes either through death, destruction, separation (either partner forms an outside match with strictly positive surplus) or via excess break-up. Steady state thus implies M satisfies

$$\frac{\overline{\lambda}^2 p\,[G(0)-M]^2}{\overline{\Lambda}} = M\left\{\delta+\delta_0+2\overline{\lambda}(1-p)\int_0^1 P(\theta)F'(\theta)d\theta+\mu\right\}.$$

Substituting out $\overline{\Lambda}$ using the above and $F'=1$ yields the equation stated. This completes the proof of Lemma 4. ■

B Additional Figures

Figure B.1: Probability of separation by age gap



Note: Estimates from the model in column 5 of Table 7.

Appendix A Imputing Productivity Changes from Qualitative and Banded Quantitative Survey Responses

As discussed in Section 3, and compared to wave 3 of the UKHLS Covid module survey (June 2020), waves 5, 7 and 9 ask two additional quantitative questions regarding productivity changes, for all interviewees who have reported productivity changes in the qualitative question. Specifically, for those who have reported gains in productivity the survey asks:

> *"Thinking about how much more you get done these days, would you say that what you can do in an hour now would previously have taken you:"*

Then interviewees are supposed to select one choice from following:

> *1 - Up to an hour and a quarter;*
>
> *2 - Between an hour and a quarter and an hour and a half;*
>
> *3 - More than an hour and a half*

Similarly, respondents who have reported declines in productivity are asked:

> *"Thinking about how much less you get done these days, would you say that what you can do in an hour now would previously have taken you:"*

Then they can select one choice from below:

> *4 - Between 45 minutes and an hour;*
>
> *5 - Between 30 and 45 minutes;*
>
> *6 - Less than 30 minutes.*

These choices directly imply percentage changes in productivity. For example, choosing "*1. Up to an hour and a quarter*" translates into what can be done in 60 minutes now would have previously taken up to 75 minutes. Thus, the upper threshold of percentage productivity change $\Delta prod$ during lockdown can be computed as:

$$\Delta prod = \frac{\frac{1}{60} - \frac{1}{75}}{\frac{1}{75}} = \frac{1}{4} = 25\%.$$

Therefore, choices 1 to 6, together with respondents answering their productivity stays the same as before the lockdown, imply the frequencies shown in the left hand labels column of Table A.1.

Table A.1: Response Frequencies of Productivity Change Variables

	June 2020	Sept. 2020	Jan. 2021	Sept. 2021
Quantitative question				
$>$ +50%		4.18	3.42	4.61
+25% to +50%		8.93	8.29	10.42
below +25%		9.71	9.74	10.45
no change		61.92	55.66	64.28
above -25%		6.47	8.62	4.55
-25% to -50%		5.65	8.06	3.69
-50 % to -100%		3.14	6.21	2.00
Qualitative question				
Much more	12.43	11.51	10.66	13.6
Little more	14.79	11.51	10.94	12.36
Same	43.05	61.32	54.73	63.62
Little less	19.72	9.92	13.75	7.45
Much less	10.00	5.74	9.92	2.87

Note: This table presents the response frequencies of the productivity questions in the UKHLS Covid waves. In each of the waves presented individuals are asked to qualitatively compare their current productivity per hour to their productivity in Jan/Feb 2020 (bottom half of table). From Sept 2020 onwards individuals that indicated their productivity changed in the qualitative question are also asked to quantify that change. Specifically they are asked how much time it would have taken them to get done what they previously achieved in an hour. Response options are specified in the row labels. See the text for more details. Sample weights are used throughout.

We fit a flexible Pearson type VII distribution to these quantitative responses. The survey questions provide 2 pairs of symmetric cutoffs for productivity change at -50%, -25%, +25% and +50%, respectively. In addition, we assume there exists a response interval $[a_1, a_2]$ such that any productivity change that falls within this interval is recorded as "same". Figure A.1 plots the Pearson distribution of (quantitative) productivity change, which is divided into 7 areas (A to G) by these thresholds. Let q^A, q^B,

$q^C, ..., q^G$ denote the size of area A, B, C, ..., G, respectively in the figure and $\Omega(\frac{(x-\mu)}{S}, \nu)$ denote the Pearson distribution with three distribution parameters: μ represents a shift in the distribution, S is the scaling parameter and ν is the parameter controlling kurtosis. This implies a system consisting of 7 equations corresponding to the size of each area in Figure A.1, with 5 unknown parameters (the distribution parameters plus a_1 and a_2). Then we solve the system of equations by selecting a combination of the parameters that minimize the sum of errors, weighted by the inverse of the actual size (fraction) of each area.

Figure A.1: Distribution of Productivity Change Quantitative Measure

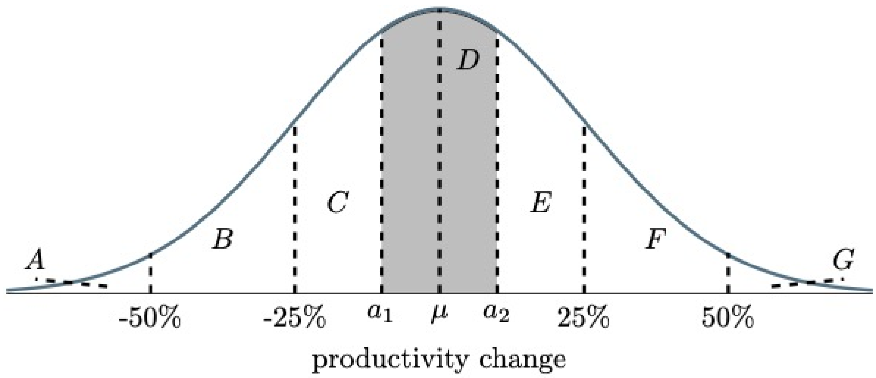


Table A.2 shows results for the three relevant waves. The computed response interval for reporting "same" is $[-0.15, 0.16]$ in September 2020, where a_1 and a_2 are sufficiently close in absolute values. For comparison, we also fit the data with Gaussian distributions, as displayed on the right hand side of Table A.2. The goodness of fit measure on the bottom row shows that the Pearson distribution fits the data much better in all waves.

Our analysis also makes use of the qualitative information from June 2020. To illustrate the data structure for these, Figure A.2 plots the distribution for qualitative answers to productivity change, with two thresholds that distinguish answers of "a little less productive" from "much less productive", and "a little more productive" from "much more productive", respectively and the response interval $[a_1', a_2']$ for reporting "same". We use these data by first comparing the qualitative and quantitative

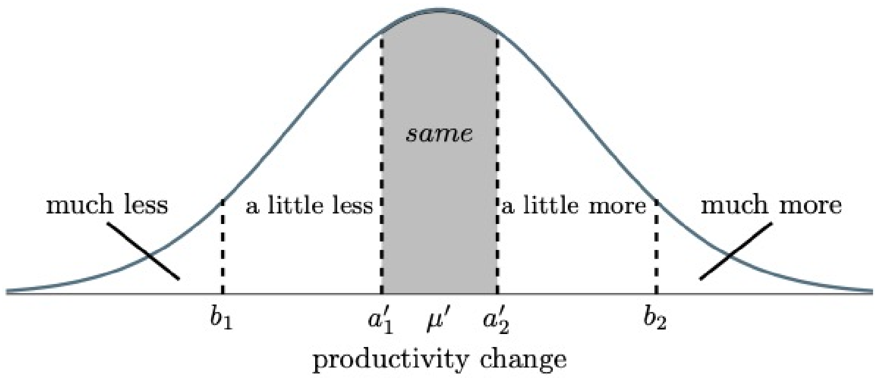
Table A.2: Imputing Productivity Changes from Banded Questions

	Pearson VII			Gaussian		
	Sept. 2020	Jan. 2021	Sept. 2021	Sept. 2020	Jan. 2021	Sept. 2021
Parameters						
Location (μ)	3.56	-0.61	7.99	2.04	-0.01	4.22
Scale (σ)	13.52	15.33	11.39	18.94	20.82	18.16
Shape (ν)	1.87	1.77	1.70			
Cut-off 1 (a_1)	-15.21	-15.69	-14.21	-17.49	-16.73	-18.21
Cut-off 2 (a_2)	16.14	14.20	17.55	16.69	15.89	17.06
Cell means						
$>$ +50%	76.19	78.08	76.16	56.06	56.90	55.83
+25% to +50%	34.17	34.6	33.78	33.04	33.48	32.98
below +25%	20.07	18.95	20.86	20.55	20.12	20.76
no change	1.70	-0.70	4.67	0.19	-0.36	0.75
above -25%	-19.51	-19.86	-18.86	-20.94	-20.6	-21.31
-25% to -50%	-34.5	-34.55	-34.71	-32.61	-33.49	-32.07
-50 % to -100%	-77.69	-77.84	-79.66	-55.72	-56.91	-55.16
Goodness of fit	4.24E-04	0.0073	0.0025	0.0208	0.0271	0.0197

Note: To impute the percentage change values for each band of the productivity change responses we assume a continuous underlying distribution and minimize the squared distance between the simulated density and observed density for each of the Pearson VII distribution and the Gaussian distribution. Figure A.1 and Figure A.2 illustrate how the bands make up the continuous distribution. The top half of the table presents the parameters from the resulting distributions. The bottom half of the distribution provides the estimates of the mean percentage change in productivity within each band. The goodness of fit displays the sum of squared distances. See text above for more details.

responses from *September* 2020 as follows: From the fitted distribution above, we impute the two pairs of threshold, b_1, b_2 and a'_1, a'_2, to match the distribution of responses to the September wave qualitative question. b_1 and b_2 are found to be -34.17% and 27.29%, respectively, and $[a'_1, a'_2]$ is near identical to $[a_1, a_2]$.

Figure A.2: Distribution of Productivity Change Qualitative Measure



Finally, to operationalize the June 2020 data, we assume that the thresholds b_1, b_2, a'_1 and a'_2 are identical across June and September. All that remains is to fit another Pearson distribution (i.e. mean, variance and kurtosis parameters) to match the distribution of responses in *June* 2020. Therefore, we solve a system of 5 equations in terms of the size of each areas in Figure A.2, with three unknown distribution parameters. Based on this fitted distribution the imputed average productivity changes to answers of "*much less productive*", "*a little less productive*", "*a little more productive*" and "*much more productive*", for June 2020, are −44.9%, −22.4%, 22.3% and 40.8%, respectively.

To validate the data and our imputation we carry out basic analyses to test internal consistency. These are shown in Table A.3. The first two columns show logit regressions of the responses to the qualitative questions in waves 5, 7, and 9, on the responses to the banded quantitative questions. They show, for example, that when someone responds "much more" to the qualitative question, they are far more likely to also provide the strongest response to the quantitative question than the less strong response. The third column uses ordered logit regressions to show correlations over time. It shows that those who responded "much more" in the previous wave respond with far stronger responses in the current wave. While there could be many reasons for this pattern, including individual fixed effects in the nature of responses, this column does show convincingly that the survey responses are not just random noise. Finally the last column shows a similar pattern using the imputed quantitative questions as a continuous measure. The R^2 indicates that the correlation of the responses across waves is around 0.35.

Table A.3: Internal Consistency of Productivity Questions

	Logit	Logit	Ordered Logit	OLS
	"Much more"	*"Much less"*	Qual. cat.	$\Delta Prod$
	(1)	(2)	(3)	(4)
Quantitative category				
(base α_2 to -25%)				
+25% to +50%	0.62			
	(0.07)			
>+50%	1.56			
	(0.10)			
Quantitative category				
(base α_1 to -25%)				
-25% to -50%		0.80		
		(0.11)		
-50% to -100%		2.46		
		(0.12)		
Lagged qualitative category				
(base *"Much less"*)				
"Little less"			0.42	
			(0.09)	
"Same"			1.22	
			(0.08)	
"Little more"			2.29	
			(0.09)	
"Much more"			3.13	
			(0.095)	
$\Delta Prod_{t-1}$				0.31
				(0.01)
Constant	-0.55	-1.51		0.07
	(0.07)	(0.10)	(0.01)	
N	4,046	2,605	10,818	10,631
(pseudo) R^2	0.05	0.15	0.07	0.13
Wave dummies	Yes	Yes	Yes	Yes

Note: Table shows results of four exercises to examine the properties of the productivity change data. Columns 1 and 2 show correlations of the qualitative question and quantitative question within period. Columns 3 and 4 show correlations within the question type over time. Specifically, Column 1 shows results of a logit regression comparing responses *"Much more"* and *"Little more"* with the three possible associated quantitative responses, treated as categorical outcomes. Column 2 shows a parallel logit regression for responses "Much less" and "Little less" with binaries for the associated quantitative responses. Column 3 shows an ordered logit of the response to the qualitative question in wave t on the lagged qualitative question. Column 4 shows a parallel OLS regression the qualitative question, here treated as a continuous variable. See text for more details.

Appendix B Additional Information on Supplementary Data Sources

B.1 Cross-walk between SOC2000 and O*NET Occupation

Table B.1 shows the cross-walk this paper adopts to convert the Standard Occupational Classification (SOC) 2000 to the Occupational Information Network (O*NET) codes, taken from 2020. Specifically, we assign each 3-digit SOC (sub-major occupation groups) into 2-digit O*NET codes (major occupation groups) by first matching 4-digit SOC (sub-sub-major occupation groups) codes with the most appropriate 2-digit O*NET category. Then, we assign each 3-digit SOC, based on the matching outcomes of 4-digit SOC to 2-digit O*NET code using an employment-weighted majority rule.

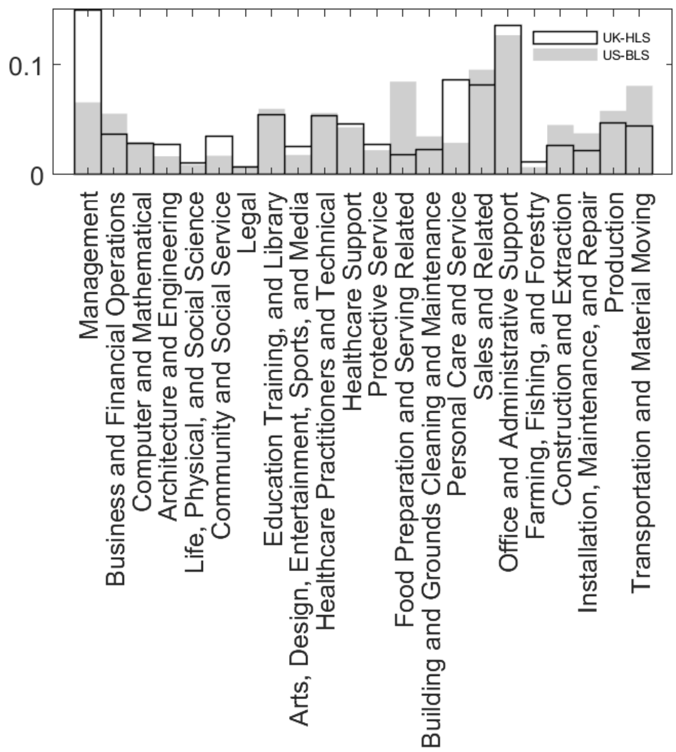
Although in most cases the overwhelming majority of 4-digit SOC codes are assigned to the same 2-digit O*NET code, this is not always the case. As a result, some matches between SOC 2000 and O*NET codes are necessarily imprecise. For instance, SOC 231 'Teaching Professionals' is classified into O*NET 25 'Education, Training, and Library Occupations', yet under it, SOC 2317 'Registrars and senior administrators of educational establishments' is more appropriate to be put into 2-digit O*NET 11 'Management Occupations', according to O*NET description. Due to the unavailability of 4-digit SOC information in the UKHLS, we are unable to specifically subtract sub-sub-major occupation group SOC 2317 from sub-major occupation group SOC 231. [11] In one case, we use industry information to split SOC 922 'Elementary Personal Services Occupations', which is mainly lined up with O*NET code 39. In this case, however, several food preparation related occupations are listed, such as 'Kitchen and catering assistants', 'Waiters and Waitresses'. These occupations belong to the industry related to food. Therefore, we move these respondents into O*NET 35 'Food Preparation and Serving Related Occupations'. Table B.1 shows the full assignment.

To show the quality of the match, Figure B.1 plots occupation distributions of respondents from wave 9 and the Covid module of UK Household Longitudinal Study (UKHLS), based on the imputed O*NET employment shares, together with national employment statistics from 2019 US Bureau of Labor Statistics (BLS). In the figure, white columns represent occupation percentages in UKHLS and grey columns represent occupation percentages in US-BLS. The correlation coefficient between both is around 0.7. The occupation categories showing largest differences are Management

[11] As an additional example, we would ideally move SOC 5241 'Electricians' out of O*NET 49 'Installation, Maintenance, and Repair Occupations' and into O*NET 47 'Construction and Extraction Occupations' if we had the 4-digit measures.

and Food Preparation and Serving Related. The sign of these differences is, at least, very likely genuine. The UK is reported to be particularly intensive in managers Blundell et al. (2022). Similarly, the US is more intensive in Food Serving (waiting). If we exclude these occupations, the correlation coefficient between UK and US occupation percentage rises to around 0.8.

Figure B.1: Occupation Percentage Distributions, UKHLS and US-Bureau of Labor Statistics (BLS)



B.2 Aggregate Production Data from the ONS

Figure 2 in Section 4 shows a comparison of the UKHLS-Covid productivity data with aggregate information from the UK Office for National Statistics ONS. As discussed in the main text, the ONS data are presented at a much coarser industry division level of aggregation. For example the Covid survey has 13 sub-industries within the single ONS category of ‘manufacturing’. The ONS categories are (with rough shortened titles): Agriculture; Mining and Quarrying; Manufacturing; Energy; Water supply and

Sewage; Construction; Wholesale and Retail Trade; Transportation and Storage; Accommodation and Food; Information and Communication; Finance; Real Estate; Professional Services; Administrative Services; Government Services; Arts; and Other Services. We use data from the quarter which contains the month of the UKHLS wave. However, for baseline data we use those from 2019 Q4. We consider this as providing a better fit with the January/February 2020 baseline in the Covid survey, because 2020 Q1 data are affected by the start of the pandemic.

The more complicated aspect of the comparison is that comparing individual productivity changes to aggregate data is non-trivial. We show the relevant calculation below. To simplify the computation somewhat we align our data to a measure of aggregate production change from labour inputs at the industry level as follows:

$$\begin{aligned}\Delta \ln Y_t &\approx \frac{\sum_i y_{it} - \sum_i y_{it-1}}{\sum_i y_{it-1}} \\ &= \frac{1}{\bar{Y}^{S+L}_{t-1}} \left[p^S \bar{Y}^S_{t-1} \sum_{i \in S} w^S_{it-1} \left(\Delta \ln prod_{it} + 1 \right) \frac{h_{it}}{h_{it-1}} + p^E \bar{Y}^E_t - p^L \bar{Y}^L_{t-1} \right] \end{aligned} \tag{5}$$

where we decompose the industry-level workforce into three groups: stayers, S; industry leavers, L, and industry entrants E. Then $\bar{Y}^X_t$ is average output at time t for group X (e.g. stayers), n^X is population size of group X and $p^X \equiv \frac{n^X}{n^{S+L}}$. h_{it}, h_{it-1} are hours of individual i at times t and $t-1$, and y_{it-1} is output/earnings of individual i at time $t-1$. Finally, and importantly, we calculate weights, $w^S_{it-1} \equiv \frac{1}{n^S} \sum_{i \in S} \frac{y_{it-1}}{\bar{Y}^S_{t-1}}$ that sum to 1 and capture relative position in the earnings/output distribution.

Almost all of the elements in (5) are observable. In particular, individual-level industry codes are observed in each of waves 3, 7 and 9. The only component we do not directly observe is earnings y_{it} in the Covid period. Here we assume that average earnings for this group $\bar{Y}^E_t$ are equal to baseline earnings for the stayers. The calculation is robust to altering this assumption because for most industries the proportion of entrants p^E is small, and so the contribution to the overall calculation is also small.

On the side of the aggregate data we use the percentage change in gross value added. In terms of national accounting concepts, this quantity includes not only change in contribution of workers, but change in profits. In effect therefore, we assume that these components move in parallel.

Figure C.7 shows this computation in each wave of data: June and September 2020, and January and September 2021.

To complete the discussion we return to the comparison of aggregate productivity. Aggregate productivity can be expressed in terms of individual level variables as follows:

$$
\begin{aligned}
\Delta\frac{\ln Y_t}{\ln H_t} \approx & \frac{\frac{\sum_i y_{it}}{\sum_i h_{it}} - \frac{\sum_i y_{it-1}}{\sum_i h_{it-1}}}{\frac{\sum_i y_{it-1}}{\sum_i h_{it-1}}} \\
= & p^S \frac{\bar{Y}^S_{t-1}}{\bar{Y}^{S+L}_{t-1}} \left(\frac{\dot{Y}^S_t}{\dot{H}_t} \sum_{i\in S} w^S_{it} \Delta \ln relhours_{it} + \sum_{i \in S} w^S_{it-1} \Delta \ln prod_{it} \right) + p^E \frac{\bar{Y}^E_t}{\bar{Y}^{S+L}_{t-1}\dot{H}_t} - p^L \frac{\bar{Y}^L_{t-1}}{\bar{Y}^{S+L}_{t-1}}
\end{aligned}
$$

where we use the same notation as that used in (5), and additionally $\dot{X}_t \equiv \bar{X}^{E+S}_t / \bar{X}^{S+L}_{t-1}$ is the growth in the average of variable X, $\dot{X}^S_t \equiv \bar{X}^S_t / \bar{X}^S_{t-1}$ is the growth for stayers only, and $\Delta \ln relhours_{it} \equiv \frac{h_{it} - h_{it-1}\dot{H}_t}{h_{it}}$ is a measure in change of hours share: Intuitively, if relative hours go down for low-wage workers, then aggregate productivity goes up.

Table B.1: Cross-walk from 3-digit SOC 2000 to 2-digit O*NET Classification

3-digit SOC	SOC title	2-digit O*NET	O*NET title
111	Corporate managers and senior officials	11	Management
112	Production managers	11	Management
113	Functional managers	11	Management
114	Quality and customer care managers	11	Management
115	Financial institution and office managers	11	Management
116	Managers in distribution, storage and retailing	11	Management
117	Protective service officers	11	Management
118	Health and social services managers	11	Management
121	Managers in farming, horticulture, forestry and fishing	11	Management
122	Managers and proprietors in hospitality and leisure services	11	Management
123	Managers and proprietors in other service industries	11	Management
211	Science professionals	19	Life, Physical, and Social Science
212	Engineering professionals	17	Architecture and Engineering
213	Information and communication technology professionals	15	Computer and Mathematical
221	Health professionals	29	Healthcare Practitioners and Technical
231	Teaching professionals	25	Education, Training, and Library
232	Research professionals	19	Life, Physical, and Social Science
241	Legal professionals	23	Legal
242	Business and statistical professionals	13	Business and Financial Operations
243	Architects, town planners, surveyors	17	Architecture and Engineering
244	Public service professionals	21	Community and Social Service
245	Librarians and related professionals	25	Education, Training, and Library
311	Science and engineering technicians	17	Architecture and Engineering
312	Draughtspersons and building inspectors	17	Architecture and Engineering
313	IT service delivery occupations	15	Computer and Mathematical
321	Health associate professionals	29	Healthcare Practitioners and Technical
322	Therapists	29	Healthcare Practitioners and Technical
323	Social welfare associate professionals	21	Community and Social Service
331	Protective service occupations	33	Protective Service
341	Artistic and literary occupations	27	Arts, Design, Entertainment, Sports, and Media
342	Design associate professionals	27	Arts, Design, Entertainment, Sports, and Media
343	Media associate professionals	27	Arts, Design, Entertainment, Sports, and Media
344	Sports and fitness occupations	27	Arts, Design, Entertainment, Sports, and Media
351	Transport associate professionals	53	Transportation and Material Moving
352	Legal associate professionals	23	Legal
353	Business and finance associate professionals	13	Business and Financial Operations
354	Sales and related associate professionals	41	Sales and Related
355	Conservation associate professionals	45	Farming, Fishing, and Forestry
356	Public service and other associate professionals	21	Community and Social Service
411	Administrative occupations: Government and related	43	Office and Administrative Support
412	Administrative occupations: Finance	43	Office and Administrative Support
413	Administrative occupations: Records	43	Office and Administrative Support
414	Administrative occupations: Communications	43	Office and Administrative Support
415	Administrative occupations: General	43	Office and Administrative Support
421	Secretarial and related occupations	43	Office and Administrative Support
511	Agricultural trades	45	Farming, Fishing, and Forestry
521	Metal forming, welding and related trades	47	Construction and Extraction
522	Metal machining, fitting and instrument making trades	51	Production
523	Vehicle trades	49	Installation, Maintenance, and Repair
524	Electrical trades	49	Installation, Maintenance, and Repair
531	Construction trades	47	Construction and Extraction
532	Building trades	47	Construction and Extraction
541	Textiles and garments trades	51	Production
542	Printing trades	51	Production

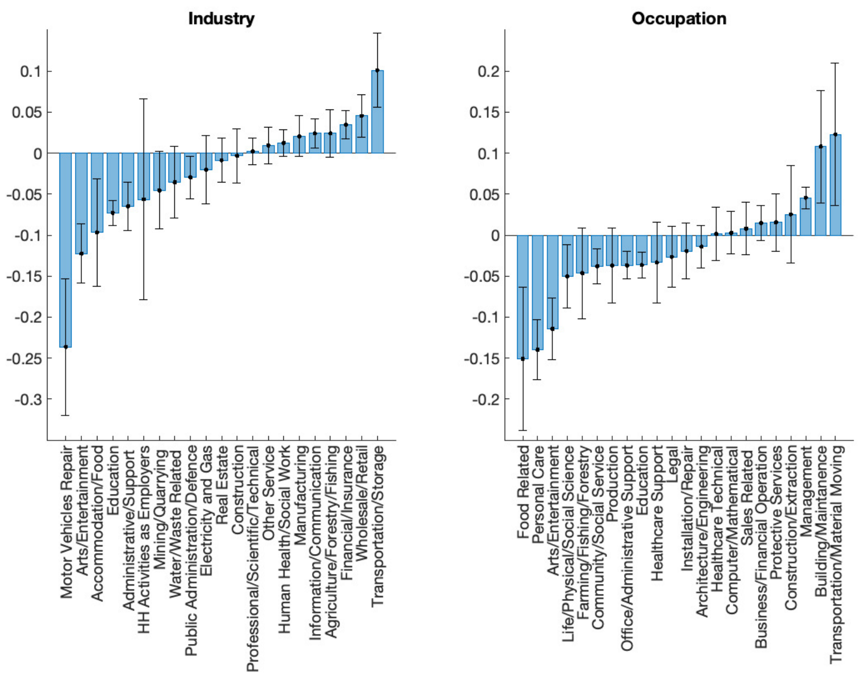
Cross-walk from 3-digit SOC 2000 to 2-digit O*NET Classification (continued)

3-digit SOC	SOC title	2-digit O*NET	O*NET title
543*	Food preparation trades	35	Food Preparation and Serving Related
549	Skilled trades	51	Production
611	Healthcare and related personal services	31	Healthcare Support
612	Childcare and related personal services	39	Personal Care and Service
613	Animal care services	39	Personal Care and Service
621	Leisure and travel service occupations	39	Personal Care and Service
622	Hairdressers and related occupations	39	Personal Care and Service
623	Housekeeping occupations	37	Building and Grounds Cleaning and Maintenance
629	Personal services occupations N.E.C.	39	Personal Care and Service
711	Sales assistants and retail cashiers	41	Sales and Related
712	Sales related occupations	41	Sales and Related
721	Customer service occupations	43	Office and Administrative Support
811	Process operatives	51	Production
812	Plant and machine operatives	51	Production
813	Assemblers and routine operatives	51	Production
814	Construction operatives	47	Construction and Extraction
821	Transport drivers and operatives	53	Transportation and Material Moving
822	Mobile Machine Drivers And Operatives	53	Transportation and Material Moving
911	Elementary Agricultural Occupations	45	Farming, Fishing, and Forestry
912	Elementary construction occupations	47	Construction and Extraction
913	Elementary process plant occupations	51	Production
914	Elementary goods storage occupations	53	Transportation and Material Moving
921	Elementary administration occupations	43	Office and Administrative Support
922	Elementary personal services occupations	39	Personal Care and Service
923	Elementary cleaning occupations	37	Building and Grounds Cleaning and Maintenance
924	Elementary security occupations	33	Protective Service
925	Elementary sales occupations	41	Sales and Related

Note: Part of occupation 922 is allocated to O*NET occupation 35 Food Preparation and Serving Related. See text for more details.

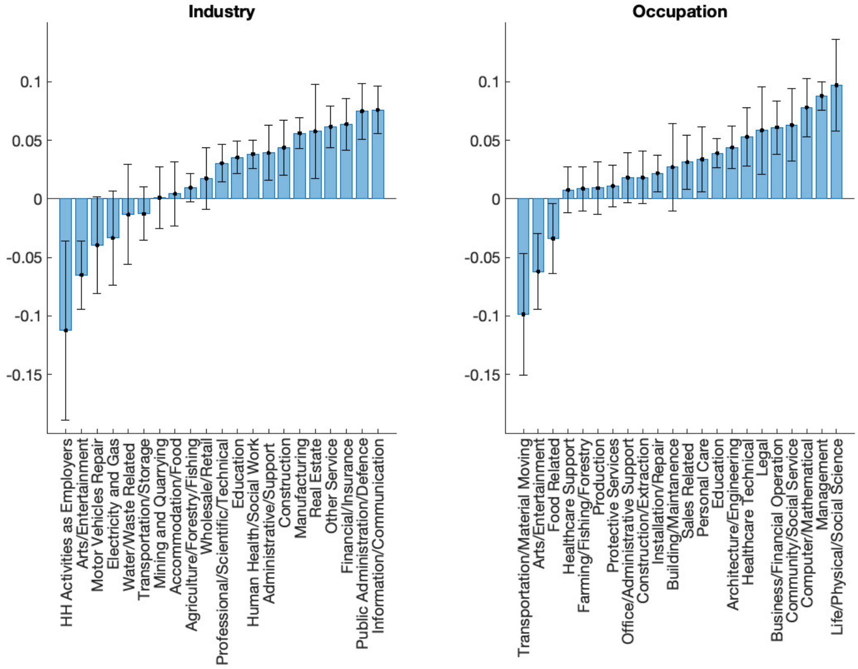
Appendix C Additional Figures and Tables

Figure C.1: Mean Productivity Change in June 2020, by Industry and by Occupation



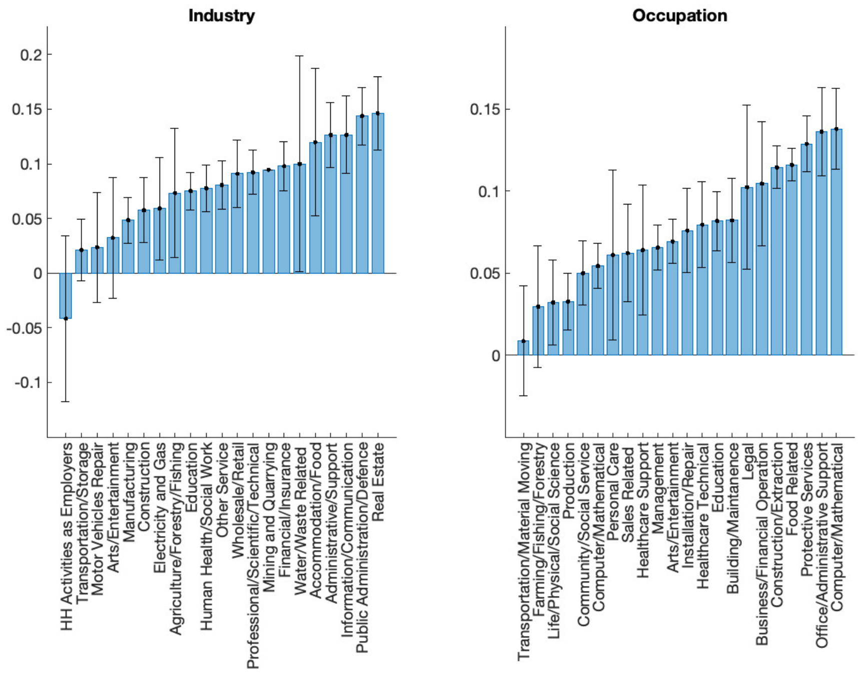
Note: This figure depicts the mean semi-standardized productivity change by industry (left) and occupation (right) from January/February 2020 to June 2020 using UKHLS Covid-19 module data. The lines correspond to the 95% confidence interval. Occupation information is taken from the 2019 UKHLS main survey responses and is converted into the 2-digit O*NET codes. See Appendix B.2 for additional details.

Figure C.2: Mean Productivity Change in September 2020, by Industry and by Occupation



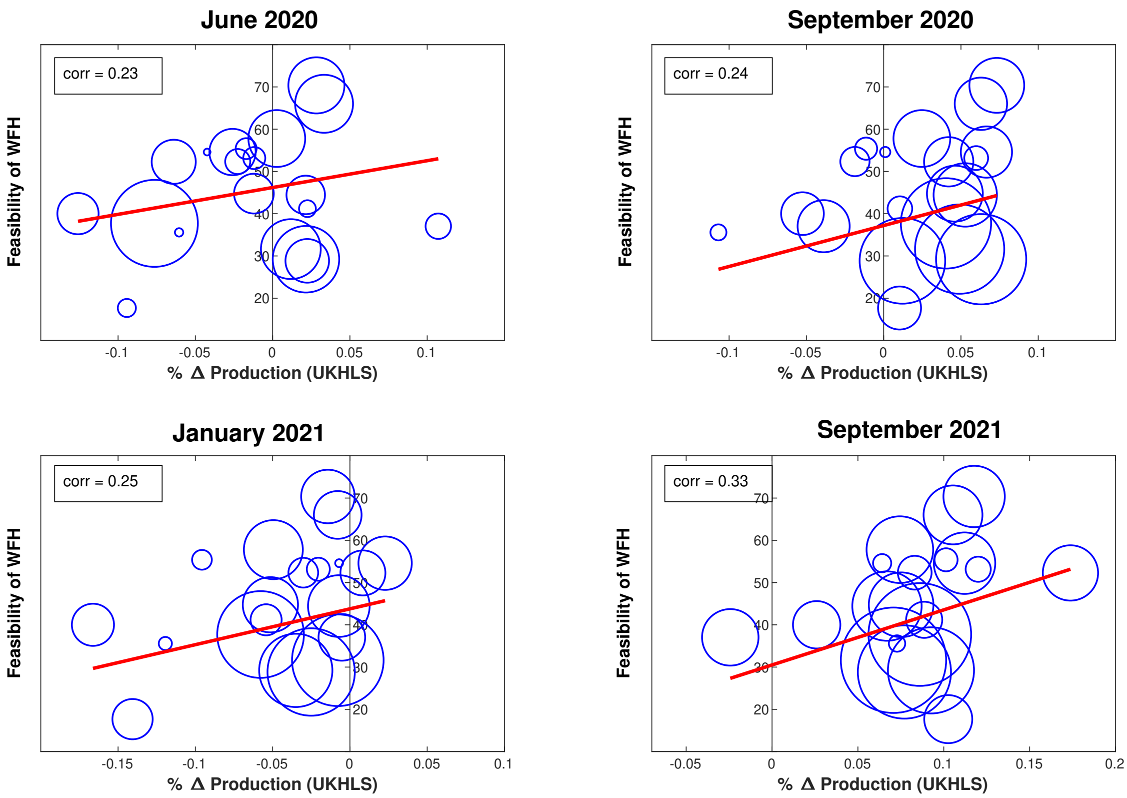
Note: This figure depicts the mean semi-standardized productivity change by industry (left) and occupation (right) from January/February 2020 to September 2020 using UKHLS Covid-19 module data. The lines correspond to the 95% confidence interval. Occupation information is taken from the 2019 UKHLS main survey responses and is converted into the 2-digit O*NET codes. See Appendix B.2 for additional details.

Figure C.3: Mean Productivity Change in September 2021, by Industry and by Occupation



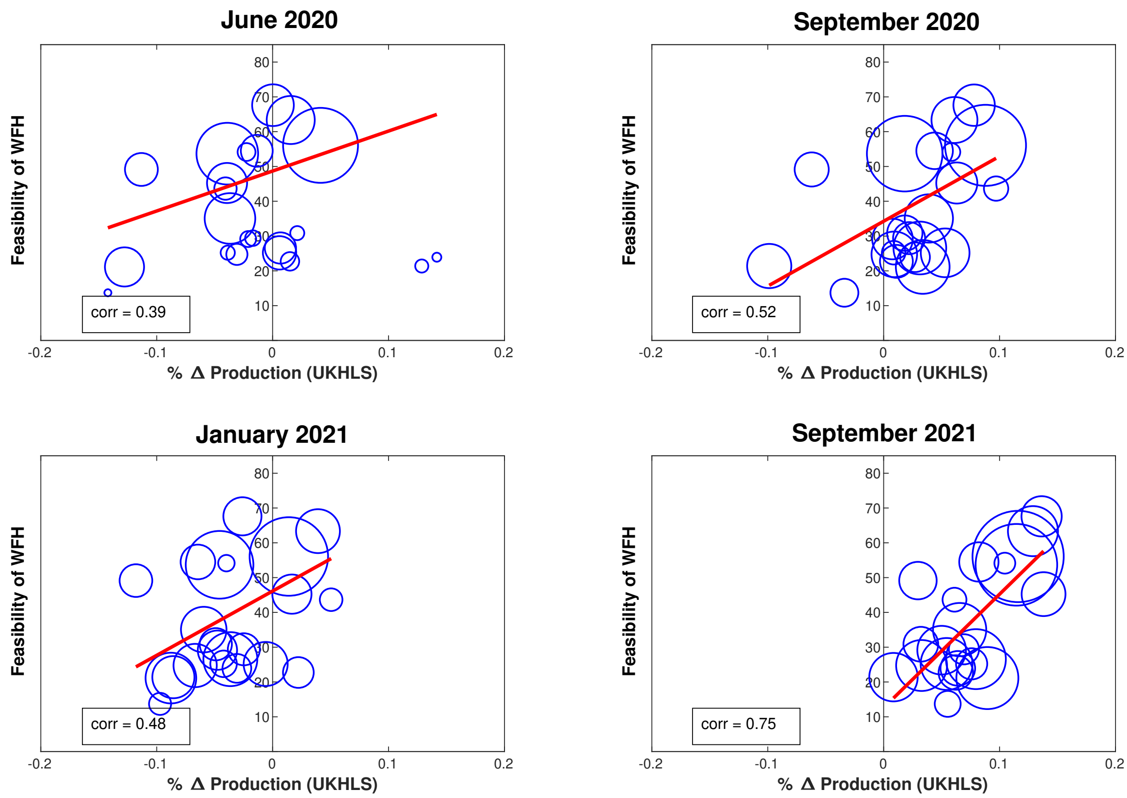
Note: This figure depicts the mean semi-standardized productivity change by industry (left) and occupation (right) from January/February 2020 to September 2021 using UKHLS Covid-19 module data. The lines correspond to the 95% confidence interval. Occupation information is taken from the 2019 UKHLS main survey responses and is converted into the 2-digit O*NET codes. See Appendix B.2 for additional details.

Figure C.4: Productivity Changes and WFH Feasibility by Industry



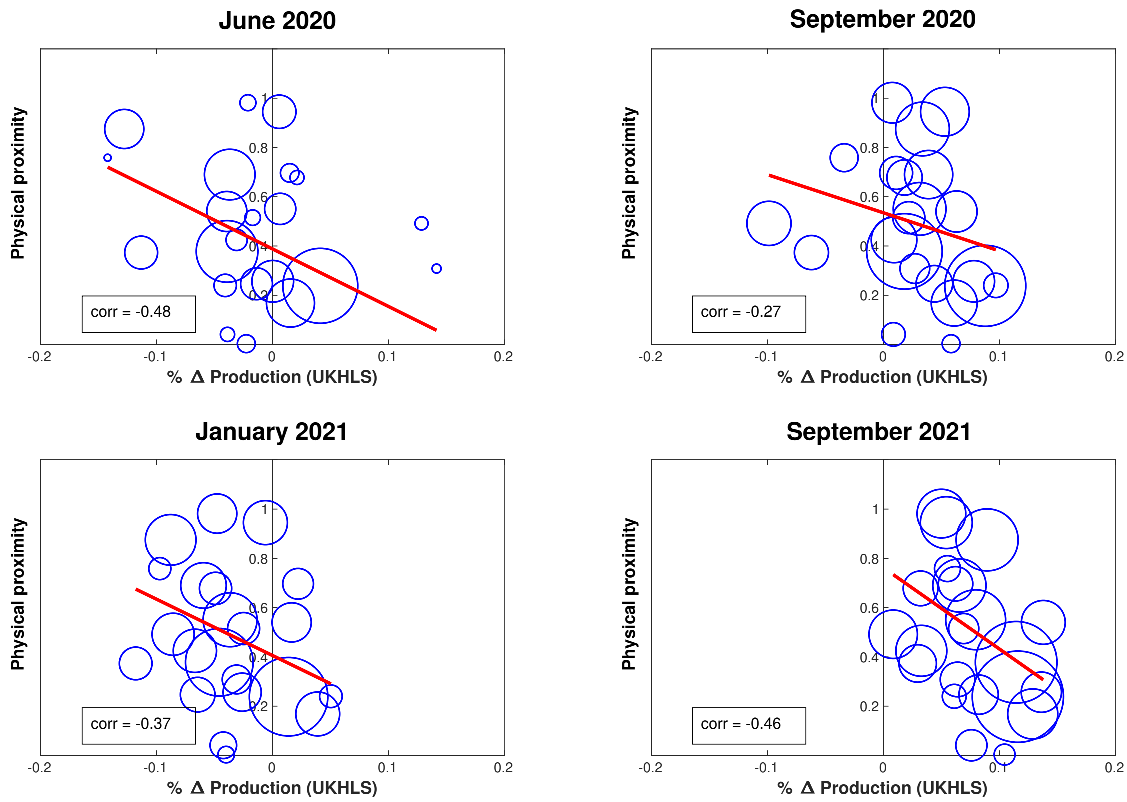
Note: Figure shows scatter plots of productivity changes against the measure of feasibility of WFH from Adams-Prassl et al. (2022), by baseline industry measured in the Covid survey, and by survey wave. Bubble sizes are proportional to industry employment. The solid line is the line of (weighted) best fit. See text for more details.

Figure C.5: Productivity Changes and WFH Feasibility by Occupation



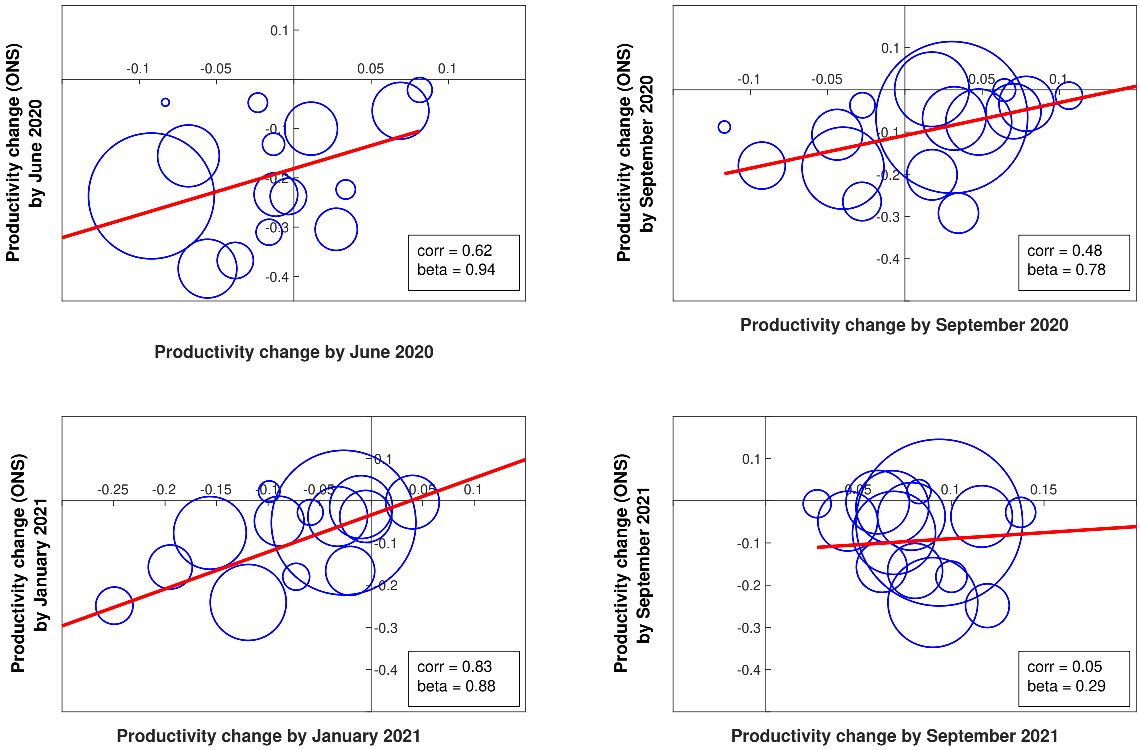
Note: Figure shows scatter plots of productivity changes against the measure of feasibility of WFH from Adams-Prassl et al. (2022), by baseline occupation measured in the Covid survey, and by survey wave. Bubble sizes are proportional to occupation employment. The solid line is the line of (weighted) best fit. See text for more details.

Figure C.6: Productivity Changes and Physical Proximity Needed for Job by Occupation



Note: Figure shows scatter plots of productivity changes against measure of physical proximity in job from Mongey et al. (2021), by occupation and by survey wave. Bubble sizes are proportional to occupation employment. Solid line is a line of (weighted) best fit. Occupation is from 2019 Covid Survey, converted to 2-digit O*NET code. See Appendix B.1 for fuller discussion and main text for further details.

Figure C.7: Production Changes: Self Reported vs ONS aggregate



Note: Data from Office for National Statistics and Covid module of UKHLS. Figure shows scatter plots of aggregate production changes estimated using the UKHLS Covid data against external ONS aggregate data by industry. See Appendix B.2 for detailed details of the computation.

Table C.1: Summary Statistics

	N	Min	Max	Mean	Standard deviation
Demographic					
Male[ℵ]	19,293	0	1	0.47	0.50
Age[ℵ]	19,293	17	65	43.24	12.14
Degree[*]	19,293	0	1	0.42	0.49
Children in hh[ℵ]	19,293	0	1	0.35	0.48
Region of residence[*]	19,293	1	12	6.27	3.00
London	19,293	0	1	0.11	0.31
South East	19,293	0	1	0.14	0.35
Married[*]	19,293	0	1	0.65	0.48
Race[ℵ]	19,293	1	4	1.14	0.51
White	19,293	0	1	0.92	0.27
Covid Work					
Working from home[ℵ]	19,293	1	4	2.68	1.32
Always	19,293	0	1	0.32	0.47
Never	19,293	0	1	0.44	0.50
Baseline period wfh[♣]	19,293	1	4	3.53	0.80
Always	19,293	0	1	0.04	0.20
Never	19,293	0	1	0.68	0.47
Productivity change (qualitative)[ℵ]	19,293	1	5	3.17	0.94
Imputed productivity change (quantitative)[ℵ]	19,293	-0.80	0.78	0.04	0.25
Other Employment					
Baseline monthly net earnings[♣]	19,293	125	17,200	1,901	1,225
Self-employed[*]	19,293	0	1	0.05	0.22
Managerial duties[*]	19,293	0	1	0.48	0.50
Size of firm[*]	19,293	1	12	5.80	2.84
1000 + employees	19,293	0	1	0.17	0.37
Industry[ℵ]	19,293	1	22	13.23	5.53
Occupation[*]	19,293	11	55	30.02	13.90
Commuting to work					
Distance to work[*]	18,557	1	100	11.26	14.80
Commuting mode[*]	18,557	1	3	1.31	0.58
Car	18,557	0	1	0.75	0.43
Difficulties travelling to work[*]	18,557	0	1	0.48	0.50
Housing					
People in household[*]	19,293	1	11	3.03	1.30
Number of rooms in home[*]	19,293	2	10	5.03	1.69
Own home[*]	19,293	0	1	0.74	0.44
Home has internet access[*]	19,293	0	1	0.98	0.12
All who wfh have desk space[ℵ]	19,293	0	1	0.79	0.41
Individual Traits					
Agreeableness[*]	13,552	1	7	5.51	1.01
Conscientiousness[*]	13,552	2	7	5.50	0.99
Extraversion[*]	13,552	1	7	4.53	1.27
Neuroticism[*]	13,552	1	7	4.57	1.18
Openness[*]	13,552	1	7	3.67	1.36

Note: * - Underlying data comes from UKHLS main survey waves. ℵ - Underlying data comes from Covid survey waves. ♣ - Underlying data comes from Covid survey and refers to Jan/Feb 2020. The sample contains working age individuals (17-65) who report being employed or self-employed and are not on furlough. If the individual reports being in work but works 0 hours (less than 5 hours), they are presumed to be on furlough (from wave 4 on wards). Individuals are considered to be married is they are legally married, in a civil union or are cohabiting with a partner. Productivity change variables ask individuals to compare their current productivity to the baseline period Jan-Feb 2020. Difficulties travelling to work are recorded for those who travel by private transport or by public transport. The latter is only asked in UKHLS main survey wave 10. Individual skills information was collected in the third wave of the UKHLS main survey and corresponds to the question about agreeableness. Earnings, the total number of rooms in the house and distance to work have been winsorized at the 99th percentile. Missing variables are imputed for the desk space variable by estimating a probit regression of desk space on individual controls, employment controls and housing controls and obtaining predicted values. If the predicted value was above 0 the individual was assumed to have enough desk space in their household. Survey weights are used throughout.

Table C.2: Proportions WFH By Industry

	Jan/Feb'20	April '20	Change April to June '20	Change June to Sept '20	Change Sept'20 to Jan'21	Change Jan 20 Sept '21
Agriculture/Forestry/Fishing	0.25***	0.30***	-0.07**	0.06	0.02	-0.04
	(0.07)	(0.08)	(0.04)	(0.04)	(0.04)	(0.04)
Mining and Quarrying	0.15	0.50***	-0.07	-0.03	0.06	0.00
	(0.09)	(0.15)	(0.07)	(0.03)	(0.06)	(.)
Manufacturing	0.19***	0.36***	-0.03	-0.00	0.05***	-0.05***
	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Electricity and Gas	0.31***	0.57***	-0.01	-0.09*	0.06*	-0.03
	(0.06)	(0.08)	(0.05)	(0.05)	(0.04)	(0.11)
Water/Waste Related	0.24***	0.47***	0.06	-0.04	-0.00	-0.01
	(0.06)	(0.09)	(0.04)	(0.06)	(0.05)	(0.04)
Construction	0.22***	0.35***	-0.05**	0.03	0.02	-0.08
	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.06)
Wholesale/Retail	0.14***	0.21***	-0.01	0.00	0.03*	-0.03
	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)
Motor Vehicles Repair	0.20***	0.23***	-0.14	-0.04	0.24*	-0.08
	(0.07)	(0.07)	(0.13)	(0.07)	(0.13)	(0.07)
Transportation/Storage	0.12***	0.20***	-0.02	0.04	-0.02	-0.02
	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)
Accommodation/Food	0.14***	0.15***	0.04*	0.05	0.00	-0.05
	(0.03)	(0.03)	(0.02)	(0.04)	(0.03)	(0.08)
Information/Communication	0.63***	0.84***	-0.05*	0.03	0.03	-0.01
	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
Financial/Insurance	0.47***	0.84***	0.01	-0.05**	0.04*	-0.04**
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Real Estate	0.44***	0.70***	-0.02	0.10	-0.03	-0.09
	(0.06)	(0.06)	(0.04)	(0.07)	(0.05)	(0.06)
Professional/Scientific/Technical	0.53***	0.80***	-0.04	0.01	0.04	-0.07**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Administrative/Support	0.32***	0.65***	0.00	-0.04	0.12***	-0.06**
	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.03)
Public Administration/Defence	0.37***	0.69***	-0.00	-0.02	0.03	-0.04**
	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Education	0.31***	0.71***	-0.02	-0.26***	0.25***	-0.29***
	(0.02)	(0.02)	(0.01)	(0.02)	(0.03)	(0.03)
Human Health/Social Work	0.25***	0.39***	0.00	0.00	0.05***	-0.03
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)
Arts/Entertainment	0.50***	0.61***	-0.05	0.07	0.15**	-0.10**
	(0.05)	(0.05)	(0.05)	(0.04)	(0.07)	(0.05)
Other Service	0.29***	0.42***	-0.04**	0.02	-0.01	-0.03
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
HH Activities as Employers	0.15**	0.21**	-0.06	0.05	-0.04	0.16
	(0.06)	(0.08)	(0.08)	(0.04)	(0.04)	(0.12)
Missing	0.25***	0.40***	0.02	0.09	-0.05	-0.01
	(0.01)	(0.02)	(0.02)	(0.08)	(0.05)	(0.05)
Observations	10,408	9,839	7,023	6,086	5,327	5,016
Adjusted R^2	0.333	0.560	0.007	0.066	0.074	0.086

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Note: Current estimates from a regression with a dummy variable which is set equal to 1 if any WFH is reported. All possible observations of working age individuals are used; no restrictions are currently placed on the sample.

Table C.3: Proportions WFH By Occupation

	Jan/Feb'20	April '20	Change April to June '20	Change June to Sept '20	Change Sept'20 to Jan'21	Change Jan 20 Sept '21
Management	0.48***	0.66***	-0.01	0.03**	0.01	-0.04*
	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)
Business/Financial Operation	0.55***	0.90***	-0.04**	-0.02	0.02	-0.06**
	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)
Computer/Mathematical	0.61***	0.87***	0.03	-0.00	0.01	0.03
	(0.04)	(0.02)	(0.02)	(0.03)	(0.03)	(0.03)
Architecture/Engineering	0.35***	0.71***	-0.04*	-0.05	0.05	-0.10***
	(0.03)	(0.04)	(0.02)	(0.04)	(0.03)	(0.03)
Life/Physical/Social Science	0.34***	0.73***	-0.03	0.01	-0.00	0.01
	(0.05)	(0.05)	(0.07)	(0.04)	(0.03)	(0.03)
Community/Social Service	0.50***	0.80***	-0.03	0.02	0.06**	-0.07**
	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)
Legal	0.47***	0.82***	0.01	0.04	0.05	-0.10**
	(0.06)	(0.04)	(0.02)	(0.04)	(0.05)	(0.05)
Education	0.51***	0.88***	-0.01	-0.30***	0.32***	-0.33***
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)
Arts/Entertainment	0.60***	0.75***	-0.07	0.05	0.09**	-0.08*
	(0.04)	(0.03)	(0.06)	(0.05)	(0.04)	(0.04)
Healthcare Technical	0.25***	0.38***	-0.01	-0.01	0.02	-0.01
	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)
Healthcare Support	0.11***	0.16***	0.00	-0.00	0.04*	0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Protective Services	0.12***	0.24***	-0.05*	-0.08*	0.08**	-0.03
	(0.03)	(0.04)	(0.03)	(0.05)	(0.03)	(0.02)
Food Related	0.09***	0.08***	-0.00	0.08**	-0.00	-0.02
	(0.03)	(0.03)	(0.03)	(0.04)	(0.02)	(0.04)
Building/Maintenance	0.12***	0.10***	-0.02	0.04	-0.03*	0.08
	(0.03)	(0.03)	(0.04)	(0.03)	(0.02)	(0.06)
Personal Care	0.14***	0.33***	0.01	-0.17***	0.18***	-0.23***
	(0.02)	(0.03)	(0.02)	(0.03)	(0.05)	(0.04)
Sales Related	0.15***	0.27***	-0.04**	0.03	0.04**	-0.07**
	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.03)
Office/Administrative Support	0.21***	0.51***	0.00	-0.05***	0.08***	-0.11***
	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.03)
Farming/Fishing/Forestry	0.21***	0.30***	-0.04	0.00	0.07	-0.04
	(0.05)	(0.07)	(0.03)	(0.02)	(0.04)	(0.03)
Construction/Extraction	0.11***	0.17***	-0.11**	0.05	-0.04	0.03
	(0.02)	(0.03)	(0.05)	(0.03)	(0.03)	(0.04)
Installation/Repair	0.19***	0.34***	-0.10	0.02	0.10**	-0.12
	(0.05)	(0.06)	(0.07)	(0.06)	(0.04)	(0.10)
Production	0.16***	0.21***	-0.03	-0.02	0.04*	-0.01
	(0.03)	(0.03)	(0.04)	(0.02)	(0.02)	(0.01)
Transportation/Material Moving	0.08***	0.08***	-0.01	0.01	0.00	-0.00
	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
Missing	0.15***	0.31***	-0.04	-0.02	0.11**	-0.09
	(0.03)	(0.05)	(0.04)	(0.05)	(0.06)	(0.07)
Observations	10,408	9,839	7,023	6,086	5,327	5,016
Adjusted R^2	0.391	0.613	0.007	0.053	0.068	0.086

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Note: Current estimates from a regression with a dummy variable which is set equal to 1 if any WFH is reported. All possible observations of working age individuals are used; no restrictions are currently placed on the sample.

Table C.4: Changes in Productivity During Covid-19 by Characteristics - Age of Children

	(1) June'20	(2) Sept.'20	(3) Jan.'21	(4) Sept.'21	(5) June'20	(6) Sept.'20	(7) Jan.'21	(8) Sept.'21	(9) June'20	(10) Sept.'20	(11) Jan.'21	(12) Sept.'21
Children 0-15												
Parent × female	-5.01***	6.51***	-3.46***	7.68***								
	(1.26)	(1.18)	(1.30)	(1.15)								
Parent × male	0.36	5.24***	1.46	8.49***								
	(1.34)	(0.91)	(2.09)	(1.31)								
No children × female	-1.48	5.06***	0.87	10.22***								
	(1.30)	(0.87)	(1.05)	(0.69)								
No children × male	2.05*	5.21***	0.50	8.77***								
	(1.09)	(0.77)	(0.94)	(0.91)								
Children 0-4												
Mother					-6.46***	6.00**	-2.49	9.05***				
					(2.33)	(2.68)	(2.27)	(2.30)				
Father					2.83	4.08***	-4.77**	12.09***				
					(2.27)	(1.12)	(1.99)	(1.60)				
Children 5-15												
Mother									-5.89***	6.71***	-3.87***	7.90***
									(1.35)	(1.20)	(1.42)	(1.14)
Father									-0.03	5.26***	2.18	8.15***
									(1.43)	(0.98)	(2.37)	(1.47)
N	3,498	5,533	4,753	5,509	3,498	5,533	4,753	5,509	3,498	5,533	4,753	5,509

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses.

Note: Same specification as Table 2. See Table 2 notes and text for further details.

Appendix D Further Details on the Selection Model Presented in Section 5

Section 5 presented our selection model concisely. We now lay out the empirical framework in further detail. In what follows, in the spirit of French and Taber (2011) we discuss identification non-parametrically. It should be borne in mind that, equally in the spirit of French and Taber (2011), we estimate the model in our empirical application in a simple linear setting.

We first recap the basic ingredients of the model presented in the main text. Productivity is as follows:

$$
\begin{aligned}
prod_{it}^{h} &= g^{h}\left(X_{it}\right) + \epsilon_{it}^{h} \\
prod_{it}^{f} &= g^{f}\left(X_{it}\right) + \epsilon_{it}^{f}
\end{aligned}
\tag{6}
$$

We allow for utility, V_{it}^{h}, of costs or benefits of WFH compared to being located in the standard workplace. This is specified as follows:

$$
V_{it}^{h} = k\left(z_i, X_{it}\right) + \nu_{it} \tag{7}
$$

Given this set-up the decision rule is simple, and specified as follows:

$$
j_{it}^{*} = \begin{cases} h & \text{if } prod_{it}^{h} - prod_{it}^{f} + V_{it}^{h} > 0 \\ f & \text{otherwise} \end{cases} \tag{8}
$$

The fundamental identification problem that we need to address is that $\mathbb{E}\left[\epsilon_{t}^{j} | X_t, j^{*}\right]$ is likely not equal to zero for $j^{*} \in \{f, h\}$. i.e. individuals are selected by idiosyncratic productivity in their observed location. As such, properties of $g^{j}\left(\right)$ cannot be identified immediately. However, we maintain the standard argument of 'identification at infinity', and suppose that at extreme values of z, utility-based preferences for each location are so strong that productivity no longer plays a role. Suppose that, as $z \to \infty$, then individuals prefer home, and as $z \to -\infty$ individuals prefer the workplace, then formally, and dropping some subscripts, we use:

$$\lim_{z \to \infty} \mathbb{E}\left[\epsilon_t^h | X, j^*, z\right] = \lim_{z \to -\infty} \mathbb{E}\left[\epsilon_t^f | X, j^*, z\right] = 0 \tag{9}$$

Next consider the baseline period 0, before the pandemic. We use a simpler production function and location choice:

$$\begin{aligned}
prod_{i0}^j &= l^j\left(X_{i0}\right) + \epsilon_{i0}\ , j = h, f \\
V_{i0}^h &= m\left(X_{i0}\right) + \nu_{i0} \\
j_{it}^* &= \begin{cases} h & \text{if } prod_{i0}^h - prod_{i0}^f + V_{i0}^h > 0 \\ f & \text{otherwise} \end{cases}
\end{aligned}$$

where $l^j\left(\right)$ may differ from $g^j\left(\right)$ because production may differ during the pandemic from before. Further note two simplifications of this pre-pandemic model compared to (6), (7) and (8): the idiosyncratic component ϵ_{i0} does not depend on location, and we do not require any variable to affect $m\left(\right)$ that is excludable from the production function. In practice, the first assumption ensures that idiosyncratic productivity is exogenous of observed location, and so that location at time 0 can be treated as 'given'. This ensures that an additional instrument is not required. Formally:

$$\mathbb{E}\left[\epsilon_0 | X_0, j_0^*\right] = 0 \tag{10}$$

As discussed in the main text quasi-differences in productivity are defined as follows:

$$\begin{aligned}
\tilde{\Delta} prod_{it}^j &\equiv prod_{it}^j - prod_{i0}^{j^*} \\
&= g^j\left(X_{it}\right) + \epsilon_{it}^j - \left(l^{j_0^*}\left(X_{i0}\right) + \epsilon_{i0}\right)
\end{aligned} \tag{11}$$

which, importantly, captures the change in productivity at time t in each location j compared to the *observed* location j_0^* at time zero.

Building on (8) since

$$
\begin{aligned}
& prod_{it}^{h} - prod_{it}^{f} + V_{it}^{h} > 0 \\
\iff & \left(prod_{it}^{h} - prod_{i0}^{j^*}\right) - \left(prod_{it}^{f} - prod_{i0}^{j^*}\right) + V_{it}^{h} > 0 \\
\iff & \tilde{\Delta} prod_{it}^{h} - \tilde{\Delta} prod_{it}^{f} + V_{it}^{h} > 0.
\end{aligned}
$$

It's the case that,

$$
j_{it}^{*} = \begin{cases} h & \text{if } \tilde{\Delta} prod_{it}^{h} - \tilde{\Delta} prod_{it}^{f} + V_{it}^{h} > 0. \\ f & \text{otherwise.} \end{cases} \tag{12}
$$

Finally we come to identification. We observe j_{it}^* , $\Delta prod_{it} \equiv \tilde{\Delta} prod_{it}^{j^*}$ and the full array of covariates. Exploiting orthogonality conditions (9), (10) and the definition of quasi-differences in (11) then we observe the following regression functions:

$$
\begin{aligned}
\lim_{z \to \infty} \mathbb{E}\left[\Delta prod_{it} | X, j_{i0}^{*} = \bar{j}, z\right] &= g^{h}\left(X_t\right) - l^{\bar{j}}\left(X_0\right) \\
\lim_{z \to -\infty} \mathbb{E}\left[\Delta prod_{it} | X, j_{i0}^{*} = \bar{j}, z\right] &= g^{f}\left(X_t\right) - l^{\bar{j}}\left(X_0\right)
\end{aligned}
$$

Intuitively, we can both condition on baseline location as given, and condition on pandemic location using the exclusion restrictions.

The model therefore permits identification of key parameters. First, and using economical notation, average treatment effects are identified as follows:

$$
\lim_{z \to \infty} \mathbb{E}\left[\Delta prod_{it} | ...\right] - \lim_{z \to -\infty} \mathbb{E}\left[\Delta prod_{it} | ...\right] = g^{h}\left(X_t\right) - g^{f}\left(X_t\right)
$$

In our empirical application we focus on marginal effects on the production function for different characteristics. To use a concrete example, we want to examine the effect of having adequate home desk space (say $D = 1$) compared to inadequate desk space ($D = 0$) on the pandemic productivity change for those WFH. We identify this as follows:

$$\lim_{z \to \infty} \mathbb{E}\left[\Delta prod_{it} | D = 1, ...\right] - \lim_{z \to \infty} \mathbb{E}\left[\Delta prod_{it} | D = 0, ...\right] = \left(g^h\left(D = 1\right) - l^{j_0^*}\left(D = 1\right)\right)$$
$$- \left(g^h\left(D = 0\right) - l^{j_0^*}\left(D = 0\right)\right)$$

If we are willing to push this further, and maintain the assumption that desk space at home should not affect productivity at work, then we can impose that $l^f\left(D = 1\right) = l^f\left(D = 0\right)$, and then identify $g^h\left(D = 1\right) - g^h\left(D = 0\right)$.

References

AASSVE, A., F. C. BILLARI, S. MAZZUCO, AND F. ONGARO (2002): “Leaving home: A comparative analysis of ECHP data,” Journal of European social policy, 12, 259–275.

ADAMS-PRASSL, A., T. BONEVA, M. GOLIN, AND C. RAUH (2020): “Inequality in the Impact of the Coronavirus Shock: Evidence from Real Time Surveys,” Journal of Public Economics, 189.

ADAMS-PRASSL, A., T. BONEVA, M. GOLIN, AND C. RAUH (2022): “Work that can be done from home: Evidence on variation within and across occupations and industries,” Labour Economics, 74, 102083.

AKSOY, C. G. (2016): “The effects of unemployment on fertility: evidence from England,” The BE Journal of Economic Analysis & Policy, 16, 1123–1146.

AKSOY, C. G., J. M. BARRERO, N. BLOOM, S. J. DAVIS, M. DOLLS, AND P. ZARATE (2022):“Working from home around the world,” Tech. rep., National Bureau of Economic Research.

ALIPOUR, J.-V., O. FALCK, AND S. SCHULLER (2023): “Germany’s capacity to work from home,” European Economic Review, 151, 104354.

ALMLUND, M., A. L. DUCKWORTH, J. HECKMAN, AND T. KAUTZ (2011): “Personality psychology and economics,” in Handbook of the Economics of Education, Elsevier, vol. 4, 1–181.

ALON, T., S. COSKUN, M. DOEPKE, D. KOLL, AND M. TERTILT (2022): “From mancession to shecession: Women’s employment in regular and pandemic recessions,” NBER Macroeconomics Annual, 36, 83–151.

ANDERSON, A. AND L. SMITH (2010): “Dynamic matching and evolving reputations,” The Review of Economic Studies, 77, 3–29.

ANDREW, A., S. CATTAN, M. COSTA DIAS, C. FARQUHARSON, L. KRAFTMAN, S. KRUTIKOVA, A. PHIMISTER, AND A. SEVILLA (2020): “How Are Mothers and Fathers Balancing Work and Family under Lockdown?” IFS Briefing Note BN290.

ANDREW, M. AND G. MEEN (2003): “Housing transactions and the changing decisions of young households in Britain: the microeconomyic evidence,” Real Estate Economics, 31, 117–138.

ANGRIST, J. D. AND J.-S. PISCHKE (2009): Mostly harmless econometrics: An empiricist’s companion, Princeton university press.

ATKIN, D., A. SCHOAR, AND S. SHINDE (2023): “Worker sorting, work discipline and development," Tech. rep., Technical report, MIT Sloan working paper.

AUTOR, D., D. DORN, AND G. HANSON (2019): "When work disappears: Manufacturing decline and the falling marriage market value of young men," American Economic Review: Insights, 1, 161–178.

AUTOR, D., A. DUBE, AND A. MCGREW (2023): "The unexpected compression: Competition at work in the low wage labor market," Tech. rep., National Bureau of Economic Research.

BAGGER, J. AND R. LENTZ (2019): "An empirical model of wage dispersion with sorting," The Review of Economic Studies, 86, 153–190.

BAQAEE, D. AND E. FARHI (2022): "Supply and demand in disaggregated Keynesian economies with an application to the Covid-19 crisis," American Economic Review, 112, 1397–1436.

BARKER, K. (2006): Barker review of land use planning: final report, recommendations, The Stationery Office.

BARRERO, J. M., N. BLOOM, AND S. J. DAVIS (2021a): "Internet access and its implications for productivity, inequality, and resilience," Tech. rep., National Bureau of Economic Research.

BARRERO, J. M., N. BLOOM, AND S. J. DAVIS (2021b): "Why working from home will stick," Tech. rep., National Bureau of Economic Research.

BARRERO, J. M., N. BLOOM, S. J. DAVIS, B. H. MEYER, AND E. MIHAYLOV (2022): "The Shift to Remote Work Lessens Wage-Growth Pressures," Tech. rep., National Bureau of Economic Research.

BAYRAKDAR, S. AND R. COULTER (2018): "Parents, local house prices, and leaving home in Britain," Population, Space and Place, 24, e2087.

BEAUJOUAN, É. AND M. NÍ BHROLCHÁIN (2011): "Cohabitation and marriage in Britain since the 1970s," Population trends, 145, 35–59.

BECKER, G. S. (1973): "A theory of marriage: Part I," Journal of Political Economy, 81, 813–846.

BECKER, G. S. (1974): "A theory of marriage: Part II," Journal of political Economy, 82, S11–S26.

BECKER, S. O., S. BENTOLILA, A. FERNANDES, AND A. ICHINO (2010): "Youth emancipation and perceived job insecurity of parents and children," Journal of population Economics, 23, 1047–1071.

BELLOC, F. (2009): "Are we really mama's boys? How incomes affect Italians' leaving home decisions," Journal of Income Distribution, 18, 63.

BERRINGTON, A. AND J. STONE (2014): "Young adults' transitions to residential independence in the UK: The role of social and housing policy," in Young people and social policy in Europe: Dealing with risk, inequality and precarity in times of crisis, Springer, 210–235.

BERRINGTON, A. AND J. STONE (2015): "Cohabitation trends and patterns in the UK," .

BICK, A., A. BLANDIN, AND K. MERTENS (2021): "Work from home before and after the Covid-19 outbreak," Available at SSRN 3786142.

BLEEMER, Z., M. BROWN, D. LEE, W. VAN DER KLAAUW, ET AL. (2014): "Tuition, jobs, or housing: What's keeping millennials at home," Staff Reports, 700.

BLOOM, N. (2020): "How Working from Home Works out," Tech. rep.

BLOOM, N., J. LIANG, J. ROBERTS, AND Z. J. YING (2015): "Does Working from Home Work? Evidence from a Chinese Experiment," The Quarterly Journal of Economics, 130, 165–218.

BLUNDELL, R., D. A. GREEN, AND W. JIN (2016): "The UK wage premium puzzle: How did a large increase in university graduates leave the education premium unchanged?" Tech. rep., IFS Working Papers.

BLUNDELL, R., D. A. GREEN, AND W. JIN (2022): "The UK as a technological follower: Higher education expansion and the college wage premium," The Review of Economic Studies, 89, 142–180.

BONADIO, B., Z. HUO, A. A. LEVCHENKO, AND N. PANDALAI-NAYAR (2021): "Global supply chains in the pandemic," Journal of international economics, 133, 103534.

BONILLA, R. AND K. BURDETT (2010): "On-the-job search and labor market equilibrium," The BE Journal of Macroeconomics, 10.

BOWMAKER, S. W. AND P. M. EMERSON (2015): "Bricks, mortar, and wedding bells: does the cost of housing affect the marriage rate in the US?" Eastern Economic Journal, 41, 411–429.

BRIEN, M. J., L. A. LILLARD, AND S. STERN (2006): "Cohabitation, marriage, and divorce in a model of match quality," International Economic Review, 47, 451–494.

BRINKLEY, I., B. WILLMOTT, M. BEATSON, AND G. DAVIES (2020): "Embedding new ways of working: Implications for the post-pandemic workplace," .

BROWNING, M. AND P.-A. CHIAPPORI (1998): "Efficient intra-household allocations: A general characterization and empirical tests," Econometrica, 1241–1278.

BRUECKNER, J. K., M. E. KAHN, AND G. C. LIN (2023): “A new spatial hedonic equilibrium in the emerging work-from-home economy?” American Economic Journal: Applied Economics, 15, 285–319.

BRUZE, G., M. SVARER, AND Y. WEISS (2015): “The dynamics of marriage and divorce,” Journal of Labor Economics, 33, 123–170.

BRYNJOLFSSON, E., J. J. HORTON, A. OZIMEK, D. ROCK, G. SHARMA, H.-Y. TUYE, AND A. O. UPWORK (2020): “COVID-19 and Remote Work: An Early Look at US Data,” .

BURDETT, K. AND M. G. COLES (1997): “Marriage and class,” The Quarterly Journal of Economics, 112, 141–168.

BURDETT, K., R. IMAI, AND R. WRIGHT (2004): “Unstable relationships,” The BE Journal of Macroeconomics, 1.

BURDETT, K. AND D. T. MORTENSEN (1998): “Wage differentials, employer size, and unemployment,” International Economic Review, 257–273.

BURDETT, K. AND R. WRIGHT (1998): “Two-sided search with nontransferable utility,” Review of Economic Dynamics, 1, 220–245.

CAHUC, P., F. POSTEL-VINAY, AND J.-M. ROBIN (2006): “Wage bargaining with on-the-job search: Theory and evidence,” Econometrica, 74, 323–364.

CALVO, P. (2022): “The Effects of Institutional Gaps between Marriage and Cohabitation,” .

CAMERON, D. (2012): “Welfare Speech,” Remarks by David Cameron, Prime Minister of the United Kingdon, at Bluewater Kent.

CHADE, H., J. EECKHOUT, AND L. SMITH (2017): “Sorting through search and matching models in economics,” Journal of Economic Literature, 55, 493–544.

CHEN, Y., P. CORT ́ES, G. KOS ̧AR, J. PAN, AND B. ZAFAR (2023): “The Impact of COVID-19 on Workers’ Expectations and Preferences for Remote Work,” Tech. rep., National Bureau of Economic Research.

CHESHIRE, P. AND S. SHEPPARD (2002): “The welfare economics of land use planning,” Journal of Urban economics, 52, 242–269.

CHETTY, R., L. S ́ANDOR, AND A. SZEIDL (2017): “The effect of housing on portfolio choice,” The Journal of Finance, 72, 1171–1212.

CHIAPPORI, P.-A. AND B. SALANI ́E (2023): “Mating Markets,” chap. 2.

CHIURI, M. C. AND D. DEL BOCA (2010): "Home-leaving decisions of daughters and sons," Review of Economics of the Household, 8, 393–408.

CHRISTENSEN, B. J., R. LENTZ, D. T. MORTENSEN, G. R. NEUMANN, AND A. WERWATZ (2005): "On-the-job search and the wage distribution," Journal of Labor Economics, 23, 31–58.

CHU, C. C., J.-C. LIN, AND W.-J. TSAY (2020): "Males' housing wealth and their marriage market advantage," Journal of Population Economics, 33, 1005–1023.

CISCATO, E. (2019): "The changing wage distribution and the decline of marriage," Unpublished manuscript.

CLAIR, A. (2022): "The effect of local housing allowance reductions on overcrowding in the private rented sector in England," International Journal of Housing Policy, 22, 119–137.

CORLETT, A. AND F. ODAMTTEN (2021): "Hope to buy: The decline of youth home ownership," Resolution Foundation, London. Retrieved, 1, 2022.

CORNELIUS, T. J. (2003): "A search model of marriage and divorce," Review of Economic Dynamics, 6, 135–155.

CROSSLEY, T. F., P. FISHER, P. LEVELL, AND H. LOW (2021): "A year of COVID: the evolution
of labour market and financial inequalities through the crisis," Available at SSRN 3955892.

DAVIDOFF, T. (2015): "Supply constraints are not valid instrumental variables for home prices because they are correlated with many demand factors," Available at SSRN 2400833.

DEOLE, S. S., M. DETER, AND Y. HUANG (2023): "Home sweet home: Working from home and employee performance during the COVID-19 pandemic in the UK," Labour Economics, 80, 102295.

DETTLING, L. J. AND M. S. KEARNEY (2014): "House prices and birth rates: The impact of the real estate market on the decision to have a baby," Journal of Public Economics, 110, 82–100.

DEVEREUX, K. AND L. TURNER (2016): "Divorce, Remarriage, and Fertility with On-the-Marriage Search," Tech. rep., Working Paper.

DEY, M. S. AND C. J. FLINN (2005): "An equilibrium model of health insurance provision and wage determination," Econometrica, 73, 571–627.

DINGEL, J. I. AND B. NEIMAN (2020): "How Many Jobs Can Be Done at Home?" Journal of Public Economics, 189.

D'ACUNTO, F., D. HOANG, M. PALOVIITA, AND M. WEBER (2022): "IQ, Expectations, and Choice," The Review of Economic Studies, rdac075.

EECKHOUT, J. AND P. KIRCHER (2010): "Sorting and decentralized price competition," Econometrica, 78, 539–574.

ELSBY, M. W. AND A. GOTTFRIES (2022): "Firm dynamics, on-the-job search, and labor market fluctuations," The Review of Economic Studies, 89, 1370–1419.

EMANUEL, N. AND E. HARRINGTON (2023): "Working" Remotely? Selection, Treatment, and the Market Provision of Remote Work,"

ERMISCH, J. (1981): "Economic opportunities, marriage squeezes and the propensity to marry: an economic analysis of period marriage rates in England and Wales," Population Studies, 35, 347–356.

ERMISCH, J. (1999): "Prices, parents, and young people's household formation," Journal of urban economics, 45, 47–71.

ERMISCH, J. (2002): "Trying again: Repartnering after dissolution of a union," Tech. rep., ISER Working Paper Series.

ERMISCH, J. (2016): "An economic analysis of the family,"

ERMISCH, J. AND P. DI SALVO (1997): "The economic determinants of young people's household formation," Economica, 64, 627–644.

ETHERIDGE, B., Y. WANG, AND L. TANG (2020): "Worker productivity during lockdown and working from home: Evidence from self-reports," ISER Working Paper Series.

EUROFOUND (2020): "Living, Working and COVID-19 - First Findings," Tech. rep.

FARNHAM, M., L. SCHMIDT, AND P. SEVAK (2011): "House prices and marital stability," American Economic Review, 101, 615–619.

FELSTEAD, A. AND D. REUSCHKE (2020): "Homeworking in the UK: Before and during the 2020 Lockdown," Tech. rep.

FELSTEAD, A. AND D. REUSCHKE (2021): "A flash in the pan or a permanent change? The growth of homeworking during the pandemic and its effect on employee productivity in the UK," Information Technology & People.

FERNANDEZ, R., N. GUNER, AND J. KNOWLES (2005): "Love and money: A theoretical and empirical analysis of household sorting and inequality," The Quarterly Journal of Economics, 120, 273–344.

FETZER, T. (2022): “Subsidising the spread of COVID-19: Evidence from the UK’s Eat-Out-to-Help-Out Scheme,” The Economic Journal, 132, 1200–1217.

FORREST, R., P. KENNETT, AND P. LEATHER (1999): “Home ownership in crisis? The experience of negative equity in Britain,” .

FRENCH, E. AND C. TABER (2011): “Identification of models of the labor market,” in Handbook of Labor Economics, Elsevier, vol. 4, 537–617.

FUJITA, S. AND G. RAMEY (2012): “Exogenous versus endogenous separation,” American Economic Journal: Macroeconomics, 4, 68–93.

GARROTE SANCHEZ, D., N. GOMEZ PARRA, C. OZDEN, B. RIJKERS, M. VIOLLAZ, AND H. WINKLER (2021): “Who on Earth can work from home?” The World Bank Research Observer, 36, 67–100.

GAUTIER, P. A., M. SVARER, AND C. N. TEULINGS (2009): “Sin city? Why is the divorce rate higher in urban areas?” Scandinavian Journal of Economics, 111, 439–456.

GAUTIER, P. A., M. SVARER, AND C. N. TEULINGS (2010a): “Marriage and the city: Search frictions and sorting of singles,” Journal of Urban Economics, 67, 206–218.

GAUTIER, P. A. AND C. N. TEULINGS (2015): “Sorting and the output loss due to search frictions,” Journal of the European Economic Association, 13, 1136–1166.

GAUTIER, P. A., C. N. TEULINGS, AND A. VAN VUUREN (2010b): “On-the-job search, mismatch and efficiency,” The Review of Economic Studies, 77, 245–272.

GEMICI, A. AND S. LAUFER (2011): “Marriage and cohabitation,” New York University, mimeo.

GENSOWSKI, M. (2018): “Personality, IQ, and lifetime earnings,” Labour Economics, 51, 170–183.

GHOLIPOUR, H. F. AND M. R. FARZANEGAN (2015): “Marriage crisis and housing costs: Empirical evidence from provinces of Iran,” Journal of Policy Modeling, 37, 107–123.

GIBBS, M., F. MENGEL, AND C. SIEMROTH (2023): “Work from Home and Productivity: Evidence from Personnel and Analytics Data on Information Technology Professionals,” Journal of Political Economy Microeconomics, 1, 7–41.

GLAESER, E. L. AND J. GYOURKO (2002): “The impact of zoning on housing affordability,”

GOTTFRIES, A. (2018): “Bargaining with renegotiation in models with on-the-job search,” Available at SSRN 3058870.

GOTTLIEB, C., J. GROBOVŠEK, M. POSCHKE, AND F. SALTIEL (2021): “Working from home in developing countries,” European Economic Review, 133, 103679.

GOUSSÉ, M., N. JACQUEMET, AND J.-M. ROBIN (2017): “Marriage, labor supply, and home production,” Econometrica, 85, 1873–1919.

GUPTA, A., V. MITTAL, J. PEETERS, AND S. VAN NIEUWERBURGH (2021): “Flattening the curve: pandemic-induced revaluation of urban real estate,” Journal of Financial Economics.

GYOURKO, J., A. SAIZ, AND A. SUMMERS (2008): “A new measure of the local regulatory environment for housing markets: The Wharton Residential Land Use Regulatory Index,” Urban Studies, 45, 693–729.

HAGEDORN, M., T. H. LAW, AND I. MANOVSKII (2017): “Identifying equilibrium models of labor market sorting,” Econometrica, 85, 29–65.

HANSEN, S., P. J. LAMBERT, N. BLOOM, S. J. DAVIS, R. SADUN, AND B. TASKA (2023): “Remote Work across Jobs, Companies, and Space,” Tech. rep., National Bureau of Economic Research.

HAURIN, D. R., P. H. HENDERSHOTT, AND D. KIM (1993): “The impact of real rents and wages on household formation,” The Review of Economics and statistics, 284–293.

HECKMAN, J. J. (1979): “Sample selection bias as a specification error,” Econometrica: Journal of the econometric society, 153–161.

HILBER, C. A. AND W. VERMEULEN (2016): “The impact of supply constraints on house prices in England,” The Economic Journal, 126, 358–405.

HILL, M. J. (2014): “Homes and husbands for all: Marriage, housing and the baby boom,”

HOLZNER, C. AND B. SCHULZ (2019): “Marriage and Divorce under Labor Market Uncertainty,” .

HORRACE, W. C. AND R. L. OAXACA (2006): “Results on the bias and inconsistency of ordinary least squares for the linear probability model,” Economics letters, 90, 321–327.

INSTITUTE FOR SOCIAL AND ECONOMIC RESEARCH (2020): “Understanding Society: COVID-19 Study,” SN: 8644, 10.5255/UKDA–SN–8644–1.

JOHN, O. P., S. SRIVASTAVA, ET AL. (1999): “The Big-Five trait taxonomy: History, measurement, and theoretical perspectives,” .

JOVANOVIC, B. (1982): “Selection and the Evolution of Industry,” Econometrica, 649–670.

KALAI, E. AND M. SMORODINSKY (1975): “Other solutions to Nash’s bargaining problem,” Econometrica, 513–518.

KAPLAN, G. (2012): “Moving back home: Insurance against labor market risk,” Journal of Political Economy, 120, 446–512.

KAPLAN, G., K. MITMAN, AND G. L. VIOLANTE (2020): “Non-durable consumption and housing net worth in the great recession: Evidence from easily accessible data,” Journal of Public Economics, 189, 104176.

KAPLAN, G. ET AL. (2009): “Boomerang kids: labor market dynamics and moving back home,” Federal Reserve Bank of Minneapolis. Working Paper, 675.

KLEIN, J. (2017): “House price shocks and individual divorce risk in the United States,” Journal of Family and Economic Issues, 38, 628–649.

KNIPE, E. (2017): “Families and households: 2017,” Office for national statistics.

KUENZEL, R. AND B. BJØRNBAK (2008): “The UK housing market: Anatomy of a house price boom,” ECOFIN Country Focus, 5, 1–10.

LAFORTUNE, J. AND C. LOW (2020): “Collateralized marriage,” Tech. rep., National Bureau of Economic Research.

LAUSTER, N. T. (2006): “A room of one’s own or room enough for two? Access to housing and new household formation in Sweden, 1968–1992,” Population Research and Policy Review, 25, 329–351.

LEHMANN, H. AND J. WADSWORTH (2000): “Tenures that shook the world: worker turnover in Russia, Poland, and Britain,” Journal of Comparative Economics, 28, 639–664.

LENTZ, R. (2010): “Sorting by search intensity,” Journal of Economic Theory, 145, 1436–1452.

LI, L. AND X. WU (2019): “Housing price and intergenerational co-residence in urban China,” Journal of Housing Economics, 45, 101596.

LISE, J. AND J.-M. ROBIN (2017): “The macrodynamics of sorting between workers and firms,” American Economic Review, 107, 1104–35.

LOW, H., C. MEGHIR, L. PISTAFERRI, AND A. VOENA (2018): “Marriage, labor supply and the dynamics of the social safety net,” Tech. rep., National Bureau of Economic Research.

LUNDBERG, S., R. A. POLLAK, AND J. STEARNS (2016): “Family inequality: Diverging patterns in marriage, cohabitation, and childbearing,” Journal of Economic Perspectives, 30, 79–102.

MANACORDA, M. AND E. MORETTI (2006): “Why do most Italian youths live with their parents? Intergenerational transfers and household structure,” Journal of the European Economic Association, 4, 800–829.

MAS, A. AND A. PALLAIS (2017): “Valuing alternative work arrangements,” American Economic Review, 107, 3722–59.

MARIMON, R. AND F. ZILIBOTTI (1999): “Unemployment vs. mismatch of talents: reconsidering unemployment benefits,” The Economic Journal, 109, 266–291.

MARINESCU, I. (2016): “Divorce: what does learning have to do with it?” Labour Economics, 38, 90–105.

MASTERS, A. (2008): “Marriage, commitment and divorce in a matching model with differential aging,” Review of Economic Dynamics, 11, 614–628.

MARTINS, N. C. AND E. VILLANUEVA (2006): “Does limited access to mortgage debt explain why young adults live with their parents?” Banco de Espana Research Paper No. WP-0628.

MCELROY, M. B. (1985): “The joint determination of household membership and market work: The case of young men,” Journal of labor Economics, 3, 293–316.

MCFALL, S. (2013): “Understanding Society: UK household longitudinal study: Cognitive ability measures,” Institute for Social and Economic Research, University of Essex.

MCKINNISH, T. G. (2007): “Sexually integrated workplaces and divorce another form of on-the-job search,” Journal of Human Resources, 42, 331–352.

MENZIO, G. AND S. SHI (2011): “Efficient search on the job and the business cycle,” Journal of Political Economy, 119, 468–510.

MIAN, A., K. RAO, AND A. SUFI (2013): “Household balance sheets, consumption, and the economic slump,” The Quarterly Journal of Economics, 128, 1687–1726.

MIAN, A. AND A. SUFI (2009): “The consequences of mortgage credit expansion: Evidence from the US mortgage default crisis,” The Quarterly journal of economics, 124, 1449–1496.

MILES, D. AND V. MONRO (2021): “UK house prices and three decades of decline in the risk-free real interest rate,” Economic Policy, 36, 627–684.

MONDRAGON, J. A. AND J. WIELAND (2022): “Housing demand and remote work,” Tech. rep., National Bureau of Economic Research.

MONGEY, S., L. PILOSSOPH, AND A. WEINBERG (2021): “Which workers bear the burden of social distancing?” The Journal of Economic Inequality, 19, 509–526.

MORTENSEN, D. (2003): Wage dispersion: why are similar workers paid differently?, MIT press.

MOSCARINI, G. AND F. POSTEL-VINAY (2008): “The timing of labor market expansions: New facts and a new hypothesis,” NBER macroeconomics annual, 23, 1–52.

MOSCARINI, G. AND F. POSTEL-VINAY (2012): “The contribution of large and small employers to job creation in times of high and low unemployment,” American Economic Review, 102, 2509–2539.

MOSCARINI, G. AND F. POSTEL-VINAY (2013): “Stochastic search equilibrium,” Review of Economic Studies, 80, 1545–1581.

MOSCARINI, G. AND F. POSTEL-VINAY (2016): “Did the job ladder fail after the Great Recession?” Journal of Labor Economics, 34, S55–S93.

MUELLER, G. AND E. PLUG (2006): “Estimating the effect of personality on male and female earnings,” Ilr Review, 60, 3–22.

MULDER, C. H. (2006): “Population and housing: a two-sided relationship,” Demographic Research, 15, 401–412.

MULHEIRN, I. (2019): “Tackling the UK housing crisis: is supply the answer,” UK Collaborative Centre for Housing Evidence.

NEELS, K., M. MURPHY, M. N. BHROLCH ́AIN, AND ́E. BEAUJOUAN (2017): “Rising educational participation and the trend to later childbearing,” Population and Development Review, 43, 667.

OFFICE FOR NATIONAL STATISTICS (2023): “Characteristics of homeworkers, Great Britain,”
Tech. rep.

ONS (2019): “Divorces in England and Wales: 2018,” Tech. rep., Office of National Statistics.

PELIKH, A., J. MIKOLAI, AND H. KULU (2022): “Make up or break up? Partnership transitions among young adults in England and Wales,” Advances in Life Course Research, 52, 100475.

POSTEL-VINAY, F. AND J.-M. ROBIN (2002): “Equilibrium wage dispersion with worker and employer heterogeneity,” Econometrica, 70, 2295–2350.

POSTEL-VINAY, F. AND J.-M. ROBIN (2004): "To match or not to match?: Optimal wage policy with endogenous worker search intensity," Review of Economic Dynamics, 7, 297–330.

PREVOO, T. AND B. TER WEEL (2015): "The importance of early conscientiousness for socio-economic outcomes: Evidence from the British Cohort Study," Oxford Economic Papers, 67, 918–948.

PROTO, E. AND A. ZHANG (2021): "Covid-19 and Mental Health of Individuals with Different Personalities," Proceedings of the National Academy of Sciences, 118.

QUERCIOLI, E. (2005): "Training, turnover, and search," International Economic Review, 46, 133–143.

RAINER, H. AND I. SMITH (2010): "Staying together for the sake of the home?: house price shocks and partnership dissolution in the UK," Journal of the Royal Statistical Society Series A: Statistics in Society, 173, 557–574.

REIS, S. (2019): "A Home of Her Own: Housing and Women," Tech. rep., Women's Budget Group.

REUBEN, E., P. SAPIENZA, AND L. ZINGALES (2015): "Taste for competition and the gender gap among young business professionals," Tech. rep., National Bureau of Economic Research.

REUSCHKE, D. AND A. FELSTEAD (2020): "The Effect of the Great Lockdown on Homeworking in the United Kingdom," Tech. rep.

ROSENZWEIG, M. R. AND K. I. WOLPIN (1993): "Credit market constraints, consumption smoothing, and the accumulation of durable production assets in low-income countries: Investments in bullocks in India," Journal of political economy, 101, 223–244.

ROSENZWEIG, M. R. AND K. I. WOLPIN (1994): "Parental and public transfers to young women and their children," The American Economic Review, 1195–1212.

RUBIN, O., A. NIKOLAEVA, S. NELLO-DEAKIN, AND M. TE BROMMELSTROET (2020): "What Can We Learn from the COVID-19 Pandemic about How People Experience Working from Home and Commuting?" Tech. rep.

SAIZ, A. (2010): "The geographic determinants of housing supply," The Quarterly Journal of Economics, 125, 1253–1296.

SALCEDO, A., T. SCHOELLMAN, AND M. TERTILT (2012): "Families as roommates: Changes in US household size from 1850 to 2000," Quantitative Economics, 3, 133–175.

SEITZ, S. (2009): "Accounting for racial differences in marriage and employment," Journal of Labor Economics, 27, 385–437.

SHEPHARD, A. (2019): “Marriage market dynamics, gender, and the age gap,” .

SHIMER, R. (2006): “On-the-job search and strategic bargaining,” European Economic Review, 50, 811–830.

SHIMER, R. AND L. SMITH (2000): “Assortative matching and search,” Econometrica, 68, 343–369.

SMITH, L. (2006): “The marriage model with search frictions,” Journal of Political Economy, 114, 1124–1144.

STONE, J., A. BERRINGTON, AND J. FALKINGHAM (2011): “The changing determinants of UK young adults’ living arrangements,” Demographic Research, 25, 629–666.

STONE, J., A. BERRINGTON, AND J. FALKINGHAM (2011): (2014): “Gender, turning points, and boomerangs: Returning home in young adulthood in Great Britain,” Demography, 51, 257–276.

SVARER, M. (2004): “Is your love in vain? Another look at premarital cohabitation and divorce,” Journal of Human Resources, 39, 523–535.

TEULINGS, C. N. AND P. A. GAUTIER (2004): “The right man for the job,” The Review of Economic Studies, 71, 553–580.

WEI, S.-J., X. ZHANG, AND Y. LIU (2012): “Status competition and housing prices,” Tech. rep., National Bureau of Economic Research.

WEISS, Y. AND R. J. WILLIS (1997): “Match quality, new information, and marital dissolution,” Journal of Labor Economics, 15, S293–S329.

WRENN, D. H., J. YI, AND B. ZHANG (2019): “House prices and marriage entry in China,” Regional Science and Urban Economics, 74, 118–130.

ZOOPLA (2022): “Rents reach least affordable level in 10 years,” Zoopla.